Student Resource

# *Guide to Historical Reconstruction via the Comparative Method*

The goal of the comparative method is to use a process of deduction to posit the probable word shapes and phonemic inventory of an ancestral language based on attested forms in two or more daughter languages. The comparative method uses analytical skills and concepts from phonology, but applies them diachronically (over time), rather than synchronically (at a single point in time).

There are established steps to the comparative method. By following these steps closely, you can reconstruct proto-words, both sounds and meanings. In more advanced stages, it is possible to apply the method to grammar. This guide will take you step-by-step through the comparative method, focusing on the phonological forms of proto-words.

After you have completed an analysis using the comparative method, the problem should be written up in prose and the analysis justified with evidence. The student resource entitled "Writing for Linguistics – Historical Reconstruction" discusses the conventions for writing in this sub-field and advice on presenting an analysis.

**STEP 1: Hypothesize a genetic relationship between two or more languages based on similarity.**

Historical reconstruction really begins when a linguist notices that a set of languages have a greater number of surface similarities than can likely be attributed to chance or language contact. The following famous quote, presented by Sir Williams Jones in 1786 to the Royal Society of Bengal, reflects this stage in the study of Indo-European.

> "The Sanskrit language, whatever its antiquity, is of a wonderful structure; more perfect than Greek, more copious than Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists: there is a similar reason, though not quite so forcible, for supposing that both the Gothic and the Celtic, though blended with a very different idiom, had the same origin with the Sanskrit; and the old Persian might be added to the same family…"

In beginning linguistics classes, this step is skipped and students are given a small set from languages already hypothesized to be related.

**STEP 2: Gather words with similar meanings in each language. Discard borrowings.**

This is called creating or assembling **cognate sets**. Cognate sets are words in two or more related languages that have descended from the same word in the proto-language. Words in cognate sets almost always have semantic and phonological similarities. In introductory exercises, cognate sets are already assembled and presented to students for further analysis.



*How Languages Work: An Introduction to Language and Linguistics, Revised Ed.*

**Student Resource: Guide to Historical Reconstruction via the Comparative Method**

**STEP 3: List sets of sound correspondences and the environments in which they are found.**

Both pieces of this step are important in historical reconstruction. If you reliably do this with ALL cognate sets, you are unlikely to miss something and the analysis will be easier. Being less than exhaustive and accurate in this step (i.e., taking shortcuts) can lead to loss of critical information, hence frustration and inability to complete the problem. See Textbox 1 for advice on listing the environments.

Consider the word for 'to see' from the Central Tungusic languages:

| **Nanai** | **Ulcha** | **Udihe** | **Oroch** | **Gloss** |
|---|---|---|---|---|
| itʃə | itʃə | isə | itʃe | 'to see' |

The correspondence sets for each sound should look like this; note that the environments are also listed:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| i | : | i | : | i | : | i | | #_ |
| tʃ | : | tʃ | : | s | : | tʃ | | i_ə (see Textbox 2) |
| ə | : | ə | : | ə | : | ə | | _# |

**Textbox 1: Listing the Environments**

**Should I write the specific environment (i_ ə) or the general environment (V_V)?**

When you have an environment such as i_ə, you could list it as such or you could generalize to V_V. While it may seem obvious at first that the more general notion of "between vowels" is producing the change in Udihe *tʃ > s, this assumption may not be warranted. If you wrote the general environment V_V everywhere, it would be hard to see a pattern where a sound change was triggered by some vowels and not others. You would need to go back and rewrite the specific environments in order to proceed with the analysis.

Going from the specific to the general is much easier. You would do this in light of more data, for example, the same correspondence set also occurring in the environments o_a, e_u, etc. If you can show that any of the vowels of the language can both precede and follow the sound that has undergone change, then you can generalize to V_V.

Keeping track of the phonetic detail initially by writing the specific environment may seem tedious, but it will ensure that you always have access to the most information and it could prevent you from having to go back and redo your analysis later.

***As you list multiple words, look for patterns in the environments***. This will help in several ways with the next step. Here are some questions to ask yourself when you are looking for patterns:

- Do particular sets of sounds occur over and over in certain environments? Here we would look for more correspondences of tʃ / tʃ / s / tʃ.
- Do some sounds only occur at the beginnings of words? At the ends? Between vowels? Before sounds that share a feature (such as velar)?



*How Languages Work: An Introduction to Language and Linguistics, Revised Ed.*

- Do any of these environments bring to mind certain phonological processes, such as palatalization (e.g., a correspondence set like t / t / t / tʃ that is only found in the environment __ i, suggesting palatalization)?

**STEP 4: Reconstruct proto-sounds for each set of correspondences.**

This is the most important and interesting part of the analysis. The reconstruction must be based on sound argumentation as well as on considerations such as phonetic naturalness and overall simplicity of the analysis.

| Nanai | | Ulcha | | Udihe | | Oroch | Environment | Reconstruction |
|---|---|---|---|---|---|---|---|---|
| i | : | i | : | i | : | i | #_ | *i |
| tʃ | : | tʃ | : | s | : | tʃ | i_ə | *tʃ |
| ə | : | ə | : | ə | : | ə | _# | *ə |

The sets i / i / i / i and ə / ə / ə / ə are **identity sets**, meaning the same sound is found in all languages. For these, it is simple to reconstruct that sound in the proto-language. It allows for the simplest analysis, since we say that no change occurred in any of the languages regarding this sound. All other things being equal, we always reconstruct the sound that will give us the simplest analysis.

