

Chapter 7

Extensions and applications: historical linguistics, sociolinguistics and psycholinguistics

7.1. Introduction

In the preceding chapters you have encountered a host of language phenomena from various domains (phonetics, phonology, morphology, syntax, semantics, pragmatics), and you have learned how these phenomena can be investigated in order to find out more about the structure and use of language. Apart from these areas of linguistic research, scholars (and laypersons) are interested in many more questions concerning language and languages. In this chapter we will deal with three of such sets of questions. First, there are historical questions. How do languages develop and where do individual languages come from? What is the historical relationship between languages, e.g. between English and German? Why and how do languages change? Another set of questions concerns the social significance of language. Why is it that after listening to only a few words a speaker has uttered, we seem to know a lot about this speaker's social background, for example her education, her social status, in which region she probably grew up, etc.? What exactly is it that is so telling? Third, many people are curious about how humans store and process language, and what language can reveal about our cognitive capacities in general.

We will deal with each set of questions in turn, looking at important studies in these fields in an exemplary fashion. As you will notice, trying to meaningfully answer such questions necessitates a general understanding about how language 'works', i.e. what kinds of structural entities language consists of and how these entities interact with each other to create that highly complex system we call language. Having worked through the preceding chapters, you are now equipped with this kind of understanding.

7.2. Historical linguistics: how languages develop

In the late 18th century, William Jones made a discovery that was seminal for the study of the historical development of languages, i.e. for the academic field that later came to be known as **historical linguistics**. Studying the ancient Indian language Sanskrit he found that

[t]he Sanskrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them *a stronger affinity*, both in the roots of verbs and in the form of grammar, *than could possibly have been produced by accident*; so strongly indeed, that no philologist could examine them all three, without believing them *to have sprung from some common source which, perhaps, no longer exists*: there is a reason, though not quite so forcible, for supposing that both the Gothic and the Celtic, though blended with a very different idiom, had the same origin as the Sanskrit; and the Old Persian might be added to the same family. (Jones 1786, emphasis added)

At the time little was known about the ‘common source’ of the languages mentioned, and Jones was mostly speculating. However, his hypothesis was so strong and attractive that many people started investigating the alleged ‘affinity’ in a systematic fashion. Let us follow in their footsteps by looking at data from English and German, two languages which also show a remarkable affinity that does not look accidental. In (1) I have listed a number of word forms from English and German that show some potentially non-accidental correspondences in both meanings and sounds. Before reading on, try to figure out all correspondences between English and German sound, concentrating on the consonants (you may find it useful to differentiate between word-initial, word-medial, and word-final positions). For the benefit of the reader I also give the phonemic transcription of the German words (or roots, where pertinent).

(1) Some similar English and German words

English	German		meaning of German word
pan	Pfanne	/pfanə/	‘pan’
pole	Pfahl	/pfa:l/	‘pole’
path	Pfad	/pfa:d/	‘path’
plough	Pflug	/pflu:g/	‘plough’
plant	Pflanze	/pflantsə/	‘plant’
pipe	Pfeife	/pfaɪfə/	‘pipe’
town	Zaun	/tsaʊn/	‘fence’
toll	Zoll	/tsɔl/	‘customs (duty)’
tame	zähm	/tsa:m/	‘tame’
foot	Fuß	/fu:s/	‘foot’
hate	Hass	/has/	‘hate’
bite	beiß(en)	/baɪs-/	‘bite’

heart	Herz	/herts/	‘heart’
book	Buch	/bu:x/	‘book’
grip	Griff	/grif/	‘grip’
make	mach(en)	/max-/	‘make’
dream	Traum	/traum/	‘dream’
shape	schaff(en)	/ʃaf-/	‘create’
break	brech(en)	/brɛx-/	‘break’
reckon	rechn(en)	/rɛxn-/	‘calculate’
open	offen	/ɔfn/	‘open’
ship	Schiff	/ʃɪf/	‘ship’

Although the corresponding words do not all mean exactly the same thing in both languages, the meanings are sufficiently close so that the similarities in both meaning and sound cannot be accidental. Let us look at the sound correspondences of these words more systematically. In (2) I have listed the correspondences of consonants evidenced by the data in (1):

(2) Systematic sound correspondences

	English	German	Example
a.	/b/	/b/	break — brechen
	/f/	/f/	foot — Fuß
	/n/	/n/	pan — Pfanne
	/l/	/l/	plant — Pflanze
	/ʃ/	/ʃ/	shape — schaff(en)
	/r/	/r/	dream — Traum
	/h/	/h/	hate — Hass
b.	/p/	/pf/ or /f/	pipe — Pfeife, grip — Griff
	/t/	/ts/ or /s/	town — Zaun, hate — Hass
	/k/	/x/	book — Buch

The correspondences in (2a) might look rather boring since the corresponding consonants in the two languages are the same. However, such a close correspondence of so many sounds may be taken as a strong indication of a close historical relationship. In (2b) things look more intricate, since each English sound systematically corresponds to a different, but phonetically very similar, sound in German. A closer look at this set reveals an interesting generalisation: what is a voiceless plosive in English corresponds systematically to a fricative or an affricate in German. Fine, you might be tempted to say, but what do such systematic correspondences tell us about language history?

