Language Families of the Native Americas

Language families are established by comparing structural similarities between languages through using a number of different strategies commonly grouped under the term *comparative method*. The classification of languages into families is a large topic that covers a broad range data. I focus on only a small number of issues essential to the topic. I highly recommend one of the introductory texts to historical linguistics for more details.

Basics of Language Families

Proposals for Native American language families have a long and rich history of controversy and success. The controversies stem, in some part, because of the poor documentation of many native languages and their increased rate of extinction. It is safe to say that many native languages of the Americas will cease to be spoken in the following years, just as many native languages have gone extinct in the last one-hundred years.

Originally, starting around the late 1800's, many people thought the languages of the Americas (from Canada to South America) belonged to one large family; but as work and interest in these languages grew a different picture emerged. It is now hypothesized that Native America is home to roughly 180 language families comprising over 2,000 languages (this is approximately 1/3, or 30%, of all languages on earth!). However, there are still attempts to classify Native American languages into one large family or a small group of families, but these classifications are highly speculative and very controversial – they are not typically accepted by professional comparative linguists because the

evidence is not very convincing. See suggested further reading, specifically Lyle Campbell, Joseph H. Greenburg, and Merritt Ruhlen.

In determining the classification of language families there is one major assumption about languages that is shared by all linguists: All modern languages are the descendents of languages from the past. Embedded in this assumption is the crucial thesis that languages change over time, and that the speakers of one common language can split into different groups that also change over time. The accumulated evidence for this is overwhelmingly positive. For an accessible example, languages such as French, Italian, and Spanish are all descendents of a common mother language: Latin. This is captured in the model of a family tree that expresses genetic relationships – with older languages called mother languages and descendents called daughters. In establishing language families one looks for structural similarities that can be used to *reconstruct* what are called Proto-languages. Latin is the Proto-language, and mother, of French, Italian, and Spanish.

How to define a Language Family

Evidence used for defining a language family is most commonly found by looking at cognates. A cognate is a word that has similar pronunciation and similar meaning in two different languages. The most reliable types of cognates used for comparison are words that belong to basic vocabulary. This is vocabulary that refers mainly to kinship terms such as 'mother,' 'brother,' 'father;' to body part terms such as 'hand,' 'head,' 'arm;' to low numbers such as 'one' through 'five;' and to important aspects of the natural environment. Cognates make good evidence because they tend to be resistant to borrowing. But relying only on cognates of basic vocabulary is not itself sufficient

evidence for proposing family relations. A disciplined understanding of sound patterns (phonology and phonetics) and sound change is necessary to make good cognate comparisons. And in fact, sound correspondences found in cognates constitutes the main body of evidence for deriving reconstructions. Reconstruction refers to the process of discovering sound correspondences in different languages that are used to *reconstruct* one sound of the hypothesized ancestral Proto-language, from which the various languages are historically derived. One can also reconstruct entire words.

For example, if we compare six languages (Kaqchikel, Tz'utujil, Poqomam, Uspanteko, and Q'eqchi') in the K'ichean branch of the MAYAN family, we find that the word for person in five languages is *winaq* and in one language, Q'eqchi', it is *kwinq*. We can see that in five of the six languages the sound correspondences are exact, but in Q'eqchi' they are not (the linear sequence *w-i-n-a-q* does not correspond exactly to the sequence *k-w-i-n-q*). One can propose that in Q'eqchi' the *a* between *n* and *q* was deleted, and also, that a *k* was added in front of the *w*. And in fact, if we could look at more data we would see that in all instances where the other five languages begin a word with *w*, Q'eqchi' has a *kw*. For this reason, the reconstruction of 'person' in Proto-K'ichean is *winaq*. In this case, we only need to propose changes in one language, Q'eqchi'. However, if we reconstructed *kwinq* as the proto-form, then we would be implying that the five other languages shared the exact same changes as each other in deriving *winaq* from *kwinq*. It is much more economical to assume that only one language changed, and therefore, we have opted for the simplest solution. Although this example has been simplified for expository purposes, it reflects general methodology.

An additional problem to proposing language families is the possibility that a language's basic vocabulary has been borrowed. This poses problems for relying only on sound

correspondences in cognates. Other structural similarities such as sentence structure (syntax) and word structure (morphology) are also used for comparison – though comparative syntax and morphology are more complicated and time consuming than comparative phonology and vocabulary.

What is the Comparative Method?

The comparative method is not a step-by-step process with precisely defined instructions. Instead, it is a group of related criteria and general methods for reconstructing the most reliable proto-forms from a set of data. I list here the general processes and criteria associated with the comparative method, taken from Lyle Campbell's introductory text on historical linguistics. See suggestions for further reading, in particular any introduction to historical linguistics, for more details on the comparative method and its history, problems, criticisms, and advances. But also, the example of the K'ichean languages given above is a short and simplified version of the comparative method in practice.

The first step is to find and assemble cognates, preferably of basic vocabulary. Second, establish sound correspondences, preferably using a large set of data. Third, reconstruct the proto-sound using four guiding principles: (i) be familiar with the way sounds change in terms of what is physiologically-acoustically possible in humans and what is common in the world's languages – this is called directionality of change; (ii) majority wins, as seen in the K'ichean example above where we reconstructed the proto-sound *w* based on five out of six descendents that had it; (iii) use as much data as can be found, the more information the better – this way common features can be compared in language descendents; and (iv) reconstructing the most economical proto-forms reduces

the likelihood of mistakes. The K'ichean example above utilizes the notion of economy by assuming that it is more economical to say that one language changed from the Proto-language than did five. Remember, extraordinary claims require extraordinary evidence.

