Chapter 8

Historical Linguistics

<Start chapter overview>

Overview

In this chapter your will learn about historical linguistics, the subfield of linguistics that studies **language change** and **past language stages.** You will

- acquire an understanding of the development of languages across time
- learn about the changes that occur in their phonology, morphology, syntax, and lexicon
- compare present language stages to past language stages
- learn about the reasons for language change
- gain knowledge of the reconstruction of unattested languages
- study and apply the methods of comparative reconstruction

<End chapter overview>

<H1> 8.1 What is historical linguistics?

Historical linguistics is the field that studies past language stages ("dead languages") and language change over time. Languages are constantly changing, and historical linguists study how and why this happens.

- Historical linguists study languages that are no longer spoken, like Old English or Old French
- Historical linguists also study changes between two language stages, for example, the changes from Old English to Middle English and from Middle English to Modern English
- Historical linguists moreover **reconstruct** even earlier language stages based on ***attested languages***

You have probably noticed that the variety of English you speak (your personal **idiolect**) is subtly different from the English of your grandparents. Over time, such miniscule changes accumulate and lead to quite profound differences between past and present stages of a language – just compare your idiolect to the English of William Shakespeare around 1600 CE, or to that of the Old English poem *Beowulf* (ca. 800 CE).

Languages change constantly, and there is nothing we can do about it. Yet there is a common notion that language change is somehow inherently "bad", a sign of deterioration or just general sloppiness on the

part of “lazy” or “uneducated” speakers or “bad parents”. Unfortunate and misguided as these notions are, they go back quite a long way. For example, the Latin *Appendix Probi* from the 3rd century AD contains a list of “wrong” pronunciation of Latin words, and the “correct” way of pronouncing them. Here are some entries from this list:

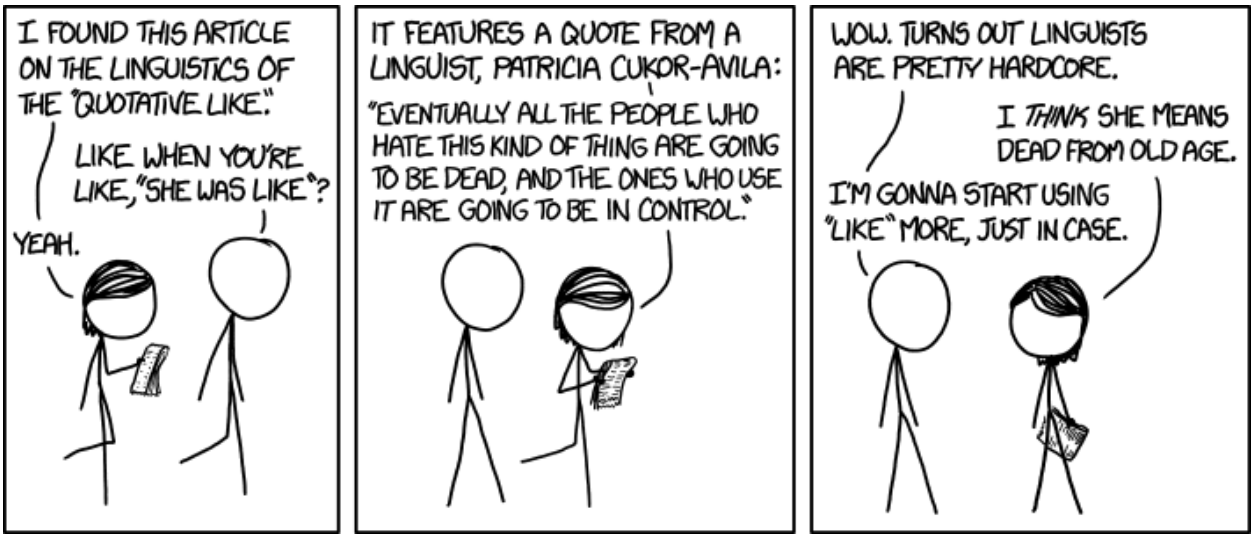
1) Appendix Probi
    a. *speculum non speclum* “say *speculum*, not *speclum*” (Italian *specchio* “mirror”)
    b. *calida non calda* “say *calida*, not *calda*” (Italian *caldo* “hot”)
    c. *auris non oricla* “say *auris*, not *oricla*” (French *oreille* “ear”)

In the third century AD, Latin was well on the way to developing into the Romance languages (Italian, French, Spanish, and many others). But the *Appendix Probi* proves that some speakers were quite unhappy about this development and tried to correct their contemporaries’ use of the Latin language: “Don’t say *speclum*, say *speculum*!”. They had a **prescriptive** approach to the use of the Latin language, whereas historical linguists prefer a **descriptive** approach. That is, as historical linguists we want to *describe* how speakers use language and how their grammars change over time, rather than forcing them to use an arbitrary “correct” standard language (you will learn more about descriptivism vs. prescriptivism in language in chapter 11 on sociolinguistics).

The Modern Romance languages moreover show that the speakers who said “*speclum*” eventually won out over the speakers who pronounced the word as “*speculum*”, as the latter would have developed into Italian **specolo*, rather than the actual Italian word *specchio* “mirror” (the asterisk “*” indicates that a word is reconstructed rather than attested in spoken or written form). Language change is inevitable, and historical linguists study its pathways and mechanisms.

<Start cartoon>

Don’t try to stop language change



Source: https://xkcd.com/1483/

<end cartoon>

<Begin Pause and reflect box>

Think about the way your first language(s) differ(s) from that/those of your friends, parents and grandparents. What differences in its phonology, morphology, syntax and lexicon can you think of? What differences do you notice between formal and informal/colloquial speech, and between different regional varieties?

<End box>

<H1> 8.2 Language change

<H2> 8.2.1 I-language and E-language

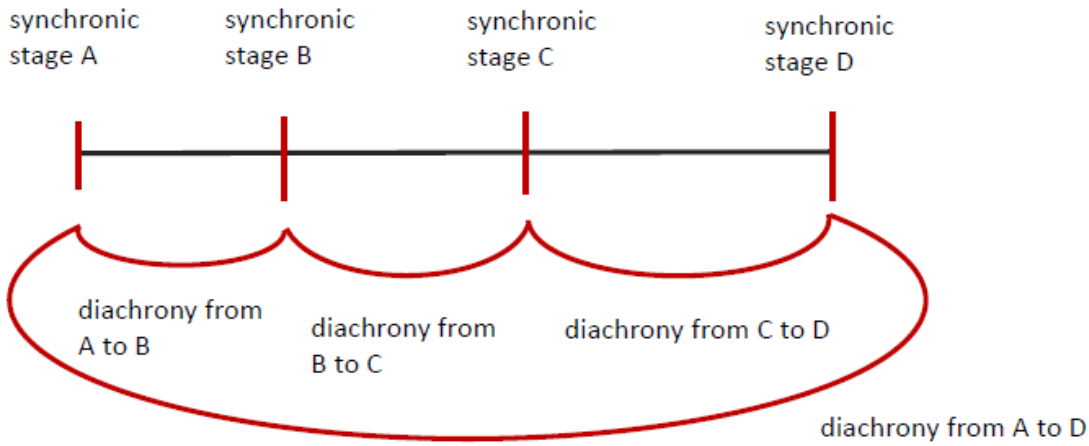
We have observed that languages change over time and that the English or French spoken in Montreal nowadays is very different from the English of Shakespeare or the French of Voltaire. But notions like "the English language" or "the French language" are sociopolitical concepts that characterise the linguistic conventions and language use of different speech communities (see chapter 11). Linguists refer to these entities as **E-language**, where "E" stands for *external* to the individual. On the other hand, you have already learned in chapter 1 that theoretical linguists refer to the knowledge of language or **mental grammar** of an individual as **I-language**, where "I" stands for the *internal* **knowledge state** of an individual. Language in this sense is part of the cognitive capacities of an individual's brain.

When we talk about language change, we must keep in mind that "language" can refer to these two very distinct notions—the habits of a speech community, and the knowledge state in the brains of each member of that speech community. I-language and E-language interact in complex ways over time. In the following, we will use "language" in the I-language sense—the mental grammar of an individual.

<H2> 8.2.2 Synchrony and diachrony

A core concern in historical linguistics is how languages develop *over time*. This is known as **diachronic** approach (Greek *diá* 'through, across', *khrónos* 'time'). Diachronic linguistics studies the changes in speakers' grammars across time (between a given stage X and stage Y), the **diachrony** of a language is its development over time. **Synchronic** linguistics, on the other hand, studies the properties of speakers' grammars at a given linguistic stage X (Greek *sýn* 'with, together' + *khrónos* 'time'). Historical linguists often work on the interaction of synchrony and diachrony in any given language. These two ways of studying language are illustrated in figure 8.1.

<Start Figure>



<Caption> Figure 8.1 Synchrony and diachrony

<End Figure>

In this illustration, each vertical bar stands for a synchronic language stage, called A, B, C, and D. You can think of these as the mental grammars of four different individuals: A speaks (a variety of) Old English, B speaks Middle English, C speaks early Modern English, and D speaks contemporary Toronto English. Under a synchronic approach to these four stages, a linguist would study the properties of each of these four stages without referring to the stages that came before or after. On the other hand, the arcs between the

vertical bars illustrate the diachrony between two given stages. Under a diachronic approach, a linguist would study the changes between A and B (Old English and Middle English), or between B and C (Middle English to Early Modern English), or between A and D (Old English to Toronto English). Of course, such a diachronic approach presupposes an understanding of the relevant synchronic properties of the two stages under study and of the intermediate stages.

<H.2> 8.2.3 Correspondence between grammars

How do we decide whether two grammars ("languages") are diachronically related? We determine whether there are correspondences between the two grammars.

- **Correspondence**: A linguistic form X (a phoneme, morpheme or word) in Grammar G_1 diachronically corresponds to a linguistic form Y in Grammar G_2 if X can be related to Y through one or more changes which transformed X into Y.

The following example illustrates a simple diachronic correspondence:

2) Old English *mūs* /mu:s/ > Modern English /maws/ 'mouse'

The symbol '>' means that a form X in G_1 (in this case Old English) changes into/turns into a form Y in G_2 (in this case Modern English). As a result of this change, the Modern English word 'mouse' corresponds to the Old English word *mūs* (note that in historical linguistics, it is customary to use the macron sign to indicate a long vowel, therefore *ū* = /u:/, etc.).

The two forms are related through a change that transformed the Old English long high back vowel /u:/ into the Modern English diphthong /aw/ (the other sounds stayed the same). This change was regular, it happened to all Old English long high back vowels on the way to Modern English. By observing regular correspondences such as this one, we establish that Modern English is in a **lineal descent** relationship with Old English. Lineal descent means that the grammar G_2 (Modern English) corresponds diachronically to the grammar of G_1 (Old English) through a series of intermediate changes. The following figure illustrates a simplified lineal descent relationship (based on Hale 2007: 28).

<Start Figure>

<Caption> Figure 8.2 Lineal descent relationship between G_1 and G_2

<End figure>

There are also **non-lineal descent** relationships, which we'll explore in section 8.8.

<H2> 8.2.4 Change and diffusion

Now that we have established how grammars are related over time, we can sharpen our definition of "language change". Remember that G_1 and G_2 represent two different I-languages which can be related across time through a series of intermediate steps. These intermediate steps happen piecemeal, as grammars are transmitted from one generation to the next. Change means that the transmission of a given grammar during language acquisition is flawed with respect to some feature of that grammar. In other words, language change is **imperfect transmission**.

- **Imperfect transmission**: "Change results when transmission is flawed with respect to some feature. When transmission is not flawed (with respect to some feature), there has been no change in the strict sense." (Hale 2007: 36)

Under this view, language change happens during **language acquisition** when language learners acquire a grammar G_2 that is different from the input grammar G_1 with respect to some feature(s) (you will learn more about language acquisition in chapter 9). For our purposes, it suffices to know that imperfect transmission happens when language learners misanalyse some aspect(s) of the input data they receive and end up constructing a slightly different grammar than the input grammar(s).

This view of language change leads to a paradox: If change happens piecemeal and grammar by grammar through "transmission errors" during language acquisition, why do we have the impression that language change affects whole speech communities? After all, the changes from Old English to Middle English to Modern English (and their respective varieties) affected large speech communities, not just individuals. The solution proposed by Hale and others is to distinguish between **change** and **diffusion**. Change means that a language learner maps a string of input (Primary Linguistic Data, PLD) to a grammar G_2 that is

different from the grammar G_1 that generated the input data. This is misanalysis or **imperfect transmission**. Diffusion, on the other hand, is the spread of a change in an individual speaker's grammar through a linguistic community. That is, change is flawed transmission, diffusion is successful transmission throughout a community. Another way of thinking about this is the distinction between I-language and E-language discussed above. Change affects the I-language of an individual, while diffusion is an E-language phenomenon that affects a linguistic community. Diffusion can and must be studied using the tools of sociolinguistic analysis (see chapter 11), while change in the narrow sense must be studied with the analytical tools used in theoretical linguistics to understand language as a faculty of the human brain.

