“Making the Best Use of Bad Data”

Phonological Variation
and
Historical Sociolinguistics

Master Thesis

Language, Mind and Society
29 June 2012

Moragh Gordon
3342336
Supervisor: Dr Anita Auer

Table of Contents

List of Maps

List of Tables

1. Introduction

In one of his seminal works Labov (1994: 11) stated about historical linguistics that it can "be thought of as the art of making the best use of bad data." He refers to the "limitations of [historical] data that can not be compensated for (11)." This is certainly true and even more so when the quantitative sociolinguistic approach is being applied to historical data, since this requires information about the social position and structure of a language user but also a sufficient amount of data to make relevant observations. In the case of historical data researchers cannot set up controlled experiments and sample a controlled group of informants. Research into the field of historical sociolinguistics is dependent on the material that has been preserved, and this has serious implications for the results that can be retrieved from the data. Nevertheless, these challenges have not stopped historical linguists from applying Labov's and other sociolinguistic theories to diachronic data. In the past few decades the field of historical sociolinguistics has developed into a whole new branch. Suzanne Romaine (1982) was one of the first who took the step to apply sociolinguistic methods to historical data and showed that certain problems with regard to bad data are not necessarily insurmountable. One of the great advantages of the present time is the increasing availability of large language corpora containing (written) language of a wide variety of genres and styles and language varieties. Through the availability it has become relatively easy to study large amounts of data in a short amount of time. Terttu Nevalainen and Helena Raumolin-Brunberg (2003) have shown that the *Corpus of Early English Correspondence* that was specially compiled to test Labovian theories on historical data is a fruitful source for sociolinguistic research and they were able to establish variants that could be related to social variables. To date research in the field of historical sociolinguistics has primarily dealt with syntactic variation and change. In addition to that the kind of language use that has hitherto been studied within the field of historical sociolinguistics is the language use of the higher orders of past societies. One of the

obvious reasons that phonological variation has received less attention is that the only source of historical language use is written language. Then again, as for instance Wyld (1920) Milroy (1992) and (Beal 2004) have shown it is possible to reconstruct English speech of past times through the means of spelling variation and comments by contemporaries, although it becomes more difficult to use spelling variation from the time that spelling was standardized, i.e. c. 1476, onwards. One of the reasons for why the language use of the lower social classes has not received much attention to date might be because this kind of data is simply not available. After all, since up until the eighteenth and nineteenth centuries literacy was a privilege reserved for the higher orders of society. As historical sociolinguistics is dependent on written data this of course implies that higher class language use is studied more. Another factor might be that languages in the past decades tended to be studied from the perspective of language ideologies (Milroy1992; Elspaß 2007). I will return to this issue and discuss it in more detail in Chapter 5. The collection of *Letters of Artisans and the Labouring Poor* that will be used in this case study may be an excellent source to study both phonological features and language use of the lower classes of Later Modern English Society. First, because the collection contains language of the lower classes of society, and second, the spelling often strongly reflects speech and thus provides an opportunity to study phonological features in more detail.

The aim of this thesis is to see if quantitative sociolinguistic theory is applicable to historical phonological data. Since this has not been done so extensively before[1], the first objective is to see what problems can be expected, what challenges are involved and whether we can find possible solution to these problems. The second objective is to test the theory on historical data by means of a small case study, so that possible problems can be identified and

[1] Thomas Toon (1983) has used quantitative methods to study change in Early Old English vowels but mainly focused on the internal linguistic factors. Tony Fairman (2003) has looked at the reflection of speech but not in a systematic way.

the feasibility of a phonological approach will be assessed. The first chapter of this thesis will deal with the challenges and solutions as found in previous studies in the field of historical linguistics. The following chapters will focus on the case study; first will be chapter 3 on the (social) historical context of the language use that will be studied in the case study. Chapter 4 will be a corpus description and the possible limitations but also prospects with regard to the material will be discussed. Chapters 5 and 6 will be devoted to a case study on *h*-dropping and the implications the results of this study have on the theories and problems posed in the first chapters.

2.1 Sociolinguistic Theory and the Application to Historical Data: An Overview

Up until the 1960s linguists tended to treat languages as homogenous entities. It had been assumed that each language could be defined by a clear set of systems and features (McColl Millar 2007). The fact that languages in reality were highly heterogeneous was largely ignored. Although variation in languages had been observed, the variation that could not be related to linguistic factors was referred to as free variation, which meant that this variation was "arbitrary and of no significance" (McColl Millar 2007: 336). As Aitchison (2012: 49) points out, those linguists "unwittingly omitted the evidence that was needed to study change in progress." Chomsky, according to Wardhaugh (2010: 3), too, argues that in order to find out what constitutes a language, it is important to focus on a speaker's competence of a language and not his performance. According to Chomsky, a speaker's knowledge about a language gives insight into the underlying system of a language, whereas the actual performance appears to be unsystematic and redundant in constructing a language's grammar. However, many linguists, among whom the leading figure William Labov, started to argue that language variation was not arbitrary but patterned and socially determined (Wardhaugh 2010: 3). Moreover, they discovered that variation appears to be the vehicle of language change. In the 1960s William Labov was one of the first to find correlations between the degree of variation in an individual and extra-linguistic, or more precisely, social, factors (McColl Millar 2007: 335-337). Labov also established that this seemingly arbitrary variation was a sign of change in progress. In his groundbreaking study of the island of Martha's Vineyard he studied the "frequency and distribution of phonetic variants of /ay/ and /aw/ in the several regions, age levels, occupational and ethnic groups" (Labov 1972: 1). Labov (1972: 2) found that he could correlate the linguistic pattern of variation with social patterns and in doing so it was "possible to isolate the social factors which bear directly upon the linguistic process". Until Labov's quantitative approach to variation had been successfully

applied, it was only possible to explain how a variant came into being. Change of a word, for instance, may have been triggered by internal linguistic processes, by borrowing or by analogy (Labov 1972: 161). How such a variant spread and finally replaced an older variant was more difficult to explain though. There are internal, structural forces that partly explain the spread of a change and this is one of the tasks that historical linguistics deals with. By studying changes that have already been completed in the past, it is possible to explain into which direction a change spread, and what general linguistic constraints were involved (Labov 1972: 161). However, in order to address questions about how a change proceeded and what factors triggered it, a language needs to be considered in its social context (Labov 1972: 161). Labov states that "the question of the mechanism of change, the inciting causes of change, and the adaptive functions are best analyzed by studying in detail linguistic changes in progress" (Labov 1972: 161). An important notion in this line of reasoning is the uniformitarian principle which presupposes that the linguistic forces that operate today are no different from the forces that operated in the past. In other words, the processes or forces that can be observed at present are no different from the processes and forces that were at play in the past (Labov 1972; Romaine 1982; Nevalainen & Raumolin-Brunberg 2003: 8). This notion moved the study of language change from a synchronic to a diachronic perspective, using the present to explain the past.

Suzanne Romaine (1982) was one of the first to move the mainly synchronic variationist study into a diachronic dimension. She wanted to test how the modern Labovian sociolinguistic model could deal with historical data. Until then, variation theory had mainly been tested on synchronic speech data and historical studies had not necessarily taken modern sociolinguistic theories into consideration. Romaine's objective was to find out whether this theory was also applicable to both historical data and syntactic data, as opposed to synchronic phonological data. In her study she examined relativization in Middle Scots as found in

different text types. Variation between wh-forms, *that* and the omission of relatives is typical of sixteenth-century Scots texts. Romaine hypothesized that the variation would correlate with linguistic factors as well as with social factors. Moreover, in texts too there would be variation in a patterned way, as had already been attested in present-day speech. In other words, the writer's or the speaker's choice of a variant is never completely free (Romaine 1982). Choices will always be constrained by (extra)linguistic factors. Sociolinguistic theory could measure this variation objectively (Romaine 1982). As pointed out, one of the main challenges of working with historical data is that it concerns written language instead of spoken language. Romaine (1982: 14) challenges the view that speech is the only true manifestation of a language. She rather regards speech and writing as independent "linguistic behaviours or events which may be realized in different channels". She does not consider writing as a function to record speech but as something that has an "independent existence" (1982: 15). Romaine defines the notion of variation as alternations that may be linguistically meaningless; but that carry stylistic or social meaning. This approach might be more straightforward in the case of variation on a phonological level. Syntactic variation may pose a challenge because syntactic variants tend to carry meaning, whereas phonological variants mainly differ in appearance; they do not carry a different meaning. When it comes to syntactic features, it might be more difficult to establish if a supposed syntactic variant really is a semantic equivalent (Auer and Voeste 2012: 329). In other words, the question is if syntactic variation can solely function as alternate ways to express exactly the same thing. According to Romaine (1982: 35), "a variant must be understood as an alternative realization of an element on the next or some higher level of abstraction in the grammar". Romaine (1982) argues that the notion of the same meaning can be related to different levels of abstraction in the grammar and should therefore not be restricted to the surface structure level of an utterance. Sameness of meaning should therefore be interpreted as functional sameness and not necessarily as

semantic sameness. As Romaine (1982) points out, sociolinguistics has been mainly focused on speech data and the classic sociolinguistic model as developed by Labov was only designed to deal with phonological and phonetic variation. Thus, another problem Romaine had to face was that she investigated syntactic variation, while the theory was not developed in and for that area. In sociolinguistics, written language was often considered to belong solely to historical linguistics (Romaine 1982: 15). The question remains, however, whether and in what way the different text types relate to the social levels as found in spoken language by Labovians. Romaine argues that isolating textual styles is comparable to how Labov tried to isolate different styles within the context of a linguistic interview. Labov had discovered that within the sociolinguistic interview, an individual varied in the use of a variant along a continuum of different settings starting from informal to formal. The degree of the use of a certain variant in each setting could be related to a social continuum. Romaine defines difference in style as a difference in the frequency of the use of alternate features to say the same thing. She points out that Labov's studies show that differences in social dialects differ in quantity and not in quality. This means that this can also be measured in texts. In Romaine's case the quantity of a certain variant of the relativizer can be used to identify different text styles (1982: 117). These styles can then be related to the social dimension (1982: 116). In other words, by measuring stylistic differences Romaine could assign the texts along a stylistic continuum ranging from colloquial style to formal style. She takes departure from the uniformitarian idea and states that current sociolinguistic findings are therefore applicable and useful for studying the past as well. In other words, the relationship between the stylistic continuum and the sociolinguistic continuum is also valid in historical studies. Because this relationship "gives the theory a great predictive power" (Romaine 1982: 123), it should be possible to test the theory on data where one of the important elements is missing. In the case of historical research this is essential because in contrast to a synchronic researcher

a historical researcher cannot rely on his or her own intuitions about the social structure in society. Romaine's line of reasoning is that if certain linguistic variables really are socially diagnostic, they must also be stylistically diagnostic. This means that if variables are stylistically diagnostic they should also be socially diagnostic, which in turn implies that the social context can be reconstructed from historical text styles (1982: 124). The question remains, however, if present-day social class hierarchy really is comparable to that of the past. Records of social historians could be a solution here and although Romaine acknowledges this, she does not deal with this problem. Instead, she mainly focuses on the issues of the quantitative model and how the model could be applied to textual styles. However, she shows that the quantitative approach can be applied to assign texts along a stylistic continuum by determining the frequency of certain variants of a grammatical variable.

Nevalainen and Raumolin-Brunberg (1996; 2003), on the other hand, are more concerned about the issue of sociohistorical reconstruction. In contrast to Romaine (1982), they aim to study language that approaches spoken language as closely as possible because they believe that informal spoken language is the setting in which most linguistic changes take place. In addition to that they do not focus on the textual variation but on variation as found in the individual. Thus their approach is closer to how present-day sociolinguistic research is carried out. Nevalainen and Raumolin-Brunberg studied morphological and syntactic changes that took place in the course of Late Middle and Early Modern English (c.1410-1681) by examining personal correspondence of that time. They have chosen to study the genre of personal letter writing because in the continuum of different writing styles this genre seems to be closest to informal spoken language. Their particular focus lies on social factors such as the author's age, social status, gender, residence, and relationship with the correspondent (2003: 2). In synchronic studies these are established variables that correlate with language change and variation (2003: 10). Nevalainen and Raumolin-Brunberg's

objective is to see if generalizations based on synchronic research also apply to historical data. Moreover, because these researchers have access to data that cover a relatively long period, it is possible to carry out both real time studies and apparent time studies and compare the two different approaches. The great advantage of historical research is that, according to Raumolin-Brunberg (1996: 19), "[w]e know, at least approximately, when a particular change took place and what the outcome was." In this way it is possible to examine the validity of apparent time approaches. One of the advantages is that the social conditions of fifteenth-, sixteenth- and seventeenth-century England are fairly well-documented and that it is possible to reconstruct the social structure of this period by for instance using information from the field of historical sociology (Nevalainen and Raumolin-Brunberg 2003: 2). An important source of information with respect to the use of language variants are comments by contemporaries on language use of that time. From the seventeenth century onwards comments on language variation started to appear more and more frequently in the form of written grammars. This kind of data is not without its problems, however. The further one goes back in time, the amount of available sources decreases and if it is available at all, the information may be difficult to interpret, or based on stereotypes and tainted by attitudinal remarks (Nevalainen and Raumolin-Brunberg 2003: 7). Another valuable source from around 1500 onwards are public debates about the issue of spelling reform, also referred to as the Inkhorn Controversy (Nevalainen and Raumolin-Brunberg 2003: 7). The discussion was based on the question if English could be considered a fully fledged national language. However, this was a debate that typically arose among intellectuals and does therefore reveal little about the illiterate majority. Indeed, a shortcoming of written data from this period is that only a small minority could write and since it concerns written data, the researcher will not have access to the language use of a large majority of the studied language community.