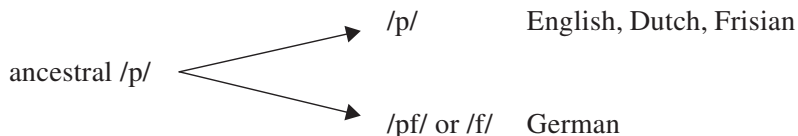
To answer that question we must return to the very basic idea that words are essentially arbitrary pairings of sound and meaning (cf. chapter 3). This is

the reason why the same form can mean entirely different things in two languages, and why two languages can have two entirely different forms to designate the same thing. For example, the string of sounds [ti] means 'not' in the West African language Yemba (spoken in Cameroon), but 'a hot brown drink made by pouring boiling water onto the dried leaves from a particular Asian bush' (*Longman Dictionary of Contemporary English*, s.v. *tea*) in English. In contrast, the numeral '2' is [tu:] in English, but [membɾja] in Yemba. In other words, if two languages happen to have very similar sound sequences to represent the same concepts it is highly unlikely that this is due to chance, since there is such a huge number of possible combinations for the sounds available in a language. Excluding chance, two possibilities remain, borrowing or common ancestry. 'Borrowing' means that either one language took over the words from the other, or both languages took them from a third. Common ancestry means that both languages developed from a common ancestor language, preserving the words in question, with minor changes in form or meaning in one or the other language.

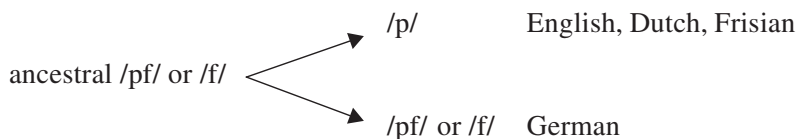
Now, how do we know whether the words in (1) are a case of borrowing? We cannot be 100 percent sure, but we know from many studies of language contact that everyday vocabulary items such as the ones in (1) are not easily borrowed from one language into another. We have therefore good reason to assume that both English and German developed from a common ancestral language. Which language could that be and how did that language look like? In order to answer that question, we would have to check many more languages which may be potential offspring candidates from that ancestral language. If you have learned any Dutch or Frisian, it may have occurred to you that these languages have many words that are very similar to German or English words. If you have ever learned some Finnish, Japanese, or Thai, your feeling might be quite the opposite. Finnish, Japanese, or Thai words hardly ever seem to be even remotely similar to English or German words (unless they have been borrowed from one of these languages). Given this intuitive feeling of similarity, we would rather start looking for a historical relationship between English, German, Dutch, and Frisian than for a relationship between English, German, Finnish, Japanese and Thai. And indeed, the data do suggest a systematic relationship between English, Dutch, and Frisian on the one hand, and German on the other. Thus, Dutch and Frisian, like English, often have voiceless plosives, such as /p/, where German has affricates or fricatives (such as /pf/ or /f/) in corresponding words (cf. English *path*, Frisian *paad*, Dutch *pad*, German *Pfad*). To account for such facts, it seems most likely that there was one ancestral language from which all four languages developed. Either one of the languages, i.e. German, underwent some systematic sound

change, or the other three together underwent some change, and German preserved the old sounds. The two possibilities are given in (3), using /p/ as an example:

(3) a. Hypothesis A



b. Hypothesis B



The obvious question is, which hypothesis is more likely? Two arguments can be used. First, it may seem more likely that one language undergoes a sound change, as in (3a), than that three languages undergo the same sound change independently, as in (3b). This favours (3a). We have to be cautious, however. It could also be the case that the split resulted in only two languages, let us call them ‘Germanic 1’ and ‘Germanic 2’, and only later ‘Germanic 2’ split up into the three languages English, Dutch, and Frisian, while ‘Germanic 1’ became German. Under this assumption, the scenarios in (3) would involve only two daughter languages and our first argument against hypothesis B would be less compelling.

The second argument is that the development of plosives into affricates (or fricatives) is a very common change that is very well motivated by articulatory considerations. Thus, the strong aspiration of a plosive is articulatorily very similar to the fricative part of an affricate. You may try this out by producing, for example, an aspirated [t^h] very slowly, which makes the [t^h] rather sound like a [ts]. Over time, strong aspiration may thus turn phonetically into frication, which turns the plosive into an affricate. A good example of this is Italian, in which the Latin [k] has turned into [tʃ] in certain environments (through some intermediate stages that also affected the place of articulation), as in *pa[tʃ]e* ‘peace’ or [tʃ]ento ‘hundred’. As a second step in the development from plosive to fricative, the first element of an affricate may be weakened, which eventually may lead to a fricative (e.g. [p^h] > [pf] > [f]). These facts strongly favour hypothesis A. At the same time, it is highly unlikely that an existing affricate and an existing fricative would develop into a plosive.

In sum, we have an argument against hypothesis B, and two arguments for hypothesis A. Hence we can posit that an ancestral sound /p/ developed into an affricate or fricative in German, and remained a /p/ in the other languages.

What we have just done is apply one of the oldest and still well-respected methods in linguistics, **comparative reconstruction**. Following this method, the systematic comparison of corresponding words (so-called **cognates**) in very many languages can be used to establish genetic relationships between these languages. In the case at hand, we have reconstructed a sound of a common ancestral language of English, Frisian, Dutch, and German. Such ancestral languages are called **proto-languages**, and the common ancestral language of our four languages (but also of Icelandic, Gothic, Danish, Norwegian and Swedish) is called 'Proto-Germanic'. Of course one would need a lot more data and many more potentially related languages to figure out the detailed genetic relationships among them, but our little investigation above should have shown how this could be done. Genetic relationships are usually represented in the form of tree diagrams, just like family trees. The proto-language and its descendants together are called a **language family**. Branches of a larger language family can be referred to as subfamilies.