So, what do historical linguists study? Change or diffusion? The answer must be: both. When we deal with past language stages, we only have access to features of grammars that were successfully transmitted and are attested in one way or another. Grammars (or features thereof) that are unattested (either directly or through **daughter languages**) cannot, by definition, be studied. Our knowledge of change (in the I-language sense) therefore always depends on successful transmission (diffusion). However, we should be careful to distinguish between the two, as they depend on fundamentally different mechanisms.

<Begin Pause and Reflect box>

Some varieties of Canadian English are currently undergoing a "change" called the Canadian Shift, in which the lax vowels in words like *bet* /bɛt/ and *bat* /bæt/ are retracted towards the back of the mouth and/or lowered (depending on the variety), so that *bet* either sounds closer to (Standard English) *but or bat*, and *bat* sounds closer to *bought* (Labov et al. 2006). In some varieties, the lax vowel in words like *bit* /bɪt/ is also affected, so that *bit* sounds more like *bet*.

- Pay attention to how your friends and relatives pronounce these words. Which "shifts" do you detect?
- How could we distinguish between change and diffusion in this shift? Hint: are your informants consistent in their shifts? Did they move around the country a lot? Where did they live, and for how long?

<End box>

<H2> 8.2.5 What can change?

We have defined language change from the I-language perspective as change in speakers' grammars. Language change in this sense is usually divided into four separate domains, corresponding to the domains of grammar that are generally also used for synchronic linguistic analysis:

- **Phonological change** or **sound change** affects the phonological system of a given language. An example is the change from Old English /m**u:**s/ to Modern English /m**aw**s/ 'mouse', in which a long vowel was turned into a diphthong. This change was **regular**: it applied to all instances of the affected sound (i.e., Old English /h**u:**s/ became Modern English /h**aw**s/ 'house', etc.). The precise mechanisms of this will be discussed in section 8.3.

- **Morphological change** alters a language's inventory of functional items, usually inflectional and derivational morphemes. For example, the Modern English adjectival/adverbial suffix *-ly* (in *friend-ly*, *easi-ly*, etc.) is related to Modern English *like* and the Old English noun *līc* 'body' (probably originally 'shape'). In compounds like Old English *freond-līce* 'friend-like, friend-shaped' > Modern English *friend-ly*, this noun developed into a suffix with adjectival or adverbial meaning. We will discuss morphological change in section 8.4.

- **Syntactic change** primarily affects the word order and distribution of functional and lexical items in a given language. Common syntactic changes include changes in the distribution and movement properties of verbs, negation markers, interrogatives ("*wh*-words") and pronouns. Syntactic change is discussed in section 8.5.

- **Semantic change** falls into two broad categories: 1) changes in the semantics of **lexical categories**, such as the change in meaning of English *dog*, which used to have a more specific meaning than it does nowadays (it used to refer to a specific breed of dog, cp. German *Dogge* 'mastiff'), and 2) changes in **functional categories** like personal and demonstrative pronouns, modal verbs and adverbs. For example, third person pronouns such as English "he", "she", and "they" have almost always developed diachronically from anaphoric demonstratives ("this one" or "that one"). Both types of semantic change are discussed in section 8.6.

Keep in mind that these different domains of language change tend to interact (as you have probably noticed in the examples above). While it is useful to discuss these domains separately, most research in

historical linguistics spans two or more of these domains for any given topic, making it a truly multidisciplinary field!

<H1> 8.3 Phonological change

<H2> 8.3.1. Sound laws and the Neogrammarian Hypothesis

In studying the changes in the phonological systems between different language stages, historical linguists rely on the **Neogrammarian hypothesis**: the observation that sound change is **regular** and **exceptionless**.

3) The Neogrammarian hypothesis: "Sound laws suffer no exceptions"

<Begin Linguistic tidbits box>

Linguistic tidbits: The Neogrammarians

The Neogrammarian hypothesis is named after a group of 19th century linguists from Germany who held what was considered quite a radical view at the time: that sound change follows general and law-like principles similar to the general laws of physics. These scholars were young(ish) at the time, and the German term *Junggrammatiker* translates more accurately as "young grammarians".

<End box>

What is a **sound law**? It is a phonological change that occurs between two grammars which are in a lineal descent relationship (or, more informally, "two stages of a given language"). Sound laws can be described using phonological rules like the ones you encountered in chapter 3 on phonology, for example:

4) A template for sound laws:
    a. A > B
       "A turns into B" or "A becomes B"
    b. A > B/_C
       "A turns into B before C"

We have already seen a sound law like (4a) above: The change of Old English /u:/ to Modern English /aw/. This was an across-the-board change: it happened to all instances of Old English /u:/, independent of the environment they were in. That is, this change did not depend on a particular phonetic context. Moreover, this change was part of a broader change that affected the entire vowel system of English in the Middle English period (ca. 1300-1500). This change is known as the **Great Vowel Shift**. The stage of this shift that

turned the long high vowels /u:/ and /i:/ into the diphthongs /aw/ and /aj/ is called **diphthongisation**, illustrated in the following table.

<Start table>

Middle English	Modern English	meaning	Middle English	Modern English	meaning
/mu:s/	/maws/	'mouse'	/mi:s/	/majs/	'mice'
/hu:s/	/haws/	'house'	/ri:də/	/rajd/	'ride'
/hlu:d/	/lawd/	'loud'	/ti:mə/	/tajm/	'time'
/ku:/	/kaw/	'cow'	/hwi:t/	/wajt/	'white'

Table 8.1: Diphthongisation in the history of English

<End table>

<Begin pause and reflect box>

Look at the Middle English words below. Based on the sound change rule you just learned (diphthongisation of high vowels), what are the predicted Modern English versions of these words after the Great Vowel Shift took place? Note that final *-ə* is also lost on the way to Modern English.

- /bi:tə/
- /pru:d/
- /lu:s/
- /li:k/
- /u:t/

<End box>

While this sound law applied independently of a particular context, most sound changes are context-sensitive: they apply in a particular phonetic context only. Rule (4b) above illustrates the format of such sound laws. A common instance of context-sensitive sound changes are **palatalisation** changes, in which non-palatal consonants (usually velar or dental plosives) develop a "front" (palatal) co-articulation or affrication before front vowels. Such a change happened in Old English: the Old English velars /k/ and /g/ became the postalveolar affricate /ʧ/ and the palatal glide /j/, respectively, before front vowels. Before back vowels and consonants, these sounds did not change. The effects of this change are still seen in

Modern English—it's the reason you say ***c**ool* and ***c**limb* with a /k/ and ***g**lad* with a /g/, but ***ch**urch* and **ch**eek with a /ʧ/ and ***y**ard* with a /j/. The following table illustrates the difference between palatalised and non-palatalised velars in Modern English and their early Old English predecessors (before palatalisation took place).

<Begin table>

Palatalisation		No palatalisation	
Early Old English	Modern English	Early Old English	Modern English
cirice /**k**irike/	*church* /**ʧ**ərʧ/	*climban* /**k**limban/, later /**k**l:imban/	*climb* /**k**lajm/
cīdan /**k**i:dan/	*chide* /**ʧ**ajd/	*cōl* /**k**o:l/	*cool* /**k**u:l/
geard /**g**eard/	*yard* /**j**ard/	*gōs* /**g**o:s/	*goose* /**g**u:s/
geolu /**g**eolu/	*yellow* /**j**ɛlo/	*glæd* /**g**læd/	*glad* /**g**læd/

Table 8.2 Palatalisation in the history of English

<End table>

The velars in the Old English column on the left are before front vowels (Old English /ea/ and /eo/ are diphthongs whose first element is [+front]), while the velars in the Old English column on the right are before back vowels or before consonants. Old English /k/, /g/ correspond to Modern English /ʧ/, /j/ before front vowels, but to /k/, /g/ everywhere else (this is called an **overlapping correspondence set**, see section 8.8.4.). We can now formalise this as a diachronic rule of the type A > B/_C:

5) Old English *k*, *g* > *ʧ*, *j*/_V[+front]

Note that the **phonetic environment** is crucial to the operation of this rule—if the velars are not before a front vowel, the rule does not apply. A more abstract way of writing a palatalisation rule is the following:

6) C[+velar] > C[+palatal]/_V[+front] "a velar consonant becomes palatal before a front vowel"

This rule is very general—as the Old English case shows, „palatalisation" affects different sounds in different ways (e.g., /k/ becomes an affricate, but /g/ becomes a glide).

<Begin pause and reflect box>

Look at the early Old English words below. Based on the sound change rules you just learned (palatalisation, diphthongisation), what are the predicted Modern English versions of these words? (Assume that the final *-an* of the first two words is lost; further hints are given below).

- /ke:osan/ (the sequence *e:o* developed into Modern English /u/, /s/ became /z/)
- /gieldan/ (the sequence *ie* developed into Modern English /i/)
- /gold/
- /kɪn/
- /ki:ld/

<End box>

While the template for sound laws we used above is similar to that used by phonologists for synchronic phonological rules, it is important to note that our *diachronic* sound change rules have a different status than synchronic phonological rules. Synchronic rules (e.g., "oral vowels are pronounced as nasal vowels before a nasal stop" in English) are part of a speaker's mental grammar. Sound laws, or diachronic phonological rules, describe the start and end points of a change, or sequence of changes, *between* two different grammars at two different diachronic stages.

<H2> 8.3.2. Rule ordering

An important observation in studying sound change is that sound laws are *ordered* with respect to one another. Consider our English sound laws above, diphthongisation and palatalisation. Dipthongisation turns the high front vowel /i:/ into the diphthong /aj/, whose first element /a/ is [-front] (it also turns /u:/ into /aw/, but this is not relevant to the palatalisation rule). Palatalisation, on the other hand, takes place before [+front] vowels, but not before [-front] vowels. We can now determine whether palatalisation happened before or after diphthongisation. If it happened *before* diphthongisation, we'd expect velars to be palatalised before an Old English /i:/ before that /i:/ turns into /aj/. If palatalisation happened *after* diphthongisation, we would not expect velars to become palatalised before an Old English /i:/, since this sound would turn into /aj/ before the palatalisation rule took place. Table 8.3. illustrates these two options for the Old English word *cīdan* /ki:dan/ 'chide':

<Start table>

Ordering a)	1. Palatalisation	2. Diphthongisation	Expected Mod.Engl.
Input: /**ki:**dan/	/**ʧi:**dan/	/**ʧaj**dan/	/**ʧaj**d/ ✓
Ordering b)	1. Diphthongisation	2. Palatalisation	
Input: /**ki:**dan/	/**kaj**dan/	n/a	/**kaj**d/ ✗

Table 8.3. Rule ordering

<End table>

In the first ordering (odering a), palatalisation takes place *before* diphthongisation, so we get /ki:dan/ > /ʧi:dan/ > /ʧajdan/. Since both palatalisation and diphthongisation take place, we expect the Modern English word to be /ʧajd/, and this is correct (*chide*; the Old English ending *-an* is lost). What happens if we change the order, as in ordering b) above? If the input is the same and diphthongisation happens first, we get /ki:dan/ > /kajdan/. So far, so good, but now we cannot apply the second rule, palatalisation. Palatalisation happens before front vowels, but the initial /k/ of /kajdan/ is now before /a/, which is not a front vowel. Since the environment for palatalisation is no longer there, the palatalisation rule is not applicable in this ordering. The expected Modern English form is /kajd/ (which would probably be spelled *kide*), which does not exist. We have shown that the ordering must have been ordering a), with palatalisation happening before diphthongisation. The ordering of two changes with respect to each other is called their **relative chronology** (relative to each other).

<Begin pause and reflect box>

Consider the following data from Hawaiian and its ancestor (the language from which it developed), Proto-Polynesian (data from Hale & Kissock 2013):

Proto-Polynesian	Hawaiian	English
ʔau	au	‘current’
peka	peʔa	‘bat’
ʔumu	umu	‘oven, earthen’
ika	iʔa	‘fish’
waʔe	wae	‘leg’

- Which sound changes took place from Proto-Polynesian to Hawaiian?
- How are these sound changes ordered with respect to one another? Give a rule ordering/relative chronology and explain why it's the correct one.

