The future of historical sociolinguistics?

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

Abstract

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 will 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

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 is 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 multidisciplinarity 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 multidisciplinarity 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 those 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 multidisciplinarity, 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.

Baker, Brezina & McEnery (this volume) provide insights from historical discourse analysis. Like Fitzmaurice et al., they analyse big data (the *Hansard Corpus*) in its historical and social contexts; here the collaboration is between a historian and corpus linguists, with the aim of

uncovering changes in discourse. Their approach is based on tracking collocational change. The innovative methods they present include Meaning Fluctuation Analysis, which can pinpoint periods and collocations of interest for closer analysis, and sparklines, small line graphs that can be embedded in text to convey a great deal of information at a glance. These methods, too, are readily applicable to other materials and could also be used to compare different social groups (in the *Hansard Corpus*, this could include members of different political parties).

While Nevala & Sairio (this volume) position themselves within the "third wave" of variation studies with a qualitative focus on individual agency, they also study representations of discord quantitatively at the communal level. Taking their inspiration from Nevalainen & Tissari's (2010) study of cultural keywords, their innovative combination of methods draws on historical discourse analysis, emotion studies and corpus linguistics. They analyse concepts of *discord* by studying rich data, including the *Corpus of Early English Correspondence Extension* and the *Bluestocking Corpus*, with extensive reference to the socio-cultural background of the eighteenth century and of the individuals in question. Their methodological approach complements the big-data approaches taken by Fitzmaurice et al. and Baker et al., and together these three studies provide a helpful set of models to follow for future semantics- and discourse-oriented research in historical sociolinguistics.

In line with Nevalainen's (2015) thoughts on layered simultaneity, the commonalities between the chapters in the methods section of the volume include an emphasis on a combination of quantitative work, contextualisation and close reading. As for the principle of informational maximalism, multidisciplinarity is essential to all of the studies, to an even greater extent than has been customary before in the inherently multidisciplinary field of historical sociolinguistics. Two of the contributions focus on big data, which is also handled in a methodologically advanced manner by Laitinen et al. (this volume), while the other two contributions showcase the advantages of rich sociolinguistic metadata. Future studies will benefit from the ideas of crossing boundaries and bridging gaps as well as from the specific methods and baseline evidence introduced in these chapters.

2.2 New data for historical sociolinguistic research

As pointed out by Nevalainen (2015: 245), one main challenge faced by the historical sociolinguist are the "material constraints on the reconstruction of past usage". That is, written data has survived randomly and to varying depths of time from different languages, different strata of society and different contexts of use. The survival of written data is patchy, and the more carefully preserved documents tend to represent the upper echelons of society writing in highly formal styles about things of religious, political,

scientific or cultural importance (cf. the previously discussed bad-data problem). Spoken data has to be approximated for the earlier periods, whether through drama or other fictional dialogues, trial proceedings and other forms of recorded spoken language, or the informal registers of language, such as personal correspondence between intimates. It is only in the late nineteenth century that sound recordings became available, and the earliest data only seldom represent the spoken vernacular of the "man in the street", more often recording the voices of the inventors and their intimates.

The development of historical sociolinguistics has gone hand in hand with the variety of materials available to researchers – or recognised as such. In her seminal work in the field, Romaine (1982) engaged more in genre analysis than in historical sociolinguistics as such, as her extralinguistic variables were type of text (verse vs. prose) and style (legalistic vs. literary, for example). Such stylistic comparisons were evident in the next steps of historical sociolinguistics as well, as evidenced by e.g. Tieken-Boon van Ostade (1987), who compared the linguistic practices of sixteen informants over a number of genres, including correspondence, prose fiction and drama. Her work was focused on approaching genres which represented the more informal types of written language and the representations of speech, but her informants were all members of the polite society and published authors.

The data sets of these early scholars allowed the study of high-frequency items in focused contexts, but it was only with the arrival of

computerised corpora that a more extensive field of research has come into existence. While the *Helsinki Corpus of English Texts* is coded for genres and a number of variables for the author's background and the text's setting (Kytö 1996), in practice the author and setting parameters only appear in a very limited number of texts and do not provide much of a starting point for historical sociolinguistic analysis. The first corpus specifically compiled to explore the feasibility of applying the methods of present-day sociolinguistics into the historical study of language, and later, to make such study feasible, was the *Corpus of Early English Correspondence* (CEEC), based on the idea by Terttu Nevalainen and Helena Raumolin-Brunberg. The corpus has been in use since the mid-1990s and has expanded, as a corpus family was built around it, making the systematic study of stratificational sociolinguistics a possibility. (For descriptions of the compilation principles of the corpus and the reasoning behind them, see, e.g. Nevalainen & Raumolin-Brunberg 1994, 1996b and 2003; for a more recent description, see Nevala & Nurmi 2013.)