Examples of Language Families

I list here some language families which have a sufficient amount of support and are generally accepted by professional comparative linguists, focusing more detail on families in North America. It should be pointed out that the characterization of the size of a family by referring to a specific number of languages reflects my own opinion. There is no widely accepted numerical definition for size of language family. I merely want to give a sense of the relative proportion of "large" and "small" by highlighting a numerical range. For more details on language families see suggestions for further reading, particularly the work of Lyle Campbell, Ives Goddard, Marianne Mithun, and references cited in those works.

The Eyak-ATHAPASKAN family is a large family (over 45 languages) and extends from Alaska to Mexico.

SALISHAN is a medium sized family (over 20 languages) in British Columbia, Vancouver Island, Washington, Oregon, Idaho, and Montana.

Wakashan is a fairly small family (less than 10 languages) in British Columbia, Vancouver Island, and Washington.

HAIDA is a small family (with 2 languages) located on the Queen Charlotte Island.

WASHOE, or Washo, is an isolate family (1 language). This means there are no known relatives to the language Washo, for which the family isolate is named. Isolates do not

appear to belong to other families nor do they seem to contain any other languages except the one for which it is named. Other known isolates include Alsea, CHITIMACHA, Kutenai or KOOTENAI, TLINGIT, and ZUNI.

Uto-AZTECAN is a large family in terms of both number of languages and number of speakers. It extends from Oregon in North America to Panama in Central America.

IROQUOIAN is a medium sized family and was probably the first Native American language recorded by Europeans in North America. It is found from parts of Canada – Ontario, Quebec – down into New York, Pennsylvania, Wisconsin, Virginia, North Carolina, and Oklahoma.

Algic, made up of the two major branches ALGONQUIAN and Ritwan, is a large family that covers considerable geographical ground. The Ritwan branch is found in the northern California coast, while the Algonquian branch extends from the west in Wyoming to the Atlantic seaboard in the east, and from the north in the subarctic to the south in Coahuila of northern Mexico.

Other language families include CHINOOKAN, COOSAN, KIOWA-Tanoan, KLAMATH-MODOC, MAIDUAN, MIWOK-Costanoan, SIOUAN, TAKELMAN, and TSIMSHIAN in North America; Otomanguean, Mixe-Zoquean, and Mayan in the Middle Americas; Maipurean/Arawakan – which is the largest family in the Americas with over 60 languages, including ARAWAK –, Puinavean, Tucanoan, Witotoan, QUECHUAN, Panoan, Tupian, and Cariban in South America.

Controversial Language Families

There is one major reason that controversial family proposals are made: lack of supporting evidence. This lack of evidence can be the result of using superficial cognate

similarities – which stems from an undisciplined use of the comparative method. But lack of evidence can also be the result of a lack of documentation of the languages concerned. The latter case is the more problematic. While improper use of the comparative method can always be restrained by criticism and reformed by better training, there is no way to document extinct languages that did not leave written records. Generally, where no documentation exists, no proposal can be made.

It may be interesting to speculate on possible families using distant genetic relations or linguistic pre-history, but it should be recognized that such proposals are problematic because there is either a lack of documentation of languages or an undisciplined use of the comparative method. However, this is not to say that controversial proposals may not some day be proved correct, just that extreme caution should be taken in accepting controversial proposals.

Probably the most controversial language family for the native Americas is the Amerind family, which consists of almost all 180 Native American families except Eskimo-ALEUT and the controversial Na Dene. Other problematic groupings include Eskimo-Uralic, Hokan, Na-Dene, Maya-Chipayan. CAYUSE is poorly attested and extinct. There are many more.

See Aleut, Algonquian, Arawak, Athapaskan, Aztec, Cayuse, Chinook, Chitimacha, Coosan, Haida, Inuit (Eskimo), Iroquoian, Kiowa, Klamath, Kootenai, Maya, Maidu, Miwok, Modoc, Quechuan, Salishan, Siouan, Takelma, Tlingit, Tsimshian, Washoe, Zuni

Suggested further reading:

Lyle Campbell, *American Indian Languages: The Historical Linguistics of Native America* (Oxford, Oxford University Press, 1997); Lyle Campbell, *Historical Linguistics: An Introduction* (Cambridge, MIT Press, 2004); Ives Goddard (ed.), *Handbook of North American Indians* (Washington, DC, Smithsonian Institution, 1997); Joseph H. Greenburg, *Language in the Americas* (Stanford, Stanford University Press, 1987); Hans Henrich Hock, *Principles of Historical Linguistics* (The Hague, Mouton de Gruyter, 1991); Roger Lass, *Historical Linguistics and Language Change* (Cambridge, Cambridge University Press, 1997); Marianne Mithun, *The Languages of Native North America* (Cambridge, Cambridge University Press, 1999); Ian Roberts, *Diachronic Syntax* (Oxford, Oxford University Press, 2007); and Merritt Ruhlen, *On the Origin of Languages: Studies in Linguistic Taxonomy* (Stanford, Stanford University Press, 1994).

Joshua Bowles