<end box>

<H2> 8.3.3 Why is sound change regular?

The Neogrammarian hypothesis is one of the fundamental insights of historical linguistics, without which we would not be able to posit sound laws and work on **comparative reconstruction**. But why *is* sound change regular?

We have seen above that sound changes usually depend on a particular phonetic environment (there are also across-the-board changes, which are trickier). The fact that sound change is regular is precisely *because* it occurs only in a particular phonetic environment. That is, it is *conditioned* by some phonetic property of that environment. Consider the palatalisation of Old English /g/ discussed above. /g/ is a voiced velar plosive, but before front vowels, it became a palatal glide, /j/. Why did this happen? /j/ is pronounced closer to the front of the mouth than /g/, it is more "front". With respect to the manner of articulation, a glide is fairly similar to a front vowel in terms of the closure of the articulators. What happened, therefore, is that the plosive /g/ **assimilated** to the following front vowel by taking on or adapting to some of its phonetic properties, notably place and manner of articulation. This is an **anticipatory** change, since the sound that undergoes the change seems to anticipate some of the properties of the following sound. More precisely, the change happens because language learners misanalyse the [+front] feature of the vowel as belonging to the preceding consonant.

If sound change is phonetically conditioned, this explains why it is regular and exceptionless: it occurs because the sounds undergoing the change pick up some feature(s) of the conditioning environment (or rather, because language learners unconsciously assign some feature of that environment to the sound in question). If the conditioning environment is absent, the change does not occur. For example, in the word *cool* /kul/, the sound following the velar is [+back], rather than [+front]. We therefore do not expect to see the fronting of the velar associated with palatalisation, since there is nothing in the phonetic environment that would trigger such a fronting. This explains why the velars in the right column of table 8.3. above stayed the same in Old English while the ones in the left column became more "front" in their articulation.

<Begin tidbits box>

Linguistic tidbits: Taboo! Exceptions to the Neogrammarian hypothesis?

Are there exceptions to the regularity of sound change? **Taboo words** are often exceptional: they are words that speakers avoid or consciously modify because there is a social inhibition or constraint on their use, e.g., words for body parts (especially genitalia), swearwords, dangerous animals, and deities. For example, you might say "shoot" in public instead of… you know. This is not a sound change because it is not conditioned by the misanalysis of a *phonetic* environment, and therefore not an exception to the Neogrammarian hypothesis.

<End box>

<H2> 8.3.4 Common types of sound changes

In this section, some common types of sound changes are briefly introduced.

- **Assimilation**: a segment takes on some or all phonetic features of a preceding or following segment (it becomes "more similar" to a preceding or following segment)

Assimilation is an umbrella term for several types of changes which involve "feature spreading". Palatalisation, for example, is a type of assimilation: a consonant takes on the phonetic feature [+front] from the following segment (a front vowel). Another common type of assimilation is consonant assimilation in clusters: when two consonants are adjacent to each other, they tend to assimilate in voicing, place, or manner of articulation.

7) a. Latin *se**pt**em* 'seven' > Italian *se**tt**e*

If a sound assimilates to a following sound, like in (7), it's called a **regressive** (or **anticipatory**) assimilation. Assimilation to a preceding sound is called **progressive** assimilation.

Another common type of assimilation is nasalisation, which is usually regressive.

- **Nasalisation**: an oral vowel becomes nasalised before (or, more rarely, after) a nasal stop

Nasalisation took place in the history of French. In Old French, vowels were nasalised before nasal stops:

8) a. Latin *bonus* 'good' > French *bon* /bɔ̃/
   b. Latin *līnum* 'flax' > French *lin* /lɛ̃/ (contrast this vowel with the non-nasal vowel in *laid* /lɛ/ 'ugly')

While the French writing system obscures this change a bit, the nasal stop has been lost completely in the words in (8), and only the nasalisation on the preceding vowel tells us that it was once there.

<Linguistic tidbits>

Linguistic tidbits: Sound change and orthography

Orthographic conventions aren't always sensitive to sound changes that have taken place: English orthography reflects the way words were pronounced around 1400 CE! For example, the spelling of the word *chide* shows the linguistic stage of English *after* palatalisation took place (hence the *ch-* at the beginning instead of *c-* or *k-*), but *before* diphthongisation (hence *-i-*) and *before* word-final vowels were lost (you don't pronounce the *-e*). Irksome though they are, conservative orthographic conventions are of great value to historical linguists!

<End box>

Other types of sound change involve the loss of a segment. The loss of a segment in the middle of the word is called **syncope** (Greek *synkopē* 'cutting short'). The loss of a segment at the end of a word is called **apocope** (Greek *apokopē* 'cutting off').

- **Syncope**: a segment (usually an unstressed or weakly stressed vowel) is lost in the middle of a word

Syncope has applied in English words like *family* and *laboratory*, which are usually pronounced / fæmli/ (two syllables) and /læbrətɔri/ (four syllables), respectively. Syncope also occurred in many of the Romance languages as they developed from Latin:

9) Latin *po**pul**us* 'people' > French *peu**pl**e*, Spanish *pue**bl**o* (but: Italian *po**pol**o*)

- **Apocope**: a segment (usually an unstressed or weakly stressed vowel) is lost at the end of a word

We have already seen that apocope occurred in the history of English, where final consonants and unstressed final vowels were lost, for example in Old English *climb**an*** (two syllables) > Modern English *climb* (one syllable), or *mōn**a*** (two syllables) > *moon* (one syllable). Apocope also occurred in the development of the Modern Indic languages such as Hindi, Bengali and Gujarati from Sanskrit (via Middle Indic).

10) a. Sanskrit *e:k**ah*** 'one' > Hindi *e:k*

b. Sanskrit *sapt**a*** ‘seven’ > Hindi *sa:t*

Syncope and apocope usually target vowels. When a consonant is lost, on the other hand, a preceding vowel is often lengthened, as if to „compensate“ for the loss of a segment. This is called compensatory lengthening.

- **Compensatory lengthening**: a vowel is lengthened when a following consonant is lost

Compensatory lengthening occurred in those varieties of English that lost /r/ in codas (so called “non-rhotic” varieties of English). In those varieties, *farm* is pronounced as /fa:m/ and *car* is pronounced as /ka:/. You may have noticed that compensatory lengthening also took place in the Hindi example in (10b) above. Here are more examples from Hindi:

11) a. Sanskrit *bh**ak**tam* ‘cooked rice’ > Hindi *bh**a:**t*
    b. Sanskrit *p**aɲ**ʧa* ‘five’ > Hindi *p**ã:**ʧ***

Note that the nasal in (11b) triggered **nasalisation** before it was lost with **compensatory lengthening**!

- **Merger**: a phonemic contrast between two segments is lost and they merge as one and the same segment

In some varieties of North American English, the vowels /ɛ/ and /ɪ/ merge before nasals, so that the words *pen* and *pin* have the same vowel. Consonants can also merge: in Sanskrit, the contrast between /r/ and /l/ was given up and the two segments were merged as /r/:

12) a. Proto-Indo-European **sa**l*** ‘run, jump’ > Sanskrit *sa**r***
    b. Proto-Indo-European **b^{h}e**r*** ‘carry’ > Sanskrit *bha**r***

Note that the vowels /a/ and /e/ also merge in the example above!

- **Monophthongisation**: a diphthong becomes a monophthong

We’ve already encountered diphthongisation, by which a monophthong (usually a long vowel) becomes a diphthong, like in Old English /m**u:**s/ > Modern English /m**aw**s/. The reverse happens in monophthongisation, for example in Modern Greek:

13) a. Classical Greek *grap^{h}**ei*** ‘he/she writes’ > Modern Greek *graf**i***
    b. Classical Greek *erk^{h}om**ai*** ‘I come, go’ > Modern Greek *erxom**e***

<Begin Pause and Reflect box>

Analyse and classify the sound changes you observe in the following examples (some may not fall under any of the categories discussed above). Note that in most words more than one change has taken place!

- Proto-Semitic **damiqum* 'good' > Akkadian *damqum*
- Latin *fabulāre* 'to speak' > Spanish *hablar* /ablar/
- Latin *aurum* 'gold' > French /ɔr/
- Sanskrit *agni* 'fire' > Pāli *aggi*
- Proto-Algonquian **eθkwe-* 'woman' > Ojibwe *ikkwe-*

<End box>

<H1> 8.4 Morphological change

<H2> 8.4.1. Types of morphological change

Morphological change affects the morphological categories of a given language, in particular its inflectional and derivational morphology. We can distinguish between changes that result in the *loss* of morphological material (e.g., case endings on nouns or agreement markers on verbs) and changes that result in the *development* of new morphological material. Another useful distinction, introduced by Andersen (1973), is between **deductive change** (rule extension) and **abductive change** (creation of a new morphological rule).

<H2> 8.4.2. Deductive change

Deductive change means extending an already existing morphological rule or pattern to an environment where it did not previously apply. **Analogy** is a very common form of deductive change and involves the generalisation of a formal relationship from one form (or set of forms) to another form (or set of forms). Simply put, analogy makes words more similar to one another. **Proportional analogy** copies a relationship between one set of forms to another set of forms and can be formalised as follows (based on Campbell 1998):

- Proportional analogy: **a : b = c : x**
  "a is to b as c is to x, where x is ..." (x = the new, **analogical form**)

In proportional analogy, a part (or all) of the relation between a and b (in terms of phonological and/or morphological features) is copied to the relation between c and x, where c shares some salient features

with a. For example, the past tense of English *dive* used to be *dived*, but this was replaced by *dove* in analogy with past tense forms like *strove*, *drove*, etc., through proportional analogy:

<Begin table>

Present		Past
strive	:	*strove*
ride	:	*rode*
drive	:	*drove*
dive	:	x, x = *dove*

Table 8.4. Proportional analogy

<End table>

Note that this type of analogy extends *unproductive* morphology to new contexts (the synchronically productive way of forming the simple past tense is by adding the suffix /-(ə)d/ to a verb).

Another morphological change that is usually analysed as analogy extends *productive* morphology to a new context. This happened to the plural of the word *cow*. In Old English, the singular was *cū* /ku:/, while the plural was *cȳ* /ki:/ (this form lives on in Scottish Engl. *kye* /kaj/). This plural was later changed to *kīne /ki:nə/* in Middle English by adding the plural marker *-n(e)* (an archaic plural ending also seen in *childr-en* and *ox-en*), so that the Early Middle English pair was /ku:/, pl. /ki:ne/. After the Great Vowel Shift, these forms should have become Modern English /kaw/, /kajn/ (compare table 8.1. above for the development of long high vowels during the Great Vowel Shift). Instead, Modern English has /kaw/, pl. /kaw-z/. We already know that /ki:ne/ cannot have turned into /kawz/ by regular sound change: Early Middle English /i:/ should have turned into /aj/, not /aw/, and Middle English /n/ did not become Modern English /z/. We therefore need a morphological explanation. First, we need to understand the status of regular plural forms such as *cows*, *dogs*, *hats*, *linguists* vs. irregular plural forms such as *feet*, *mice*, *children*, *cacti*. The latter need to be stored in your mental **lexicon**, since they are not derivable by a productive morphological rule (see chapter four on morphology). The former, on the other hand, do not need to be stored in the lexicon. They can be generated by applying the productive rule of plural formation to any singular noun—even nouns that you've never heard before, like *zlorp* or *wug*, whose plural forms must be *zlorps* /zlorps/ and *wugs* /wəgz/. So how do we lose irregular plurals like *cacti* or Middle English /ki:ne/? All that needs to happen is that speakers fail to store the irregular plural form in their mental lexicon. This may happen

because they never hear them pronounced, or because they don't hear them often enough, or because they assign them a new meaning (so that *kīne* comes to mean 'cattle', for example). Now, a speaker who has never heard the rare and highly irregular plural *cacti* will probably pronounce the plural of *cactus* as *cactuses* (which is what many English speakers actually say). The same seems to have happened to Middle English /ki:ne/, which was replaced by the regular form /kawz/ because English speakers at some point failed to store the irregular form, for one of the reasons mentioned above.

While this kind of replacement of an unproductive pattern with a productive pattern is usually called analogy, it does not depend on a relation of similarity between two forms in the way that the proportional analogies discussed above do. All that needs to happen is for speakers to lose (or never acquire) the irregular form. The regular form then comes for free.

<Begin Pause and reflect box>

- In some varieties of Quebec French, the plural of *cheval* /ʃval/ 'horse' is /ʃval/ 'horses' (as if it were spelled *chevals*) instead of the irregular plural *chevaux* /ʃvo/. What kind of morphological change is this and why do you think it happened?
- Form the English plurals of the words *octopus* and *rhinoceros* and ask your friends to do the same. How many variants do you get? Why is the plural of these words difficult?