Different languages face partly different problems with regard to materials. English has a long written history with correspondence by identifiable individuals in robust numbers surviving from the fifteenth century onwards, but other languages, such as Dutch, German or Finnish (to mention a few), tend to have either a shorter written history or less surviving data. All languages share a particular dearth of material written by people representing the lower strata of society and women. In corpus compilation

relying on edited volumes, the long-term bias favouring the important writings of famous men has further acerbated the historical sociolinguist's task. Recent discoveries of previously unknown or unused data caches such as the "Letters as loot" (see e.g. Rutten & van der Wal 2014) or emigrant letters (Elspaß 2012; Laitinen & Nordlund 2012) have provided opportunities of observing less well-educated writers of various European languages expressing themselves in familial surroundings.

There has also been a movement towards embracing "a broader view of evidence than is customary in present-day studies of spoken interaction" (Nevalainen 2015: 250; see also Nevalainen & Rutten 2012: 261–262). One such avenue pursued are prescriptive grammars which, in addition to providing evidence of the prestigious forms of language evident at the time of writing, also give clues on the frequent non-prestigious forms written or spoken (e.g. Tieken-Boon van Ostade 2008; Anderwald 2016).

The questions regarding materials this volume sets out to answer are the following: What kinds of genres can be used for historical sociolinguistic research? What are the challenges and benefits of new and uncharted data sets? How can we study old data sets in new ways? How can we best employ both big and rich data in historical sociolinguistics? To answer these questions, new Finnish materials are brought to the fore by Nordlund & Pallaskallio (this volume). Their data is proof of the benefits of painstakingly searching all kinds of archives for yet untapped sources of writing: the discovery of rural news letters from all over Finland allows

them to trace processes of standardisation over the course of the nineteenth century. Similar archive finds are included in the *Corpus of Nineteenth-Century Scottish Correspondence*, which consists of business correspondence but nevertheless sheds light on the usage patterns of a wide range of people (see e.g. Dossena 2010).

The *Going Dutch Corpus* by Krogull, Rutten & van der Wal (this volume) continues the tradition of making use of ego documents, that is, first-person writings. There is a plethora of genres that represent this type of document, from letters to memoirs, diaries and travelogues, all focused on the first-person experience of the writer (van der Wal & Rutten 2013: 1). These documents can be regarded as being closer to the spontaneous language production of speech than most more polished and formal texts. As mentioned above, documents of this type are also typically a locus where the authentic language of lower strata of society has potentially survived. Krogull et al. take the use of ego documents to the next level through systematically compiling a data set allowing for a diachronic study of changes. They add another dimension to their corpus by taking into account space in a more complex manner than usual, i.e. not only including data from various areas of the country but also building in the opportunity to observe a distinction between centre and periphery. Furthermore, the corpus allows the study of register variation, as Krogull et al. (as well as Nordlund & Pallaskallio in their contribution) make use of both public (newspapers) and private writings to investigate language change from the perspective of

standardisation in the wake of nationalism and the concomitant language ideologies.

The problems of edition-based corpora, such as CEEC, have been solved by smaller, sophisticated digital editions of correspondence into corpora. Examples of such carefully selected and transcribed collections are e.g. the *Corpus of Scottish Correspondence* (see e.g. Meurman-Solin 2007) or the *Bluestocking Corpus* (see e.g. Sairio 2009). The latter, along with two other smaller corpora, the *Bess of Hardwick Corpus* and the *Cocks Corpus*, are used by Kaislaniemi et al. (this volume) to evaluate the usability of the CEEC in the study of orthography. While the practice of evaluating existing materials for suitability for new uses is one overarching trend of historical sociolinguistics, the novelty of the approach of Kaislaniemi et al. lies in taking advantage of sophisticated editions by the authors themselves to evaluate the usability of a large standard corpus of English historical sociolinguistics. With the surviving data – or the corpora already prepared – it is helpful to systematically assess their usability for further studies, providing new insights into usage patterns of the past.