Nonetheless, Nevalainen & Raumolin-Brunberg were able to consolidate present-day generalizations by studying historical data. One interesting topic they focused on is what is called the gender paradox. In present-day sociolinguistics it has been firmly established that women opt for prestige forms more than men do and that they use innovative forms more frequently than men do. In general, women seem to be in the lead when it comes to the propagation of a change. Nevalainen and Raumolin-Brunberg (2003) found in their data that with respect to a number of the linguistic variables they studied, the women were indeed in the lead and used the forms more frequently than the men. This is an interesting observation because the social differences between the sexes were certainly different from the social differences at present. One major difference between present-day patterns and those found by Nevalainen and Raumolin-Brunberg is the way in which males seem to have taken the lead when a language change was consciously promoted. For instance, the use of multiple negation was commented upon by learned men and, indeed, this feature also disappeared in the writing of male professionals. The explanation for the difference between the pattern of present-day gender differentiation and the way gender differentiation seems to be patterned in Middle and Modern English is the fact that women from late medieval times to early modern times did not have access at all to the professional and intellectual world. Thus, this difference might be explained by the different place that women had in society in the past. What also may play a role is that the majority of the women was illiterate at the time and it is therefore questionable if the sample of women's language from the corpus is representative of the time.

Another factor that is considered to be of major influence in present-day sociolinguistics is social class. One of the challenges is to single out groups that can be assigned to a certain social class. In the case of Nevalainen and Raumolin-Brunberg's study, it was necessary to reconstruct the social structures of Tudor and Stuart England which is fundamentally different from present structures, to come up with a social hierarchy and to

assign individuals to the different ranks accordingly. In doing so they were able to observe that language changes sometimes started in the lower social ranks and spread to higher ranks and vice versa.

2.2. Historical Sociolinguistics; the Challenges

One of the pivotal issues of historical sociolinguistics is what Labov called the problem of 'bad data'. Historical data "are rich in so many ways are impoverished in others. Historical documents survive by chance and not by design and the selection that is available is the product of an unpredictable series of historical accidents" (1994: 11). However, as Romaine's and Nevalainen and Raumolin-Brunberg's studies have shown, it is possible to apply the theories to historical data and more importantly, it offers an opportunity to test the applicability of the theories and the generalizability of assumed present-day universals. It cannot be denied that historical data have their limitations that need to be addressed or dealt with in some way. Hernandez-Campoy (2012: 108-112) summarizes the major issues as found in historical sociolinguistics in seven points, six of which are relevant to the discussion in this chapter: representativeness, empirical validity, invariation, authenticity, authorship, social and historical validity.

1. The Issue of Representativeness:

Historical sociolinguistic researchers will always be dependent on the written sources that have been preserved and they can therefore not pick and sample a group of individuals that can represent a certain social group. In other words, they cannot set up a controlled experiment like a synchronic researcher can. Instead, they will have to spend a considerable amount of time to trace background information in order to establish with what kind of informant they are dealing and they will have to base themselves on the material that is

available. What Hernandez-Campoy does not mention is the problem of literacy. The only historical material available is written data. This inevitably implies that it mainly concerns material written by people from the higher social ranks of society. Up until the nineteenth century the people of lower ranks had largely no access to education and could therefore not be able to read and write. Consequently, a large majority is automatically underrepresented in historical sociolinguistic research (Nevalainen and Raumolin-Brunberg 2003; Beal 2004; Hebda 2012).

2. The Issue of Empirical Validity

The amount of data that has been preserved is often limited and it is at times difficult to carry out a reliable quantitative analysis. The limitations often also affect the information available about the social structure or the social background of an informant.

3. Invariation

As written material is often more formal and normative than spoken language, this implies that written language will inevitably contain less variation than the spoken language at a particular time. Raumolin-Brunberg (1996: 19) points out that

> we inevitably have to work on a relatively general level of analysis. For instance, we can classify all our informants according to their social status, but only some of them offer material for stylistic variation. This means that we analyse social indicators rather than markers, but, on the other hand, we are sometimes able to refer to stereotypes on the basis of contemporary comments.

4. Authenticity

A problem related to the normative character of texts is that the written language of an author does not necessarily reflect his spoken language because the author is likely to make an attempt to write in a normative style, resulting in hypercorrection or dialect mixture. However, scholars such as Romaine (1982), Nevalainen and Raumolin-Brunberg (1996, 2003), and Hernandez-Campoy (2012) rightfully point out that this is a problem that is not restricted to written data. In speech one will also find hypercorrect forms and dialect mixture. Studies carried out within the framework of sociolinguistic accommodation theory show that it is very difficult to eliminate the individual's tendency to accommodate to the interlocutor's expectations (Wardhaugh 2007: 113). This is what Labov (1972: 209) the *observer's paradox*: "The aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation." Accommodation theories show us, however, that speech will always be monitored in some way, regardless of it being spoken or written language.

5. Authorship

Another problem that arises is that the written material may not always be autographic. Due to widespread illiteracy it was very common that someone other than the sender wrote the letter. In that case the linguistic variants might be related to the wrong independent variables and their correlation might be misinterpreted. For instance, the researcher may think that the text is written by a female of a certain age and social rank, while it has actually been written by a male scribe.

6. Social and Historical Validity

Another one of the challenges that researchers face is the limited access to knowledge regarding the social structure and order of a particular time in the past. This implies that the standard paradigm of social hierarchy of today may not be the same as that in the past. Indeed, Nevalainen and Raumolin-Brunberg (2003) seemed to have overcome this problem by reconstructing a time-specific social hierarchy paradigm and they were able to relate the amount of variation to the different social ranks. It is vital, however, that there is enough social historical information available to produce a reliable social reconstruction.

2.3. Historical Sociolinguistics and Phonology

Not much has been done to study phonological variation in the framework of historical sociolinguistics. Although phonological change has been related to socio-historical events by historical linguists (Hebda 2012: 306-320), few have approached the data within the quantitative variationist framework. As pointed out earlier, both Romaine (1982) and the team Raumolin-Brunberg and Nevalainen (1996; 2003) focussed their historical sociolinguistic studies on syntactic and morphological features. One of Romaine's motivations to do so was that she wanted to extend the hitherto phonology-dominated sociolinguistic theory to other linguistic levels and features, but there are probably also practical reasons for this. After all, it is easier to establish grammatical variants in written language than to discover sound changes because those are rarely reflected in spelling, especially from the time that spelling standardization (c. 1400) started onwards. Moreover, nowadays sounds can be singled out and differences in sound can be objectively measured by means of highly advanced auditory equipment. A Possible historical source that could provide more information about phonological variation is private correspondence because, according to textual historians, personal correspondence may be a text type that is closest to oral language along the

continuum of oral and literate text types and thus might be more likely to contain reflections of speech than the other text types, such as sermons (Nevalainen and Raumolin-Brunberg 2003: 29). Another source might be poetry or drama where the author intended to portray a character speaking a certain dialect (Nevalainen and Raumolin-Brunberg 2003; Beal 2004; Hebda 2012). The latter source, i.e. drama, can hardly be considered a faithful source, however, and is more likely to be subject to extreme stereotyping. Comments on variation by contemporaries might be a more useful source but they are also tainted by value judgments (Nevalainen and Raumolin-Brunberg 2003; Beal 2004; Hebda 2012). On the other hand, comments of contemporaries could give insights into the social evaluation of certain features. Comments on variation in pronunciation become particularly common from the sixteenth century onwards since this was the time when there was a general tendency to purify language in order to maintain a standard variant that resisted corrupting changes (Beal 2004). Toon (1983) has looked at the spread of phonological features of the Mercian dialect during the Old English period and has used the quantitative approach by examining spelling variation in Old English texts from the time of the Mercian hegemony. He explains the variant replacement of <ae> by <e> and <a> by <o> in Kentish texts as evidence of a raising process which could be related to the Mercian 'second fronting process' and therefore was a direct consequence of the Mercian political dominance. Toon (1983) has been criticized on his lack of a thorough social reconstruction though. In addition to that, Hogg (1988) raised doubts as to whether the Kentish raising process and the Mercian second fronting process are related at all (Hebda 2012: 314). Lodge and Milroy both studied phonological features and reconstructed possible socio-historical events that may have been involved in the spread of those features. Milroy (1983) has examined the phenomenon of *h*-dropping and also linked this to socio-historical events (Hebda 2012). A more detailed account on his findings will be discussed in chapter 5. Lodge (2004) has carried out a similar study on variation in fourteenth-century Paris French

and has tried to relate this variation to socio-historical events such as industrialization, urbanization and general attitudes. However, the social factors are not correlated quantitatively but more used as general complementary explanations to dialectal variation. Despite the issues that need to be overcome to be able to carry out historical sociolinguistic research on phonological features, it should be possible, at least to a certain extent, to both reconstruct the social structure in which the studied phonetic variant might be embedded and to reconstruct possible phonetic variation by means of, for instance, spelling variation, contemporary comments and rhyme.

3.1 The Socio-historical Context

As Nevalainen and Raumolin-Brunberg (2003: 8-15) have pointed out, in order to be able to correlate social factors with language variation, it is necessary to reconstruct the social context of the language that is studied. The researcher should not automatically assume that the social structure is identical to that of the present. This may sound paradoxical with regard to the basic assumption of the uniformitarian principle. However, it may be argued that the uniformitarian principle operates on a more abstract level and that basic human needs play a universal role in physical, psychological and social domains (Nevalainen & Raumolin Brunberg 2003: 30).The language variety that will be examined in this study is that used during the period c.1750-1835. It roughly falls within what historical linguists refer to as the Modern English period, which is subdivided into a Early and Late Modern period, but it all depends on what should be taken as the end of the Early modern period and the starting point of the Late Modern Period. This is a much debated issue and while it is beyond the scope of this work to come up with a strict definition, I will briefly point out different suggestions. In the *Cambridge History of the English Language* series, for instance, the date of America's Independence (1776) has been taken as the cut off point. Then again, Görlach (1998: 463) takes historical events as a guideline and defines the beginning of the Industrial Revolution, i.e. 1776-1800, as a possible end point to the Early Modern English period and the start of the Late Modern period. Due to the Industrial Revolution the structure of society changed, which may therefore be seen as an appropriate starting point for the Late Modern period with regard to a sociohistorical perspective since "the prestige and stigmatization of linguistic varieties, and in consequence, frequency of use of individual 'styles' depends on the type of society speakers live in"(Görlach 1998: 463). The period that is covered by the letters of this study can be characterized as the period where the Industrial Revolution began and it was also the threshold of an era that saw a significant increase in the spread of literacy (Görlach 1999: 6).

In her monograph *English in Modern Times,* Beal (2004) describes the period that includes the beginning of the eighteenth century and the end of World War II as the modern period of English, avoiding a strict boundary between early and later modern English. Since it is the aim of this study to reconstruct the social and external factors that might have influenced language use, it might be relevant to look at important movements and events that may be seen as the historical foundations of the period from around 1700-1850 and that might have affected the way in which people viewed language. Beal (2004: 2) considers the Restoration of the monarchy in 1660 as an important turning point, since this time was marked by a major change in the political structure of Britain. In 1689 the Bills of Rights brought an end to the absolute and divine power of the monarch; a sovereign could no longer decide on laws or raise taxes without the consent of parliament. This fits in with the Enlightenment movement that was characterized by a rational way of viewing the world; people no longer placed their faith solely in the hands of God and started to explore the strength of rational thinking. Divine right of a monarch was no longer accepted as a given and replaced by a more or less constitutional monarch. This new way of thinking contributed to new discoveries and the world of science thrived. It heralded the beginning of the Industrial Revolution which opened the gateway to capitalism and self-made men. It also caused a shift in the way society was ordered. Aristocracy gradually started to be replaced by plutocracy. Social esteem was no longer determined by birth but by "those who had gained influence through wealth" (Beal 2004: 5), and who were in a position to receive education. Education and especially the attainment of a prestige variety of English was an important means to show class distinction. As Beal (2004: 5) points out, "'new money' was no guarantee of class, but money could buy access to the public schools." Furthermore, with the Industrial Revolution an industrial working class was born (Romaine 1998: 11). The Industrial Revolution but also the improvement of the infrastructure started a surge of migration and urbanization. Travelling

became easier and faster, new and improved roads were constructed between the 1730-70s, and industries created new job opportunities. In 1750 London was the only town in Britain that had a relatively large population but by 1801 there were seven cities that were populated with more than 50.000 citizens. By that time London had approximately 1.1 million inhabitants (Beal 2004: 2-7).

3.2 Literacy and Education

Education was largely a privilege reserved for the higher classes and the affluent (Stone 1969, Cressy 1980). However, although the First Education Act, which introduced elementary compulsory schooling, was not passed until 1870, and a real revolution in literacy was yet to begin, there was already a growing sense among the people that reading and writing were valuable skills for everyone (Cressy 1980: 1). This view was mainly held by religious men. The reformation called for a new approach to the profession of faith. Being able to read the Bible was one of the great virtues of Christianity and would make a person a better Christian (Cressy 1980: 1). Being able to read and to memorize the sacraments secured a person's way to salvation. Literacy was believed to be an important means to spread religion and Christians were encouraged to teach their children and their neighbours to read. What also needs to be considered is the influence of the so-called 'Age of Reason'; a movement that can be "characterized by reliance on science and reason and concern for humanity" (van Gelderen 2006: 204). Literacy meant access to a world of written information and it was also a means to improve upon yourself (Cressy 1980). Although this may not have been felt immediately among the lower classes, society became more and more literacy driven (Cressy 1980). Gradually, the lower classes may have been pressed to acquire some literacy skills for practical reasons:

> People involved in trade, specialized manufacturing and farming for the market increasingly found themselves confronted by print or script, and more and more of them maintained written records of their transactions […] One who could write his own letters and dispose of his own affairs in writing would be free of the awkwardness, expense and possible untrustworthiness of the scrivener or writing-man (Cressy 1980: 12).