Below you find the family tree for the Germanic languages. As you will see, the genetic relationship between our four Germanic languages is much more complex than our simple example might have suggested.

The Germanic languages in turn belong to the larger family of Indo-European, of which I only give the main branches in figure 7.1. below:

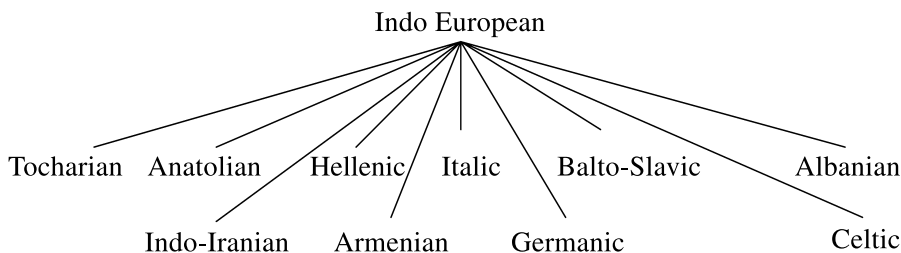


Figure 7.1. The Indo-European language family

Let us now return to our data in (2). Apart from having reconstructed a genetic relationship between some languages (i.e. English and German), we have also found a systematic sound change German has undergone, namely the development of voiceless plosives into affricates and fricatives, as illustrated by the correspondences in (2b). This change, which probably started between the 3rd and the 5th centuries AD and ended in the 9th century, is known as the

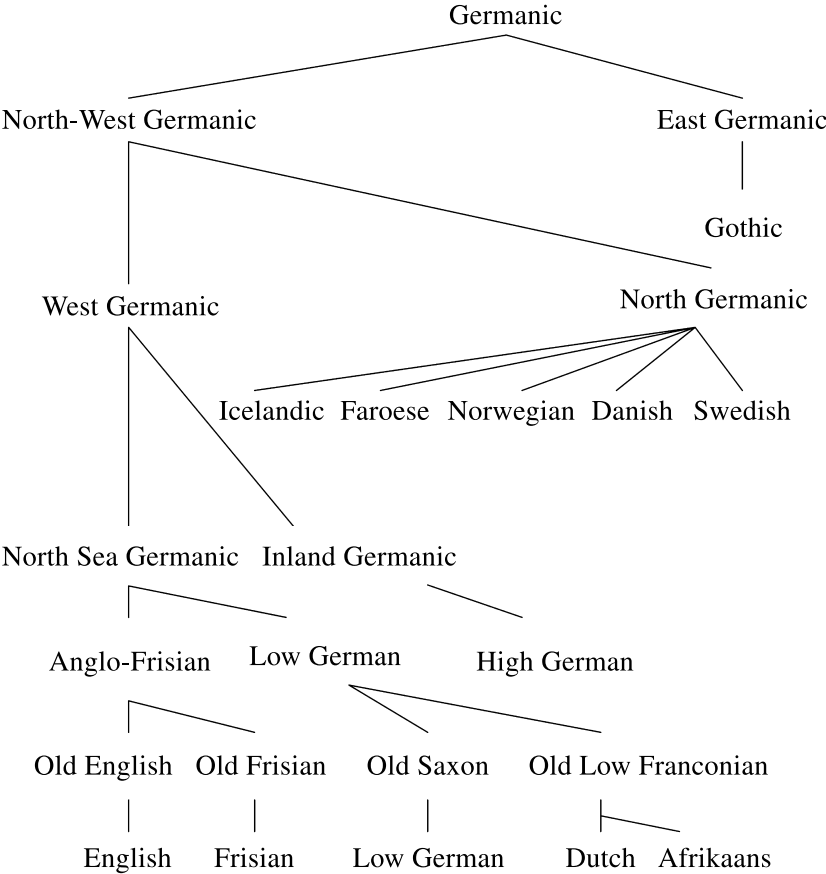


Figure 7.2. The Germanic language family

Second Germanic Sound Shift. As we have seen, the other languages under discussion (English, Frisian, and Dutch) did not undergo that change. If you now take a look again at figure 7.2., you can see that it is the Second Germanic Sound Shift that resulted in the split between what is called ‘North Sea Germanic’ and ‘Inland Germanic’ in the tree. Out of these two proto-languages the modern languages developed: English, Dutch, and Frisian on the one hand, and German on the other.

Note that, curiously, German still has voiceless plosives, which should be impossible if the Second Germanic Sound Shift had turned them all into affricates or fricatives. The answer to this apparent puzzle is that the German voiceless plosives in today’s language are either those that survived the sound

shift (because the shift in fact did not uniformly affect all voiceless plosives in all positions in all words), or they were re-introduced with words that were borrowed into the language after the shift was completed.

