

1

What Is Historical Sociolinguistics?

KEY TERMS

- language variation
- variants
- variable
- inherent variability
- speaker choice
- social meaning
- principle of accountability
- observer’s paradox
- real and apparent time
- bad data problem
- literacy
- social history
- accommodation
- convergence and divergence
- speech/language community
- structured/ordered heterogeneity
- social and linguistic embedding
- internal and external factors
- S-curve
- S-curve stages
- social rank/class
- social order/ hierarchy
- social mobility
- social aspirers
- change from above/ below socially
- linguistic insecurity
- hypercorrection

Language in the Social World

We do not expect we are telling you anything new when we say that languages are far from homogeneous. Some speakers of German use the word *Samstag* for Saturday; others prefer *Sonnabend*. Both words have the exact same meaning. In English, words such as *car*, *four* and *floor* with a so-called postvocalic *r* can be pronounced with or without the final *r*. Independent of the absence or presence of the *r*, however, the word *car* denotes an automobile. Likewise, Romance languages such as French and Spanish often allow a speaker to choose between the indicative and the subjunctive mood, even if prescriptive grammars would allow only one option. In all these cases – and many more examples from any language around the world could be given – language displays diversity. Some speakers prefer one form, while others prefer another option. In other words, in every language around the globe, meanings can often be expressed by different linguistic forms.

In sociolinguistics, the technical term for this kind of diversity is *language variation*, or simply *variation*. Note that the word *variation* does not simply refer to any kind of difference in language. Rather, sociolinguistic variation implies a specific semantic claim. *Variation* is only used to refer to cases where you find

multiple forms within one language that express the same meaning. These forms are called *variants*. Thus, *Samstag* and *Sonnabend* are two variants with the same meaning. The meaning involved is easily determined in the case of lexical variants, as in our German example. But the meaning can also be a grammatical function or a specific phonetic context, such as the consonant *r* in postvocalic position. Words such as *car* thus also have two variants: with and without the final consonant (and in fact, there are even more variants, since, for example, the *r* can be pronounced in different ways). For sociolinguists, the exact description of the meaning involved is less important than the empirical observation that a different variant does not alter the semantic meaning of the utterance.

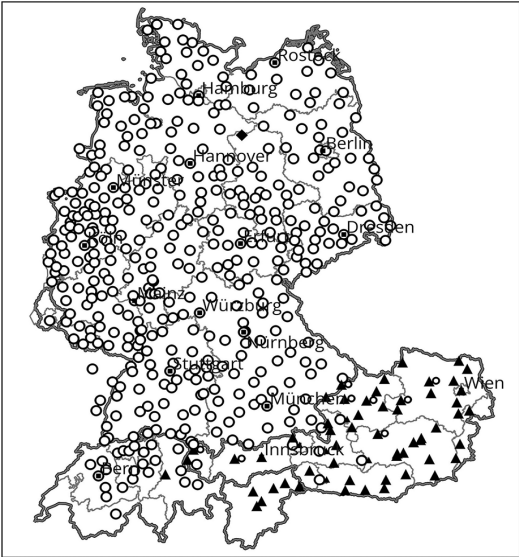
Instead of *meaning*, another term is commonly used to refer to the abstract category represented by multiple variants, namely, the *variable*. Thus, German has the variable ‘Saturday’ with two variants, and English has the variable ‘postvocalic *r*’ with multiple variants. An important principle in sociolinguistics is that every language is characterised by this particular type of variability; that is, in every language you will find variants instantiating a particular variable. This is usually referred to as the *inherent variability* of language.