It is not only the papers in the dedicated materials section of the volume that explore issues to do with data. While their focus is more on data-driven methodology, Fitzmaurice et al. (this volume) discover patterns and trends in printed sources, and show that the range of materials which can be approached when inquiring into the social patterning of language use in the past is stunningly broad. Similarly, Baker et al. (this volume), while

developing innovative methodological approaches, tackle the records of

spoken language in the *Hansard Corpus* and bring the object of study for

historical sociolinguistics all the way to the twenty-first century in a long

sweep of diachrony, covering trends of usage and change in the British

Parliament. Laitinen et al. (this volume) explore the uses of present-day

social media language in the form of Twitter messages as a source for

tracking and investigating the spread of ongoing changes in English today.

While their focus is on the theoretical aspects of historical sociolinguistics,

they are at the same time innovative in terms of material selection, aiming to

reach usage patterns of language variation and change in the most informal

of genres.

2.3 Theory: bridging gaps, new challenges

Historical sociolinguistics today are characterised by the understanding of

language as a complex cognitive and social phenomenon situated in

multilayered contexts and several parallel processes. This kind of an

approach calls for a plurality of paradigms as well as new combinations of

diverse approaches on linguistic phenomena. To achieve a more nuanced

sociolinguistic understanding of language variation and change and the

interplay of diachrony and synchrony, multidisciplinary perspectives, such

as input from and influence on social and economic history, social sciences

and data sciences, are needed. Although it is difficult to combine all

possible aspects in one study, current sociolinguistic research is often trying to bridge gaps. In practice, this type of activity may consist, for instance, of combining knowledge produced in different paradigms and employing this to answer sociolinguistic questions (see e.g. Nevalainen 2015; Nevalainen & Raumolin-Brunberg 2012, for recent overviews of the field of historical sociolinguistics).

One of the important theoretical aspects that (historical) sociolinguists have recently focused on includes bridging the gap between approaches highlighting linguistic practices either as individual or group phenomena, i.e. the micro and the macro. Several scholars (including Nevalainen 2015; but see also e.g. Wood 2004; Auer & Hinskens 2005; Palander-Collin et al. 2009; Culpeper & Nevala 2012; Palander-Collin 2012) have discussed the multilayered nature of context and provided models for contextualising language use on several simultaneously operating levels. They have also shown with various situated examples that individual linguistic acts are dependent on broader cultural and societal phenomena, such as who has access to which resources, who can do what, and what is possible or considered appropriate in a specific context and situation.

Another gap we sometimes notice emerges between historical sociolinguistics and sociolinguistics of present-day languages. Historical sociolinguists have tended to rely on models and hypotheses formulated in present-day sociolinguistics as a starting point for their studies, but perhaps

the flow of influence towards the other direction has not been quite so extensive in the past. There are, of course, fundamental differences between studying the present and the past as some standard methodologies of present-day sociolinguistics are not available for historical studies, including data elicitation and the use of spoken language as data. It seems, however, that this gap is narrowing as historical linguists are moving forward in time to the twentieth century and are then able to employ early speech recordings like Jucker & Landert (2015) did in their study. Or, conversely, present-day sociolinguists like Jankowski & Tagliamonte (in this volume) may reach to the (near) past for data and more comprehensive timelines.

As the final gap to be bridged in this context we point out the tendency of historical sociolinguistics to focus primarily on the social and situational constraints conditioning the use of linguistic variables. Even if the significance of linguistic factors is pointed out, this aspect tends to deserve far less attention. The articles in this section of the book in various ways fill in these gaps from their own perspectives and add novel insights to the theoretical understanding of historical sociolinguistics. In this section we ask the following questions on sociolinguistic theory formation: How can insights from other fields inform and challenge theoretical assumptions in historical sociolinguistics? What can new sources of data reveal about the nature of language variation and change? How can historical sociolinguistics be used to inform present-day sociolinguistics?

Hilpert (this volume) provides answers to the first question as he discusses new combinations of theoretical approaches, in this case construction grammar and historical sociolinguistics. In the approach that he suggests, the gap between attention to form, on the one hand, and social factors, on the other, is bridged as language is viewed both as a cognitive and a social system. Although the traditional approaches adopted in construction grammar and sociolinguistics seem almost mutually exclusive, this combination can provide tools and insights that allow researchers to deal simultaneously with linguistic forms, the meanings of their parts, the way they are used, their grounding in particular social practices, as well as distribution across different communities of speakers. The two paradigms can potentially enrich each other as e.g. issues of multilingualism and cross-linguistic phenomena are not usually accounted for in construction grammar but are at the heart of sociolinguistics; and, conversely, sociolinguistics can benefit from attention to speaker-internal processes. Such an approach is one answer to the philosophy of informational maximalism called for by Nevalainen (2015).