The term ‘Second Germanic Sound Shift’ suggests the existence of another such shift, the **First Germanic Sound Shift**, and indeed, there is such a thing. Historical linguists have provided overwhelming evidence that Proto-Germanic underwent a change in its consonant system that separates this branch from all other branches of Indo-European. The shift presumably happened around 1000 BC and is quite complex, so that the details need not concern us here. The major developments of the First Sound Shift are schematised in (4). The shift is also known as **Grimm’s Law**, since it was the German scholar Jakob Grimm who first formulated the generalisations in this way in 1822:

(4) Grimm’s Law (partial representation, in phonemic transcription)

Proto-Indo-European	Proto-Germanic	Generalisation
b ^h	> b	} aspirated voiced plosives > voiced plosives
d ^h	> d	
g ^h	> g	
b	> p	} voiced plosives > voiceless plosives
d	> t	
g	> k	
p	> f	} voiceless plosives > voiceless fricatives
t	> θ	
k	> x	

Apart from reconstructing genetic relationships between languages, historical linguists are of course also interested in the development of individual languages. Given the lack of speakers or tape recordings from a few centuries back, they try to find as many texts from earlier stages of the language as possible, such as medieval manuscripts, or, in the case of older Germanic languages, runic inscriptions on stone crosses, bones, or gravestones. From such historical attestations, the development of lexicon, pronunciation, and grammar of these languages can be reconstructed.

Let us take a brief look at the history of English, which began in 449 AD with the invasion of Britain by the Saxons, Angles, and Jutes. The invaders spoke various Germanic dialects, and out of this set of dialects grew the language that we know as English today. In the course of its history, English came under heavy influence of various other languages, such as (in chronological order) Celtic, Scandinavian, French, and Latin. Scholars usually distinguish four major periods in the history of English: Old English (440–1100), Middle English (1100–1500), Early Modern English (1500–1700), and Modern English (1700 until today). To get an idea about how English looked like at earlier stages, consider the Bible verses 21 and 22 from three different English versions of the Parable of the Lost Son, straddling about a millennium:

Old English version (Anglo-Saxon, 10th century)

²¹Ðā cwæð his sunu, “Fæder, ic syngode on heofon and beforan ðē. Nū ic ne eom wyrðe þæt ic þīn sunu bēo genemned.” ²²Ða cwæð se fæder tō his þēodum, ...

Middle English version (Wyclif Bible, ca. 1380)

²¹And the sone saide to him, “Fadir, Y haue synned in to heuene and bifor thee; and now Y am not worthi to be clepid thi sone.” ²²And the fadir saide to hise seruautis, ...

Modern English version (New International Version, 1978)

²¹And the son said to him, “Father, I have sinned against heaven and in your sight, and am no longer worthy to be called your son.” ²²But the father said to his servants, ...

As can be easily seen, changes have taken place on all levels of the language and its written representation. I will mention only very few here to give you a taste of what kinds of phenomena could be further investigated. Perhaps most striking upon your first encounter with Old English is the fact that on the way to Modern English, the writing system has changed. Some letters were lost altogether (e.g. <ð>, <þ>, or macrons above vowel letters, as in <ō>), and some letters had a different function (<u> was used to represent also what is now <v>, as in *seruautis* ‘servants’). Furthermore, the pronunciation has changed (e.g. the plosive /d/ in *fæder* is now a dental fricative /ð/, or some final vowels are lost, as in *sone* > *son*). On the lexical level, some words have fallen out of usage (e.g. *þēod* ‘people’, or *cwæð* ‘said’, but compare archaic *quoth*, which we still read occasionally in 19th century literature). Other words have

changed their meaning. Thus, in Old English *heofon* still meant both ‘heaven’ and ‘sky’, while at later stages its meaning became narrowed down to ‘heaven’. With regard to morphology, we can see that some inflectional affixes are now obsolete (e.g. the past participle prefix *ge-* in Old English *genemned*, or the dative plural suffix *-um* in Old English *bēodum*). The word order has also changed, as illustrated by Old English object-verb order in *þīn sunu bēo genemned* ‘your son be called’ becoming verb-object *be clepid thi sone* ‘be called your son’ in Middle English. In sum, it is rather obvious that English has undergone massive changes in the course of its history.

Seeing all these changes attested in documents of earlier times, and having gained some insight into even earlier developments by comparative reconstruction, the question arises why language should change at all. And, more interestingly perhaps, why do certain changes occur in one language, but not in others? And can we predict the types of changes a language will undergo? Obviously, these questions are much too complex to be dealt with in this rather brief section on historical linguistics and language change. But some of the mechanisms at work in language change can be understood much better if we relate them to the set of questions concerning the social significance of language. To these questions we now turn.

7.3. Sociolinguistics: the social significance of language

As mentioned already in section 7.1., the way we speak reveals quite a lot about our social background, for example how educated we are, where we come from, etc. In the 1960s linguists began to systematically investigate the relationship between language and society by looking at the variable use of certain linguistic features by different groups of speakers. This branch of linguistic research is called **sociolinguistics**. Take, for example, the familiar variability of how speakers of English pronounce the final sound in the following words: *surfing*, *walking*, *sleeping*. They either use the velar nasal [ŋ] or the alveolar nasal [n]. This is illustrated in (5):

(5)	surfing	[sɜ:fɪŋ]	[sɜ:fɪn]
	walking	[wɔ:kɪŋ]	[wɔ:kɪn]
	sleeping	[sli:pɪŋ]	[sli:pɪn]

Is this possible with any word ending in <ing>? Let us see:

(6)	thing	[θɪŋ]	*[θɪn]
	sing	[sɪŋ]	*[sɪn]
	Beijing	[beɪdʒɪŋ]	*[beɪdʒɪn]