The inherent variability of German implies that speakers of German can choose which word they want to use to identify the day between Friday and Sunday. Such *speaker choice* depends on various factors. We will return to this issue many times throughout this book, as one of the main goals of sociolinguistics is to find out which speakers choose which forms and why. For now, the crucial thing to understand is that speaker choice does not depend solely on grammatical factors. Our German example clearly demonstrates this basic tenet of sociolinguistics. There are no grammatical principles or rules that tell you that *Sonnabend* is better or more grammatical than *Samstag*, or the other way around. To a large extent, the choice for one of these forms depends on the place where you were born, or where you live or with which you identify. If you grew up in eastern or northern Germany, it is probable that you will use *Sonnabend*. If you live in southern Germany, Austria or Switzerland, you are very likely to prefer *Samstag*. The same line of reasoning holds for postvocalic *r* in English. For example, there can certainly be phonetic or phonological explanations for the occurrence of postvocalic *r* in particular contexts, and its absence in other contexts. In addition, however, there are also other explanations. British English, for example, is in general less rhotic than American English. Therefore, you will often not hear the *r* in *car* when you are in England, but you will hear it in the US – unless you are in Boston, which is famous for its non-rhoticity.

GERMAN EGGPLANTS

Another lexical variable in the German language area is the word used for eggplant (see Figure 1.1). The map (taken from the online *Atlas zur deutschen Alltagssprache*, Elspaß and Möller 2003) shows the distribution of three variants *Aubergine*, *Melanzani* and *Eierfrucht*. *Aubergine* is the most common variant in

GERMAN EGGPLANTS (cont.)



Aubergine / Melanzani

- Aubergine
- ▲ Melanzani
- ◆ Eierfrucht

Zweitmeldungen kleiner

Atlas zur deutschen Alltagssprache
www.atlas-alltagssprache.de

Figure 1.1 *Aubergine*, *Melanzani* and *Eierfrucht* in the *Atlas zur deutschen Alltagssprache* (Elspaß and Möller 2003, www.atlas-alltagssprache.de/r10-f3k/?child=runde).

Germany and Switzerland, while *Melanzani* is more common in Austria and Süd-Tirol (the German-speaking areas in northern Italy) and in some eastern parts of Switzerland. The third variant, *Eierfrucht*, is quite infrequent. Note that the map, like any dialect map, does not show categorical differences. It only shows the most frequent and the second most frequent forms in certain areas so that it is, for example, by no means impossible for speakers of Austrian German to produce *Aubergine* or even *Eierfrucht*.

In the examples discussed, region appears to be a relevant factor: eastern and northern Germany versus southern Germany, Austria and Switzerland, and British English versus American English. However, speakers have a choice, even if they have grown up in a place where one of the variants is clearly dominant. Speakers originating from eastern Germany, where *Sonnabend* is by far the most frequently occurring variant, are by no means obliged to use this variant. If they choose to use *Samstag* while in eastern Germany, they may be perceived as someone from a different region in the German language area. If they are known to originate from the east, however, they may also be seen as people who try to dissociate themselves from eastern Germany. If they use *Sonnabend* in southern Germany, they may be identified as eastern or northern. Using *Samstag* could remain unnoticed there, but could also be perceived as a way to accommodate to southern German usage, or to distance oneself from eastern or northern usage. Thus, choosing a particular variant

4 What Is Historical Sociolinguistics?

carries all kinds of social connotations. This is what we mean by *social meaning*. Although variants are neutral with respect to their propositional or referential meaning – Saturday is Saturday, a car is a car – they do carry this different kind of meaning.

To sum up, sociolinguistics is the study of language as a social object, as a phenomenon that is not just a cognitive competence or a communicative tool but also an integral part of the social world in which human beings live. Languages display inherent variability, and sociolinguists are interested in the social meanings attached to particular variants, as these can help explain the distribution of variants across speakers and communities. By choosing a particular variant, speakers can identify with a social group within the language area, such as ‘people originating from eastern Germany’. As said, this also implies that they can choose *not* to identify with a particular social group. Finally, the concept of speaker choice entails that speakers can move back and forth between variants, thus invoking different social meanings at different times. As one of the founders of the discipline of sociolinguistics, Bill Labov (1972: 208) famously said: ‘there are no single-style speakers’.