Historical sociolinguistics are empirical and data is of vital importance for the formation of theories and testing of hypotheses as Jankowski & Tagliamonte (this volume) show in their study of two vernacular universals, verbal -*s* with third-person plural noun phrases and preterite *come*. The study employs a new combination of unique data sets that enables a trend study with two samples collected in 1975–1981 and

2012–2013, allowing access to language use of individuals born in the 1880s to the late 1960s. The under-explored variety of Ottawa Valley English provides a window to a dialect in the process of recession, yet with strong roots in the local community. In this case, the verbal -*s* is slowly dissipating and becoming restricted to fossilized contexts that were the most favoured locations for the variant in the past. Moreover, the dissipating plural -*s* is no longer socially constrained. Preterite *come* behaves in quite the opposite way as it is maintained and shows age-grading as older people perhaps find community values important and mark this with the increasing use of a dialectal form. These novel results have implications for models of dialect dissipation and concentration and as such they are valuable baseline evidence that can be tested in other time periods, languages and contexts to establish more general sociolinguistic tendencies, as we do not yet have a clear picture as to how social variation corresponds with different stages of change (for social variation of linguistic features in different stages of change in eighteenth-century English, see Nevalainen et al. forthcoming).

One aspect of historical sociolinguistics that research in the field has paid particular attention to is the impact of norms and standardisation, notably alongside a strong focus on language as used by individuals from various social backgrounds. The interest in norms and language standardisation is reflected in the publications of an accumulating body of comparative knowledge from languages such as Dutch, English, French and German as to how norms affect use and language change (e.g. Deumert and

Vandenbussche 2003; Rutten et al. (eds.) 2014; Tieken-Boon van Ostade 2014 and her earlier work). In this volume, Anderwald employs a newly composed data set containing both British and American grammar books of the nineteenth century to explore the role of metalinguistic discourse in language change and the development of vernacular universals. In particular, she focuses on the historical evolution of stigmatisation as she describes the "nonstandardisation" of features that used to be "normal" in earlier English, but are now highly stigmatised, such as multiple negation, *you was* or adverbs without -*ly*. Language change may happen below the level of consciousness or even contrary to prescription, but when processes of overt social conditioning such as stigmatisation take place, they can be very powerful determinants of change. The nineteenth century examples of nonstandardisation discussed in this chapter show how specific linguistic forms were linked to strong social class awareness and used as a tool of social distinction. Anderwald's study produces a model that she uses to analyse how individual features were stigmatised with negative epithets referring either to education, mistake, social evaluation, or logic. Her results show that the features were stigmatised differently and that British and American prescriptivism differed from each other.

Finally, Laitinen et al. (this volume) test the social network model with an entirely new type of data, which is both big and rich as well as written and present-day, namely tweets. They show that methodologies and theories can and should be developed across disciplinary boundaries and

that historical and present-day sociolinguistics enrich each other. Present-day big data is still more easily available, but their study shows that digital archives and digitalised data is of vital importance for historical studies just as well. Furthermore, present-day studies can be enriched and developed by focusing on entirely new type of data, in this case modelled on sociohistorical corpora with metadata enabling the identification of individuals and the construction of their network structures. With their data Laitinen et al. can ask new questions such as whether the distinction between loose- and tight-knit networks disappears if network size is increased. They argue that innovations spread in weak-tie networks but networks with slightly stronger ties can also spread innovations provided that they are large enough, at least c. 100–130 individuals. This is a macro approach to non-native English in the Nordic region, but again the understanding of the Nordic context and the role of English in these countries is vital.