As the data in (5) and (6) show, the variation is restricted to the verbal suffix *-ing*. We can therefore say that the verbal suffix *-ing* has two variants: [ɪŋ] and [ɪn]. For that reason, the suffix *-ing* can be seen as a ‘variable’, i.e. a linguistic entity which varies in its manifestations in speech. The concept of variable should remind you of the concept of allomorphy, where we also found an abstract linguistic entity (i.e. the morpheme) to have differing manifestations in speech. While morphologists study which linguistic factors determine the distribution of the different variants, sociolinguists are interested in the social factors that may be responsible for the choice between different possible variants. In such cases we then speak of a **sociolinguistic variable**. Sociolinguistic variables are notationally given in normal orthography, but in parentheses, e.g. (ing). In the case of our suffix *-ing* the morphologist would find that there are two allomorphs [ɪŋ] and [ɪn], whose distribution cannot be explained by linguistic factors. Speakers can use both variants with any given verb. The obvious question now is: if there are two variants and no strictly linguistic rules that determine their usage, what makes a speaker choose between one or the other variant in a given situation? Is this entirely arbitrary or at least to some extent systematic? We could venture, for example, the hypothesis that the variable (ing) is distributed in such a way that one variant, [ɪŋ], is the variant used in the standard language, while [ɪn] rather occurs in non-standard, colloquial English. By ‘standard’ we mean a variety of English which is used in formal settings (such as educational institutions, business, and the media) and which enjoys the highest social prestige.

A word is in order here on the notion of standard and correctness. Speakers of the standard variety often tend to regard non-standard varieties, including regional dialects, as incorrect or sloppy forms of English. From a scientific point of view this does not make sense. The speakers of a regional dialect are native speakers of their dialect and their dialectal speech involves the correct application of the grammar (i.e. the system of rules) of their dialect. That these rules differ from the rules of the standard variety is obvious, because otherwise we could not make the distinction between that dialect and the standard. The difference between dialect and standard is thus a difference between two different linguistic systems, just like the difference between English and German is a difference between two different linguistic systems. Of course, a regional dialect of English is more similar to standard English than German is, but the dialect still constitutes its own system with its own words, rules of pronunciation, morphology, and syntax. In view of these facts it is nonsensical to say that a given regional dialect is ‘incorrect (standard) English’ (we would also not say that German is incorrect English). We are simply dealing with two different varieties of English. The crucial point is, however,

that the two varieties are socially not equal. While the standard provides the norm for a whole country and is socially highly prestigious, the regional dialect is restricted to its local setting and is not supposed to be used in official or formal social contexts. So if a dialect speaker uses her dialect in a formal situation, it is not the case that her grammar is 'incorrect', or she is too sloppy to apply the rules of English grammar. It is simply that her choice of language variety is inappropriate.

Returning to the problem of [ɪŋ] and [ɪn] and its potential social significance, we can now take a look at a classic study of the variable (ing) in the English town Norwich, in which Trudgill (1974) discovered that the distribution of this variable is indeed determined by social factors. He collected data from various speakers from different social classes and counted the occurrences of the two variants. **Social class** is a construct borrowed from sociology, according to which the members of a society can be categorised. The basis for this categorisation is usually income, occupation, and education, but other parameters (e.g. occupation of parents, neighborhood in which a person lives, social status, etc.) are also sometimes taken into consideration. In his study, Trudgill distinguished between five different classes: lower working class (LWC), middle working class (MWC), upper working class (UWC), lower middle class (LMC), and middle middle class (MMC). He also distinguished between four different speaking styles, ranging from the reading out of a word list, over the reading out of a passage of prose text, to formal speech, and finally casual speech. The assumption here was that reading out a word list is the most formal and most standard-like type of speech, while casual conversations represent the least formal type of speech. Here are his results:

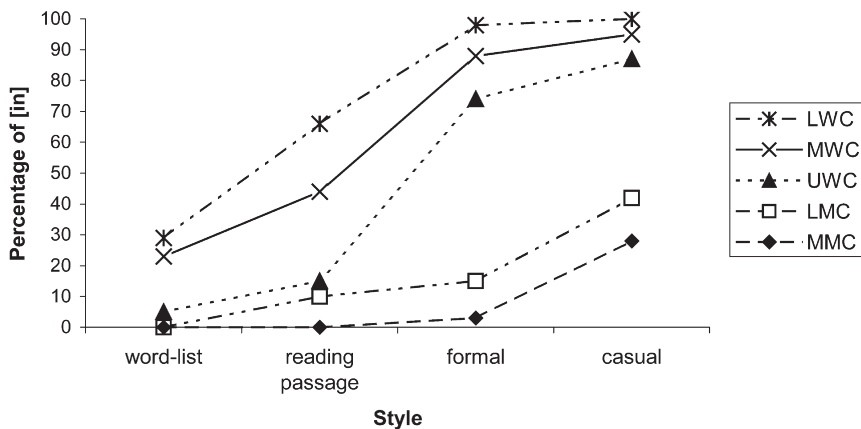


Figure 7.3. The variable (ing) in Norwich (adapted from Trudgill 1974)

On the y-axis we find the percentage of the variant [ɪn], and on the x-axis we find the four different styles, ranging from the most formal on the left to the least formal on the right. The graph shows that there are indeed important differences in the distribution of the two variants between the different styles and between the different classes. Thus there is a general trend that we find fewer [ɪn]'s (and hence more [ɪŋ]'s) in more formal styles. The second trend is that, in general, the lower the class of a speaker, the more often does the speaker use [ɪn] and the less often does he or she use [ɪŋ]. In other words, we now have firm evidence for our above hypothesis that [ɪŋ] is the more formal, standard variant and tends to occur more often towards the upper end of the social ladder, whereas [ɪn] is non-standard and indicative of the speech of people from the lower end of the social ladder.