<End box>

<H2> 8.4.3. Abductive change

Abductive change means creating a new (morphological) rule based on a misanalysis of the available data during language acquisition. Like in section 8.3. above, "misanalysis" means that language learners interpret a given pattern in their input differently than the previous generation, resulting in a slightly different grammar than that of the previous generation. The changes usually subsumed under the term **grammaticalisation** can be characterised as abductive changes. Grammaticalisation means that lexical items (nouns, verbs, adjectives) develop into functional items (for example, classifiers, determiners, auxiliaries, modal verbs, etc.); while functional syntactic categories (auxiliaries, determiners, complementisers, etc.) tend to become reduced to affixal material.

- **Grammaticalisation** (Hopper and Traugott 2003: xv): "the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalised, continue to develop new grammatical functions."

In other words, grammaticalisation makes lexical items more abstract and functional and thereby creates new functional items, especially inflectional and derivational morphology—as the linguist Talmy Givón put it, "today's morphology is yesterday's syntax".

Grammaticalisation usually involves several steps:

- **Phonological reduction** ("weakening"): functional elements are weakly stressed or unstressed compared to lexical elements. If a lexical element becomes unstressed, this can lead to its **reanalysis** as a functional element and further phonological reduction.
- **Semantic bleaching**: a lexical item loses part or all of its meaning (functional categories usually have more abstract and restricted meaning than lexical categories)
- **Loss of syntactic freedom, cliticisation**: functional elements are highly restricted in their distribution and are often clitics (unstressed elements which are dependent on a preceding or following stressed word) or affixes. If a lexical item undergoes grammaticalisation, it usually becomes syntactically restricted to certain environments.

One example of grammaticalisation is the development of the Modern English adjectival/adverbial suffix *-ly* (in *friend-ly*, *easi-ly*, etc.) from the Old English noun *līc* 'body' (< Proto-Germanic **līk(a)-* 'body, shape'). This noun lost phonological content (it underwent phonological reduction), semantic content, and syntactic freedom on its way to developing into an adverbial affix. This development started in compounds like Old English *freond-līce* 'friend-like, friend-shaped' > Modern English *friend-ly* through a **reanalysis** of the second part of the compound. Reanalysis means that speakers assign a different underlying structure to a string (of words or morphemes) than the previous generation of speakers. This process was called "misanalysis" above, implying that speakers "miss" the target grammar G_1 (cp. the discussion of **flawed transmission**). In this case, the noun "shape" in the second part of the compound in (14a) must have been reanalysed as an adjective (14b) at some point (we know that this happened independently to Engl. *like*).

14) Reanalysis

a. $[\textit{freond}]_N$-$[\textit{līce}]_N$ (early Old English)

b. $[\textit{freond}]_N$-$[\textit{līce}]_A$ (late Old English)

c. $[[\text{friend}]_N$-*ly*$]_A$ (Modern English)

Note that nothing has changed on the surface yet: (14a) and (14b) look exactly alike, but the subscript letters indicated different underlying structures ("noun" vs. "adjective"). From (14b) to (14c), another

reanalysis took place: while [-*līce*] in (14b) could still stand by itself in a clause, -*ly* in (14c) is an affix and must attach to something else, as indicated by the bracketing.

<Begin Pause and reflect box>

You can test the "grammaticalisation diagnostics" yourself by comparing the Modern English suffix -*ly* to its distant predecessor, the Old English noun *līc*:

- -*ly* is phonologically reduced (note that the regular development of Old English *līc* is Modern English /lajk/, as in *like*)
- -*ly* is semantically bleached: while you can easily describe the meaning or referent of most nouns (except maybe very abstract ones), the only way to describe the meaning of -*ly* is by making reference to its adverbial *function*
- -*ly* is morphosyntactically restricted: while (Old English and Modern English) nouns can occur on their own and in different positions in a sentence, affixes like -*ly* must attach to a base.

<end box>

Reanalysis depends on the possibility of more than one analysis of a given construction. Another example is the development of auxiliaries from main or "lexical" verbs. English has two future auxiliaries:

15) English future auxiliaries
    d. I'**m going to** see Cora tomorrow
    e. I'**ll** see Cora tomorrow

Both (15a) and (15b) refer to a single event: the event of seeing Cora, which will take place in the future. *Going to* and *will* are **future auxiliaries** (they express the future tense in (15a-b)). However, *go* is also a lexical verb that refers to a particular type of motion. Does it refer to motion in (15a)? The answer is no: when you utter (15a), you do not need to be physically walking towards Cora in order for it to be true. This indicates that semantic bleaching has taken place: the lexical verb *go* has lost its meaning as verb of motion and instead expresses the functional category "future" in sentences like (15a)—but note that *go* is of course still a verb of motion in sentences like *I'm going to the store*.

This suggests that English has (at least) two verbs *go*: 1. *go* verb of motion, 'walk, move, advance', 2. *go* AUX, future auxiliary. The development of the second one from the first one is an instance of grammaticalisation through **reanalysis** of the underlying structure of sentences like (15a). This is illustrated in the following example:

16) Grammaticalisation of *going to*:

f. [I am going$_V$ [to see$_V$ Cora]$_{TP}$]$_{TP}$

g. [I am going$_T$ to see$_V$ Cora]$_{TP}$

(16a) is a *biclausal* structure, meaning that there are two separate events: the event of going and the event of seeing Cora (*go* and *see* each head a separate **verb phrase**, or **VP**). You might utter this sentence as you are walking towards Cora. In (16b), on the other hand, there is only one event: the event of seeing Cora, which takes place in the future. That is, (16b) is *monoclausal*: there is only one event, expressed by the verb *see*. *Going* has been reanalysed as a functional category expressing future tense. Functional categories expressing tense are usually thought to head a functional **tense phrase** (**TP**) above the verb phrase. Like in (14), **rebracketing** has taken place: the underlying structure of (16b) is different from that of (16a), even though they are the same on the surface.

<Begin Pause and reflect box>

Apply the grammaticalisation diagnostics discussed above to the English *go*-auxiliary. We have already seen that semantic bleaching has applied. What about the other diagnostics? Specifically, note that:

- In many varieties of English, *going to* is reduced to *gonna*: *I'm gonna see Cora tomorrow* (while it's ungrammatical to say **I'm gonna the store*)
- Is *going to*/*gonna* as restricted in its distribution as the suffix *-ly* discussed above? How does it differ? In which context(s) is it found?

<End box>

<H1> 8.5. Syntactic change

<H2> 8.5.1. Parameters of syntactic change

Syntactic change is often tied to morphological and semantic change. However, while morphological change can affect individual lexical items (e.g., the words *līc-* 'body' or *go* in section 8.4.), syntactic change affects syntactic categories at the sentential level, usually across the board. For instance, the change of object-verb (OV) to verb-object (VO) word order affects *all* instances of verbs and their objects, not just one particular verb and one particular object.

In the study of syntactic change, the **Principles and Parameters** approach has been especially successful. **Principles** are abstract properties of grammars that are shared by all languages—they are part of Universal

Grammar (UG). Basically, principles are what all languages have in common. **Parameters** are "options" of implementing a principle, or "switches". Basically, parameter settings are what distinguishes languages from one another (at the syntactic level—languages are of course also distinct in terms of their phonology, morphology, and lexicon). In the following, we will discuss one particular parameter and its change in the history of English.

<H2> 8.5.2. Null subjects

Consider the examples in (17), which illustrate variation in the expression of subject pronouns in Italian, French and English. In standard Italian (as in many other languages around the world), subject pronouns such as "he", "she", "it", "they", etc., can remain unexpressed (indicated by the symbol "Ø" in ex. 17a).

17)

a. Ø Parla italiano (Italian)
   speak.3sg.pres Italian
   "He/she speaks Italian"

b. * Ø Parle italien (French)

c. * Ø Speaks Italian (English)

The sentence (17a) is a grammatical sentence of Italian and means "she speaks Italian" or "he speaks Italian" (depending on context), even though the element that means "he" or "she" is "dropped". This phenomenon is known as **pro-drop** (for "pronoun dropping"). Spanish, Italian, Mandarin Chinese, Japanese and Arabic are pro-drop languages, while English, French and German are not. That is, in these languages dropping a subject pronoun makes a sentence ungrammatical, as illustrated in (17c) for English. Italian allows null subjects, whereas French and English do not. Roberts (2007) formalises this as the "null-subject parameter:

- **The null-subject parameter** (Roberts 2007: 25): *Does every finite clause require an overt subject?*
  YES: non-null-subject languages (French, English …)
  NO: null-subject languages (Italian, Spanish, Greek, Japanese, Navajo …)

Parameters are like switches that can be turned on or off for any given language. If the null-subject switch is turned on, the result is a language in which all pronominal subjects are overtly expressed (like in English). If the null-subject switch is turned off, the result is a language like Italian in which pronominal subjects

can be dropped. Crucially, the parameter must be set to on or off for any given language. That is, a child acquiring Italian or English cannot simply *not* set the parameter—it *must* be on or off. However, the setting can change over time. If the setting of the null-subject parameter changes in the course of the history of any given language, it will affect *all* pronominal subjects, and that is indeed what we see. Old English was a subject pro-drop language. Here are some examples of null subjects in Old English (from Van Gelderen 2013a):

18) Nu ___ scylun hergan hefaenricaes uard
now Ø must praise heaven.kingdom's guard
'Now **we** must praise the lord of the heavenly kingdom.'

19) Nearwe ___ genyddon on norðwegas
anxiously Ø hastened.3pl on north.ways
'Anxiously, **they** hastened north.'

These sentences are not grammatical with null subjects in Modern English, so the parameter must have switched from NO to YES at some point. How might this have happened? Note that the definition of the parameter does not *exclude* the possibility of overt subject pronouns. In fact, Old English did also use overt subject pronouns (as do all other null-subject languages, under varying circumstances), as in the following examples (from Van Gelderen 2013a):

20) **hi** cwædon him betweonan þæt **hi** woldon bugan to þæra apostola geferrædene
they said them between that **they** would bend to the apostles' fellowship
'they said between themselves that they wanted to join the fellowship of the apostles.'

21) Nolde **ic** sweord beran …
not.would **I** sword bear
'I would not bear a sword …'

It is conceivable that the null-subject parameter switched from NO to YES because some children who were acquiring Old English as their first language had more sentences with subject pronouns, like (20)-(21), in their input than sentences without subject pronouns, (18)-(19).

<Begin Pause and reflect box>

Compare the following Old English sentences from the Anglo-Saxon Chronicle to their Modern English counterparts. Which syntactic changes do you observe? Pay attention to word order changes, especially the position of subject, object, verbs, participles, etc.

22) *And hi hæfdon heora cyning aworþanne* …
And they had their king overthrown
"and they had overthrown their king"

23) *Her for se ilca here innan Myrce to Snotingham* …
Here went the same army inside Mercia to Nottingham
"Here (in that year) the same army went inside Mercia to Nottingham …"

<End box>

<H1> 8.6 Semantic change

<H2> 8.6.1. Types of semantic change

Semantic change is change in meaning over time. Traditionally, historical linguists distinguish between semantic change in **lexical categories** (nouns, lexical verbs, adjectives) and semantic change in **functional categories** (complementisers, determiners, auxiliary verbs).

<H2> 8.6.2. Semantic change in lexical categories

Change in lexical categories (nouns, verbs, adjectives) is very varied. A common change is that a given word comes to refer to either a *subset* or *superset* of its original denotation. If it refers to a superset, it is called **semantic broadening**. If it refers to a subset, it is called **semantic narrowing**.

- **Semantic broadening**: a word refers to a superset ("broader set") than its previous meaning. Examples include Engl. *dog*, which originally referred to a specific breed of dog (cp. German *Dogge* 'mastiff') and now refers to the superset 'dogs'.
- **Semantic narrowing**: a word refers to a subset ("narrower set") than its previous meaning. For example, the word *deer* originally meant 'animal' (cp. German *Tier* 'animal'), but its meaning 'narrowed' to refer to only a particular type of animal.

Semantic change of lexical items often builds on the **metaphorical** use of nouns, verbs, and adjectives. A **metaphor**, or "figure of speech" draws a comparison between two related concepts. The metaphorical use of a lexical category can become its primary function over time. Here are some examples:

- *to grasp*: to understand
- *to break up*: to end a romantic relationship
- *dope*: awesome, great
- *river bed*: the bottom of a stream or river, *not* an actual bed that contains a river

Examples of metaphorical use of language can be multiplied almost indefinitely and is a major source of *lexical* semantic change.