Stone (1969: 81) argues that especially the rivalry between the different Christian churches in early eighteenth-century England led to a surge in the establishment of Charity schools to educate the poor and lowest classes of society. The Anglicans and non-conformist were driven by a desire to "root out popery" and they did so by instilling their religion through the practice of teaching reading and writing (Stone 1969: 81; Lawson and Silver 1973: 182).The Society for Promoting Christian Knowledge was dedicated to this purpose and sponsored schools for the most poor (Lawson and Silver 1973: 184). However, even in the eighteenth century, educating the lower class mass was considered potentially dangerous by many of the ruling classes because it would make the lower class majority too articulate and insurgent. This could have threatened the "social stability" of society (Stone 1969: 85). The curriculum of the Charity schools seemed to have been designed to constantly remind the children of their "low estate and the duty and respect they owed to their betters" (Lawson and Silver 1973: 184). Yet, the growth of literacy could no longer be stopped. Especially from the 1780s onwards literacy was connected with new ideals of social change (Lawson and Silver 1973: 229). It needs to be added, however, that the lower classes still lagged behind when it comes to literacy rates and if they received schooling at all, it was often very restricted (Stone 1969; Lawson and Silver 1973; Cressy 1980). It may come as no surprise that the writing and spelling skills of the lower classes were therefore often very limited. Moreover, it is very

likely that it was a skill that they rarely needed in their daily lives since, especially the labourers, mainly carried out manual work (Cressy 1980).

3.3 Language in Society: The Emergence of Standard English

One of the developments that had already started in the fifteenth century but that started to attain its full development in the eighteenth century was the emergence of English as a standard language in written form. The introduction of the printing press in 1476 definitely contributed to the acceleration of the process of the codification of a supra-regional standard (Görlach 1998). The fast and extensive spread of printed Standard English reading material "almost automatically devalued the use in writing of all forms that were locally or otherwise deviant" (Görlach 1998: 460). Before 1476 there was not really a particular prestigious variety that competed with other varieties (Lass 1998: 6). The sense that a standard language was desirable fitted in with the Enlightenment and late Renaissance too (Lass 1998: 8). Like many phenomena in society, such as social order, politics and science, language too was approached as something that could be analysed and designed by human reason. Standard English could be fixed and designed in the shape of an aesthetic language that would match the elegance and esteem of Latin. This automatically implied that the language had to be rid of features that were perceived as vulgarities, and other features that were considered good had to be preserved (Lass 1998: 8). In the course of the fifteenth and the eighteenth centuries a standard language became more and more defined. With the firmer establishment of a standard language, a sense of the existence of non-standard language grew (Görlach 1998: 532). In the eighteenth and nineteenth centuries, with the establishment of class distinctions by wealth rather than by birth, upward social mobility became a possibility and being able to use the standard variety became a way to display this climb on the social ladder (Romaine 1999: 1). Where the previous two centuries were marked by the discussion whether and how

English as a standard language could fulfil the function of a national language, the eighteenth century was the time during which the standard was fixed and protected from 'corrupting changes' (Romaine 1999: 1). The nineteenth and twentieth centuries marked an imperialist era and a general sense of fear of foreign influences and threat contributed to an even greater desire to preserve an English standard language (Romaine 1999: 1). The eighteenth century, however, may be seen as the heyday of spelling books and prescriptive grammars. Moreover, during that time a standard pronunciation to consolidate English as a standard language even more firmly was developed (Beal 2004). From the second half of the eighteenth century onwards scholars, usually referred to as orthoepists or elocutionists, started to devote themselves to the writing of pronouncing dictionaries that instructed people about what may be considered to be good and bad pronunciation, and from that time onwards a spoken standard was developed next to the written standard (Beal 2004: 127).

3.4. Later Modern English Speech

The description of Late Modern English speech is far from comprehensive when it comes to its social evaluation, the existence of different varieties and the specific features that might be associated with it (Görlach 1998, Beal 2004). Moreover, in terms of sociolinguistic description, information about the speech of lower and middle classes is close to non-existent (Görlach 1998: 495). The only kind of evidence that might reveal more about the speech of the lower classes often comes in the form of negative comments made by contemporary grammarians or orthoepists. Below is an example of such a comment from Walker's list of 'faults' in his *Critical Pronouncing Dictionary*:

> The letter *s* after *st*, from the very difficulty of its pronunciation, is often sounded inarticulately. The inhabitants of London, of the lower order, cut the knot and, pronounce it in a distinct syllable, as if *e* were before it. (Walker 1791: xii)

The amount of information provided by contemporaries with regard to features that corrupt the standard language is overwhelming but, as Görlach (1998) rightfully points out, these descriptions merely show how the educated higher class interpreted the speech of the lower classes and it does not necessarily show a "sociolinguistically realistic picture" (Görlach 1998: 495). Nonetheless, the evidence provided by orthoepists is the most reliable information available to date. Due to the far advanced spelling standardization even the spelling of most personal letters is highly standardized, except for informal letters written by writers who were not classically schooled (Beal 2004: 127). Even rhymes are no longer reliable because at this time words did not necessarily rhyme by sound but by the eye (Beal 2004: 127). Then again, letters written by writers that were not classically schooled, and in the case of this study by lower class people, might serve as a corroboration of what language guardians referred to as low or commoners' speech. Works written during the Late Modern period, by so called elocutionists, scholars who aimed to educate people about 'proper' speech, are amply available and reflect the view that accent was considered an important way to distinguish oneself socially (Mugglestone 1995: 4). Elocutionists who might provide evidence of Late Modern English speech are Kenrick (1773), Sheridan (1780) and Burn (1786), Spence (1814) (Beal 132: 2004). The difficulty with meta-linguistic comments by contemporary elocutionists is that one cannot be sure whether the latter were aware of socially determined differences in register, how well they were informed about phonological theory, and if they had schooling in phonology (MacMahon 1999: 378). Furthermore, the descriptions often lack the precision and sophistication of present-day phonetic transcriptions, which makes it sometimes difficult to interpret the material (MacMahon 1999: 378). It also remains questionable whether the linguistic features the prescriptivists refer to as lowly speech may safely be connected with lower class speech. Mugglestone (1995) points out that many of the proscriptions found in pronunciation guides may also have been used by the higher classes but while the particular

proscribed features may have been perceived as wrong for various other reasons, they were often automatically associated with the lower classes. It might well be the case that some of those proscribed features were used in the higher classes albeit less frequently and depending on the context and the associated register. Mugglestone illustrates this problem by addressing the variation between /ɪn/ and /ɪŋ/ as the realization of present participle *–ing*. Although the use of /ɪn/ came to be strictly associated with the less well educated and the use of /ɪŋ/ became strictly associated with the educated, reality is more complex and it is possible that "the respective percentages of each being stratified alongside social variables such as status and style, context or gender" (Mugglestone 1995: 155) may have been a more likely pattern.

Nonetheless, the works of the elocutionists are indispensable in the attempt to what Labov (1994: 11) refers to as "the art of making the best use of bad data," since other evidence is scant. The author of the *Critical Pronouncing Dictionary* (1791), John Walker, was a very successful elocutionist in the late eighteenth century and his work remained popular up until the early twentieth century (Beal 2004: 129). Walker's works are, according to Beal (2004: 130), prescriptive and normative, but he also seems to give a detailed description of language as it was actually used by the higher social classes, giving insight into the prestigious variant, and there are also proscriptions that give insight into non-standard variants. The terminology Walker uses to describe the articulatory processes that are involved in the production of particular sounds is quite similar to present-day phonetic descriptions (Beal 2004: 130).

Although Late Modern Standard English is believed not to differ from present-day Standard English to a great extent, there are a fair amount of differences to be found. Below is a brief account of the differences and the social implications of some of the contemporary variants as listed by Mugglestone (1995) and Beal (2004) who based themselves on the

accounts of contemporaries as Kenrick (1773), Spence (1775), Sheridan (1780), and Burn (1786).

3.4.1 Vowels

FACE and GOAT Sets

As opposed to the present-day diphthongal realizations, these vowels were monophthongs throughout most of the eighteenth century. The monophthongal variants are still found in regional, i.e. mostly Northern dialects. The earliest evidence that the FACE vowel was realized as a diphthong is provided by Batchelor (1809). The first evidence about the diphthongization of the GOAT vowel dates from 1795. However, frequent manifestations of the use of the diphthongal variants do not occur until the second half of the nineteenth century.

BATH and CLOTH sets

The general trend of this set was that both vowels tended to be lengthened, backed and/or lowered, whereas they used to have shorter realizations in earlier times. In Middle English there were particular phonological environments where these vowels would undergo the following process: The shorter variant of the BATH vowel was lengthened before pre-consonantal /r/, /s/, before /θ/, before /ð/ followed by /s/ or syllabic /r/, as for instance in paths or father. Both the CLOTH and the BATH vowel were lengthened before pre-consonantal /n/, /lf,lm,lv/. In Later Modern English the environments were extended to final /r/ and before all fricatives. Interestingly, the lengthening of the BATH vowels in most BATH words, except some individual words, were considered to be vulgar in the eighteenth and nineteenth centuries, for instance according to the elocutionist Walker. The lengthening of the CLOTH

vowel was also labeled vulgar by Walker (1791), especially when followed by fricatives. Beal (2004: 141) argues that the acceptance of the lengthened BATH vowel in present-day Standard English may be an example of change from below, i.e. a linguistic change that has spread from the lower to higher classes (Elspaß 2007), since "semi-phonetic spellings such as *larst* and *larf* for *last*, *laugh* have been used since the nineteenth century to represent Cockney speech" (Beal 2004: 141). Thus, the lengthened vowel in BATH words in the present-day prestige variant RP might originally be a Cockney feature. The lengthening of the CLOTH vowel too was often typically associated with Cockney speech and words such as *of* were often spelled *orf* to represent the Cockney variant. In this case however the shorter variant seems to be favoured at present, whereas the longer variant is falling out of use (142).

FOOT-STRUT sets

Phonemically the distinction between the FOOT and STRUT sets was probably already present in the eighteenth century. Phonetically the descriptions of the eighteenth-century orthoepists seem to suggest that it has undergone a change; it has almost been fronted to /a/ by the twentieth century, but it was probably more central in the eighteenth century. The shorter /ʊ/ vowel did not seem to be acknowledged by all elocutionists or orthoepists. Walker (1791) generally refers to /u/ in words such as *look* and *took*, whereas Stephen Jones (1798) talks about a shorter variant of /u/ that might well have been associated with /ʊ/. Walker seems to be aware of the shorter variant too, but he seems to consider it a "recent development whose spread should be discouraged" (Beal 2004: 145).

Yod-dropping

Words from the GOOSE and CURE sets consist of a subset of words that in earlier times used to be pronounced as a diphthong /iu/. At some point the /i/ of /iu/ was replaced by a /j/ that subsequently assimilated into the preceding consonant or was dropped altogether. For instance, the /siu/ in sugar first changed into /sju/ and then /j/ assimilated with /s/, resulting in /ʃu/. This yod-dropping had already been attested in the eighteenth century and it was negatively commented on by some of the elocutionists. Walker accepts the y-dropping in some individual words but proscribes the over-extension of it to other words, especially those starting with an /s/. There is however variation between contemporaries with regard to the acceptability of it. Another matter of discussion with regard to yod-dropping in the eighteenth century was the pronunciation in unstressed syllables. Initially, the deletion of /j/ with a reduced vowel was preferred. For example, nature was to be pronounced as /neːtər/ but in the course of the eighteenth century /neːtʃər/ became the preferred pronunciation. The yod-dropping after /t/, /d/, /n/, /θ/, /z/, /s/ and /l/, as for instance in *duke* and *tune*, was stigmatized and associated with Cockney speech.

3.4.2 Consonants

Deletion of /r/

The eighteenth-century elocutionists recommended that the /r/ was pronounced whenever the spelling of a word indicated that. However, Beal points out that there must already have been a process of deletion of /r/ in certain contexts. Especially Walker gives a revealing account on the use and pronunciation of the /r/ in his time (as quoted in Beal 2004: 153):

> In England, and particularly in London, the r in *lard, bard card, regard,* is pronounced so much in the throat, as to be little more than the middle or Italian a lengthened into *baa, baad, cad, regaad* , while in Ireland , the r in these words is pronounced with so strong a jar against the fore part of the palate, and accompanied with such an aspiration or strong breathing at the beginning of the letter, as to produce that harshness we call the Irish accent. But if this letter is too forcibly pronounced in Ireland, it is often too feebly pronounced in England.