Such results are quite common in investigations of the distribution of standard and non-standard variants and have led to the conclusion that language is a marker of class membership and social identity (similar to clothing, hair style, leisure activities, etc.). In a nutshell, by speaking in a certain way we express who we are and where we belong (socially and geographically). The marking of social identity can be found on all levels of linguistic description, such as lexicon, phonology, morphology, and syntax. For instance, people vary in socially significant ways in their use of different words for the same thing (e.g. *prof* vs. *professor*), or of different morphological or syntactic constructions (cf. *She don't want no coffee* vs. *She doesn't want any coffee*).

How can this behaviour of people from different social classes be explained? How come that speakers from different classes differ in their usage of the different variants? Or, to put it more bluntly, why don't we all speak alike? To understand this phenomenon, it is instructive to look at a classic sociological study that showed that the social class system restricts communication (Bogart 1950—1). The author investigated how the news about a particular event that was important for the community disseminated in a small rural American prairie town. A local girl won an invitation to visit the New York Philharmonic Orchestra and was interviewed on the national radio during the intermission of the concert. This news was an important event for the whole town and was very much talked about in all kinds of social contexts. Three weeks after the event in question, the researchers conducted 268 interviews and found out that the knowledge of the event had spread unevenly across town. Age, sex and neighbourhood of the interviewees did not make a difference as to their knowledge of the event, i.e. the knowledge of the event had spread equally through all age groups, across both sexes and through all neighbourhoods. However, social class made a difference. Among the lower classes, only 27 percent knew of the event, while it was 77 percent among the

highest social group. As all other factors could be ruled out, the only explanation for this difference was that the social classes were not in constant or close contact. In general, we can thus say that members of a given social class may have — on average — less contact with speakers of other classes than with speakers of their own class. As a consequence, different social groups may come up with differences in their speech, which may first go unnoticed, but which may later be reinterpreted as markers of social identity.

Let us look at another interesting study that will substantiate the point that language can serve as a marker of social identity. Labov (1972) investigated the speech of the people living on Martha's Vineyard, a very popular summer vacation island off the coast of Massachusetts. Labov found that the local pronunciation of certain diphthongs differed remarkably from that of the standard language. For example, on the island one could hear islanders pronounce /aʊ/ as [ʌʊ] and /aɪ/ as [ʌɪ]. In technical-phonetic terms, islanders 'centralised' their diphthongs: [ʌʊ] and [ʌɪ] have more central initial vowels than their standard counterparts [aʊ] and [aɪ] (cf. the vowel and diphthong charts in chapter 1.4.2. for illustration). Labov found, however, that islanders showed a considerable amount of variation in their use of the local variants [ʌʊ] and [ʌɪ]. Upon closer inspection of the two variables (ai) and (au) it turned out that the attitude towards the island and life on the island determined the degree to which a given speaker centralised the diphthongs. Consider the following figure (adapted from Labov 1972):

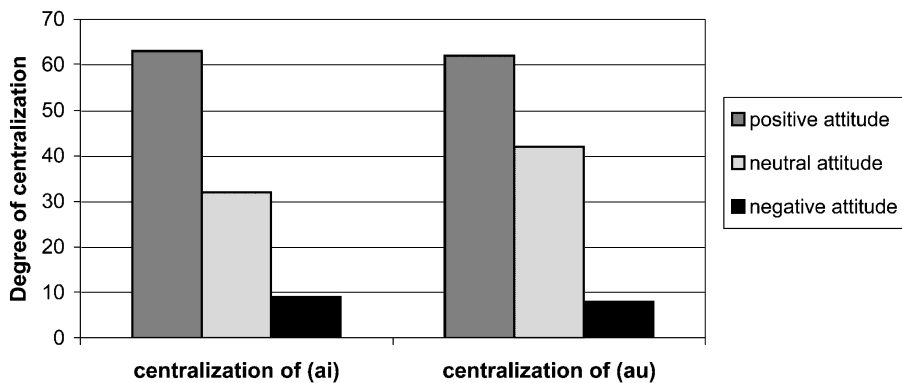


Figure 7.4. The centralisation of diphthongs on Martha's Vineyard

The height of the bars indicates how much and how often the speakers centralised, and for each diphthong the graph gives us three bars, with each bar representing one group of speakers with an either positive, neutral, or negative attitude towards the island, respectively. The graph clearly shows

that a more positive attitude correlates with a higher degree of centralisation. The local, centralised variant was favoured by those islanders who wanted to distance themselves from the incoming tourists and felt a high allegiance to their island and to traditional island life, while the standard pronunciation was favoured by people who did not feel strongly connected with the traditional life style of Vineyarders. In other words, islanders favouring the local pronunciation wanted to stress their identity as native Vineyarders (as opposed to the incoming tourists, who use mostly standard diphthongs).

Again, this example shows how language (or rather: the way we use the options the language offers us) can serve as a badge of identity. In the examples above, we only looked at variability in pronunciation, but this is by no means the only area where language use may vary. People differ according to which words they choose in a given social context, or which syntactic constructions they prefer. Wherever possible, we tend to use those linguistic variants that — apart from conveying the right message — are felt to express certain pertinent and important non-linguistic concepts, such as attitude, group identity, or level of formality.