THE PRINCIPLE OF ACCOUNTABILITY

In a sociolinguistic approach, it is not enough to establish the frequency of the variant that you are interested in. You also have to look into the linguistic contexts where it could have occurred but has not, and account for all instances of a linguistic variable in the data, not just those that are most salient or frequent. This is called the Principle of Accountability, which is well explained by Tagliamonte (2012: 9–21). She gives the fascinating example of the quotative *be like*. Comparing the grammatical contexts in which *be like* occurs, it seems that it is most frequent in direct speech as opposed to indirect speech, sounds/gestures, hypothetical speech and writing. But this is not accountable, since the contexts in which *be like* could have occurred, but that have another quotative such as *go*, *think* or *say*, have been left out. In each context, we need to consider the proportional frequency of *be like* compared to that of other variants, such as *go*, *think* or *say*. If we consider this, *be like* turns out to actually favour contexts of indirect speech. The high proportion of *be like* in the non-accountable analysis was merely an effect of the high frequency of quotatives introducing direct speech in general, and thus it would be wrong to conclude that *be like* is typical for direct speech.

Language in the Sociohistorical World

While sociolinguistics proper generally focuses on language variation in contemporary settings, historical sociolinguistics studies language variation in the past. There are many good reasons for doing this. First of all, language histories and more

traditional studies of language change over time often lack the social dimension outlined in the previous section. They describe particular changes in a language, but variation and social meaning are usually not part of such descriptions, or only to a limited extent. Naturally, people in the past also lived in a social world in which language was an important social object. There seems to be no good reason to assume that inherent variability was not part of past societies. By focusing on these sociolinguistic aspects of past societies, historical sociolinguistics aims to arrive at a richer, fuller, more complete account of language history.

Secondly, language change necessarily involves a period of variation. In written French from the sixteenth and seventeenth centuries, you will frequently find the negation pattern with *ne* and *pas*, that is, with two words expressing just one negation. This is also the pattern that is dominant in the present-day written standard, for example, *elle ne lit pas* ‘she does not read’. Throughout the centuries, however, many language users have also used a variant with only one negator, as in *elle lit pas* ‘she does not read’. The single variant is extremely common today in spoken French, but can also be found in historical sources. Thus, the single and the double option have already co-existed for centuries. The language change that we are witnessing here is from double *ne pas* to single *pas*, but the centuries-long period in which the two variants co-occur constitute an intriguing period of variability – a period, in other words, in which speakers get to choose, and in which consequently social meanings become attached to the variants. To fully understand the change from *ne pas* to *pas*, a study of their variation seems to be in place.

SINGLE NEGATION AND THE FRENCH DAUPHIN

In a thoughtful paper on the variable use of negation in historical and present-day French, Ayres-Bennett (1994) presents various seventeenth-century examples of negative statements lacking *ne*. The most well known are perhaps phrases such as *je veu pa* ‘I don’t want to’ and *je sui pa beau* ‘I am not beautiful’, presumably used by the Dauphin in 1608 and 1609 at the age of six and seven. The language use of the Dauphin – that is, the heir to the throne, in this case the later King Louis XIII (1601–1643) – was recorded by the personal physician to the Dauphin, Jean Héroard (1551–1628), in his *Journal*. It is a matter of debate to what extent the *Journal* offers an accurate rendering of spoken French of the time, but it is also clear that such *ne*-less clauses are evidence that variable use of sentence negation is an age-old phenomenon in French.

Sociolinguistics has been extremely successful in describing and explaining present-day variation in a wide variety of languages from around the world. Historical sociolinguistics sets out to extend the sociolinguistic approach into the past, although – as we will discuss in Chapter 2 – this is actually a more complicated endeavour than one might think at first glance. One of the reasons for this is that contemporary sociolinguists normally use spoken language data. They record speech by living speakers, and they prototypically analyse phonetic variation. This

What Is Historical Sociolinguistics?

poses obvious problems for historical sociolinguists, who tend to work with written data produced by language users who are no longer alive. This suggests that there can be no simple application of sociolinguistic theories and methods to past situations. On the contrary, historical sociolinguistics has its own theoretical and methodological challenges, and goes beyond merely applying principles and methods from contemporary sociolinguistics to historical data.