According to Beal (2004: 154), this description suggests that there already was an ongoing change of the loss of /r/ in pre-consonantal positions and it can again be described as a change from below. A process related to non-rhoticity is the intrusive /r/. This phenomenon was already noticeable in eighteenth-century Cockney. The /r/ was only deleted in pre-consonantal position but not between two vowels. This rule was also extended across word boundaries. In other words, if a word ended with an /r/ orthographically and it would be followed by a word beginning with a consonant, the /r/ would not be realized but if it was followed by a word beginning with a vowel, the /r/ would be pronounced. This phenomenon is at present quite common in RP and is referred to as the linking /r/. In the case of Cockney, Sheridan (1762) noted that this 'linking'/r/ even occurred where it was not orthographically represented, hence the name intrusive r, ending up with examples as *Americar is*. Non-rhoticity remained highly stigmatized up until the nineteenth century. Again, this seemed to be a feature typically associated with the lower classes. Spence (1814) for instance put this process down to the inability of the lower classes to spell (Beal 2004: 159). In Spence's view, if they had known that the words were written with an r they would have pronounced it.

H-dropping

Although the deletion of the initial *h* before vowels is a phenomenon that has been attested long before the eighteenth century, it has never been condemned before that time. Sheridan refers to *h*-dropping as one of the worst mistakes to make:

> But there is one defect which more generally prevails in the counties than any other, and indeed is daily gaining ground amongst the politer part of the world, I mean omission of the aspirate in many words by some, and in most by others. Were this to become general it would deprive our tongue of one great fund of force and expression. (Sheridan 1769: 49)

Walker associates it with the peculiarities of Cockney speech:

> A still worse habit than the last prevails, chiefly among the people of London, that of sinking the *h* at the beginning of words where it ought to be sounded, and of sounding it, either where it is not seen, or where it ought to be sunk. (Walker 1791: xii)

By the beginning of the nineteenth century *h*-dropping became more and more associated with the uneducated lower classes (Mugglestone 1995: 115) The insertion of *h* in places where it is not orthographically represented was equally condemned. Similarly, the deletion of *h* in wh-words, such as *which, where* and *wheat*, was proscribed. Although this became at some point accepted and is now an established feature of RP, whereas *h*-deletion and insertion in most words are still stigmatized. Again, Spencer (1814) attributed the tendency to insert the *h* in places where it is not orthographically represented to the lack of knowledge about spelling among the uneducated. Some words, however, were written with initial *h* but always had been

silent in speech. These were probably remnants of language contact between the Normans who did not pronounce the *h* in those words either and introduced them as such in the period after the Norman invasion in 1066. This practice seemed to have been extended to native words as well. The use of forms with and without the *h* were used variably but by the eighteenth century this variation was no longer tolerated. Probably due to hypersensitivity to *h*-dropping the *h* was reintroduced in speech, while historically the *h* never had been pronounced in England or was lost at some point at earlier stages. For instance, in Walker's list the word *hospital* occurs as one of the words that contains a silent *h*, whereas it is pronounced in most present-day varieties of English (Mugglestone 1995; Beal 2004). Charles Dickens is well known for portraying the lowly characters with features such as *h*-dropping and insertion and the use of /ɪn/ for *–ing* (Mugglestone 1995: 155). In the example below from Dickens' David Copperfield from 1850 it becomes clear that by that time the deletion of *h* in *humble* has become a social marker of lower class speech. Here *h*-dropping seems to have an ironic effect: "Oh thank you Master Copperfield!," said Uriah Heep "for that remark! It is so true! Umble as I am, I know it is *so* true!" (Dickens 1850: 168)

/n/ versus /ŋ/

The present participle *–ing* was traditionally pronounced /ɪn/. However, /ɪŋ/ became the more desirable pronunciation by the eighteenth century. Although it was not yet highly stigmatized by prescriptivists, it became an important social marker and by the second half of the eighteenth century it was definitely considered more elegant and sophisticated to pronounce it as /ɪŋ/ (Mugglestone 1995: 150). Interestingly, the /ɪŋ/ was not to be produced when the spelling of a word had two *–ing* syllables in succession. For example, in *singing* and *ringing*, Walker (1791) and Batchelor (1809) recommended speakers not to produce the sound in succession but to pronounce the last *–ing* syllable as /ɪn/; this is because they considered the

repetition of the sound as inappropriate in this context and that it “would have a very bad effect on the ear” (Walker 1790: 410 qtd. in Mugglestone 1995: 151). In the later nineteenth century the use of /ɪn/ became a stereotypical marker of the lower classes and the form was frequently used to portray lower class characters in literature.

4.1 The Corpus: Letters of Artisans and the Labouring Poor (LALP)

One of the few ways to get insight into the language use of the lower classes in Late Modern English society is either through portraits of lowly characters in plays and literature or witness depositions and accounts of trial proceedings, e.g. see the Old Bailey proceedings. One of the greatest drawbacks of this kind of material is that it is a reflection of lower class speech as perceived by people who were generally from the higher educated classes, i.e. in the case of playwrights and authors, or it is the language of a courtroom scribe who put the lower class speech on paper. While these are valuable sources, it is also questionable how representative these accounts are. Diaries and personal letters written by lower class people would probably be more reliable and at least complement the one-sided picture. The difficulty is, however, that the literacy levels are proportionally low among the lower strata of society and the lower classes are therefore automatically underrepresented in most historical linguistic studies. Considering that the lower classes formed the majority of the Late Modern English society, i.e. c. 60-70% of the population (Cressy 1980), this might have serious implications for the description of language usage at the different social levels of those days, especially when Labov's theory that the spread of a change often starts from below the level of social awareness is taken into account. The opportunity to study language as actually used by the lower classes increases as literacy increases among those groups. As stated in the previous chapter, the literacy rate was on the increase from the 1700s onwards, although mass literacy was only brought about by the first Education Act in 1870, which made elementary schooling compulsory (Stone 1969; Cressy 1980). The LALP corpus consists of a collection of letters from the period *c*. 1750-1835 and offers a unique opportunity to gain more insight into the language as used by the lower layers of society. The corpus consists of a collection of approximately 2050 pauper letters, which have been compiled over a period of 18 years by the independent researcher Tony Fairman (Maidstone, UK). He collected the letters from

archives and record offices in England. The corpus consists of letters written by artisans and the labouring poor. The reason these letters exist has to do with the Law of Poor Relief, which granted paupers the right to apply for out-relief from parish funds during the period 1795-1834. Out-relief was a form of financial aid for people in financial dire straits who resided outside their home parish. In order to receive out-relief, the people had to write a letter of application to their home parish. Since these letters were written in the period preceding the introduction of compulsory elementary schooling (1870), the degree of training in writing skills was often limited. This is reflected in most of the letters, which clearly differ from those written by educated people from the same period. The fact that most of the paupers had very limited education implies that they were often not aware of standard spelling and grammar and thus reveal more about possible innovations or variation in the language, whereas the educated people of this period mostly adhered to standard spelling and grammar, even in the more informal settings. The paupers, on the other hand, had to rely on their knowledge of "translating" speech into orthographic units. Hence their writings often strongly reflect speech, which makes it possible to observe speech variation at different social levels as well as according to different geographical areas. The social variation, as well as the speech reflection, is illustrated below in two letters taken from the LALP collection. The first letter is an application letter written by a pauper who probably received less education than the author of the second letter. The second letter is written by an overseer of a parish. The overseer was responsible for granting out-relief and administered the applications. When comparing the two letters the differences in spelling and punctuation clearly stand out. The first letter lacks punctuation and the spelling varies greatly. The spelling of the first letter also seems to reflect speech, as for instance the spelling of *Sir* in the first line. Even sociolinguistic speech phenomena such as hypercorrection are reflected in the spelling. In the third line an *h* is

inserted before *am*, whereas the *h* is dropped in *had* in line 7. The spelling of the second letter, on the other hand, closely resembles modern standard spelling and punctuation.

(1) Letter 1: Written by Elizabeth Howell, Cheltenham, 26 December, c. 1835.

1. Cheltenham Dcember the 26
2.Sur I have Taken the leberty of senden
3. To you as I hame in much Nede of
4. my Hilth Contino verey Hell and my
5. Hies is so Bad I cannote Hardley find my
6. waye I cannote Doo eney in ploy for wante
7. of my site I have ad Hall The advise I Can have
8. I have noboday to fle To Bute God and you
9. for my Helpe
10. Sur I hame youre moste Humbel servante
11. a leasabeth Howell wedo

(2) Letter 2: Written by A.J. Jellico, Wandsworth, 13 October 1826

1. Wandsworth 13 Oct 1826
2. Sir
3. Application has been made to me by George
4. Newman, who states himself to be settled in your
5. Parish. to request of you some Parochial aid for
6. himself and family, under the following circumstances
7. He has been confined to his bed for the last
8. fortnight and is now, by illness. His Wife
9. has just arisen from Confinement with her
10. second child. Her other child is 2 years old.

11. They are greatly in Debt, both for necessaries
12. and Rent, and have pawned most of their
13. articles of Dress and furniture. You will be
14. good enough, I do not doubt to take this
15. Statement into your Consideration, and afford
16. them such Relief, as your discretion shall
17. dictate
18. I remain
19. Sir
20. Your obedient Servant
21. A. J. Jellicoe
22. V. C. & Sol.

Examples of spellings that might reveal more about variation in pronunciation are for instance spellings such as *payn* for *paying.* This suggests that the *–ing* form was pronounced with /n/ and not with velar/ŋ/. The spelling of *ands* for *hands* is a clear example of *h*-dropping.

4.2 Limitations of the Corpus Material with Regard to Sociolinguistic Research

In order to investigate language variability and change in the corpus material, the sociolinguistic background of each informant is needed. Because the corpus mainly contains letters of the lower classes, it is more difficult to find background information about the letter writers, i.e. in particular in comparison to educated letter writers (see for instance Nevalainen and Raumolin-Brunberg's 2003 corpus study, Network of Eighteenth-century English texts (NEET) by Fitzmaurice 2007, the Bluestockings by Sairio 2009, Robert Lowth by Tieken-Boon van Ostade 2011). Because of the relatively high status of the latter, their life histories have often been recorded. It is therefore easier to trace independent sociolinguistic variables

such as the informant's sex, age, social status, geographical information, occupation, and sometimes it is even possible to establish the informant's network (see Network of Eighteenth-century English texts (NEET) by Fitzmaurice 2007, the Bluestockings by Sairio 2009). In the case of the labouring poor, it is sometimes just a single letter that can be regarded as the only attestation of that person's existence. This implies that the sociolinguistic information needs to be extracted from what is mentioned in the letter, e.g. sex can only be derived from what is mentioned in the letters or in one of the few records kept by overseers of a parish, or any correspondence concerning an applicant. In most cases the name with which the letter is signed reveals the author's gender, or sentences such as *my husband* or *my wife.* However, in some cases letters are not signed or only signed with initials, or signed with more than one name, which makes it impossible to determine the author's gender. As for the informant's age, it is only possible to ascertain this if the author mentions this in the letter, or, in rare cases, from what is mentioned in related correspondence between overseers of parishes. As regards the geographical background of informants, this is often indicated in the letters because the applicants had to apply for relief to their home parish and in order to receive their relieve they also had to indicate their place of residence. However, the home parish was not always the place of birth of the applicant. The home parish was where an applicant had so-called settlement rights. The first way in which settlement rights were established was by birth but it could also be established by, for instance, marriage, apprenticeship or by renting a property in that parish for a certain length of time (Auer and Fairman forthc.). It is therefore not always safe to assume that the home parish to which the application was sent was also the applicant's place of birth.

Another issue of concern is what Hernandez-Campoy (2012) refers to as the authenticity of the language. Although most applicants lacked significant education and were therefore more likely to use non-standard language, there is a chance that they perceived the

context of the application as a formal setting and hence they might also have used a different register from what they might have used in their daily lives and social surroundings. This is also suggested by the hyper-correct forms, such as *h*-insertion, that are frequently found in the letters. On the other hand, according to Labov's theory (1972: 122), the frequency of possible prestigious forms are still relatively low compared to what is used in the middle and higher classes. Furthermore, one issue that is difficult to overcome is the question of whether a letter is an autograph or not. In this corpus there are examples of letters signed by the same name but written in different hands. Because many of the paupers did not have any schooling, or had very limited schooling, it is likely that they asked another family member or a neighbor who was (more) literate to write the letter of application for them. In these cases it is impossible to be sure if all social variables can be correlated to the language that is used. Some of the letters need, therefore, to be treated with caution and it also needs to be considered that there are letters that did not raise any questions with regard to authorship while they in reality may not be autographs.

Also, caution is called for when interpreting orthographs and the relation to possible pronunciation. The spelling is sometimes highly idiosyncratic and the same word can be spelled in various ways within one single letter by one author. Interpretations with regard to pronunciation might thus be highly speculative in those cases. Some applicants consistently write words such as *hoping* as *hoaping* and *wrote* as *roat.* The question if this can be considered an early attestation of the diphthongization of the GOAT vowel, or whether the authors derived their spelling from, for example, the word *cloak* or *road.*

Despite the challenges, the material of the corpus contains valuable information. After all, it could serve as a direct source containing language as used by the lower class that might complement and corroborate the findings obtained from indirect sources.