7.4. Psycholinguistics: how do we store and process language?

The field of study which investigates how human beings store and process language in the brain is called **psycholinguistics**. Psycholinguists investigate, for example, how we discriminate sounds, how we store words, how we produce sentences, how we segment sentences into meaningful units, or how we acquire language. In the following we will focus on one field of inquiry of psycholinguistics, the organisation of words in the mental lexicon. As already mentioned in chapter 5, ‘mental lexicon’ is a metaphor for the repository of words in our mind. Evidence for its structure and configuration can not only be indirectly gathered through experiments, but also through the careful analysis of deviant speech. ‘Deviant’ refers here in a non-technical sense to everyday speech errors, or to the speech of patients with brain damage who suffer from speech and language disorders (so-called ‘aphasics’). In this section we will concentrate on experiments and speech errors.

In chapter 5 you already saw a kind of experiment which showed that the words in the mental lexicon are stored in a systematic way. Association tests demonstrate that semantically related words are more strongly linked to each other than to words that are semantically unrelated.

Another standard experiment to tap the organisation of words in the mind is the so-called ‘lexical decision task’. In such an experiment the subjects see

a word on a computer screen (or hear the word over headphones, or both) and must press a button as quickly as they can, deciding whether the word that they see or hear is a word of their language. If we then measure their reaction times (so-called 'latencies') we find a strong negative correlation between frequency and reaction time. The more frequent the word, the shorter the reaction times. In other words, the more often we read, hear, or use a given word, the easier it is to retrieve that word from our mental lexicon. That this effect also holds for second language speakers is familiar to most learners of foreign languages, who generally have a much harder time to memorise rare words (e.g. *purchase*) than words that they come across or use more frequently (e.g. *buy*). Psycholinguists have created the metaphor of 'resting activation' to account for this phenomenon. The idea is that words are sitting in the lexicon, waiting to be called up, or 'activated', when the speaker wants to use them in speech production or perception. If a word is retrieved at relatively short intervals, it is thought that its activation never completely drops down to zero in between. The remaining activation is called 'resting activation', and this resting activation becomes higher the more often the word is retrieved. As mentioned above, it can be observed that more frequent words are more easily activated by speakers. Such words are therefore said to have a higher resting activation, while less frequent words have a lower resting activation.

Getting deeper into the nature of the frequency effect, scholars have recently detected another factor influencing response latencies, the 'morphological family'. The morphological family of a word is a set of derived words or compounds that contain the same root. For example, *derivation*, *derivational*, *trans-derivational*, *non-derivation*, *derivative*, *zero-derivation*, and *cross-derivation* all belong to the morphological family of the root *derive*. Baayen and Schreuder (1997) have shown that simplex words (such as *derive*) which occur as constituents in many complex words (such as *derivation*, *derivative*, etc.) are processed faster than words with only a few morphological family members. No matter how we would account for that effect, we have to admit that the morphological family must have some psychological reality, and that, therefore, morphological relatedness and morphological structure must somehow play a role in lexical organisation.

When speakers search for a word in their mental lexicon, not only the target word is activated but also phonologically, morphologically, semantically, and syntactically similar words. One experimental paradigm used to test such effects is 'priming', which works as follows. Subjects hear or see a certain word, the so-called 'prime', before they read out, or make a lexical decision on, a different word, the so-called 'target'. If the prime is orthographically, phonologically, morphologically, semantically, or syntactically similar to the

target, subjects tend to be faster in their reading out or their decision on the target word. Thus, a prime that rhymes with a target word (e.g. *mask* — *task* vs. non-rhyming *mask* — *beach*) will facilitate access to the target. The same effect occurs with words that have the same onset in their first syllable (cf. related *train* — *try* vs. unrelated *train* — *fly*). Facilitation of access can also be expected for pairs such as *train* — *car*, in which the semantically related prime facilitates access to the target, while this is not the case with semantically unrelated primes, as in *tree* — *car*.

Based on such evidence we can conclude that the structure of the lexicon can be conceptualised quite adequately in the form of a highly complex network (with many sub-networks) in which similar bits of information are linked to each other. Along the same lines, lexical search (i.e. the search for words when we want to speak) can be modeled as activation spreading through networks of related words. Usually only the target item is (successfully) retrieved from memory when we speak, which means that the activation of the target must have been strongest. Sometimes, however, things go wrong when we try to retrieve words from our memory and of two or more competing words or structures, the wrong one is picked and uttered. Given that we produce speech at an enormous speed (at a regular rate of about six words per second), the occurrence of such mistakes seems rather expectable. In cases where something goes unintentionally wrong in our speech production, we speak of ‘speech errors’. Such speech errors can be an important kind of data for psycholinguists. Let us look at some of them and see what they can tell us about the representation and processing of linguistic structure in the brain. On the left hand side I give the intended utterance, on the right side the one that came out of the speaker’s mouth:

- | | | | |
|-----|------------------------------------|---|--------------------------------|
| (7) | intended | | produced |
| | a. instantaneous/momentary | → | momentaneous |
| | b. [k]lear [b]lue sky | → | [g]lear [p]lue sky |
| | c. Are my tires touching the curb? | → | Are my legs touching the curb? |

In (7a) the two words *instantaneous* and *momentary* are blended into one erroneous word **momentaneous*. What does this tell us about the mental lexicon? First of all, the two words must have been simultaneously activated in order to become blended. A major reason for their being activated simultaneously is of course their close semantic relationship (both refer to time concepts and both involve short periods of time) and their being both adjectives. Second, we see that the morphological structure must have played a role, since the suffix of one word, *-aneous*, has been attached to the stem of the other word, *moment*. This can be taken as evidence for the fact that the words

momentary and *instantaneous* are not only stored as whole words, but that they also have very strong links to their constituent morphemes. Third, morphological blending can be interpreted as evidence for the psychological reality of theoretical constructs such as the stem or the morpheme. In sum, the example shows relatedness effects on the level of meaning, syntax (i.e. word class), and morphology.