If you think working with written data by deceased people is much less ideal than making your own recordings of a well-selected group of people, you may be right to some extent. Nonetheless, a number of methodological issues that have been debated in the sociolinguistic literature have been argued to be much less problematic in historical sociolinguistics. In sociolinguistics, the *observer's paradox* is a well-known phenomenon. It entails researchers' desire to observe language users in a situation that is as natural as possible, in which they speak as spontaneously as possible. After all, more formal situations often trigger more formal and generally more standardised speech, so that language users start choosing the standard variants, thus hiding the inherent variability of the language. The paradox is that the natural, informal situation needs to be consciously created by a researcher. Informants in historical sociolinguistics – that is, past writers of textual sources – did not know that their language use would be analysed for linguistic purposes centuries later. Therefore, they could not accommodate their linguistic choices to the situation created by the researcher.

Finally, historical sociolinguists have access to *real-time* data. In contemporary sociolinguistics, it is common to compare different age groups. If a relatively new variant, such as the quotative *be like* – for instance, in the utterance *I was like yeah right* – occurs frequently among teenagers, much less frequently in the speech of people in their thirties and forties and hardly at all among the elderly, then a logical prediction would be that this new form, established in young speech, will steadily rise over the coming years and decades, as people who use the form infrequently will generally pass away sooner than the more frequent users. Sociolinguists thus seek to model language change by comparing age groups at one point in time. The approach is commonly referred to as the *apparent-time* construct, since you do not actually compare different points in time but rather use the roundabout route of looking at different age groups. In historical sociolinguistics, on the contrary, data can easily be gathered from various moments in time: when the change just starts, during the change and when the change is almost complete. We call this *real-time* data. If available, real-time data are to be preferred over apparent-time data, as the apparent-time construct is based on the assumption that variant use remains stable over time: we assume that youngsters using quotative *be like* now will continue to do so when they get older, which would allow us to project the pattern of change into the future. However, whereas this may be the case for some times of variables, it may not be the case for others. Lexical items such as youth slang may very well be used by people in their teens, but they will abandon these choices once they get older; in the same way that dress or hairstyles are not stable across the lifespan, neither are all linguistic choices. To be able to really map changes in language over time, it is thus safer to rely on real-time data than on apparent-time data.

Bad Data?

Another well-known dictum by Bill Labov (1994: 11) says that historical linguistics consists in ‘the art of making the best use of bad data’. One aspect of the so-called *bad data problem* refers to the already-mentioned fact that most data in historical linguistics are written, and were originally also meant to be written; that is, they are not transcripts of spoken interaction. Whereas the strong focus on spoken language data in modern sociolinguistics may make you think written data are inferior, another view would be that historical sociolinguistics simply faces different challenges. Writing traditions, genres, orthographical variation and change, and the relationship between the spoken and the written medium emerge as focal areas for a sociolinguistic analysis of past situations. These topics are usually not addressed in modern sociolinguistics, testifying to the independent status of historical sociolinguistics. In Chapter 6, we will discuss a number of the issues mentioned here.

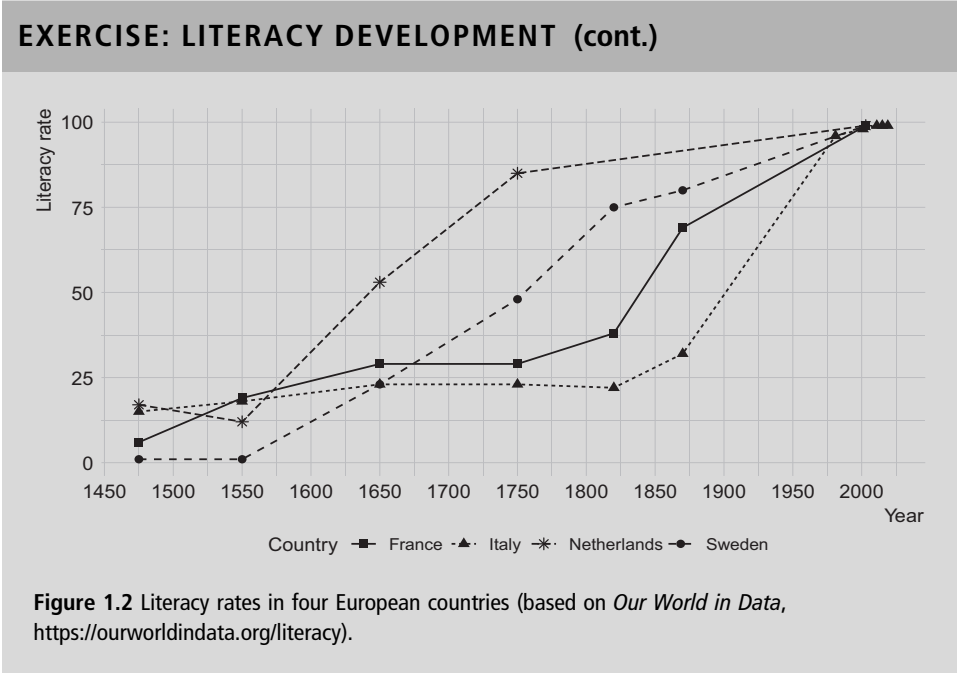
Furthermore, the bad data problem entails that historical linguists, including historical sociolinguists, depend on the textual sources that history happens to have left. Systematic data collection is still possible but obviously to a more limited extent than in present-day situations. While it will often be relatively easy for a modern sociolinguist to assemble a similar amount of data from, say, male and female speakers, in historical situations this is completely different. There is a good chance that more data will be preserved written by men. Throughout this book, we will show that there are actually quite large and reliable datasets in historical sociolinguistics, which make it possible to acquire a thorough understanding of the sociolinguistics of past situations.