5.1 Historical Sociolinguistics and the Problem of Bad Data: a Case Study

Applying sociolinguistic theories on historical data, in the quantitative sense, leads to many challenges. As synchronic sociolinguistic studies have convincingly shown, it is important to consider (social) external factors next to internal factors to explain the distribution of language variation and the propagation of linguistic change. As shown by the studies of Romaine (1982), Milroy (1992) Nevalainen and Raumolin-Brunberg (2003), the success of these approaches on present-day data has inspired historical linguists to test these theories on historical data The studies carried out by Romaine, Nevalainen and Raumolin-Brunberg show that it is possible to approach historical data from a sociolinguistic perspective. These studies mainly focus on syntactic features, however. The LALP corpus as described in the previous chapter is unique in that it provides an insight into a group of language users that has long been left out of consideration in historical language descriptions. Until recently the focus was on what Milroy (1992: 50) refers to as the "standard ideology." Milroy (1992: 50) argues that before the sixteenth century the description of the English language tended to provide a picture of how the language diverged, whereas from the onset of language standardization process in the sixteenth century onwards the focus shifted to language convergence. In other words, the focus seems to have shifted to a description of diversity in language to a description that is mainly concerned with the changes that contributed to the genesis of a "homogenous" standard language. The traditional history of Modern English sometimes seems to "suggest that [it] was consciously directing itself towards modern RP" (Milroy 1991: 126). Elspaß (2007: 3) refers to this as "language history from above." European languages in general were often studied from the perspective of a language ideology and it was primarily the language use of role models of the standard language that was studied (Elspaß 2007: 3). Thus, language description was mainly concerned with (literary) texts that were good representations of that standard language. The language description primarily concerned

language that was produced by a small educated minority of society. Non-standard forms and more informal (spoken) language was largely ignored (Elspaß 2007: 3).This over-emphasis on standard language histories makes it appear as if "modern standard languages sometimes seem not to have changed at all for 200 years" (Elspaß 2007: 4). Both Milroy and Elspaß argue that such a unidirectional approach is not sufficient in terms of a language historiography or to fully explain language change. As Milroy (1992:50) aptly puts it, "[…] there is no such thing as a uniform language or dialect (and standardization implies uniformity), and as sound-changes do not proceed in straight lines, this cannot possibly be an adequate conceptualization of English phonological history." In order to obtain a more complete picture of a language and its history, it is vital to also include informal everyday and non-standard language in a description. Within the written genre letters are believed to be the most informal type of text and closest to natural speech (Nevalainen and Raumolin-Brunberg 2003). The awareness of these issues is growing and this is reflected in the increase of (historical) letter corpora that are being created (Auer and Fairman forthc.). However, although most of these corpora represent more informal written material that is closer to spoken language and thus may reveal more about linguistic innovations, it is still often language produced by the higher layers of society (Auer and Fairman forthc.). To obtain a more complete multidimensional view of actual language, it is essential that all layers of society and their language use are taken into consideration. Elspaß (2007) therefore advocates to also include language history "from below." This implies that language history should focus on language as produced by the majority of the population, namely the lower and middle classes, but it also implies that language should be viewed and studied as being a close reflection of spoken language. As pointed out in the previous chapters, this view "from below" is often not possible because the majority of the population was illiterate and the further one goes back in time, the lower the rate of literacy is. The LALP corpus makes it

possible to look at the Modern English language "from below." First, it is material written by the lower layers of society and second, the language that can be found in the letters is often strongly speech-reflected and thus close to spoken language. Hence it also offers an opportunity to see what challenges are involved in carrying out phonological research as opposed to syntactic investigations on historical data.

5.2 A Case Study: *H*-deletion

One of the phonological phenomena that can be observed in the language of the letters contained in the LALP corpus is *h*-deletion and *h*-insertion. The feature *h*-deletion will be examined more closely in this case study. The purpose of this study is to see if and how a sociolinguistic approach could be applied to historical phonological data. It will be an evaluative and exploratory case study that is by no means exhaustive but that might open up new perspectives on socio-historical research that focuses on phonological features. The aim of this study is to include sociolinguistic factors that are associated with variation and that are well-established in synchronic sociolinguistic research. The choice of the phenomenon of *h*-deletion is twofold: (1) It is seems to occur relatively frequently and it is also to a lesser degree subject to speculative interpretations than for instance vowels. As shown in the example of *wroat* versus *wrote* in chapter 4, it is sometimes hard to determine whether a certain spelling has been created through analogy, or whether it is truly an indication of diphthongization. In the example given above it might well be an analogy of the spelling of *road*. In the case of *h*, the graph is either absent or present. Caution is called for, however, as it is likely that the language user may have known the spelling of the word and used an *h* according to the standard spelling, while the realization may still have been an *h*-less variant in the speaker's mind. A clear example is *humble*. This word occurs frequently in the corpus

data and is often part of a formulaic salutation, e.g. *your humble servant.* In this particular case, the word is often written with an *h*, even in letters where most *h* graphs are deleted. This word is of French Norman origin and at that time still had the historical pronunciation with a silent *h* [2](Mugglestone 1995; Beal 2004). The spelling of this word was probably so well established that most authors obeyed standard spelling conventions in this case. In other cases the spelling may have been less well known and it is in those words in which the author's reflection of pronunciation may be present. On the other hand, there are letters that are *h*-full throughout the text except for the word *humble*. This might be an indication that the author writes according to pronunciation and probably is a speaker who pronounces the *h* where it is historically pronounced.

(2) In addition, the LALP corpus provides material as produced by the majority of the population that has often been referred to and commented upon by eighteenth- and nineteenth-century grammarians (Walker (1791), Batchelor (1809)), and that has been portrayed by literary authors and playwrights but that hitherto never had a voice of its own. In the case of *h*-deletion this feature was often associated with 'lowly' speech and as the examples from Dickens show, this was how lower class people were portrayed as being *h*-droppers. In a sense this case study is also a chance to change the view on *h*-deletion from 'above' to a view from 'below.'

5.3 Background

As for the deletion of the *h*, Wyld (1920: 295) makes a distinction between the loss of [*h*] in stressed and unstressed syllables. The loss of *h* in unstressed syllables seems to be a

[2] Although this story is slightly more complex because Elphinston (1781) lists it as a word that should be realized with an [h], while Walker (1791) and others still list it as an h-less word. See Mugglestone (1995) for a detailed discussion.

relatively early phenomenon; auxiliaries and pronouns are often in an unstressed position within a sentence and the loss of *h* in these types of words has been attested in spellings of thirteenth-century manuscripts (Wyld 1920: 295). The loss of *h* in more prominently stressed words in the sentence poses a more problematic picture. A complicating factor is the influence of Norman French. The Norman French words were often spelled with an orthographic *h*, but had, in many cases, lost the *h* pronunciation before they entered the English language. The words of English origins with an orthographic *h*, on the other hand, were probably still realized with *h*, except for the unstressed cases as explained above. Another complicating factor is that due to the stigmatization of *h*-dropping during the eighteenth and nineteenth centuries some hs started to be pronounced in words of Norman French origin that were originally never pronounced when they entered the English language (Beal 2004: 159). Words as *humble* and *hospital* were initially realized without *h* and by some eighteenth-century grammarians these are indeed listed and commented upon as words that had silent hs. The loss of *h* seemed to be associated with low status to such a degree that, as a consequence of hypercorrection, even these words became *h*-pronounced (Mugglestone 1995; Beal 2004).

It is still unclear when and how the *h* was lost in prominent words, e.g. nouns, of English origin, such as *ope* for *hope* and *ealth* for *health*. Although examples of this kind are attested in manuscripts as early as the thirteenth century, Wyld (1920) discards them as convincing evidence for *h*-deletion as a variant because they occur very infrequently (295). Traditionally it has been assumed that the earliest solid evidence on *h*-deletion in stressed native words can only be found from the end of the eighteenth century onwards. Interestingly, the phenomenon of *h*-loss does not seem to be as widespread in most Colonial Englishes, Scotland, Ireland and North America. This is often given as evidence that *h*-deletion could not have been a widespread feature before the end of the eighteenth century (Milroy 1992: 137). However, Milroy (1992) gives reason to raise questions regarding this assumption. First,

today the phenomenon is so well established across all of England and considering that this feature is believed to have its earliest manifestation in the late eighteenth century, it is hard to explain how this could become so widespread in a relatively short time. Today, the deletion of initial *h* in stressed words is a widespread phenomenon that can be observed across England and Wales (Milroy 1992: 137). The map below shows that only a few rural areas retain the *h*. It needs to be pointed out, however, that *h*-loss has been attested in the urban vernaculars while they are situated in the *h*-retaining area on the map (Milroy1992: 137). Trudgill (1974) has found that the variable *h* is deleted by the lower classes in Norwich. This is striking because it is found within an area that is known to be largely *h*-retaining, although it needs to be pointed out that the retaining of *h* is primarily reported in older speakers and the degree of *h*-loss may be even higher than suggested by the map (Trudgill 1974: 84).

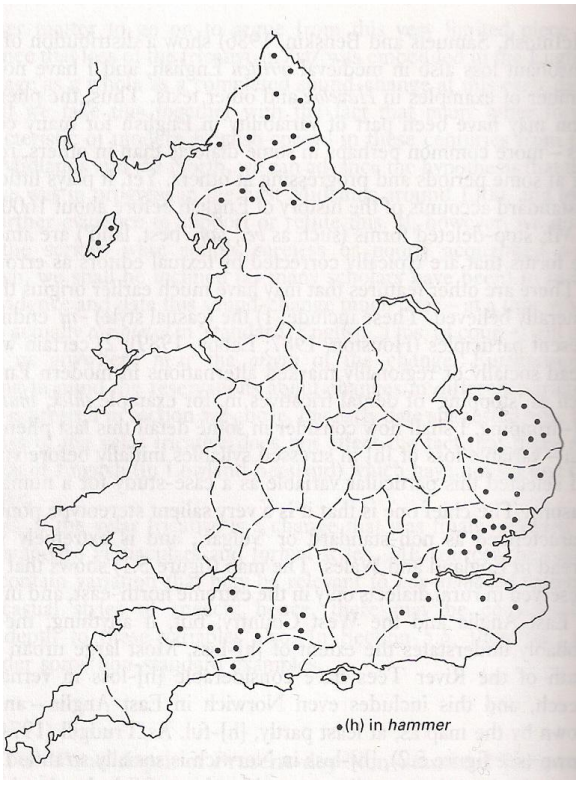
Second, there is a possibility that *h*-full and *h*-less varieties of today are "derived from varieties in which *h*-loss was *variable*-not categorical absent or categorically present" (Milroy 1992: 138-9). In this view the earlier attestations in texts and manuscripts might be relevant, since they do occur, albeit not categorically. Milroy (1992) observes that there is a high degree of instability with regard to *h*-use before vowels in stressed syllables from (early) Middle English texts onwards. The letter *h* is omitted in places where it should historically be present and inserted in places where it was historically not present. This instability might be an indication that the *h* was not present in the speech of the regions where this variability is found (Milroy 1992: 141). Since the feature was always present in orthographic tradition, it is very likely that "the scribes would omit it on some occasions and use it hypercorrectly on others" (Milroy 1992: 141), depending on the degree in which the scribe felt it was necessary to obey orthographic conventions. Variability of this kind is very common and found in texts of c.1200 in southern, East Midland/East Anglian texts, in texts of the late fourteenth century (*Norfolk Gilds)*, in the *Paston letters* of fifteenth century and the *Diary of Henry Machyn* of

the mid-sixteenth century (Wyld 1920; Milroy 1992). Milroy admits that the earliest attestations could be attributed to the direct influence of Anglo-Norman scribes. However, this cannot be said about the later texts. Moreover, the fact that the feature was highly stigmatized and frequently commented upon in the eighteenth century also suggests that it was already widespread at that time. The instability in the use of *h* in texts over a long period of time suggests that the *h* has at least been a variable that in some particular varieties resulted in a categorically *h*-less variant, whereas "general speech communities" (Milroy 1992: 143) varied in the use of *h* and probably have been using it as social and stylistic marker for centuries. Milroy (1992: 144) tentatively hypothesizes that *h*-deletion in native words might well be a contact-induced change under the influence of Norman French, since the instability occurs from the Norman French time onwards and English seems to be the "only Germanic language that is widely subject to *h*-loss."

