FILE 13.7

Internal Reconstruction and Comparative Reconstruction

13.7.1 Reconstruction

One of the goals of historical linguistics is to document and examine how languages change over time. In order to do this, linguists must know both what languages today look like and how they used to look. Unfortunately, of course, we do not have a time machine that would allow us to go back in time to study earlier states of languages directly. Therefore, linguists have come up with a number of ways of looking at older states of language.

The most useful tools for a historical linguist are direct samples of older language: recordings of speakers from the late nineteenth century, for example, or transcripts of speech from eras before sound-recording was possible. In the absence of such transcripts, other early written descriptions of a language, or documents in the language, can help linguists see how a given language used to be. But even when few (or no) written sources exist, linguists can often determine both how a single language used to look and how several languages might have derived from a common source historically. These tasks are accomplished using methods of **reconstruction.** There are two primary methods of reconstruction: internal reconstruction and comparative reconstruction. **Internal reconstruction** involves the analysis of data from a single language in order to make hypotheses about that language's history. **Comparative reconstruction** involves the systematic comparison of multiple related languages in order to make hypotheses about the common protolanguage they descended from. We will consider each in turn in the sections that follow.

13.7.2 Internal Reconstruction

As we have seen from our survey of sound changes that have occurred in the history of English, one of the effects of conditioned sound change is the creation of alternate pronunciations for the same morpheme, which is usually called morphological **alternation.** For example, early in the history of English, fricatives became voiced intervocalically. As a result, the plural form of the word *wife* changed from [wi:fas] to [wi:vas]. In the singular form [wi:f], however, the fricative [f] did not become voiced because it did not occur intervocalically. The net result of this sound change was to create alternate pronunciations for the different forms of the stem of 'wife': [wi:f] in the singular but [wi:v] in the plural. The alternation, which we can consider to be the "trace" of the completed sound change, is still evident in Modern English today, as is evident in the forms *wife/wives*.

When morphological alternations are created by sound change, we can often examine the phonetic context of the alternate pronunciations and infer what sound change(s) caused the alternations in the first place. This type of analysis, whereby the linguist examines data available from one language and one language only and makes hypotheses about that language's history, is what we mean by **internal reconstruction.** Using the internal reconstruction method, a linguist may learn much about a language's history, even if for some reason there are no known related languages to compare it with. This is not unlike seeing puddles in the street and reasoning that it rained, or seeing glass shards and reasoning that

a glass or a window has been broken. In each case we are hypothesizing about what happened to cause what we see as the current state of things.

English can provide us with a very straightforward example of the recovery of an earlier sound change via morphological alternation. In English the voiced velar stop [g] is not pronounced when it precedes a word-final nasal, for example, *sign* [sɑɪn], but it is pronounced in related words if this nasal is not word-final, for example, *signal* [sɪgnəl]. As a result, morphological alternations occur between morphemes with and without the voiced velar stop, for example, *dignity* [dɪgnəɾi], *deign* [deɪn]; *paradigmatic* [pɛɹədɪgmæɾɪk], *paradigm* [pɛɹədɑɪm]. On the basis of these alternations we can make some inferences about the history of English. Specifically, we can assume that at an earlier period the morphological alternation did not exist—that there was only one pronunciation for morphemes that had the sound sequence [gn] or [gm], and that at some point there was a sound change whereby voiced velar stops were lost when they occurred before a word-final nasal.

Sometimes, however, it is impossible to detect the sound change(s) that have created the morphological alternations that exist in a language. This is usually the case when later sound changes take place that obscure the original cause of the alternate pronunciations. Consider the following example from the history of English. At present in English the past tense of the verb *sleep* is [slɛpt] and not [slipt] as we might expect. It is only natural to wonder why the word *sleep* has forms with alternate pronunciations [slɛp] and [slip]. Unfortunately we can arrive at no satisfactory answer just by considering the evidence that exists in Modern English. We cannot say that the alternation is due to the fact that the vowel is followed by two consonants in the past tense form, because other verbs that form the past tense in a similar manner do not have alternate pronunciations, for example, *freak* [fɹik], *freaked* [fɹikt] and *peak* [pik], *peaked* [pikt]. Since we have words that form the past tense regularly and words that have an alternate pronunciation in the past tense and we can determine nothing from the phonetic contexts, it is impossible to attempt internal reconstruction in the way we did with *sign* and *signal*. In cases such as this, internal construction does not give us the answer.