We will return to the issue of sources in several chapters, and particularly in Chapter 7 on language history ‘from below’. Here, we want to mention that *literacy* is a crucially intervening factor. By definition, only textual data produced by literate people survive. In addition, literacy has been sensitive to social differentiation throughout history. In general, the male population of a certain language community was more likely to be literate than the female population, and privileged social ranks were more likely to be literate than socioeconomically lower ranks. Such shortcomings can be overcome by incorporating as much data as possible from relatively lower social ranks and from women – and historical sociolinguistics has shown that these data are in fact available for many languages.

EXERCISE: LITERACY DEVELOPMENT

The *Our World in Data* website (<https://ourworldindata.org>) has historical data on literacy for a number of countries. Figure 1.2 shows the literacy rate in France, Italy, Sweden and the Netherlands between 1475 and 2015.

Describe the differences in terms of literacy development for these countries. What possible reasons could explain the marked differences between these countries in the seventeenth and eighteenth centuries?



Importantly, the desire to unearth the social meaning of linguistic forms in past societies requires detailed knowledge of the relevant social categories in those historical situations. Here, Labov (1994: 11) was quite pessimistic with respect to historical studies: ‘We usually know very little about the social position of writers, and not much more about the social structure of the community.’ For modern sociolinguists, sociology offers a lot of information on the social make-up of present-day societies. For example, the old distinction between men and women in terms of biological sex has given way to a more nuanced view in which *gender* has replaced *sex*, confirming that gender roles are largely culturally constructed and often linked to biological sex in a partially indirect manner. Likewise, new categories of gender neutrality have emerged in recent years. Such sociological changes impact sociolinguistics, where gender and sexual orientation have become central research topics. Similarly, *social history* is of crucial importance to historical sociolinguistics. We will argue throughout this volume that Labov was too pessimistic, and that social history can help historical sociolinguists tremendously in reconstructing the social structure of past societies. Of course, there is no reason to assume beforehand that social categories such as class and gender were equally relevant in historical periods, and, even if they were, they may need to be operationalised differently. The danger of anachronism will be discussed in Chapter 2.

Case Studies

Instead of talking about the advantages and challenges of historical sociolinguistics, we now want to show you what historical sociolinguistics can contribute to our

understanding of language variation and change. One important insight from modern sociolinguistics is that social meaning can often be seen as a group phenomenon. This means that language users express group membership by adopting particular variants, often unconsciously, and by non-adopting other variants. These processes are also referred to as *convergence* and *divergence*, particularly within the social-psychological approach of *accommodation theory*. The basic idea here is that accommodation, that is, adjusting linguistic choices to your peers, leads to linguistic group behaviour, and may thus result in linguistic variants typically associated with specific social groups.