Literature on possible internal linguistic factors that are involved in *h*-deletion in English stressed syllables is scant. In a present-day account of Bahamian English it is also clear that the loss of *h* is generally not categorical and it is therefore mainly a matter of frequency that is socially and stylistically dependent (Childs and Wolfram 2008: 248). There is, however, an indication that *h*-deletion occurs most frequently at the beginning of an utterance or when the *h* is preceded by a consonant (Childs and Wolfram 2008: 248). In other words, CV sequences are preferred above CCV sequences. The same preference may influence [*h*]-insertion before vowels which has a tendency to occur in intervocalic positions or at the beginning of an utterance (Childs and Wolfram 2008: 249). *h*-insertion seems to be a complicated matter in Bahamian English because it is not clear whether it is a hypercorrect phenomenon or whether it is simply lexically determined. It co-occurs with *h*-deletion but it seems to take place in casual speech as well, which seems to suggest that it is not used as a hypercorrect form. Then again, there is no clear evidence of a lexical pattern either (Childs

and Wolfram 2008: 249). According to Wyld (1920: 310), the insertion of *h* occurred very frequently in words that "have extra-strong stress in the sentence."

h-deleting and *h*-retaining Regions



Map1 taken from Milroy 1992 (Adapted from Orton et al. 1963-9)

5.4 The Selection of the Material

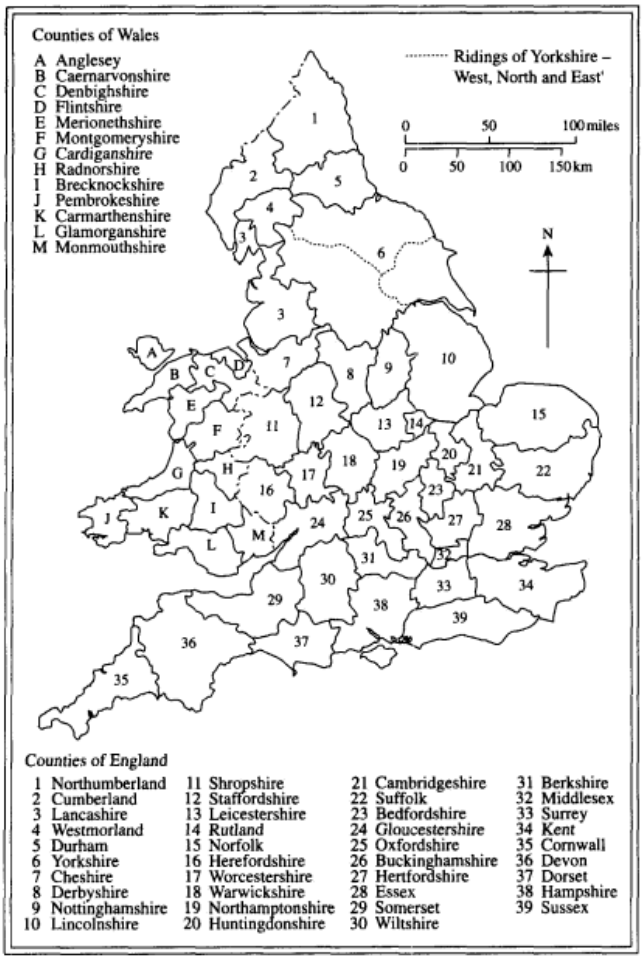
The material used in this case study is taken from the LALP corpus. As the corpus is not yet suitable for electronic searches, it had to be searched manually and with the help of a simple word-search programme. Since this is very time-consuming, it was only possible to look at a limited amount of data. Because it is the aim to also consider the geographical background of the informants, it was decided to focus on letters coming from counties that lie

in the *h*-retaining areas of today and the counties from *h*-deletion area of today. These choices are based on the map as provided by Milroy (1992). The counties in the present day *h*-deletion areas are Cornwall, Devon, Lancashire, Yorkshire, Cheshire, Derbyshire, Nottinghamshire, Lincolnshire, Shropshire, Staffordshire, Leicestershire, Herefordshire, Worcestershire, Warwickshire, Northamptonshire, Huntingdonshire, Bedfordshire, Gloucestershire, Buckinghamshire

The counties that are today mainly *h*-retaining regions are Northumberland, Durham, Cumberland, Norfolk, Suffolk, Wiltshire and Surrey.

The respective counties are shown in map 2below:

"The counties of England and Wales in the nineteenth century"
in Williams (2004: vi)



Map 2

The selection of these counties makes it possible to see how the Modern English data might relate to more recent data. In total the selected material consists of approximately 300 letters, produced by around 100 females and 150 males. Although the age of informants was sometimes provided by the header information of the letters, this factor was not used as a selection criterion, or as a possible independent variable because in this sample the amount of informants whose age was known was negligible. The date of the letters ranges from 1741 to 1837.

Some of the methodological issues as discussed in chapter 2 came to the fore in the selection of the material. The first two issues involve the problem of representativeness and empirical validity as illustrated by Hernandez-Campoy (2012). In the case of historical sociolinguistics, researchers have to base themselves on the material that is available and it is therefore often not possible to pick and sample equal representative groups. Moreover, it is not always possible to trace all background information of the informants. The material in this corpus is no exception and consequently quite some material had to be discarded because it simply did not provide enough social information with regard to gender, or origin, or date, or none of these factors was provided. The third issue is what Hernandez-Campoy (2012) refers to as the problem of invariation; since it concerns written language that was written in a rather formal setting, the language use is inevitably more formal and normative and thus contains less variation than spoken language. In other words, it does not necessarily mean that all informants supply the variation one is looking for. This problem also arises in this corpus, especially because it concerns the study of phonological features. The word *humble* exemplifies this problem. The degree of speech reflection seems to depend on the individual and the degree of schooling. A more schooled person is more likely to adhere to standard spelling than a person who does not know so well how to spell according to the standard. Moreover, it is debatable and difficult to measure when the language used in a letter should be

considered less or more representative of speech. In a purely descriptive approach this might be less problematic. However, in terms of frequency the more normatively written letters will obscure the significance of the innovations that can be found in the more speech-reflected letters. The question is what criteria should be adopted in the selection of speech-reflected written language. For this case study, letters that had no or just a single spelling that deviated from Standard English were discarded in the statistic calculations. The issue of authorship is also a problem here. A helpful feature of the corpus is that the header information that is attached to each letter gives an indication of the authenticity of a letter and whether there is reason for doubt in this respect. This could be the case when several letters signed with the same name are written in different hands, or when the same hand was found under different names. In those cases the letters had to be discarded because it is impossible to be certain about the social information of the author.

5.5 Method

In order to establish the distribution of *h*-deletion in initial position before vowels, the *h* was approached as a variable having two variants: ø and *h*. This implies that the material had to be searched on every occurrence of initial *h* and every word beginning with a vowel. This was done by using the search programme in word. The search could potentially also be carried out with a concordance programme but the problem with that was that it was not easy to relate each instance to the social variables as provided by the headers of the letters. In the future, when the corpus is fully digitized and made suitable for concordancing it will be easier to include these factors in the search commands. For the purpose of this study the search programme with the possibility to use wildcards in Microsoft Word was sufficient. After localizing these words they had to be examined in their context. A complicating factor was the high frequency of *h*-insertion words. Since this study focuses on *h*-deletion only, the

instances of *h* before words that are historically not spelled with an *h* were not included in the analyses. Also the word *honour,* or any derivation of this word was ignored because this was in essence a word that was very likely *h*-less in any speech variant but that may be spelled with an *h*. Other words of this kind may be *heir* or *honest.* As for all the words beginning with a vowel the context was used as an indication whether the word was a ø variant of variable *h*. For instance, the word *is* can be either the verb, or the ø-variant of *his*. This was fairly easy to establish in most cases. If there was any doubt about a word, it was excluded. *H*-deletion is occasionally also found in non-initial position, for instance one informant produced *beaviour* for *behaviour* and *beind* for *behind*. For this case study, however, it was decided to focus on *h*-dropping in initial syllables only because including the other cases would complicate the search process and make it extremely time-consuming and also because this study is intended to explore the applicability of a sociolinguistic quantitative approach, and not to give an exhaustive account on *h*-dropping. Once all the instances were marked as *h*-variants and ø-variants of variable *h*, every instance was transferred into a spreadsheet. Each token was coded for dependent variable variant ø/*h*, and for the possible independent variables date, geographical origin (county) of the informant, gender of the informant and class of the *h*-word. Date was divided into two periods: 1741-1825 and 1826-1837. The geographical origin was divided into present-day *h*-retaining regions and present-day *h*-loss regions. Gender was female, male, or unknown. The tokens have not been marked for the possible linguistic factor as tentatively proposed by Childs and Wolfram (2008) since it appears to be a very weak factor. Just as in the study of Childs and Wolfram (2008), none of the informants deletes the *h* categorically in the proposed environments. The distinction of word class has been chosen because there is a strong indication that the *h* is more likely to be deleted in auxiliaries and pronouns; this is because they are often in an unstressed position within the sentence (Wyld 1920).Words such as nouns and lexical verbs are often more prominently stressed within a

sentence and the loss of *h* is less likely to occur (Wyld 1920). Class has thus been divided into auxiliary, pronoun, nouns adverbials/adjectives, content verbs, -*ing* forms. To establish possible significant effects of the independent variables on the dependent variable ø/*h*, the coded tokens were entered into SPSS 20 and analysed by means of a cross tabulation and Chi-square. This was done with 2x2 crosstabs. In the case of the factor word class it was necessary to make paired comparisons in order to establish the significance of each individual class. Since paired testing has as a consequence that the p-value levels will be more significant than they actually are (Type I error). The p values of these comparisons had to be adjusted and this has been done by using the Bonferroni adjustment (Essex University). The class noun was eventually excluded from the analyses because there was only one instance of *h*-deletion in that class. For an additional chi square analyses the classes were collapsed into a group consisting of auxiliaries and pronouns and a group constituting the rest of the classes based on the distinction made by Wyld (1920) and Milroy (1992).

5.6 Results and Discussion

The total of tokens of variable *h* amounts to 1735, including nouns. Excluding nouns this is 1511. The variant ø occurs 168 times and variant *h* occurs 1343 times; see table 1 below for percentages.

Variable *H*	Frequency	Percent	Valid Percent	Cumulative Percent
ø	168	11,1	11,1	11,1
h	1343	88,9	88,9	100,0
Total	1511	100,0	100,0	

Table 1. Variable *H* Percentages

It is striking to observe that the frequency of *h*-deletion in nouns is next to non-existent, especially considering the fact that at present the deletion of *h* is well established in nouns as well (Milroy 1992). The one example of deletion in a noun seems a genuine case.

(1) Nothing Left but to Be sent *ome* With a pas

However, the author uses the word *home* one more time and this time spells it with an *h*.

(2) If yo Wish Any In quiry you Can send to Mr George

Tiley 38 saint James street portsea for *home*

It might be an accidental case but on the other hand the author deletes the *h* in many other cases as well. It must also be pointed out that there were a few more cases of *h*-deletion in nouns in other letters by different authors but they had to be discarded because there was no or very unreliable information about the geographical origins or authorship as questionable. The three other cases are *health, house, hands* and are found in three different letters by different authors The frequency of deletion in nouns is thus very low and one can speculate as to why this is the case. The only thing that can be said is that two of the cases date from 1821 and the others from 1829 and 1835 and that the spread of deletion of *h* in nouns was only just beginning. Again this might also have to do with salience of a word within the sentence. A noun is often highly salient in a sentence and may therefore be less subject to phonological loss. However, this contrasts with the findings of Wyld (1920) and Milroy (1992) who both found instances from earlier periods such as *alpenny* (Norfolk Gilds) for *halfpenny,*and *elmet* (Machyn) for *helmet.* Regarding the factor time, there seems to be no effect. As can be seen in the table 2 below the total of tokens is lower than given above. This is because some letters had no date; hence they are excluded from this analysis. The distribution of the variable *h* is fairly consistent over the time span of 96 years. It can be pointed out that the second period shows a slightly higher percentage of *h*-deletion but this is, with a P value of 0.661, not

significant. This also applies when the periods are compared with auxiliaries and pronouns only, or when compared with other classes, excluding auxiliaries and pronouns.

Periods		**VARIABLE *H***		**Total**
		deletion	realization	
1741-1825	Count	69	620	689
	%	10,0%	90,0%	100,0%
1826-1837	Count	69	573	642
	%	10,7%	89,3%	100,0%
Total	Count	138	1193	1331
	%	10,4%	89,6%	100,0%

Table 2 Distribution of Variable *H* across Periods

As for the geographical factor, there is a very strong significant effect with a P-value of 0.000. The *h* is deleted significantly more frequently in the *h*-deleting regions of today. As can be seen in table 3, in the *h*-deleting regions 14 percent of the hs is deleted, whereas in the *h*-retaining regions this is only 3,2 percent. This effect is particularly strong with auxiliaries and pronouns with a difference of almost 14 percent (p 0.000); see table 4 below. A difference of almost 7 percent is found when auxiliaries are excluded from the analysis and this is also a statistically significant difference (p.0.003); see table 5 below.

Regions		**VARIABLE *H***		**Total**
		deletion	realization	
ø-regions	Count	155	952	1107
	%	14,0%	86,0%	100,0%
h-regions	Count	13	391	404
	%	3,2%	96,8%	100,0%
Total	Count	168	1343	1511
	%	11,1%	88,9%	100,0%

Table 3 Regional Distribution, all Word Classes

Regions		VARIABLE *H*		Total
		deletion	realization	
ø-regions	Count	103	506	609
	%	16,9%	83,1%	100,0%
h-regions	Count	6	197	203
	%	3,0%	97,0%	100,0%
Total	Count	109	703	812
	%	13,4%	86,6%	100,0%

Table 4 Regional Distribution, Auxiliaries and Pronouns only

Regions		VARIABLE *H*		Total
		deletion	realization	
ø-regions	Count	52	446	498
	%	10,4%	89,6%	100,0%
h-regions	Count	7	194	201
	%	3,5%	96,5%	100,0%
Total	Count	59	640	699
	%	8,4%	91,6%	100,0%

Table 5 Regional Distribution, Auxiliaries and Pronouns Excluded

The factor gender in general seems to play no significant role. When taking a look at the data in general from both regions in table 6, the present day *h*-less regions and the present day *h*-retaining regions, it becomes clear that the difference between men and women is small, with women only just in the lead; the men delete the *h* 10 percent of the time and the women do this 12,2 percent of the time. However, this difference is not statistically significant.

Gender			VARIABLE *H*		Total
			deletion	realization	
	female	Count	72	517	589
		%	12,2%	87,8%	100,0%
	male	Count	90	807	897
		%	10,0%	90,0%	100,0%
Total		Count	162	1324	1486
		%	10,9%	89,1%	100,0%

Table 6 Gender Differentiation across all Regions, all Word Classes

When taking a look at the data from the *h*-deleting regions only, in table 7, something else comes to light. When the auxiliaries are excluded from the analyses, the deletion of *h* is clearly led by women and this difference, with a p-value of 0.008, is significant. When looking at auxiliaries and pronouns only, in table 8, the men seem to delete the *h* more but this is not statistically significant.