Another common type of lexical change concerns the **connotation** of words. Words (or rather, the concepts or entities they refer to) can have a socially, culturally or subjectively conditioned positive or negative connotation. For example, the word *cat* may have a negative connotation to somebody with a cat allergy. If the positive or negative connotation becomes part of the core meaning or **denotation** of a word over time, we speak of **amelioration** (positive connotation) or **pejoration** (negative connotation).

- **Amelioration**: a word acquires a positive connotation as part of its meaning. For example, the word *knight* goes back to Old English *cniht* 'servant, young boy' and acquired the meaning 'nobleman' via 'military servant'.
- **Pejoration**: a word acquires a negative connotation as part of its meaning. For example, the word *villain* originally meant 'villager' (it was borrowed from Old French *vilain* 'peasant') and underwent a pejoration that's typical for terms referring to the countryside-dwelling population. The word *attitude* is currently undergoing a pejoration from its meaning 'state of mind, opinion' to the meaning 'uncooperative and antagonistic behaviour', as in "don't give me that attitude!". Words that refer to women also tend to undergo pejoration over time, such as *hussy* from Middle English *husewif* 'housewife' or German *Weib* 'unpleasant woman' from Old High German *wīp* 'woman'.

<Begin Linguistic tidbits>

Linguistic tidbits: Lexical change

Semantic change in lexical items can be quite dramatic: the Albanian word for 'sister', *motër*, comes from the reconstructed Proto-Indo-European word **mā́tēr* (from even older **méh$_2$tēr*), which means 'mother' in all other related languages: Greek *mḗtēr*, Latin *māter*, Sanskrit *mātár-*, English *mother* (from Old English

mōdor), etc. Such drastic changes in the meaning of core vocabulary are rare, however. In the Albanian case, it was probably triggered by a change in family structure, in which older sisters took care of their younger siblings.

<End box>

These developments are interesting because "they tell us a lot about past cultural and social history, and probably also about certain pervasive social attitudes" (Hock 1986: 303). However, they tell us less about *grammar* change, for which we need to look to changes in functional categories.

<Begin Pause and reflect box>

Categorise and comment on the examples of semantic change listed below:

- English *silly*: from Middle English *sely* 'blissful, blessed', Old English *gesælig* 'happy, prosperous'.
- German *Du Bauer*! 'you peasant!', one of the gravest insults in the variety of German spoken by the city-dwelling author of this chapter.
- English *head* in *head of state*, Italian *capo* 'head' in *capofamiglia* 'head of the family'
- English *bird*: from Old English *brid(d*) 'young bird'
- English *wife* (related to Modern High German *Weib* 'unpleasant woman') comes from Old English *wīf* 'woman'

<End box>

<H2> 8.6.3 Semantic change in functional categories

Semantic change in functional categories such as demonstratives, auxiliaries and pronouns is closely connected to morphological and syntactic change and to the "clustering" of changes discussed under the label **grammaticalisation** in section 8.4.3. Functional categories often change their meaning by grammaticalising implicatures, entailments or presuppositions of their original meaning. For example, it is common for demonstrative pronouns like *this (one)*, *that (one)* to develop into third person pronouns like *he*, *she*, *it*. Such demonstrative pronouns originally track *contrastive* third person discourse topics and entail non-contrastive or general third person topics. This entailment then becomes grammaticalised as the more general pronominal use. This happened to the Latin demonstrative pronouns *ille*, *illa* 'that (one)' that we see in French *il*, *elle*. The French subject pronouns *il* and *elle* do not mean 'that (one)' any more:

24) a. Latin: *ille venit* “THAT ONE is going”

b. French: *il vient* “he goes”

Lexical items (nouns, verbs, adjectives), on the other hand, can acquire functional semantics over time. Such a development is often found with container nouns like *cup*, *glass*, *sack*, *bottle*, which refer to items that can *contain* a substance. These tend to develop into measure nouns and refer to a quantity or *measure* of a substance through the grammaticalisation of an entailment. Consider the nominal phrase *a glass of water*. This phrase denotes a container, *glass*, which contains a substance, *water*. But it also entails a quantity or measure, namely the quantity of water that fits into a glass. This is exemplified in the following two sentences (the ‘#’ symbol means that something is syntactically well-formed, but semantically ill-formed):

25) a. *A glass of water <u>smashed</u> on the floor* (but: #*water smashed on the floor*)

b. *A glass of water <u>spilled</u> on the floor* (but: #*a glass spilled on the floor*)

(25a) is an example of the container reading (*smash* refers to the container *glass*), while (25b) is an example of the measure or quantity reading (*spill* refers to the substance *water*). The container reading entails the measure reading, and this entailment can become grammaticalised. For example, in some varieties of English it is possible to use measure nouns to determine quantities in recipes without using the preposition ‘of’, as in *two cups water*, *one cup rice*, etc. In this context, ‘cup’ refers to a quantity, not a container (ideally, you’re not going to throw an actual cup filled with rice into a pot of water to make dinner).

<Begin Pause and reflect box>

Modern Mandarin Chinese has a negation marker *méi* ‘not’ that is used to negate sentences with the existential verb *yǒu* ‘there is, exists’ (examples modified from van Gelderen 2008):

26) *wǒ <u>méi</u> yǒu shū*

I NEG exist book

‘I don’t have a book’

In Old Chinese, this verb meant ‘to die’ (the following example is transcribed into Modern Mandarin Chinese):

27) *Yáo Shùn jì <u>mò</u>*

Yao Shun since died

'Since Yao and Shun died, …'

Describe the semantic change that happened to this verb—why do you think did it change into a negation marker?

<End box>

<H1> 8.7. Language universals and language change

While most of the examples of phonological, morphological, syntactic and semantic change we have discussed in the previous sections came from English or French, it is striking that we observe the same types of changes in completely unrelated languages as well, suggesting that some aspects of language change are due to cross-linguistically constant **universals** and the way these interact with the (likewise constant) human interpretive and production systems (basically, "meaning" and "sound"). Given that all humans are born with the same capacity to acquire language, the human language faculty, this is not particularly surprising. An important principle of historical linguistics that results from this observation is the **Uniformitarian principle**, which states that past language stages conform to the same basic principles as contemporary language stages.

- **Uniformitarian Principle**
  "The general processes and principles which can be noticed in observable history are applicable in all stages of language history." (Hock 1986)

From the perspective of theoretical and historical linguistics, the Uniformitarian Principle states that the basic principles of the human language faculty have remained unchanged since it developed. This means that if an infant born in Montreal this year time-traveled to Bronze Age South Russia (accompanied by adults, of course—unaccompanied babies mustn't time-travel), he or she would be able to acquire Proto-Indo-European (the reconstructed language spoken in that area 6,000-5,000 years ago) without any problems. Likewise, a Proto-Indo-European time-traveler baby would easily be able to acquire any of the languages currently spoken in Montreal.

According to the Uniformitarian Principle, there is no reason to believe that Old Chinese or Proto-Indo-European were any more or less complex and expressive than any of the languages currently spoken around the globe. Moreover, it means that the possible diachronic developments that we observe as we study language change are constrained by the same principles of UG that we observe in the study of synchronic variation between different languages. For example, we have seen in section 8.5.2. that the

subject **pro-drop** property of languages varies both synchronically and diachronically. Synchronically, Italian is a pro-drop language, while English is not. But while subject pro-drop is ungrammatical in Modern English, it was perfectly fine in Old English. At some point in the development of Old English to Modern English, this pro-drop property was lost. Crucially, the **diachronic variation** we observe in the history of English is parallel to the **synchronic variation** we observe in contemporary languages.

<Beginn pause and reflect box>

Take a look at the following Old French sentences (from Roberts 2007):

a) Si chaï en grant povreté
   thus fell-1sg into great poverty
   "Thus I fell into great poverty." (Perceval, *441*)

b) Si en orent moult grant merveille.
   thus of-it had-3pl very great marvel
   "So they wondered very greatly at it." (Merlin, 1)

Based on these examples, was Old French a pro-drop language? What about Modern French?

<End box>

<H1> 8.8 The comparative method and language reconstruction

<H2> 8.8.1 Attested and unattested languages

So far we have discussed changes that we observe between two different attested language stages that are in a descent relationship. Attested means that we have some sort of historical record of a past language stage (usually manuscripts or other written evidence). What about unattested language stages of which there is no historical record whatsoever? In the preceding sections, we used the symbol "*" to indicate that a form is not **directly attested**. This symbol is called the **asterisk**, and it indicates that a linguistic form is **reconstructed**. The **reconstruction of past language** stages is one of the main subfields of historical linguistics. Why do we have to reconstruct past language stages? The answer lies in the fact that our knowledge of the past is often incomplete because of a lack of sources (this is true in other fields as well, not just in linguistics). Linguists who work on "living languages" like Modern English, Mandarin Chinese or Inuktitut can gather data by working with native speakers of these respective languages and

eliciting **grammaticality judgements**. However, historical linguists usually work on languages that are no longer spoken—"dead" languages, or, more precisely, **non-informant languages** (meaning there are no native speaker informants we can work with). These languages are only accessible through **historical records**.

<Start Eyes on World Languages box>

Eyes on World Languages: Historical records and language attestation

The records and artifacts that historical linguists work with vary widely from language to language. Here are some examples:

- Rock inscriptions: the Behistun inscription in Old Persian, 6th century BCE
- Parchment (animal hides): the Gothic *Codex Argenteus*, 6th century CE
- Tree bark: Old Russian Novgorod manuscripts, ca. 11th – 15th century CE
- Animal bones: Old Chinese, ca. 1,000 BCE
- Papyrus or paper: Old Egyptian, from ca. 2,500 BCE
- Clay tablets: Sumerian, from ca. 3,000 BCE; Akkadian, from ca. 2,500 BCE
- Wax cylinder recordings: Yurok (Algic), California, 1900 CE

Historical linguists also work with contemporary languages that are still spoken. For example, the Australian linguist Claire Bowern has worked on the reconstruction of indigenous languages of Australia and has done extensive fieldwork on these languages, especially the Pama-Nyungan family.

The study and evaluation of written texts and their origin is called **philology**. Philology is an important tool for historical linguists, and the philological analysis of a given text usually goes hand in hand with its linguistic analysis. In other words, we need to understand what a text means, and where, how and why it was written in order to analyse its phonological, morphological, syntactic and semantic features, and vice versa.

<End box>

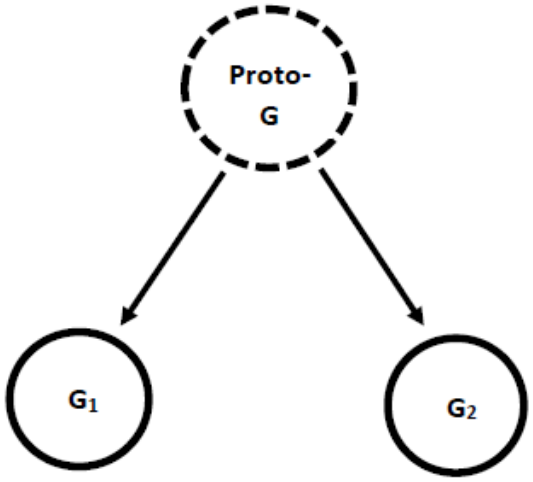
<H2> 8.8.2 Comparative reconstruction and non-lineal descent

Because our records of past language stages are often incomplete, historical linguists must "fill in the gaps", and this is done through **comparative reconstruction**. Comparative reconstruction relies on the **comparative method**, which was developed in the 19th century and establishes **genetic relationships**

between languages that belong to the same language family through **regular sound correspondences.** Languages that are genetically related share a common ancestor, or **proto-language**. For example, you might know that English, German, Dutch and Icelandic (among others) all belong to the Germanic language family. That means that they share a common ancestor language, called Proto-Germanic. French, Italian, Spanish, etc., go back to a common ancestor called Proto-Romance, and the Algonquian languages Mi'kmaq, Ojibwe and Cree go back to Proto-Algonquian. There are no written records of these proto-languages; they were reconstructed based on the attested languages that descended from them.

To understand how this works, the notion of **non-lineal descent** becomes relevant. We have defined lineal descent in section 8.2.3 above as a diachronic relationship between two grammars G_1 and G_2 in which G_2 is directly descended from G_1. However, historical linguists often deal with situations in which forms in G_1 correspond to forms in G_2 even though G_2 is not directly descended from G_1. That is, G_1 and G_2 are genetically related (they share a common ancestor), but neither is directly descended from the other. Non-lineal descent is illustrated in Figure 8.4. (based on Hale 2007: 32).