The kinds of social groups that are relevant to language users depend on the situation. In other words, this is a historical question. Gender roles and socio-economic status, for example, are significant in present-day Western Europe, but one can easily imagine social settings in which wholly different social divisions determine the social structure, such as ethnicity or race, religion or political beliefs. Likewise, social divisions irrelevant or marginal today may have been of crucial importance in the past. Literacy and differential access to schooling have been major social divides in history. In Early Modern German, northern Protestantism and southern Catholicism became associated with different spelling patterns (Rutkowska and Rössler 2012: 217). A similar divide along the same religious lines can be found in Late Modern Dutch (Vosters and Rutten 2015).

If social meaning is seen as a group phenomenon, the level of analysis is the language produced by various groups within a larger community. Modern sociolinguists sometimes speak of *speech communities*, but in historical sociolinguistics it makes more sense to use the term *language community*. These are problematic concepts for various reasons, including the simple fact that many language users are multilingual and are not members of just one community. We nonetheless contend that historical sociolinguistic studies have shown that we can reasonably hypothesise that regions where similar supra-regional forms are found can be seen as constituting one language community. For example, the geographical distance between northern England and the London area was considerable in the sixteenth century, as travelling was slower and less comfortable than nowadays. Still, the obvious resemblances between written English from the north and written English from London in this period merit comparisons between the north and London as two regions within one language community.

If the choice for certain variants is related to group identities, language variation at the level of the community is not a random phenomenon. Instead, there should be a clearly determinable order in the extensive variation often found across communities. This order, termed *structured* or *ordered heterogeneity*, is a quantifiable phenomenon at the level of the community. Importantly, it is usually a statistical tendency: language users, after all, are not just members of one social group, they can vary their linguistic choices according to circumstances and they are not forced, in a deterministic way, to follow group patterns. The way in which language variation and change relate to the social structure is usually called *embedding*. In addition to this social embedding, variants are also part of a linguistic system, in which they are thus also embedded. The *social embedding* of form entails that particular groups within the community are more likely to use it than others. *Linguistic embedding* refers to the fact that some variants are more likely to occur in particular grammatical

10 What Is Historical Sociolinguistics?

contexts than in others. If you want to find out which elements of the social structure promote a particular variant, you study the *social* or *external factors* that condition the use of the variant. Similarly, if you are interested in linguistic embedding, you study the *linguistic*, *grammatical* or *internal factors*.

Two Changes in Early English

As a first introduction to historical sociolinguistic research, we will now focus on two changes in Early English, taken from the seminal study by Nevalainen and Raumolin-Brunberg (2017). The first change is the replacement of the subject pronoun of the second-person *ye* by the form *you*, which was originally the object form. In Example (1), the writer mixes *ye* and the incoming form *you*.

- (1) I perceave **YOU** have subskrybed our bylles to Thomas Everton, and that **YOU** do looke for advise, which I do not understond, for I am sure **YE** know the parcelles that we have bawght (Ambrose Saunders 1551 – CEEC)

The second change involves the replacement of the historical suffix *-th* for the present indicative third-person singular (Example 2) by the suffix *-s* (Example 3).

- (2) that next trosty man that **COMYTH** shall bring hym (George Cely 1476 – CEEC)
- (3) who **DWELLS** at Thistleworth (John Chamberlain 1614 – CEEC)

As argued earlier, a main advantage of historical research is that you have access to real-time data. Using the *Corpus of Early English Correspondence* (CEEC) with 6,039 personal letters by 778 writers, spanning the period from c. 1410 to 1681, Nevalainen and Raumolin-Brunberg (2017) counted the variants per subperiod. They chose particular temporal intervals and established the proportion of the historical variants *ye* and *-th* and the incoming forms *you* and *-s* per interval. You can then plot the proportion of the incoming form on the y-axis and the temporal intervals on the x-axis. Figure 1.3 gives the proportion of the incoming form *you* across time.

The shape of the line in Figure 1.3 resembles the letter *s*, which is why such plots of language changes are called *S-curves*. It is a solid result of sociolinguistic research that language changes often follow this pattern, where you first witness a period of small proportions of the new variant, then a relatively fast increase of the new form and finally a period where the incoming form gradually approaches 100 per cent. Nevalainen and Raumolin-Brunberg (2017: 55) distinguish five *stages* on the slope of the S-curve:

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