The set tʃ / tʃ / s / tʃ is more complex as there are two sounds realized in the modern languages. Let's examine two hypotheses for the proto-sound: * tʃ and *s. If we reconstruct *tʃ, then we would postulate that *tʃ > s / i__ə in Udihe. Further data (not presented here) shows the same pattern occurs with any vowels, so the sound change could be generalized to *tʃ > s / V__V. This sound change is phonetically motivated, as it represents a type of **lenition**, a weakening of the sound from an affricate to a fricative, that is commonly found between vowels.

The other possible analysis would start by reconstructing *s. Then we would need to posit that *s > tʃ / V_V in Nanai, Ulcha, and Oroch. Such a change would have involved **fortition**, a strengthening of the sound from a fricative to a stop, which is unlikely to happen between vowels. It also would have the change occurring in three languages, so the analysis is more complex overall compared to the one that posits *tʃ. For all of these reasons, *tʃ is the best choice for the reconstruction.

For more advice on this important step in reconstruction, see Textbox 2.

**STEP 5: Determine the shapes of words in the proto-language.**

Using your reconstructed sounds from each correspondence set, you can now create proto-words. Based on our analysis, the word 'to see' in Central Tungusic can be reconstructed as *itʃə. Each segment was reconstructed individually, and then assembled into a proto-word.

**STEP 6: List the sound changes that each daughter language underwent.**

Once you have a list of proto-words, you can now better understand the sound changes each language underwent.

- Make a list of the sound changes that occurred in each of the daughter languages.

- You can group similar sound changes together if they share natural classes. If you had more data, you'd see that in Udihe, *dʒ > d / V_V as well. You can now write these changes together using phonological feature notation. *post-alveolar affricates > fricatives / V_V.

- Remember, you must have a sound change each time there is a difference between the proto-word and the corresponding word in the daughter language – a sound doesn't just "change here, but not there." This is a critical component of the comparative method; it works because sound change is regular.

- It is a good idea in the written portion of your analysis to also put each sound change into words, especially when sound changes apply over natural classes of sounds, e.g., "All affricates in Proto-Central Tungusic change to an alveolar stop in Udihe."

**STEP 7: Check your work!**

This is another step that is often forgotten, despite its ability to immediately diagnose problems with a reconstruction. If you check your work (in a tidy and clear fashion), there is less chance you will miss anything and someone reading your analysis can follow your argumentation.

- If your reconstruction and list of sound changes are accurate, you should be able to apply those sound changes to each proto-word and end up with the (modern) words in each respective language.

- If you find that this doesn't work, go back to your correspondence sets. You may have missed a conditioning environment or reconstructed the same phoneme for different correspondences sets that cannot be reconciled.



*How Languages Work: An Introduction to Language and Linguistics, Revised Ed.*

**Step 8: Write up your analysis**: See the student resource: Writing for Linguistics: Historical Reconstruction.

**Textbox 2. Advice for reconstructing proto-sounds**
How do you know which sound to choose? There are many guiding factors, some of which you have already learned from phonology.

- Sometimes the adage "**majority rules**" works: reconstruct the sound that occurs in the majority of daughter languages. This appears to work in the Central Tungusic example illustrated here – in three of the four languages, the reflex of *tʃ is /tʃ/. This is often a good starting point.

- However, **phonetic plausibility** is also critical. Consider the correspondence sets given below:

  n : ŋ : ŋ : ŋ / ____ [velar]
  ŋ : ŋ : ŋ : ŋ / all environments

  The two sets **overlap**, in that the same sounds are found in both sets. If we looked only at the first set, we might want to use "majority rules" to reconstruct *ŋ. However, the second set, being an identity set, also should be reconstructed as *ŋ. The key here, of course is the specific environment of the first set: the following velar consonant. Velars commonly cause preceding nasal consonants to assimilate in place of articulation (e.g., English *ink* [iŋk]). We can thus reconstruct *n for the first set, and posit the sound change *n > ŋ /__ [velar]. Phonetic plausibility is key here and trumps "majority rules."

- As the previous example showed, reconstructing proto-sounds will test your knowledge of phonology and of common sound changes. Common phonological operations are also common sound changes, such as lenition, devoicing, or palatalization. There are other changes that just happen to be very common cross-linguistically, such as *s > h > □. You will become familiar with these the more reconstruction you do and the more that you read about sound change.

- It is generally better to postulate a more common sound change (such as stops becoming fricatives between vowels) than the opposite (fricatives becoming stops between vowels), unless there is very strong and compelling evidence to the contrary. This is generally referred to as **directionality**. While *s > h is a common sound change, *h > s is nearly unattested. This means that in a correspondence set like s:h:s:h where you might be choosing between reconstructing *s or *h, *s is the best choice.