<Start Figure>



<Caption> Figure 8.4. Non-lineal descent: G_1 and G_2 are in a lineal descent relationship with an unattested Proto-G, but not with each other.

<End figure>

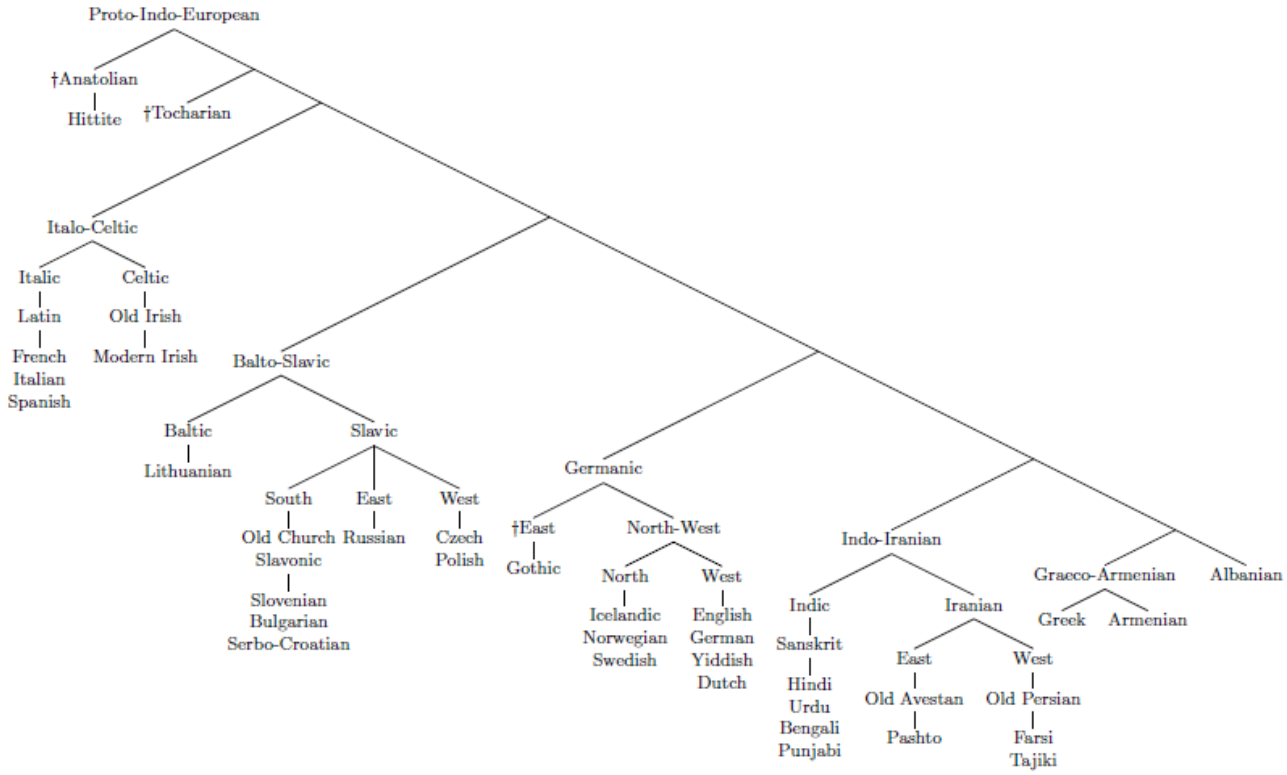
- **Non-lineal descent**: There is a regular correspondence between features of G_1 and G_2 that are NOT due to a lineal descent relationship $G_1 > G_2$

In these situations, historical linguists reconstruct a grammar "Proto-G" from which G_1 and G_2 are lineally descended.

<H2> 8.8.3 Language families

Languages that are genetically related by virtue of being descended from a common ancestor language (or "Proto-G", see above) are called **language families**. Languages that belong to the same family are often represented in tree models, like the one in figure 8.5. below. These tree models are used by linguists to represent **subgrouping** within language families, that is, to show which languages are more closely related to each other within the same family. Each node in the tree stands for a subfamily of a larger macrofamily. The tree in figure 8.5. illustrates the Indo-European language family, which descends from the reconstructed ancestor language **Proto-Indo-European** (spoken ca. 4,000 BCE).

<Start Figure>



<Caption> Figure 8.5.: The Indo-European language family

<End figure>

<Start Eyes on World Languages box>

Eyes on World Languages: Language families

Here are some more examples of reconstructed proto-languages and their descendant language families:

- Proto-Sino-Tibetan: from which, e.g., Modern Mandarin Chinese, Cantonese and Tibetan are descended
- Proto-Semitic: descendants include Aramaic, Hebrew, Arabic and Maltese and the extinct Semitic languages of ancient Mesopotamia (Akkadian, Babylonian, Ugaritic)
- Proto-Algonquian: from which the Modern Algonquian languages spoken in Canada are descended, e.g., Ojibwe, Blackfoot and Mi'kmaq
- Proto-Austronesian: e.g., Malayo-Polynesian languages (Malagasy, Malay, Oceanic languages such as Samoan and Maori)

There are many others, and the reconstruction of the world's language families is far from complete. Moreover, many of the language families mentioned above have been shown to belong to even larger **macrofamilies** that include several reconstructed proto-languages. For example, Proto-Semitic goes back to a macroframily called Proto-Afro-Asiatic which includes Egyptian and its modern descendent Coptic, the Cushitic languages (e.g., Somali), and the Berber and Chadic languages. Proto-Algonquian goes back to a macrofamily called Proto-Algic, from which the Yurok and Wiyot languages which were spoken on the North American West Coast are also descended. Other macrofamilies are more controversial: it has been claimed that Proto-Indo-European and Proto-Uralic (the reconstructed language from which Finnish and Hungarian are descended) are related through a common ancestor, sometimes called Nostratic. However, this claim has never gained acceptance due to the lack of reliable **sound correspondences** (see the next section). Whether we will ever be able to reconstruct "Proto-World" is likewise an open (and highly controversial) question.

<End box>

<H2> 8.8.4 The comparative method

The language families discussed in the previous section were established by historical linguists using the comparative method. In this section, we will work through a case study to illustrate the comparative method in action. Campbell (1998) proposes the following steps for applying the comparative method (there are more, but these will suffice to get us started):

- Assemble cognates

- Establish sound correspondences
- Reconstruct the proto-sound
- Determine the status of similar (partially overlapping) correspondence sets

We will start by establishing **correspondence sets** (cp. the discussion and definition of correspondence in section 8.2.3 above) based on **cognates**. Cognates are words that share the same **etymology**, meaning that they go back to the same (reconstructed) word in the proto-language:

- **Cognate** (Latin *co-gnātus* 'sharing ancestors, kindred, related'): a form F_1 in language L_1 and a form F_2 in language L_2 are cognate if they go back to the same proto-form *F in the reconstructed proto-language *L, where *L is the ancestor of L_1 and L_2
- **Etymology** (Greek *étymon* 'true'): the historical development in sound and meaning of a form through time; e.g., the development of F_1 and F_2 from *F

<Start Linguistic tidbits box>

Linguistic tidbits: some useful etymologies

- The word ***toe*** is related to the word ***dic****tator* (Lat. ***dic****tāre* 'to assert') and the names of the Indian states of Uttar Pra**desh** and Andhra Pra**desh** (Sanskrit ***deś****á-* 'region, country'). They all go back to a reconstructed root **deiḱ* 'to show, point to'.
- The word *lady* is from Old English *hlæf-dige*, lit. "loaf-shaper" (the person responsible for kneading dough)
- French *cher*, *chère* 'dear', Irish *cara* 'friend', English *whore* and Sanskrit *kāma-* 'love' (as in the infamous *Kāmasūtra*) all go back to the reconstructed root **kā* (< **keh₂*) 'to love, desire'.

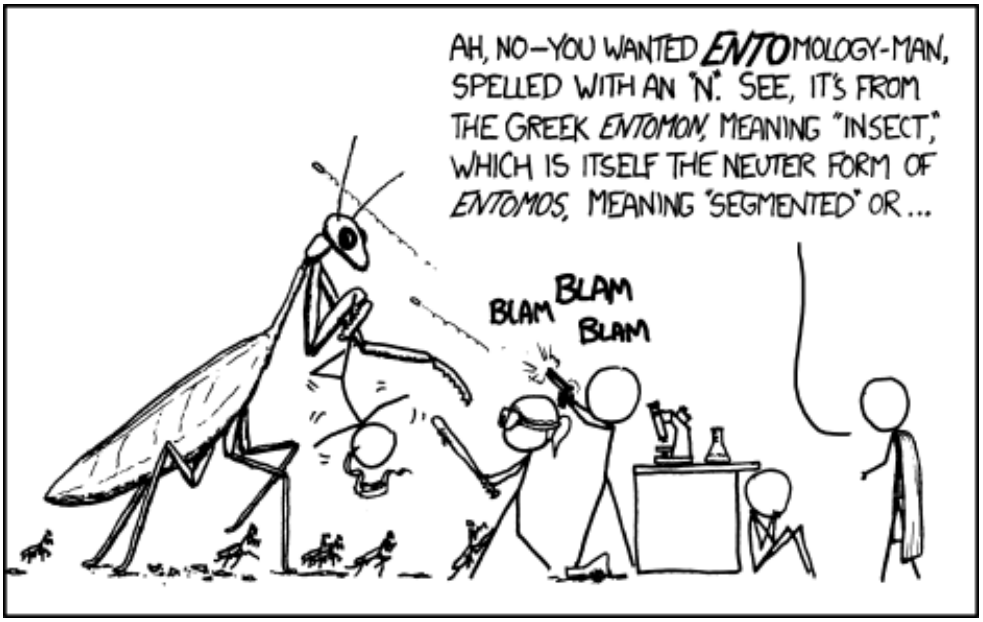
<End box>

There is a common misconception that the etymology of a word somehow reveals its "true" meaning. But the past meaning of a given word is no more "true" than the current or future meaning of that word. Each represents a specific synchronic stage in the diachronic trajectory of that word. For example, the English word "nice" meant "accurate, precise" around 1500, and "stupid, foolish" around the late 13th century. It goes back to the Latin word *nescius* "ignorant", which was borrowed into English via French in the 13th century (that is, the *etymology* of the English word "nice" is ultimately the Latin word *nescius*). However, it would indeed be foolish to claim that the word "nice" really means "accurate" or "foolish" or "ignorant"

in present-day English. Rather, the etymology of a word sheds light on past stages of its development, some of which may be relevant to understanding its present form and meaning.

<Start cartoon>

Etymology man: probably not an upcoming Marvel movie



Source: https://xkcd.com/1012/

<End cartoon>

The following table illustrates correspondences between cognates in the Romance languages, using the IPA for a more accurate rendering of the relevant sounds (see chapter two on phonetics).

<Begin table>

	Italian	Spanish	Portuguese	French	Meaning
1.	/ka**p**ɛlɔ/	/ka**β**alɔ/	/ka**b**elɔ/	/ʃ**v**ø/	'hair(s)'
2.	/ka**p**ra/	/ka**β**ra/	/ka**b**ra/	/ʃɛ**v**ʁ/	'goat'
3.	/a**p**rire/	/a**β**rɪr/	/a**b**rɪr/	/u**v**ʁiʁ/	'to open'
4.	/sa**p**ere/	/sa**β**ɛr/	/sa**b**er/	/sa**v**waʁ/	'to know'

Table 8.6. Sound correspondences in the Romance languages

<End table>

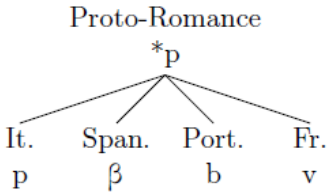
Comparing the words in table 8.6., you will notice that the bolded sounds in each language correspond exactly. That is, every time we observe a /p/ in Italian we see that Spanish has a /β/, Portuguese has a /b/ and French has a /v/ in the same word. The chances that this is a coincidence are slim. Rather, what we have discovered is a generalisation concerning the correspondence between sounds in these languages. This correspondence can be summarised as follows:

28) It. /p/ = Span. /β/ = Port. /b/ = Fr. /v/

If two or more languages share regular sound correspondences, they must be genetically related. Historical linguists use correspondence sets such as (28) to reconstruct the proto-language from which these languages descended. Sound change is regular (see section 8.3.), so the four different sounds in (28) go back to just one single sound in the ancestral language (Proto-Romance, in this case). So how do we decide what this "proto-sound" was? We can start with the assumption that this sound was [+labial], since this phonetic feature is shared by all four correspondents in (28). That is, we start by determining the **common phonetic features** found in our correspondence set. However, the four sounds differ with respect to manner of articulation (Italian and Portuguese have stops, Spanish and French have fricatives) and voicing (Italian has a voiceless sound, the other languages have voiced sounds). We could use the "majority wins" principle and just pick the feature that is most commonly found in the daughter language, but this principle is flawed for several reasons. It is not going to be helpful in determining the manner of articulation of the proto-sound, for instance: half the languages have stops, the other half has fricatives. Moreover, this principle is not sensitive to accidental gaps in our historical records. What if there were three other languages with voiceless fricatives that are simply not attested anywhere? Our reconstruction should hold independently of the number of attested languages. Therefore, we need to refer to the typology of **phonetically natural** sound changes. "Phonetically natural" means that a sound change occurs because of the properties of its phonetic environment. For example, all the labials in table 8.6. are between two voiced segments. Voiceless sounds that are between voiced sounds commonly take on the phonetic feature [+voiced] of their surrounding sounds. It is therefore more phonetically natural to assume that a voiceless sound became voiced between two voiced segments than assuming an originally voiced sound became voiceless in this position (we would not be able to motivate this change in terms of phonetic naturalness, since there is no way to explain where this [-voice] feature would have come from). So, our proto-sound was [+labial, -voice] (note that "majority wins" would mean that we would need to posit [+voice]!).