The data in (7b) are more subtle. What we see here is the exchange of a phonological feature between two sounds in adjacent words. The voiceless [k] in *clear* becomes voiced in [g]lear, while the voiced [b] in *blue* becomes voiceless in [p]lue. This shows that the pronunciation information, i.e. the phonological representation, of the word is broken down to the level of phonological features, in this case voicing. Only if the speaker has access to these features can he or she confuse them and insert them erroneously into the adjacent word's phonological information.

The final example (7c) again shows semantic and syntactic relatedness. The two competing words are both nouns. In fact, most errors where entire words are exchanged concern words of the same word-class. Furthermore, *tire* and *leg* share important semantic features. They both refer to the lower parts of an object and both are of crucial importance to transportation. Metaphorically, one could even say that the tires are a car's legs.

To summarise, we have seen that different kinds of evidence can be adduced to investigate how humans store and process language, and words in particular. The mental lexicon is a very complex place of storage in which many different kinds of information are linked with each other in a network-like structure. We have to be aware, however, that terms such as 'mental lexicon' and 'network' are metaphors that help us to make sense of what we have not yet fully understood. Much more work needs to be done in order for us to be able to explain in depth how human beings can produce and comprehend language.

7.5. Conclusion

In this chapter we have taken very brief tours into three very different fields of linguistics, namely historical linguistics, sociolinguistics, and psycholinguistics. We have seen that these areas offer fascinating research questions that are tackled using very different methodologies ranging from comparative reconstruction to psycholinguistic experiments. What all three branches of linguistics have in common is that they all involve the application of the fundamental notions of the discipline, such as phonological features, phoneme,

morpheme, suffix, word, etc. Reading this book should have enabled you to further probe into these areas and to ask interesting questions about this everyday thing that is so close to us, and at the same time so hard to investigate: language.

Further reading

Good introductions to historical linguistics are Trask (1997) and Campbell (1998). Pyles and Algeo (1993) is an accessible introductory textbook to the history of English. Another one is van Gelderen (2006), which has a great companion website (<http://www.historyofenglish.net/>). Hogg and Denison (2006) present a concise and authoritative handbook on that topic. For socio-linguistics, you may start with any of the textbooks by Chambers (2003), Holmes (2001), or Meyerhoff (2006). More advanced handbook articles can be found in Chambers (2002). Field (2003) is a text for students with no prior knowledge in the field of psycholinguistics, and Aitchison (2003) provides a nice introduction to the mental lexicon.

Exercises

Basic level

Exercise 7.1.: Comparative reconstruction

Consider the data in (8) and explain the systematic correspondences of the initial consonant between the English words on the one hand, and the Latin and Greek words on the other. Do these correspondences show that English is a daughter language of Latin and Greek?

(8)	English	Latin	Greek
	thin	tenuis	tanu-
	thou	tū	tu
	three	trēs	treis
	foot	pēs	pous
	for	per	peri
	flat	planus	platos

Exercise 7.2.: Language change

English has several native Germanic words that refer to domestic animals. When these animals end up on our plates as food, they are no longer referred to by these names, but by other names. Consider the data in (9):

- (9) pig pork deer venison
 calf veal sheep mutton
 cow beef

Check the etymologies of the words in the right columns in the *Oxford English Dictionary*, and find out what they have in common. Can you think of an explanation for this state of affairs, given your knowledge of the history of England?

Advanced level

Exercise 7.3.: African-American English

In the U.S., many African-Americans speak a variety of English, so-called African-American English, that is quite distinct from the standard. Consider the following table with figures from a study by Ash/Myhill (1986), who investigated the use of a number of non-standard African-American English forms in the speech of whites and of African-Americans in Philadelphia. These non-standard structures involved the absence of third person singular present tense *-s* (as in *she like it*), the absence of a form of BE in certain constructions where it would be required in the standard (such as *he a teacher*), the marking of possession by simple juxtaposition of the NPs denoting the possessor and the possessed (as in *John book*), and the use of the auxiliary *ain't* where the standard has *didn't* (as in *I ain't do it*).

Table 7.1. Percentage of non-standard forms in African-American and white speech in Philadelphia (from Ash/Myhill 1986, 'AfrAm' stands for 'African-American')

	3sg	BE	possessive	auxiliary
examples	<i>she like it</i>	he a teacher	John book	I ain't do it
AfrAm's with low contact with whites	73	52	79	43
AfrAm's with high contact with whites	16	4	15	8
Whites with high contact with AfrAm's	12	8	2	20
Whites with low contact with AfrAm's	0	0	0	0

What correlations can be found between linguistic behaviour and group membership? Describe the patterning of the data and relate these findings to the basic insights of sociolinguistic research described above.

Exercise 7.4.: Speech errors

Consider the following speech errors and discuss their possible origin. What do they say about the storage and processing of words in the mental lexicon, or about the processing of syntactic structure? As above, the intended utterance is given on the left of the arrow, the actual utterance on the line below it.

- (10) a. I'm not really sure / I don't really know →
I'm not really know
- b. As the individual grows older →
As the individual grows more older
- c. I want to thank all the speakers →
I want to spank all the thinkers
- d. This is the grid on the floor where the rat dug →
This is the grid on the floor where the rat digged