It is quite clear that historical sociolinguistics and its theory formation today are forward-looking and integrationist. If the cry in the past was to put more "socio" into sociolinguistics, sociolinguistic endeavours today are increasingly multidisciplinary and constantly looking for new methods and data. This in turn gives impetus to new questions and deeper, more nuanced understandings of sociolinguistic phenomena. Over the past three decades or so, we have developed a solid understanding of historical sociolinguistics and have produced a bulk of research results on

sociolinguistic variation and change. The future challenges relate to at least two issues. First, we should ensure that we accumulate knowledge and build on previous results. In this endeavour we cannot really separate historical sociolinguistics and sociolinguistics. Second, the question of layered simultaneity (Nevalainen 2015) has now been identified and taken seriously, but reaching a large variety of contexts and practices is not necessarily easy for a historical linguist and "the work in matching parallel data sources has just begun" (Nevalainen 2015: 266). In practice, answering this challenge means working even more closely with, for instance, historians who deal with various data sources and at the same time moving even closer to the emerging paradigm of digital humanities.

3. Conclusion: the future?

It was the aim of our chapter to outline some future paths for historical sociolinguistics on the basis of current, cutting-edge research and the kinds of questions many historical sociolinguists, whatever language they are working with, are pondering right now. We focused on new methodologies, data and theory in separate subsections – and have organised the rest of the book accordingly. As previously pointed out, these aspects are interlinked as, for instance, new methodologies or new type of data make new questions

possible, and new questions contribute to the development of sociolinguistic theory.

Taking Terttu Nevalainen's 2015 article as a starting point allowed us to discuss the notion of layered simultaneity in historical sociolinguistics and use the concept to reflect how the studies in this volume contribute to the idea and operationalise layered simultaneity. Such key components as multidisciplinarity, multiple data sources, crossing disciplinary boundaries and bridging gaps between different levels of analysis were identified as they add to informational maximalism that helps us to reconstruct contexts of language use as richly as possible and see what the past was like more clearly.

For some time in the past, research used to be carried out according to the traditions of each language and each separate philology, but since the explicit establishment of the field of historical sociolinguistics (cf. Romaine 1982) and ensuing collaborations, this has certainly changed in the last one to two decades (cf. Deumert & Vandenbussche 2003 and many other collaborative publications that focus on different languages). In fact, there has been a fair amount of evidence that has revealed similar patterns of development in all linguistic areas. To illustrate this with an example, Elspaß (2012) and Laitinen & Nordlund (2012) identified similar formulaic patterns in the writings of German and Finnish unschooled writers, respectively. A deeper understanding of such commonalities between patterns of use between languages and the different yet similar strategies of

writers and speakers, provided by further cross-linguistic and typological studies, will certainly continue to advance the field of historical sociolinguistics.

Although increasing multidisciplinarity and technological advancements undoubtedly take the field forward, they may also lead us to some directions that are more difficult to foresee. For one thing, historical sociolinguists who want to pursue the suggested directions will need to possess a broader set of skills and learn to work in multidisciplinary – and multilinguistic – teams in "labs", sharing resources, ideas and tasks. Such cross-fertilisation will bring anticipated insights and solutions into the problems of bad data, for example, but it will most likely also bring forth new ideas altogether. In terms of undergraduate education this often means a change of mindset on the part of teachers and students alike. Ideally, multidisciplinary and multilinguistic cooperation will increase general awareness of the kind of knowledge that historical sociolinguists produce and make it even more influential outside its own circles.

References

Anderwald, Lieselotte. 2016. *Language between description and prescription: Verb categories in nineteenth-century grammars of English*. Oxford: Oxford University Press.

Auer, Peter & Frans Hinskens. 2005. The role of interpersonal
accommodation in a theory of language change. In Peter Auer, Frans
Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and
divergence in European languages*, 335–357. Cambridge:
Cambridge University Press.

Bell, Allan. 2013. *The guidebook to sociolinguistics*. Malden & Oxford:
Wiley-Blackwell.

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.

Culpeper, Jonathan & Minna Nevala. 2012. Socio-cultural processes and the
history of English. In Nevalainen & Closs Traugott (eds.), 365–391.

Deumert, Ana & Wim Vandenbussche (eds.). 2003. *Germanic
standardizations: Past to present*. Amsterdam: John Benjamins.

Dossena, Marina. 2010. Building trust through (self-)appraisal in
nineteenth-century business correspondence. In Päivi Pahta, Minna
Nevala, Arja Nurmi & Minna Palander-Collin (eds.), *Social roles
and language practices in late modern English* (Pragmatics and
Beyond New Series 195), 191–209. Amsterdam: John Benjamins.

Dossena, Marina & Gabriella Del Lungo Camiciotti (eds.). 2012. *Letter
writing in late modern Europe* (Pragmatics and Beyond New Series
218). Amsterdam: John Benjamins.