Gender			VARIABLE *H*		Total
			deletion	realization	
Gender	female	Count	27	148	175
		%	15,4%	84,6%	100,0%
	male	Count	25	297	322
		%	7,8%	92,2%	100,0%
Total		Count	52	445	497
		%	10,5%	89,5%	100,0%

Table 7 Gender Differentiation, across Ø-regions, Auxiliaries and Pronouns Excluded

Gender			VARIABLE *H*		Total
			deletion	realization	
	female	Count	34	201	235
		%	14,5%	85,5%	100,0%
	male	Count	63	287	350
		%	18,0%	82,0%	100,0%
Total		Count	97	488	585
		%	16,6%	83,4%	100,0%

Table 8 Gender Differentiation, Ø-regions, Auxiliaries and Pronouns only

Surprisingly, when looking at the *h*-retaining regions in table 9 and 10, the *h* in auxiliaries and pronouns is significantly more frequently deleted by the women; see table 9. The men do not seem to have deleted the *h* in auxiliaries and pronouns. As can be seen in table 10, the women also delete the *h* more frequently in the other word classes but this is not significant. Thus, *h*-deletion in general seems to be led by women in the *h*-retaining area, although this is only significantly so with auxiliaries and pronouns.

Gender			**VARIABLE *H***		**Total**
			deletion	realization	
	female	Count	6	76	82
		%	7,3%	92,7%	100,0%
	male	Count	0	121	121
		%	0,0%	100,0%	100,0%
Total		Count	6	197	203
		%	3,0%	97,0%	100,0%

Table 9 Gender Differentiation across *h*-regions, Auxiliaries and Pronouns only

Gender			**VARIABLE *H***		**Total**
			deletion	realization	
	female	Count	5	92	97
		%	5,2%	94,8%	100%
	male	Count	2	102	104
		%	1,9%	98,1%	100%
Total		Count	7	194	201
		%	3,5%	96,5%	100%

Table 10 Gender Differentiation across *h*-regions, Auxiliaries and Pronouns Excluded

To summarize, what seems to be going on with regard to gender and region; in the present day *h*-retaining regions *h*-deletion in general is lower than in the present-day *h*-less regions but in the *h*-retaining area the women tend to delete the *h* more than men and significantly so in auxiliaries and pronouns. In the *h*-less region both men and women delete the *h* more in general but it is only in word classes other than auxiliaries and pronouns that

women delete the *h* significantly more than men. It could be argued that women are indeed innovators here but not in terms of the sex/prestige pattern, i.e. if we assume that the deletion of *h* was indeed perceived as a non-prestige variant. Nonetheless, as shown by Nevalainen and Raumolin-Brunberg (2003) in their study of gender differentiation in Late Middle and Early Modern English, when "[...] a conscious process was under way, the change was consistently led by men" (131), whereas changes from below the level of social awareness were led by women. As stated in chapter 2 the explanation for this difference might be the different social roles of women in Middle and Early modern English society. A closer examination of the role of women in Late Modern English society might be revealing in this respect. It should also be considered that the higher frequencies of prestige forms among women have mainly been observed in present-day lower middle classes (Labov 1972: 122) and since the material studied here is mainly from lower classes, the patterns might also be different.

As for the factor of time the deletion of *h* seems to be rather stable. It seemed to have stayed stable across the 96 years time-span that has been studied here but it also seems to have been rather stable when considering today's *h*-less and *h*-retaining areas, at least from the point of view of regional distribution. It might be argued that the deletion of *h* has spread to more word classes, including nouns and has become more frequent in general. Moreover, as the study by Trudgill (1974) has shown, the preservation of *h* seems to be a rural feature, whereas it is deleted in urban vernaculars, even within the *h*-retaining regions. Thus, *h*-deletion may be even more prevalent than is displayed by the present day map. In addition to that the present day map is based on data gathered from phonological recordings in controlled settings and by interviews, whereas the data of this study may be subject to the factors as discussed in chapter 2.2, i.e. that the issues of invariation and authenticity are very likely an interfering factor. The use of possible hypercorrect *h*-insertions might be an indication that

people tried to avoid the deletion of *h*. The comments of contemporaries as Walker and Sheridan show that the feature was becoming more and more stigmatized. It is not clear, however, how well the stigmatization was established and how strongly this affected the language users of the lower classes in the context of the application for out-relief. It also remains to be seen whether the insertion of *h* always can be described as hypercorrection, or whether it was a stylistic devise, as proposed by Wyld (1920) and as observed in Childs and Wolfram's (2008) study. A more detailed study that includes both *h*-deletion and *h*-insertion might shed some more light on these questions. In any case, the large amount of *h*-insertion might be an indication of what Milroy (1992) referred to as the instability that is associated with varieties of English that have either no variable *h* at all, or indeed variability with regard to variable *h*. The findings in this study support Milroy's (1992) hypotheses that present-day *h*-full varieties might be derived from this variability.

6. Conclusion

The main objective of this study was to see if and how a sociolinguistic quantitative approach could be applied to historical data and phonological data in particular. The main challenges with regard to the data were as expected with regard to social background information of the informants, authorship, and the authenticity of the language. An additional challenge that may be particularly challenging for phonological research in historical data might be added to the ones that were already given in chapter 2. This is the problem with regard to degree in which the informants either adhered to standard spelling or deviated from it. This probably has to do with the different levels of schooling and thus the knowledge of standard spelling or the lack thereof. It is difficult to decide when a letter is more or less representative of speech.

Nonetheless, as has been pointed out by Raumolin-Brunberg(1996: 19), one of the great advantages of historical (socio)linguistics is that very often we have some idea of when a change started and, most importantly, we know what the outcome was. The present-day map of the distribution of variable *h* corroborates what has been found in this case study. After all, the geographical factor turned out to be highly significant and the frequency of *h*-deletion could be directly related to the relevant areas of today. In other words, despite the issues of the bad data, the results regarding the geographical distribution of *h* are supported by present-day data and theories based thereupon. This is a promising prospect for the study of phonological data in the field of historical sociolinguistics and shows that the problems regarding limitations of the data are not necessarily insurmountable.

Acknowledgements

This thesis would not have been possible without the support of my supervisor Anita Auer. I feel privileged and I am grateful for the opportunities she has given me to explore the world of historical sociolinguistics. Thank you for being my guide.

I also would like to thank Tony Fairman, without whom it would not have been possible to study the unique data of the LALP corpus. It has been an honour to have met the father of the LALP 'foetus' and to be given the opportunity to work with this valuable source.

Last but not least, I would like to express my gratitude to Hans Van de Velde for guiding me through statistics and SPSS and for making me realize how important statistics are to the field of historical sociolinguistics.

References

Aitchison, J. (2012). Diachrony vs Synchrony: The Complementary Evolution of two (Ir)reconcilable Dimensions. In J. M. Hernandez-Campoy, & J. Camilo Conde-Silvestre (Eds.), *The Handbook of Historical Sociolinguistics* (pp. 47-60). Chichester: Wiley-Blackwell.

Auer, A., & Fairman, T. (forthc.). Letters of artisans and the labouring poor (england, *c.* 1750-1835). *Conference on New Methods in Historical Corpora,* Manchester. 1-14.

Auer, A., & Voeste, A. (2012). Grammatical Variables. In J. M. Hernandez-Campoy, & J. C. Conde-Silvestre (Eds.), *The Handbook of Historical Sociolinguistics* (pp. 326-346). Chichester: Blackwell.

Beal, c. J. (2004). *English in Modern Times*. London: Arnold.

Dickens, Charles (1850).*The Personal History of David Copperfield.* (1998 ed.) Bath: Bath Press

Childs, B.& Wolfram, W. (2008). Bahamian English: Phonology. In E. W. Schneider (Ed.), *Varieties of English the Americas and the Caribbean* (pp. 239-255). Berlin: Mouton de Gruyter.

Elspaß, S. (2007). A Twofold View 'from Below': New Perspectives on Language Histories and Language Historiographies. In S. Eslpaß, N. Langer, J. Scharloth & W. Vandenbussche (Eds.), *Germanic Language Histories 'from Below'* (pp. 3-9). Berlin: Walter de Gruyter.

Essex University.*A more Sophisticated Bonferroni Adjustment.* Retrieved June 18, 2012, from <http://privatewww.essex.ac.uk/~scholp/bonferroni.htm>

Fairman, T. (2003). Letters of English labouring Classes and the English Language, 1800-34, In Marina Dossena and Charles Jones (eds.) *Insights into Late Modern English*. Second edition. Bern: Peter Lang, pp.265-282.

Fairman, T., Auer, A., Gordon, M., & Laitinen, M. (Eds.).*Letters of Artisans and the Labouring Poor*. Maidstone/ Utrecht/ Jyväskylä

Görlach, M. (1999). Regional and Social variation. In R. Lass (Ed.), *The cambridge History of the English Language* (pp. 459-538)

Görlach, M. (1998). *English in Nineteenth Century England. an Introduction*. Cambridge: Cambridge University Press.

Hernandez-Campoy, J. M., & Schilling, N. (2012). The Application of the Quantitative Paradigm to Historical Sociolinguistics: Problems with Generalizibility Principle. In J. M. Hernandez-Campoy, & J. C. Conde-Silvestre (Eds.), *The Handbook of Historical Sociolinguistics* (pp. 104-120). London: Blackwell-Wiley.

Hitchcock, T., Shoemaker, R., Emsley, C., Howard, S. & McLaughlin, J. e. a. (2012). *The Old Bailey Proceedings Online, 1674-1913.* Retrieved June 18, 2012, from <www.oldbaileyonline.org>

IBM.*IBM SPSS statistics 20 for windows*

Labov, W. (1972). *Sociolinguistic Patterns*. Oxford: Blackwell.

Lawson, J., & Silver, *H*. (1973). *A Social History of Education in England*. London: Methuen & co Ltd.

McColl Millar, R. (2007). *Trask's Historical Linguistics* (2nd ed.). London: Hodder Education.

Milroy, J. (1992). *Linguistic Variation and Change*. Oxford: Blackwell.

Mugglestone, L. (1995). *'Talking Proper' the Rise of Accent as Social Symbol*. Oxford: Oxford University Press.

Nevalainen, T., & Raumolin-Brunberg, *H*. (Eds.). (1996). *Sociolinguistics and Language History*. Amsterdam: Atlanta.

Nevalainen, T., & Raumolin-Brunberg, *H*. (2005). Sociolinguistics and the History of English: A Survey. *IJES, 5*(1), 33-58.

Romaine, S. (1982). *Socio-historical Linguistics its Status and Methodology*. Cambridge: Cambridge University Press.

Romaine, S. (1998). Introduction. In S. Romaine (Ed.), *The Cambridge History of the English language* (pp. 1-56). New York: Cambridge University Press.

Sairio, A. (2009). *Language and Letters of the Bluestocking Network. Sociolinguistic Issues in Eighteenth-century Epistolary English*. Helsinki: Société Néophilologique.

Sheridan, T. (1762). *A Course of Lectures on Elocution: Together with two Dissertations on Language; and some other Tracts Relative to those Subjects*. London: printed by W. Strahan, for A. Millar, R. and J. Dodsley, T. Davies, C. Henderson, J. Wilkie, and E. Dilly.

Tieken-Boon van Ostade, I. (2011). *The Bishop's Grammar. Robert Lowth and the Rise of Prescriptivism*. Oxford: Oxford university Press.

Toon, T.(1983). *The Politics of early Old English Sound Change*. New York: Academic Press Inc.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Van Gelderen, E.(2006) *A History of the English Language.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Walker, J. (1791). *A Critical Pronouncing Dictionary and Expositor of the English language.*. London: G. G. J. and J. Robinson, Paternoster Row; and T. Cadell.

Wardhaugh, R. (2010). *An Introduction to Sociolinguistics* (6th ed.). Chichester: Wiley-Blackwell.

Wyld, *H*. C. (1920). *A History of Modern Colloquial English*. London: T. Fisher Unwin, LTD.

Appendix I: Crosstabulations and Chi-square Periods

Periods * VARIABLE *H* Crosstabulation

		VARH		Total
		deletion	realization	
Date	1741-1825 Count	69	620	689
	1741-1825 % within Date	10,0%	90,0%	100,0%
	1826-1837 Count	69	573	642
	1826-1837 % within Date	10,7%	89,3%	100,0%
Total	Count	138	1193	1331
	% within Date	10,4%	89,6%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,192[a]	1	,661		
Continuity Correction[b]	,121	1	,727		
Likelihood Ratio	,192	1	,661		
Fisher's Exact Test				,719	,364
Linear-by-Linear Association	,192	1	,661		
N of Valid Cases	1331				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 66,56.

b. Computed only for a 2x2 table

Periods * VARIABLE *H* Crosstabulation (aux. and Pron. only)

			VARH		Total
			deletion	realization	
Date	1741-1825	Count	48	315	363
		% within Date	13,2%	86,8%	100,0%
	1826-1837	Count	42	298	340
		% within Date	12,4%	87,6%	100,0%
Total		Count	90	613	703
		% within Date	12,8%	87,2%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,119[a]	1	,730		
Continuity Correction[b]	,054	1	,816		
Likelihood Ratio	,119	1	,730		
Fisher's Exact Test				,736	,409
Linear-by-Linear Association	,119	1	,730		
N of Valid Cases	703				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 43,53.

b. Computed only for a 2x2 table

Periods * VARIABLE *H* Crosstabulation (excluding aux. and pron.)