Although internal reconstruction is most often used to discover sound changes using morphophonemic alternations in the synchronic state of the language, it can also be used to recover other kinds of language change. We can hypothesize about structure at earlier stages of the language by comparing variants and patterns in the current state of the language.

Even though there are limits to what internal reconstruction can reliably tell us about the history of a language, it can be very useful when working on languages for which evidence of related languages or previous stages of the language is not available.

13.7.3 Comparative Reconstruction

Unlike internal reconstruction, comparative reconstruction relies on the existence of multiple related languages; these are compared in order to establish what language the related languages descended from and how closely related they are.

In order to use the **comparative method** of reconstruction, you must start out with related languages, using the techniques discussed in File 13.2 on language relatedness. Otherwise, you would be "reconstructing" a system that would not represent any actually occurring language. By working with related languages, you know that you can at least theoretically reconstruct an actual source language from which the languages you are working with have descended.

Another key to using the comparative method successfully is the assumption (discussed in File 13.3) that sound change is regular; that is, all the sounds in a given environment will undergo the same change, and when a language undergoes a certain sound change, that change will (eventually) be reflected systematically throughout the vocabulary of that language. For example, a language might undergo an unconditioned sound change

of [p] to [f], in which every [p] in every word is replaced by [f]. Or, for example, a language might undergo a conditioned sound change of [p] to [f] in some specific phonetic environment, such as between vowels, in which case every word with a [p] between two vowels would develop an [f] in place of the intervocalic [p]. A sound change may be conditioned by phonetic environment (e.g., it occurs only when the sound in question is between two vowels, or before a certain other sound, or after a certain sound, or at the beginning of a word, or at the end of a word, etc.), but nothing other than the phonetic environment ever limits a sound change. A sound change never randomly affects some words but not other phonetically similar words, never occurs just in words with a certain kind of meaning, and so on. That is what is meant by the regularity of sound change.

These two tendencies make it possible for linguists to establish language relationships. The arbitrary relationship between a word's form and meaning is important because it makes it highly unlikely that unrelated languages will share large numbers of words of similar form and meaning. The regularity of sound change is important because it means that two (or more) languages that are related will show regular **sound correspondences**. Let us consider an example to illustrate what we mean. Consider the forms in (1).

(1)

English	***German***	***Dutch***	***Swedish***	***Gloss***
[mæn]	[man]	[man]	[man]	'man'
[hænd]	[hant]	[hant]	[hand]	'hand'

If we compare the vowel sounds in all four languages, we can establish the following sound correspondence in the word meaning 'man': [æ] in English corresponds to [a] in German, Swedish, and Dutch. In order for this sound correspondence to be regular, it must occur in other **cognates,** words that have similar forms and meanings and that are descended from the same source. And, of course, it does, as a comparison of the words meaning 'hand' confirms.[1] Note that since this correspondence (æ—a—a—a) occurs regularly (is not unique to the word for 'man'), we have eliminated the possibility of being misled by chance similarity between words with similar form and meaning in unrelated languages.

Although comparative reconstruction is most often used to discover sound changes and reconstruct the phonetic form of words, it can also be used to recover other kinds of language change. We can hypothesize about structure at earlier stages of the language by comparing variants and patterns in two (or more) languages that are related.

The task of the comparative linguist does not end with the discussion of correspondences between languages or with the assumption that these correspondences indicate that the languages in question are related. The linguist is also interested in discovering how languages that are related developed from the protolanguage into their present forms; in other words, the linguist is interested in linguistic history.

In order to discover how languages have developed from a protolanguage, the protolanguage itself must be recoverable. And in some cases it is. For the Romance languages (French, Spanish, Portuguese, Romanian, etc.) the protolanguage (Vulgar Latin) is attested by numerous written records, for example, manuscripts, public inscriptions, funerary inscriptions, graffiti, and so on. As a result it is possible to trace the development of the various Romance languages from their parent with considerable accuracy.