Elspaß, Stephan. 2012. Between linguistic creativity and formulaic restriction: Cross-linguistic perspectives on nineteenth-century lower class writers' private letters. In Dossena & Del Lungo Camiciotti (eds.), 45–64.

Fairclough, Norman. 1992. *Discourse and social change*. Cambridge: Polity Press.

Hilpert, Martin. 2011. Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics* 16(4). 435–461.

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(2). 137–150.

Hope, Jonathan & Michael Witmore. 2010. The Hundredth Psalm to the tune of "Green Sleeves": Digital approaches to Shakespeare's language of genre. *Shakespeare Quarterly* 61(3). 357–390.

Janda, Richard D. & Brian D. Joseph. 2003. On language, change, and language change – or, of history, linguistics, and historical linguistics. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 3–180. Malden & Oxford: Blackwell.

Jucker, Andreas & Daniela Landert. 2015. Historical pragmatics and early speech recordings: Diachronic developments in turn-taking and narrative structure in radio talk shows. *Journal of Pragmatics* 79. 22–39. DOI: 10.1016/j.pragma.2015.01.010

Kytö, Merja (ed.). 1996. *Manual to the diachronic part of the Helsinki Corpus of English Texts: Coding conventions and list of source texts*, 3rd edn. Helsinki: University of Helsinki. http://clu.uni.no/icame/manuals/HC/INDEX.HTM (10 April, 2017).

Labov, William. 1994. *Principles of linguistic change, volume 1: Internal factors* (Language in Society 20). Oxford: Blackwell.

Laitinen, Lea & Taru Nordlund. 2012. Performing identities and interaction through epistolary formulae. In Dossena & Del Lungo Camiciotti (eds.), 65–88.

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_neval
ainen/ (10 April, 2017).

Lyding, Verena, Stefania Degaetano-Ortlieb, Ekaterina Lapshinova-
Koltunski, Henrik Dittmann & Chris Culy. 2012. Visualising
linguistic evolution in academic discourse. In Miriam Butt, Sheelagh
Carpendale, Gerald Penn, Jelena Prokić & Michael Cysouw (eds.),
*Proceedings of the EACL 2012 joint workshop of LINGVIS &
UNCLH*, 44–48. Stroudsburg, PA: Association for Computational
Linguistics.

Mäkelä, Eetu, Tanja Säily & Terttu Nevalainen. 2016. Khepri – a modular
view-based tool for exploring (historical sociolinguistic) data. Paper
presented at Digital Humanities 2016, Kraków, July 2016.

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.

Meurman-Solin, Anneli. 2007. Annotating variational space over time. 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/meurman-solin/
(10 April, 2017).

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 digital data* (Studies in Variation, Contacts and Change in English 14). Helsinki: Varieng. http://www.helsinki.fi/varieng/series/volumes/14/nevala_nurmi/ (10 April, 2017).

Nevalainen, Terttu. 2012. New perspectives, theories and methods: Historical sociolinguistics. In Alexander Bergs & Laurel Brinton (eds.), *English historical linguistics: An international handbook, vol. 2* (Handbooks of Linguistics and Communication Science 34.2), 1438–1457. Berlin & Boston: De Gruyter Mouton.

Nevalainen, Terttu. 2015. What are historical sociolinguistics? *Journal of Historical Sociolinguistics* 1(2). 243–269.

Nevalainen, Terttu & Elizabeth Closs Traugott (eds.). 2012. *The Oxford handbook of the history of English*. Oxford & New York: Oxford University Press.

Nevalainen, Terttu, Minna Palander-Collin & Tanja Säily (eds.). Forthcoming. *Patterns of change in 18th-century English: A sociolinguistic approach*. Amsterdam & Philadelphia: John Benjamins.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 1994. Sociolinguistics and language history: The Helsinki Corpus of Early English Correspondence. *Hermes, Journal of Linguistics* 13. 135–143.

Nevalainen, Terttu & Helena Raumolin-Brunberg (eds.). 1996a. *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence* (Language and Computers: Studies in Practical Linguistics 15). Amsterdam: Rodopi.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 1996b. The Corpus of Early English Correspondence. In Nevalainen & Raumolin-Brunberg (eds.), 39–54.