			VARH		Total
			deletion	realization	
Date	1741-1825	Count	21	305	326
		% within Date	6,4%	93,6%	100,0%
	1826-1837	Count	27	275	302
		% within Date	8,9%	91,1%	100,0%
Total		Count	48	580	628
		% within Date	7,6%	92,4%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,387[a]	1	,239		
Continuity Correction[b]	1,055	1	,304		
Likelihood Ratio	1,387	1	,239		
Fisher's Exact Test				,293	,152
Linear-by-Linear Association	1,384	1	,239		
N of Valid Cases	628				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 23,08.

b. Computed only for a 2x2 table

Appendix II: Crosstabulations and Chi-square Regions

Region * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Region	0	Count	155	952	1107
		% within Region	14,0%	86,0%	100,0%
	1	Count	13	391	404
		% within Region	3,2%	96,8%	100,0%
Total		Count	168	1343	1511
		% within Region	11,1%	88,9%	100,0%

0=*h*-deleting region 1=*h*-retaining region

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	34,831[a]	1	,000		
Continuity Correction[b]	33,748	1	,000		
Likelihood Ratio	43,046	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	34,808	1	,000		
N of Valid Cases	1511				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 44,92.

b. Computed only for a 2x2 table

Region * VARIABLE *H* Crosstabulation (aux. and pron. excluded)

			VARH		Total
			deletion	realization	
Region	0	Count	52	446	498
		% within Region	10,4%	89,6%	100,0%
	1	Count	7	194	201
		% within Region	3,5%	96,5%	100,0%
Total		Count	59	640	699
		% within Region	8,4%	91,6%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8,974[a]	1	,003		
Continuity Correction[b]	8,096	1	,004		
Likelihood Ratio	10,482	1	,001		
Fisher's Exact Test				,002	,001
Linear-by-Linear Association	8,961	1	,003		
N of Valid Cases	699				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 16,97.

b. Computed only for a 2x2 table

Region * VARIABLE *H* Crosstabulation (aux. and pron. only)

			VARH		Total
			deletion	realization	
Region	0	Count	103	506	609
		% within Region	16,9%	83,1%	100,0%
	1	Count	6	197	203
		% within Region	3,0%	97,0%	100,0%
Total		Count	109	703	812
		% within Region	13,4%	86,6%	100,0%

0= *h*-deleting region 1=*h*-retaining region

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	25,521[a]	1	,000		
Continuity Correction[b]	24,334	1	,000		
Likelihood Ratio	32,779	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	25,489	1	,000		
N of Valid Cases	812				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 27,25.

b. Computed only for a 2x2 table

Appendix III: Crosstabulations and Chi-square Word Class (paired comparisons)

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	auxiliary	Count	77	340	417
		% within Class	18,5%	81,5%	100,0%
	pronoun	Count	32	363	395
		% within Class	8,1%	91,9%	100,0%
Total		Count	109	703	812
		% within Class	13,4%	86,6%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	18,748[a]	1	,000		
Continuity Correction[b]	17,867	1	,000		
Likelihood Ratio	19,302	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	18,725	1	,000		
N of Valid Cases	812				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 53,02.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	auxiliary	Count	77	340	417
		% within Class	18,5%	81,5%	100,0%
	verb	Count	49	427	476
		% within Class	10,3%	89,7%	100,0%
Total		Count	126	767	893
		% within Class	14,1%	85,9%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	12,246[a]	1	,000		
Continuity Correction[b]	11,581	1	,001		
Likelihood Ratio	12,263	1	,000		
Fisher's Exact Test				,001	,000
Linear-by-Linear Association	12,232	1	,000		
N of Valid Cases	893				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 58,84.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	auxiliary	Count	77	340	417
		% within Class	18,5%	81,5%	100,0%
	ad	Count	7	184	191
		% within Class	3,7%	96,3%	100,0%
Total		Count	84	524	608
		% within Class	13,8%	86,2%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	24,099[a]	1	,000		
Continuity Correction[b]	22,872	1	,000		
Likelihood Ratio	29,358	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	24,060	1	,000		
N of Valid Cases	608				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 26,39.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	auxiliary	Count	77	340	417
		% within Class	18,5%	81,5%	100,0%
	ing	Count	3	29	32
		% within Class	9,4%	90,6%	100,0%
Total		Count	80	369	449
		% within Class	17,8%	82,2%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,677[a]	1	,195		
Continuity Correction[b]	1,114	1	,291		
Likelihood Ratio	1,937	1	,164		
Fisher's Exact Test				,238	,144
Linear-by-Linear Association	1,673	1	,196		
N of Valid Cases	449				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 5,70.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	pronoun	Count	32	363	395
		% within Class	8,1%	91,9%	100,0%
	verb	Count	49	427	476
		% within Class	10,3%	89,7%	100,0%
Total		Count	81	790	871
		% within Class	9,3%	90,7%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,231[a]	1	,267		
Continuity Correction[b]	,984	1	,321		
Likelihood Ratio	1,241	1	,265		
Fisher's Exact Test				,293	,161
Linear-by-Linear Association	1,229	1	,268		
N of Valid Cases	871				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 36,73.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	pronoun	Count	32	363	395
		% within Class	8,1%	91,9%	100,0%
	ad	Count	7	184	191
		% within Class	3,7%	96,3%	100,0%
Total		Count	39	547	586
		% within Class	6,7%	93,3%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,079[a]	1	,043		
Continuity Correction[b]	3,396	1	,065		
Likelihood Ratio	4,500	1	,034		
Fisher's Exact Test				,051	,029
Linear-by-Linear Association	4,072	1	,044		
N of Valid Cases	586				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 12,71.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	pronoun	Count	32	363	395
		% within Class	8,1%	91,9%	100,0%
	ing	Count	3	29	32
		% within Class	9,4%	90,6%	100,0%
Total		Count	35	392	427
		% within Class	8,2%	91,8%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,064[a]	1	,801		
Continuity Correction[b]	,000	1	1,000		
Likelihood Ratio	,061	1	,804		
Fisher's Exact Test				,738	,499
Linear-by-Linear Association	,064	1	,801		
N of Valid Cases	427				

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 2,62.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	verb	Count	49	427	476
		% within Class	10,3%	89,7%	100,0%
	ad	Count	7	184	191
		% within Class	3,7%	96,3%	100,0%
Total		Count	56	611	667
		% within Class	8,4%	91,6%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	7,789[a]	1	,005		
Continuity Correction[b]	6,950	1	,008		
Likelihood Ratio	9,019	1	,003		
Fisher's Exact Test				,005	,003
Linear-by-Linear Association	7,777	1	,005		
N of Valid Cases	667				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 16,04.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	verb	Count	49	427	476
		% within Class	10,3%	89,7%	100,0%
	ing	Count	3	29	32
		% within Class	9,4%	90,6%	100,0%
Total		Count	52	456	508
		% within Class	10,2%	89,8%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,028[a]	1	,868		
Continuity Correction[b]	,000	1	1,000		
Likelihood Ratio	,028	1	,867		
Fisher's Exact Test				1,000	,581
Linear-by-Linear Association	,028	1	,868		
N of Valid Cases	508				

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 3,28.

b. Computed only for a 2x2 table

Class * VARIABLE *H* Crosstabulation

		VARH		Total
		deletion	realization	
Class	Count	7	184	191
ad	% within Class	3,7%	96,3%	100,0%
Class	Count	3	29	32
ing	% within Class	9,4%	90,6%	100,0%
Total	Count	10	213	223
	% within Class	4,5%	95,5%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2,086[a]	1	,149		
Continuity Correction[b]	,966	1	,326		
Likelihood Ratio	1,695	1	,193		
Fisher's Exact Test				,159	,159
Linear-by-Linear Association	2,077	1	,150		
N of Valid Cases	223				

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 1,43.

b. Computed only for a 2x2 table

Appendix IV: Crosstabulations and Chi-square Class (divided into aux/pronoun and other)

Class * VARIABLE *H* Crosstabulation

			VARH		Total
			deletion	realization	
Class	auxiliary	Count	109	703	812
		% within Class	13,4%	86,6%	100,0%
	rest	Count	59	640	699
		% within Class	8,4%	91,6%	100,0%
Total		Count	168	1343	1511
		% within Class	11,1%	88,9%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9,438[a]	1	,002		
Continuity Correction[b]	8,941	1	,003		
Likelihood Ratio	9,607	1	,002		
Fisher's Exact Test				,002	,001
Linear-by-Linear Association	9,432	1	,002		
N of Valid Cases	1511				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 77,72.

b. Computed only for a 2x2 table

Appendix V: Crosstabulations and Chi-square Gender

Gender * VARIABLE *H* Crosstabulation (from all regions, all word classes)

			VARH		Total
			deletion	realization	
Gender	female	Count	72	517	589
		% within Gender	12,2%	87,8%	100,0%
	male	Count	90	807	897
		% within Gender	10,0%	90,0%	100,0%
Total		Count	162	1324	1486
		% within Gender	10,9%	89,1%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,757[a]	1	,185		
Continuity Correction[b]	1,538	1	,215		
Likelihood Ratio	1,738	1	,187		
Fisher's Exact Test				,202	,108
Linear-by-Linear Association	1,755	1	,185		
N of Valid Cases	1486				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 64,21.

b. Computed only for a 2x2 table

Gender * VARH Crosstabulation (*h*-deleting regions only, all word class)

			VARH		Total
			deletion	realization	
Gender	female	Count	61	349	410
		% within Gender	14,9%	85,1%	100,0%
	male	Count	88	584	672
		% within Gender	13,1%	86,9%	100,0%
Total		Count	149	933	1082
		% within Gender	13,8%	86,2%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,682[a]	1	,409		
Continuity Correction[b]	,540	1	,463		
Likelihood Ratio	,676	1	,411		
Fisher's Exact Test				,414	,231
Linear-by-Linear Association	,681	1	,409		
N of Valid Cases	1082				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 56,46.

b. Computed only for a 2x2 table

Gender * VARIABLE *H* Crosstabulation (*h*-deleting regions, aux. and pron only)

			VARH		Total
			deletion	realization	
Gender	female	Count	34	201	235
		% within Gender	14,5%	85,5%	100,0%
	male	Count	63	287	350
		% within Gender	18,0%	82,0%	100,0%
Total		Count	97	488	585
		% within Gender	16,6%	83,4%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,268[a]	1	,260		
Continuity Correction[b]	1,026	1	,311		
Likelihood Ratio	1,285	1	,257		
Fisher's Exact Test				,308	,156
Linear-by-Linear Association	1,266	1	,261		
N of Valid Cases	585				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 38,97.

b. Computed only for a 2x2 table

Gender * VARIABLE *H* Crosstabulation (*h*-deleting regions, aux and pron. excluded)

			VARH		Total
			deletion	realization	
Gender	female	Count	27	148	175
		% within Gender	15,4%	84,6%	100,0%
	male	Count	25	297	322
		% within Gender	7,8%	92,2%	100,0%
Total		Count	52	445	497
		% within Gender	10,5%	89,5%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	7,110[a]	1	,008		
Continuity Correction[b]	6,315	1	,012		
Likelihood Ratio	6,807	1	,009		
Fisher's Exact Test				,009	,007
Linear-by-Linear Association	7,096	1	,008		
N of Valid Cases	497				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 18,31.

b. Computed only for a 2x2 table

Gender * VARIABLE *H* Crosstabulation (*h*-retaining regions, all word classes)

			VARH		Total
			deletion	realization	
Gender	female	Count	11	168	179
		% within Gender	6,1%	93,9%	100,0%
	male	Count	2	223	225
		% within Gender	0,9%	99,1%	100,0%
Total		Count	13	391	404
		% within Gender	3,2%	96,8%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8,844[a]	1	,003		
Continuity Correction[b]	7,237	1	,007		
Likelihood Ratio	9,373	1	,002		
Fisher's Exact Test				,004	,003
Linear-by-Linear Association	8,822	1	,003		
N of Valid Cases	404				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,76.

b. Computed only for a 2x2 table

Gender * VARIABLE *H* Crosstabulation (*h*-retaining regions, aux. and pron. only)

			VARH		Total
			deletion	realization	
Gender	female	Count	6	76	82
		% within Gender	7,3%	92,7%	100,0%
	male	Count	0	121	121
		% within Gender	0,0%	100,0%	100,0%
Total		Count	6	197	203
		% within Gender	3,0%	97,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9,123[a]	1	,003		
Continuity Correction[b]	6,751	1	,009		
Likelihood Ratio	11,149	1	,001		
Fisher's Exact Test				,004	,004
Linear-by-Linear Association	9,078	1	,003		
N of Valid Cases	203				

a. 2 cells (50,0%) have expected count less than 5. The minimum expected count is 2,42.

b. Computed only for a 2x2 table

Gender * VARH Crosstabulation (*h*-retaining regions, aux. and pron. excluded)

			VARH		Total
			deletion	realization	
Gender	female	Count	5	92	97
		% within Gender	5,2%	94,8%	100,0%
	male	Count	2	102	104
		% within Gender	1,9%	98,1%	100,0%
Total		Count	7	194	201
		% within Gender	3,5%	96,5%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,559[a]	1	,212		
Continuity Correction[b]	,746	1	,388		
Likelihood Ratio	1,600	1	,206		
Fisher's Exact Test				,266	,195
Linear-by-Linear Association	1,552	1	,213		
N of Valid Cases	201				

a. 2 cells (50,0%) have expected count less than 5. The minimum expected count is 3,38.

b. Computed only for a 2x2 table