In other cases, however, written records for the protolanguage do not exist. But this does not mean that we cannot gather any information about the protolanguage; in these cases it is possible to infer what the protolanguage looked like by comparing the forms and grammars of the related languages. For example, some words in Proto-Indo-European can be reconstructed on the basis of words in the daughter languages. The lists in (2) and (3)

[1]Actually, we would want to see many more than just two words with the same correspondences, but these serve as an example.

contain sets of words having the same meaning from six Indo-European languages. The asterisk (*) means that the word is a **reconstructed form,** or a **protoform,** not one that we have ever seen attested by people who spoke the language.[2]

(2)	***'father'***	***'mother'***	***'brother'***
Proto-Indo-European	*[pəteːr]	*[maːteːr]	*[bʰraːteːr]
English	[fɑðɻ]	[mʌðɻ]	[bɹʌðɻ]
Greek	[patɛːr]	[mɛːtɛːr]	[pʰraːtɛːr]
Latin	[patɛr]	[maːtɛr]	[fraːtɛr]
Old Church Slavonic	—[3]	[mati]	[bratrə]
Old Irish	[aθɪr]	[maːθɪr]	[braːθɪr]
Sanskrit	[pɪtər-]	[maːtər-]	[bʰraːtər-]

(3)	***'mead'***	***'is'***	***'I bear'***
Proto-Indo-European	*[medʰu]	*[esti]	*[bʰer-]
English	[mid]	[ɪz]	[bɛɹ]
Greek	[mɛtʰu]	[ɛsti]	[pʰɛrɔː]
Latin	—	[ɛst]	[fɛroː]
Old Church Slavonic	[mɛdə]	[jɛstə]	[bɛrõ]
Old Irish	[mið]	[is]	[biru]
Sanskrit	[mədʰu]	[əstɪ]	[bʰəraːmɪ]

Since inferences are made by comparing words of similar form and meaning in the languages we assume to be related, the method is called the *comparative method*. Note that the comparative method is itself possible because of the regularity of sound change. If two or more languages show regular correspondences between themselves in words where the meanings are the same or similar, it means that these words have descended from a common source.

As a small preliminary example of how the comparative method works, let us return to our English-German-Dutch-Swedish example from (1). We note that the first consonant in the first word is an [m] and that the final consonant is an [n] in all four languages. Thus we can safely assume that the protolanguage had an initial *[m] and a final *[n] in the word meaning 'man,' so that at this point we can reconstruct *[m__n] in our protolanguage. With respect to the vowel sound there is some uncertainty because there is variation in the sound: English has [æ], while German, Dutch, and Swedish have [a]. However, since there are more [a] outcomes in the daughter languages than [æ] outcomes, assuming that [a] is the sound that the protolanguage possessed and that English alone has changed *[a] > [æ] allows for a simpler solution overall, with fewer changes needing to be posited. Thus we reconstruct the protoform for 'man' as *[man], and the sound change *[a] > [æ] ("*[a] changes to [æ]") for English.

13.7.4 Comparative Method Procedure

The goal of the comparative method is to reconstruct the protoforms of the protolanguage from the comparison of languages that are assumed to be related. Once the protolanguage forms have been reconstructed, it is possible to determine the changes by which the daughter languages have become distinct by comparing the protoforms with the forms present in the daughter languages.

---

[2]Note that this is the same symbol that we use for marking ungrammaticality. Generally you can tell from context which meaning is intended.

[3]There was an OCS word for 'father' [otəʁə], but it derives from a different root.

a. Compile Cognate Sets, Eliminate Borrowings. The first step is to gather and organize data from the languages in question, forming cognate sets. A cognate of a word is another word that has descended from the same source; consequently, cognates are very similar in form and are usually identical or similar in meaning. As an example of a cognate set, consider the words for 'keel' in four Austronesian languages (specifically, the Polynesian branch of Austronesian).