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. 2012. Historical sociolinguistics: Origins, motivations, and paradigms. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 22–40. Malden & Oxford: Wiley-Blackwell.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 2017. *Historical sociolinguistics: Language change in Tudor and Stuart England*, 2nd edn. London: Routledge.

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.

Nevalainen, Terttu & Gijsbert Rutten (eds.). 2012. *Comparative historical sociolinguistics*. A special issue of *Neuphilologische Mitteilungen* 113(3).

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.

Palander-Collin, Minna. 2012. Social roles, identities, and networks. In Nevalainen & Closs Traugott (eds.), 412–421.

Palander-Collin, Minna, Minna Nevala & Arja Nurmi. 2009. The language of daily life in the history or English. Studying how macro meets micro. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1400–1800)* (Pragmatics & Beyond New Series 183), 1–21. Amsterdam & Philadelphia: John Benjamins.

Romaine, Suzanne. 1982. *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press.

Rohrdantz, Christian, Annette Hautli, Thomas Mayer, Miriam Butt, Daniel A. Keim & Frans Plank. 2011. Towards tracking semantic change by visual analytics. In Yuji Matsumoto & Rada Mihalcea (eds.), *HLT*

'11: proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies: short papers – volume 2, 305–310. Stroudsburg, PA: Association for Computational Linguistics.

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.

Rutten, Gijsbert & Marijke J. van der Wal. 2014. *Letters as loot. A sociolinguistic approach to seventeenth- and eighteenth-century Dutch* (Advances in Historical Sociolinguistics 2). Amsterdam: John Benjamins.

Rutten, Gijsbert, Rik Vosters & Wim Vandenbussche (eds.). 2014. *Norms and usage in language history, 1600–1900: A sociolinguistic and comparative perspective* (Advances in Historical Sociolinguistics 3). Amsterdam: John Benjamins.

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.

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.

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.

Sairio, Anni. 2009. *Language and letters of the Bluestocking network. Sociolinguistic issues in eighteenth-century epistolary English* (Mémoires de la Société Néophilologique de Helsinki LXXV). Helsinki: Société Néophilologique.

Siirtola, Harri, Poika Isokoski, Tanja Säily & Terttu Nevalainen. 2016. Interactive text visualization with Text Variation Explorer. In Ebad Banissi, Mark W. McK. Bannatyne, Fatma Bouali, Remo Burkhard, John Counsell, Urska Cvek, Martin J. Eppler, Georges Grinstein, Wei Dong Huang, Sebastian Kernbach, Chun-Cheng Lin, Feng Lin, Francis T. Marchese, Chi Man Pun, Muhammad Sarfraz, Marjan Trutschl, Anna Ursyn, Gilles Venturini, Theodor G. Wyeld & Jian J. Zhang (eds.), *Proceedings of the 20th international conference on Information Visualisation (IV 2016)*, 330–335. Los Alamitos, CA: IEEE Computer Society.

Siirtola, Harri, Tanja Säily, Terttu Nevalainen & Kari-Jouko Räihä. 2014.
Text Variation Explorer: Towards interactive visualization tools for corpus linguistics. *International Journal of Corpus Linguistics* 19(3). 417–429.

Tagliamonte, Sali A. & Derek Denis. 2014. Expanding the transmission/diffusion dichotomy: Evidence from Canada. *Language* 90(1). 90–136.

Tieken-Boon van Ostade, Ingrid. 1987. *The auxiliary* do *in eighteenth-century English. A sociohistorical linguistic approach*. Dordrecht: Foris Publications.

Tieken-Boon van Ostade, Ingrid (ed.). 2008. *Grammars, grammarians and grammar-writing in eighteenth-century England* (Topics in English Linguistics 59). Berlin: De Gruyter Mouton.

Tieken-Boon van Ostade, Ingrid. 2014. Eighteenth-century English normative grammars and their readers. In Gijsbert Rutten, Rik Vosters & Wim Vandenbussche (eds.), *Norms and usage in language history, 1600–1900: A sociolinguistic and comparative perspective*, 129–150. Amsterdam: John Benjamins.

van der Wal, Marijke J. & Gijsbert Rutten. 2013. Ego-documents in a historical-sociolinguistic perspective. In Marijke J. van der Wal & Gijsbert Rutten (eds.), *Touching the past. Studies in the historical sociolinguistics of ego-documents* (Advances in Historical Sociolinguistics 1), 1–17. Amsterdam: John Benjamins.

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.

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.