What about manner of articulation? Again, phonetic naturalness comes to the rescue. Stops are often "weakened" (the technical term for this is **lenited**) between voiced segments. We should therefore reconstruct an original stop, and assume that this stop was lenited to a fricative in Spanish and French. This means that the Proto-Romance sound must have been *p.

<Begin figure>



<Caption> Figure 8.7. Reconstruction of Proto-Romance *p

<End figure>

<Begin Pause and reflect box>

Below is a correspondence set from the Semitic languages Akkadian (ancient East Semitic), Classical Arabic, Biblical Hebrew (both Central Semitic) and Ge'ez (ancient South Semitic). Reconstruct the bolded sound for Proto-Semitic and draw a tree like the one in figure 8.7. Give arguments for your reconstruction and explain why "majority wins" is not a helpful principle in this case.

	Akkadian	Arabic	Hebrew	Ge'ez	meaning
1.	ap**p**u	ʔan**f**	ʔa**p**	ʔan**f**	'nose'
2.	**p**e:mtu	**f**aħm	**p**eħa:m	**f**əħm	'charcoal'
3.	u**p**nu	ħu**f**nat-	ħo**p**nayim	ħə**f**n	'hollow of the hand'

<End box>

We have now reconstructed our first Proto-Romance sound. Feeling smug, we take another look at the Romance languages and find correspondence sets such as the ones in table 8.7.

<Begin table>

	Italian	Spanish	Portuguese	French	meaning
1.	/**p**ɔrto/	/**p**werto/	/**p**ɔrtɔ/	/**p**ɔʁ/	‘harbour’
2.	/**p**adre/	/**p**aðre/	/**p**aj/	/**p**ɛʁ/	‘father’
3.	/**p**ino/	/**p**ino/	/**p**iɲ(ejr)o/	/**p**ɛ̃/	‘pine’

<Caption> Table 8.7. More sound correspondences in the Romance languages

<End table>

This correspondence can be summarised as follows:

29) It. /p/ = Span. /p/ = Port. /p/ = Fr. /p/

This correspondence set evidently conflicts with our previously established correspondence set in (28), where Italian /p/ corresponded to Span. /β/, Port. /b/ and Fr. /v/, while in 29), Italian /p/ corresponds to /p/ in all the other Romance languages. We have discovered an **overlapping correspondence set**. There is, however, no need to panic. We have already established that the reason why the voiced plosives and fricatives in Spanish, Portuguese and French correspond to a voiceless plosive in Italian in table 8.6. is that they occur between two voiced sounds and have assimilated in voicing to these surrounding sounds. This did not happen in Italian, and so we reconstructed *p for these words in Proto-Romance. In table 8.7., on the other hand, the voiceless plosive /p/ is in word-initial position in all examples. Since it is not between two voiced sounds, there is no reason to expect voicing (or fricativisation, for that matter), given our previous observation concerning the words in table 8.6. In fact, the correspondence set in (29) makes it far easier for us to reconstruct the relevant Proto-Romance sound, since there is only one option: *p (as a general principle, we should not reconstruct some other sound if there is no explicit reason to do so).

To summarise, we have reconstructed Proto-Romance *p for the correspondence set in ex. (28) based on considerations of phonetic naturalness and for the correspondence set in (29) because we don’t have evidence for any other sound. But how reliable are our reconstructions? Is this methodology sound? The arguments follow from the attested data and our knowledge of the phonetic properties of the relevant sounds involved, so this reconstruction is valid in and of itself. In most cases, we are not in a position to confirm this reconstruction with actually attested data (since the reason we reconstruct past language stages is precisely that those stages aren’t actually attested!). However, there are rare cases in which we

do have external confirmation for our reconstructions. The Romance languages are one such case. While Proto-Romance is unattested, we do have written evidence from its immediate predecessor, often called "Vulgar Latin", or, more precisely, "spoken Latin". "Spoken Latin" existed as colloquial variety (or varieties, rather) beside the well-attested Classical Latin. The following table contains the Latin correspondences to the Romance words discussed above. As you can see, Latin did have a /p/ in all the positions where we reconstructed a Proto-Romance *p based on the evidence of the Modern Romance languages.

<Begin table>

	Latin	Italian	Spanish	Portuguese	French	Meaning
1.	ca**p**illus	/ka**p**ɛlɔ/	/ka**β**alɔ/	/ka**b**elɔ/	/ʃ**v**ø/	'hair(s)'
2.	ca**p**ra	/ka**p**ra/	/ka**β**ra/	/ka**b**ra/	/ʃɛ**v**ʁ/	'goat'
3.	a**p**erīre	/a**p**rire/	/a**β**rɪr/	/a**b**rɪr/	/u**v**ʁiʁ/	'to open'
4.	sa**p**ere	/sa**p**ere/	/sa**β**ɛr/	/sa**b**er/	/sa**v**waʁ/	'to know'
5.	**p**ortus	/**p**ɔrto/	/**p**werto/	/**p**ɔrtɔ/	/**p**ɔʁ/	'harbour'
6.	**p**ater	/**p**adre/	/**p**aðre/	/**p**aj/	/**p**ɛʁ/	'father'
7.	**p**īnus	/**p**ino/	/**p**ino/	/**p**iɲ(ejr)o/	/**p**ɛ̃/	'pine'

<Caption> Table 8.8. Sound correspondences between Latin and Modern Romance languages (Lat. *c* = /k/)

<End table>

While it is gratifying to know that we were right, it's rare to be able to obtain this kind of confirmation for our reconstructions. Our methods therefore need to rely on independently established principles, such as the ones discussed above, and our knowledge of linguistic universals and the typology of common sound changes.

<Begin Pause and Reflect Box>

Old English (OE), Old High German (OHG), Middle Dutch (Old Dutch isn't well attested) and Yiddish belong to the Western branch of the Germanic languages and go back to Proto-West Germanic. Using the **cognate sets** below, reconstruct the Proto-West-Germanic words for 'foot', 'apple', 'bath' and 'fife', using the methods discussed above. You may have to use the "majority wins" principle, but keep in mind that this principle is problematic.

	OE	OHG	Middle Dutch	Yiddish	meaning
1.	/foːt/	/fuos/	/vuət/	/fʊs/	'foot'
2.	/æpel/	/apfəl/	/ɑpel/	/ɛpəl/	'apple'
3.	/bæθ/	/bad/	/bat/	/bɔd/	'bath (house)'
4.	/fiːf/	/fimf/	/viːf/	/finf/	'five'

<End box>

<H1> 8.9 Summary

In this chapter, we have introduced the core concepts of historical linguistics as the field that studies language change and past language stages. We have discussed the differences between **change** and **diffusion**, **synchrony** and **diachrony**, and ways of determining **correspondences** between different diachronic language stages. Change affects all aspects of language (in the sense of "grammar", or knowledge of language), and we have discussed **phonological**, **morphological**, **syntactic** and **semantic change**. A crucial insight in the study of language change and correspondence between languages is the observation that sound change is regular (the **Neogrammarian Hypothesis**). This regularity makes it possible to establish sound correspondences, interrelationships between languages descended from the same proto-language, and **reconstructed proto-languages** themselves. Throughout this chapter, we have seen that the different components of mental grammars are interconnected when it comes to language change. This is especially evident in **grammaticalisation** changes, which usually combine changes in the phonology, syntax and semantics of a given lexical or functional item, making historical linguistics a strongly interdisciplinary field.

Finally, we have seen that language change, despite its bewildering variety of manifestations, is constrained by universal properties of the language faculty. Understanding what is constant and invariant about human language, and what is subject to variation and change is one of the most exciting challenges of modern linguistic theory. The goal of historical linguistics is to provide a theory of language change that captures those aspects that are variant.

Exercises

1. **Review: Explain the following terms and their relevance to the study of language change:**
    - Synchrony vs. diachrony
    - The Neogrammarian Hypothesis

- The comparative method
- Imperfect transmission
- Language reconstruction
- Deductive change

2. **Discussion:** "Languages change constantly"; "the language faculty has not changed since it developed in humans": there is an apparent tension between these two statements. Discuss how this tension arises, using the terms **I-language**, **E-language**, **universals** and the **Uniformitarian Principle**.

3. **Discussion:** The core tenet of the Neogrammarian Hypothesis is that sound change is regular. However, the actual texts that historical linguists work with tend to be full of "irregularities" and exceptions. For example, the Old English consonant cluster /sk/ became /ʃ/ during the time of the palatalisation changes discussed in section 8.3.1., as in the following examples:

Old English	Modern English
/**sk**(e)ort/	/**ʃ**ɔrt/ 'short'
/**sk**eakan/	/**ʃ**ek/ 'shake'
/**sk**eotan/	/**ʃ**ut/ 'shoot'

However, there are also plenty of English words which do start with /sk/, such as *skirt* /skərt/, *scar* /skar/ and *school* /skul/. Even worse, there are apparent doublets, words which seem to have both the palatalised and the non-palatalised variant: *skirt* and *shirt* are both from Proto-Germanic **skurta*- 'short garment', and *scar* is related to *share* (both go back to a root **sker*- that meant 'to cut (off)'). While the /sk/-variants at first glance look like exceptions to the Neogrammarian hypothesis, we know that they were actually borrowed from Old Norse into English at the time of the early Danish settlement in England in the 8th and 9th century CE. Old Norse was a Germanic language closely related to Old English and did not undergo palatalisation of /sk/. Discuss why such **loan words** are not counterexamples to the Neogrammarian hypothesis and refer to the note on taboo words in section 8.3.3. What problems might loan words pose for historical linguists? How could we identify them?

4. **Sound change**: The following data illustrate some of the sound changes from Sanskrit (Indo-Iranian) to Pāli, a Middle Indic language that descended from Sanskrit (data based on Oberlies 2001). The symbol 'ɴ' stands for a uvular nasal.

	Sanskrit	Pāli	Meaning
1.	sapta	satta	'seven'
2.	a:tman-	attan-	'self, soul'
3.	pa:tra-	patta-	'bowl'
4.	sakt^{h}i-	satt^{h}i-	'thigh'
5.	saɴrakta-	sa:ratta-	'impassioned'
6.	siɴha-	si:ha-	'lion'
7.	ra:trau	ratto	'at night'
8.	magna-	magga-	'immersed'
9.	paitrika-	pettika-	'paternal'
10.	pu:rɳa-	puɳɳa-	'full'
11.	ʃaikʂa-	sek^{h}a-	'to be trained'
12.	udbalika-	ubbalika-	'tax-free'
13.	ali:ka-	alika-	'lie'
14.	aurasa-	orasa-	'own; legitimate son'
15.	viɴʃati	vi:sati	'twenty'
16.	marma-	mamma-	'vulnerable point, joint'

- Make a list of sound changes from Sanskrit to Pāli and classify them according to the types of changes discussed in section 8.3.4.
- Generalise as much as possible (i.e., "Voiceless stops become voiced stops" instead of "p > b, t > d, k > g", etc.)
- Formalise as much as possible (i.e., "C[-voice] > C[+voice]" instead of "voiceless stops become voiced stops")
- Do any of the rules need to be ordered with respect to one another? Explain why.
- What are the expected Pāli outcomes of the hypothetical Sanskrit words *ʃaɴha-*, *aukta-* and *ru:tra*? You can figure out the expected outcomes by applying the relevant sound laws you found.