(4)	***Samoan***	***Māori***	***Fijian***	***Hawaiian***	
	[taʔele]	[takere]	[takele]	[kaʔele]	'keel'

Because of their semantic identity and phonetic similarity, these four words form a cognate set.

While gathering cognates, you should make sure that "suspicious-looking" forms are eliminated. Sometimes among the cognate sets you are compiling for some group of languages, there will be a cognate set with an "oddball," a form that is phonetically so different from the other members of the cognate set that it is improbable that it derived from the same source. The "oddball" may have been borrowed from some other possibly genetically unrelated language. The original form, which fit the cognate set, was probably dropped in favor of the borrowed form. When you come across one of these borrowed forms, simply ignore it for the purposes of the comparative method.

b. Determine Sound Correspondences. Next determine the sound correspondences that exist between sounds in the same positions in the words in each cognate set. The sound correspondences for our cognate set in step (a) are given in (5).

(5)	***Position***	***Samoan***	***Māori***	***Fijian***	***Hawaiian***
	1.	[t]	[t]	[t]	[k]
	2.	[a]	[a]	[a]	[a]
	3.	[ʔ]	[k]	[k]	[ʔ]
	4.	[e]	[e]	[e]	[e]
	5.	[l]	[r]	[l]	[l]
	6.	[e]	[e]	[e]	[e]

c. Reconstruct a Sound for Each Position. Given these sound correspondences, you must try to determine the earlier protoform from which the cognates have descended, following these steps **in this order:**

(i) *Total Correspondence.* If all the languages exhibit the same sound in some position in a cognate set, reconstruct that sound. In our example, in positions 2, 4, and 6, each of the languages has the same vowel, so we construct [a] for position 2, [e] for position 4, and [e] for position 6. Leaving blanks for positions that do not have total correspondence, we can collapse and write this information as *[_a_e_e].

(ii) *Most Natural Development.* For each of the remaining positions, if possible, reconstruct the sound that would have undergone the most **natural** sound change. Years of study in phonetics and historical linguistics have shown that certain types of sound changes are very common, while others almost never happen. For example, in a position between vowels, the change of a stop to a fricative at the same point of articulation is a very common change, while the reverse is much less common. Thus, if one cognate contains a stop between vowels and the other contains a fricative, the stop should be reconstructed. For each of the common sound changes listed in (6), it should be understood that the reverse direction of change is rare.

(6) Common sound changes
- Voiceless sounds become voiced between vowels and before voiced consonants.
- Stops become fricatives between vowels.
- Consonants become palatalized before non-low front vowels.
- Consonants become voiceless at the ends of words.
- Difficult consonant clusters are simplified.
- Difficult consonants are made easier (for example, voiced aspirated stops might become plain voiced stops).
- Oral vowels become nasalized before nasals.
- Fricatives other than [h] become [h], and (voiceless) stops other than [ʔ] become [ʔ].
- [h] deletes between vowels.
- Clusters of vowels are broken up by consonants.

In our Austronesian words for 'keel,' for example, we have a choice between [k] and [ʔ] in position 3. Because we know that stops other than [ʔ] often become [ʔ], we reconstruct [k] so that *[k] > [ʔ] in Samoan and Hawaiian. So at this point, we have constructed *[_ake_e].

(iii) *Occam's Razor.* This technical term refers to a guideline for evaluating competing analyses: given any pair of possible analyses, prefer the one that is simpler overall. In the case of historical linguistics, this translates into preferring a solution that requires the positing of fewer changes over one that covers the same facts but requires more changes to do so. (*Occam's Razor* is named for the medieval English philosopher William of Occam, who proposed the principle, and "razor" refers here to the way the guideline encourages the "cutting out" of extra complications.) We have already applied this principle in the English-German-Dutch-Swedish example above when it was suggested that a single change *[a] > [æ] for English was a simpler solution than having three instances of a change *[æ] > [a].

So for position 1 in our example, where we have a choice between [t] and [k], we reconstruct *[t], because this would involve a single change *[t] > [k] for Hawaiian. To choose *[k] would require us to posit three instances of the change *[k] > [t], separately in Samoan, Fijian, and Māori. Similarly, for position 5, we reconstruct *[l], since this will involve only a single change *[l] > [r] in Māori, rather than three instances of the change in the other direction in the other three languages. Using the comparative method, then, we have determined that the pronunciation of the word meaning 'keel' in the protolanguage from which Samoan, Māori, Fijian, and Hawaiian descended was most probably *[takele].

d. Check for Regularity of Sound Change. Although the procedure outlined in steps (a) through (c) can be used to reconstruct a protoform for each cognate set individually, you must check to see whether your results are consistent across the whole collection of cognate sets. We know that sound change is regular, and therefore we should be able to give for each daughter language (Samoan, Māori, Fijian, and Hawaiian in our example) a list of sound changes that applied regularly to all words in the protolanguage, resulting in the respective daughter languages. If you cannot formulate the sound changes, you must minimally modify the choices you made in step (c) so that your results conform to the regularity hypothesis.

In order to demonstrate this situation, we need to add another cognate set to our data, as shown in (7).

(7)

Samoan	***Māori***	***Fijian***	***Hawaiian***	
[taʔele]	[takere]	[takele]	[kaʔele]	'keel'
[tapu]	[tapu]	[tabu]	[kapu]	'taboo'

Confirm that steps (a) through (c) produce *[tɑpu] for the word meaning 'taboo.' We see that they do, with the further addition that *[p] > [b] (intervocalically) in Fijian, and that the sound changes listed in (8) apply regularly to both reconstructed forms, giving the correct forms in Samoan, Māori, Fijian, and Hawaiian.

(8) a. Derivation of 'keel'

	Samoan	***Māori***	***Fijian***	***Hawaiian***
Protoform:	*[takele]	*[takele]	*[takele]	*[takele]
Sound changes:	*[k] > [ʔ]	none	none	*[k] > [ʔ]
	none	none	none	*[t] > [k]
	none	*[l] > [r]	none	none
Cognate set:	[taʔele]	[takere]	[takele]	[kaʔele]

b. Derivation of 'taboo'

	Samoan	***Māori***	***Fijian***	***Hawaiian***
Protoform:	*[tapu]	*[tapu]	*[tapu]	*[tapu]
Sound changes:	none	none	none	*[t] > [k]
	none	none	*[p] > [b]	none
Cognate set:	[tapu]	[tapu]	[tabu]	[kapu]

This example was fairly straightforward, but comparative reconstruction can be challenging when more sound changes take place between the reconstructed form and the cognate set. In some cases, when multiple changes affect the same sound, or one change provides the conditioning environment for another, the order in which the changes take place is very important, as with the phonological rules discussed in File 3.3.

Another clue that you may find helpful in doing comparative reconstruction is to find a pair of words that is the same (homophonous) in language A but different in language B. When such a situation arises, you may be fairly confident in reconstructing the protoforms as they appear in B (or at least as being different from one another, unlike in A). This reconstruction follows from the fact that if you were to reconstruct both the forms as they appear in A (i.e., as identical to each other), there would be no way that they would subsequently differentiate themselves in B: no sound change can apply to only one of two homophones. This is illustrated by the data in (9) from languages in the Gbe branch of the Niger-Congo family.

(9)

Gen	***Fon***	***Gloss***
[tó]	[tó]	'ear'
[tó]	[só]	'pound'

Given the data in (9), we would have to reconstruct the protoforms *[tó] 'ear' and *[só] 'pound,' and hypothesize a sound change *[s] > [t] in Gen. If we had instead reconstructed both forms as *[tó], as they occur in Gen, it would be impossible, given the regularity of sound change, to write a rule by which one of them (but not both) changes to [só] in Fon.

The flowchart in (10) should help you work through a set of data to reconstruct earlier forms of words that are related in several languages. The rectangular boxes ask you to do something or give you some information that your working through the flowchart has revealed. The diamond-shaped boxes pose a question. Try reading through the flowchart before you attempt to solve a reconstruction problem like those found in File 13.8; it may help you understand how the whole process works.

(10) Flowchart for reconstructing word forms using the comparative method