5. **Comparative reconstruction**: Below is a list of cognate words in the Polynesian languages Hawaiian, Maori, Samoan and Tongan (from Hale & Kissock 2013) which go back to the reconstructed ancestor language Proto-Polynesian. Your task is to reconstruct these words in Proto-Polynesian and make a list of relevant sound changes (with rule ordering!) from Proto-Polynesian to the daughter languages (Hawaiian, Maori, Samoan, Tongan). You'll find some hints below.

	Hawaiian	Maori	Samoan	Tongan	Translation
1.	manu	manu	manu	manu	'bird'
2.	awa	awa	awa	awa	'channel'
3.	niu	niu	niu	niu	'coconut'
4.	pua	pua	pua	pua	'flower'
5.	peʔa	peka	peʔa	peka	'bat'
6.	muli	muri	muli	mui	'behind'
7.	kani	taŋi	taŋi	taŋi	'cry'
8.	au	au	au	ʔau	'current'
9.	kuna	tuna	tuna	tuna	'eel species'
10.	walu	waru	walu	walu	'eight'
11.	iʔa	ika	iʔa	ika	'fish'
12.	kae	tae	tae	taʔe	'excrement'
13.	lau	rau	lau	lau	'leaf'
14.	ʔuku	kutu	ʔutu	kutu	'louse'
15.	umu	umu	umu	ʔumu	'oven, earthen'
16.	walu	waru	walu	wau	'scratch'
17.	kapu	tapu	tapu	tapu	'taboo'
18.	ako	ato	ato	ʔato	'thatch, roof'
19.	lua	rua	lua	ua	'two'
20.	lua	rua	lua	lua	'vomit'

- Start with words which are identical in all four languages – in these cases you can assume that nothing has changed since Proto-Polynesian was spoken.

- You will see some overlapping correspondence sets (e.g., *n – ŋ – ŋ – ŋ* vs. *n – n – n – n* in 7. vs. 9., respectively). Make a list of these sets and compare section 8.8.4 for hints on how to deal with them.
- Some of your sound changes will need to be ordered. Look up rule ordering in section 8.3.2 above. There's also a box with Proto-Polynesian and Hawaiian data in that section, which you can use as a starting point for unlocking the changes in the plosives (hint: start with the correspondence sets *ʔ – k – ʔ – k* (e.g., 5.) and *Ø – Ø – Ø – ʔ* (e.g., 8.)).

6. Sound change: The following data illustrate some of the sound changes between Mycenaean Greek (attested ca. 1,400-1,100 BCE) and Classical Greek (Attic-Ionic, 5th century BCE). Note that k^w, g^w, k^{wh} are labialised (or rounded) velars, similar to the initial sound in ***qu**een*.

	Mycenaean Greek	Classical Greek	Meaning
1.	wanaks	anaks	'king'
2.	wetos	etos	'year'
3.	k^wetra-	tetra-	'four' (in compounds)
4.	-ok^{w}s	-ops	'-eyed' (in compounds)
5.	hek^{h}onsi	hek^{h}o:si	'they have'
6.	pansi	pa:si	'all' (dat.pl.)
7.	-k^{w}e	te	'and'
8.	-k^{wh}onta:s	-p^{h}onte:s	'-slayer' (in compounds)
9.	diwjos	dios	'of Zeus'
10.	korwa:	ko:re:	'girl'
11.	p^{h}arwos	p^{h}a:ros	'cloth'
12.	k^{w}rijato	priato	'he/she bought'
13.	ma:te:r	me:te:r	'mother'
14.	k^{wh}e:r-	t^{h}e:r-	'wild animal'

- List all the sound changes that you can find in this data set. Some changes crucially depend on the phonetic environment (especially those affecting the labialised velars).
- Be as general as possible (e.g., "voiceless stops become voiceless fricatives between vowels").
- The rules should be as formalised as possible (e.g., A > B/_C).

- For each change, explain what type of change it is, referring to the types of common sound changes discussed in section 8.3.4.
- Are there any changes that need to be *ordered* with respect to each other? Which ones, and why?

7. Comparative reconstruction: The following data illustrate cognates between Latin, Greek, and Sanskrit, which go back to the reconstructed ancestor language Proto-Indo-European (the data are in IPA). You will notice that there is an **overlapping correspondence set** in these data: the Sanskrit vowel *a* corresponds to the vowels *e*, *a*, and *o* in Greek and Latin (the relevant vowels are bolded; ignore the other vowels and the material in square brackets):

	Latin	Greek	Sanskrit	Meaning
1.	**e**st	**e**sti	**a**sti	'is'
2.	**o**kto	**o**kto:	**a**ʂʈa:	'eight'
3.	**a**ger	**a**gros	**a**ʤrah	'field'
4.	f**e**ro:	p^{h}**e**ro:	bh**a**ra:[mi]	'I carry'
5.	-**o**sj**o** (Old Latin)	-**o**j**o**	-**a**sj**a**	ending of the gen.sg. masc.
6.	n**o**kt-	nukt- (from *n**o**kt-)	n**a**kt-	'night'
7.	n**e**b[ula]	n**e**p^{h}[ele:]	n**a**bh[as-]	'fog, mist'
8.	**a**k[us]	**a**k[ros]	**a**ʃ[ri-]	'sharp(ness), point(ed)'
9.	k^{w}**o**d	p**o**(-)	K**a**d	'what, which…?'
10.	-k^{w}**e**	-t**e**	-ʧ**a**	'and' (enclitic)

- There are two ways of dealing with this set: 1) reconstruct **a* for all words based on Sanskrit and assume a **split** into *e*, *a*, *o* in Greek and Latin, or 2) reconstruct **e*, **a*, **o* based on Greek and Latin and assume a **merger** of these vowels in Sanskrit. Which solution is better?
    - Hint 1: a split must be *conditioned* by the phonetic environment (cp. the discussion of the split of Old English velars into palatalised and non-palatalised variants in section 8.3.1.). Is there a conditioning factor for a split?
    - Hint 2: 9. & 10. are crucial. Why?
- Reconstruct the Proto-Indo-European words for 1.-10. as accurately as possible (ignoring the bracketed material). You will not always be able to decide with absolute certainty, but make

use of the principles discussed in section 8.8.4. as much as possible (especially Campbell's steps of applying the comparative method).

- To figure out the last two, recall the development of *k^w in Greek from exercise 6.

8. Morphological change

In some varieties of English, the verb *bring* has acquired a past tense form *brang* and a participle *brung* (if you don't believe me, google the traditional folk song "the Hangman's Song". Not the Led Zeppelin version). Compare this change to the morphological changes discussed in section 8.4.2. and explain what kind of change it is and how it came about.

9. Morphological change/grammaticalisation:

Latin used the feminine noun *mēns* (genitive *mentis*) 'mind' in adverbial phrases such as the following:

a) *clār-ā mente*
clear-FEM.ABL mind.FEM.ABL
"of/with a clear mind, clear-mindedly"

ABL stands for ablative case, a noun case marker that means 'of' or 'from'. This construction gave rise to the French adverbial suffix -*ment* (e.g., *claire-ment* 'clearly', *clair* masc./*claire* fem. 'clear'; *franche-ment* 'frankly', *franc* masc./*franche* fem. 'frank', etc.), as in:

b) *elle parle **franchemen**t à son père*
she speaks frankly to her father

- Using the grammaticalisation diagnostics discussed in section 8.4.3., explain how the French adverbial suffix developed from its Latin ancestor
- Why are the French adverbs made from the feminine rather than masculine form of the adjective (*franchement*, not **francment*)?

10. Grammaticalisation: Latin has a preposition *ante* 'before' (ex. *ante eum* 'before him'), which corresponds to the Ancient Greek preposition *antí* 'against, instead of, opposed to' (ex. *antì*

gámoio ‘instead of a wedding’). In Hittite, an Indo-European language spoken in Anatolia in the 2nd millennium BCE, we find a noun *ḫant-* ‘forehead, front’ (the *ḫ* stands for a velar or glottal fricative). It is generally assumed that the Hittite word is related to the Latin and Greek prepositions.

- Describe the semantic and morphological differences that set the Hittite word apart from the Greek and Latin words
- Which function/meaning is the original one, the Hittite one or the Latin and Greek ones? Explain why, and describe how it changed into the other meaning.

11. Semantic change/reanalysis: English, French and many other languages use the verb ‘to be’ (the copula) to link or equate a noun with another noun or with an adjective, as in *Lisa **is** tall/an artist*. In many other languages (e.g., Russian, Turkish, Arabic), the copula can be “dropped” and remains unexpressed. In Old Chinese, for example, there was no copula in clauses such as (1) (examples from Li and Thompson 1977, cited after Lohndal 2009; the element glossed ‘DCL’ is a declarative or emphatic particle).

(1) Wáng-Tái wù zhě yě
Wang-Tai outstanding person DCL
‘Wang-Tai is an outstanding person.’

Note that there is nothing glossed as ‘be’ in (1). Modern Mandarin Chinese, on the other hand, uses the copula *shì* in copular sentences, e.g.:

(2) nèi-ge rén **shì** xuéshēng
that-CLASS man **COP/‘BE’** student
‘That man is a student.’

Where does *shì* come from? In Old Chinese, it was a demonstrative pronoun, as in the following examples:

(3) fū-zǐ zhì yù **shì** bāng yě
Confucius arrive at **this** nation DCL

'Confucius arrived at this nation.'

(4) jì yù qí shēng yoù yù qí sǐ **shí** huò yě
already wish him live also wish him die **this** indecision DCL
'wishing him to live while wishing him to die, that is indecision.'

- Using the concept of **reanalysis** discussed in section 8.4.3, explain how *shì* developed from a demonstrative pronoun into a copula. The anaphoric use of *shì* in sentence (4) is crucial—think about what possible **misanalysis** of this sentence could have changed the function of *shì* from the perspective of a language acquirer.

12. Syntactic change: Consider the following Middle English sentences, especially the bolded parts (from Ringe and Eska 2013): (from Ringe and Eska 2013):

a. … spoile him of his riches by sondrie frauds, which he **perceiueth not**.
b. Quene Ester **looked never** with swich an eye.
c. How great and greuous tribulations **suffered the Holy Appostyls**…?

These sentences are ungrammatical in Modern English. Here are the relevant Modern English correspondences:

d. … which he **did not perceive**.
e. Queen E. **never looked** with …
f. … **did the holy apostles suffer**?

- Which syntactic changes do you observe between Middle English and Modern English with respect to the position of the lexical verb?
- What is the function of the verb *do*/*did* in the examples above?
- Try to formulate a parameter that could capture the changes between Middle English and Modern English (see section 8.5. on parameter change). How might the parameter setting have changed over time?

References

Andersen, Henning. 1973. Abductive and deductive change. *Language* 49: 567–595.

Campbell, Lyle. 1998. *Historical Linguistics: An Introduction*. 2nd ed., Cambridge, Mass.: MIT Press.

Gelderen, Elly van. 2008. Negative cycles. *Linguistic Typology* 12: 195–243.

Gelderen, Elly van. 2013a. Null Subjects in Old English. *Linguistic Inquiry* 44/2: 271–85.

Gelderen, Elly van. 2013b. The linguistic cycle and the language faculty. *Language and Linguistics Compass* 7/4: 233–250.

Hale, Mark. 2007. *Historical Linguistics: Theory and Method*. Wiley-Blackwell.

Hale, Mark and Madelyn Kissock. 2013. *Introduction to Linguistic Science.* Ms., Concordia University.

Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. 2nd ed., Cambridge University Press.

Hock, Hans Henrich. 1986. *Principles of Historical Linguistics*. Berlin/New York: De Gruyter.

Labov, William, Sharon Ash and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin/New York: de Gruyter.

Li, Charles and Sandra A. Thompson. 1977. A mechanism for the development of copula morphemes. In *Mechanisms of Syntactic Change*, 419–444. Austin: University of Texas Press.

Lohndal, Terje. 2009. The copula cycle. In Cyclical Change, 209–242. Amsterdam/Philadelphia: John Benjamins.

Oberlies, Thomas. 2001. *Pāli: A Grammar of the Language of the Theravāda* Tipiṭaka. Berlin/New York: de Gruyter.

Ringe, Don and Joseph F. Eska. 2013. *Historical Linguistics: Toward a Twenty-First Century Reintegration*. Cambridge University Press.

Roberts, Ian. 2007. *Diachronic Syntax*. Oxford University Press.