SES no.98; Introduction

Senri Ethnological Studies 98: 1–8 ©2018
Let's Talk about Trees: Genetic Relationships of Languages and Their Phylogenic Representation
Edited by Kikusawa Ritsuko and Lawrence A. Reid

1. Introduction

Kikusawa Ritsuko
National Museum of Ethnology, Japan
The Graduate University for Advanced Studies, Japan
Lawrence A. Reid
University of Hawai'i
National Museum of the Philippines

1.1. Background

This volume is a collection of papers presented at the International Symposium "Let's Talk about Trees" organized by Kikusawa and hosted at the National Museum of Ethnology, Osaka, Japan on February 10, 2013. The purpose of this symposium was to evaluate and examine what it means to apply a "tree model" to express linguistic relationships, and what the advantages and potential pitfalls are in doing so. Specialists of other disciplines such as cladistics, biology, and genetics, where diagrams are also used to express the "relationships" of targeted objects were also invited to participate.

A tree diagram has been the major means for expressing genetic relationships among languages in the field of Historical Linguistics. The model appears to efficiently reflect the results of the Comparative Method, the traditional method applied to clarify the genetic relationships of languages. At the same time, however, it has long been recognized that there are certain aspects of language change that cannot be expressed using a tree diagram, with such phenomena as contact relationships and the position of "hybrid" languages (such as creole languages) being examples. Recently, as new notions such as "dialect linkage" being introduced and many more detailed synchronic descriptions of languages becoming available, linguists have become more aware of the limitations that a tree diagram imposes. Attempts are now being made to "remedy" the situation by, for example, inserting different diagrams into the tree diagram, giving alternate analyses (Kikusawa 2015: 659), or completely replacing it and/or supplementing it with something else (François 2015: 183).

The family tree model (*Stammbaum*) has been used in the field of Historical Linguistics since it was first proposed by Schleicher (1853) and it has been there as a given, supposedly modeling language transmission from parent to child, with gradual changes, primarily in phonology. The model implies splits between groups, based on accumulated differences in the correspondence sets of proposed cognates, forms that had supposedly been transmitted from an earlier parent language, or proto-language. There are multiple problems to this view, however. One problem is that a family tree reflects

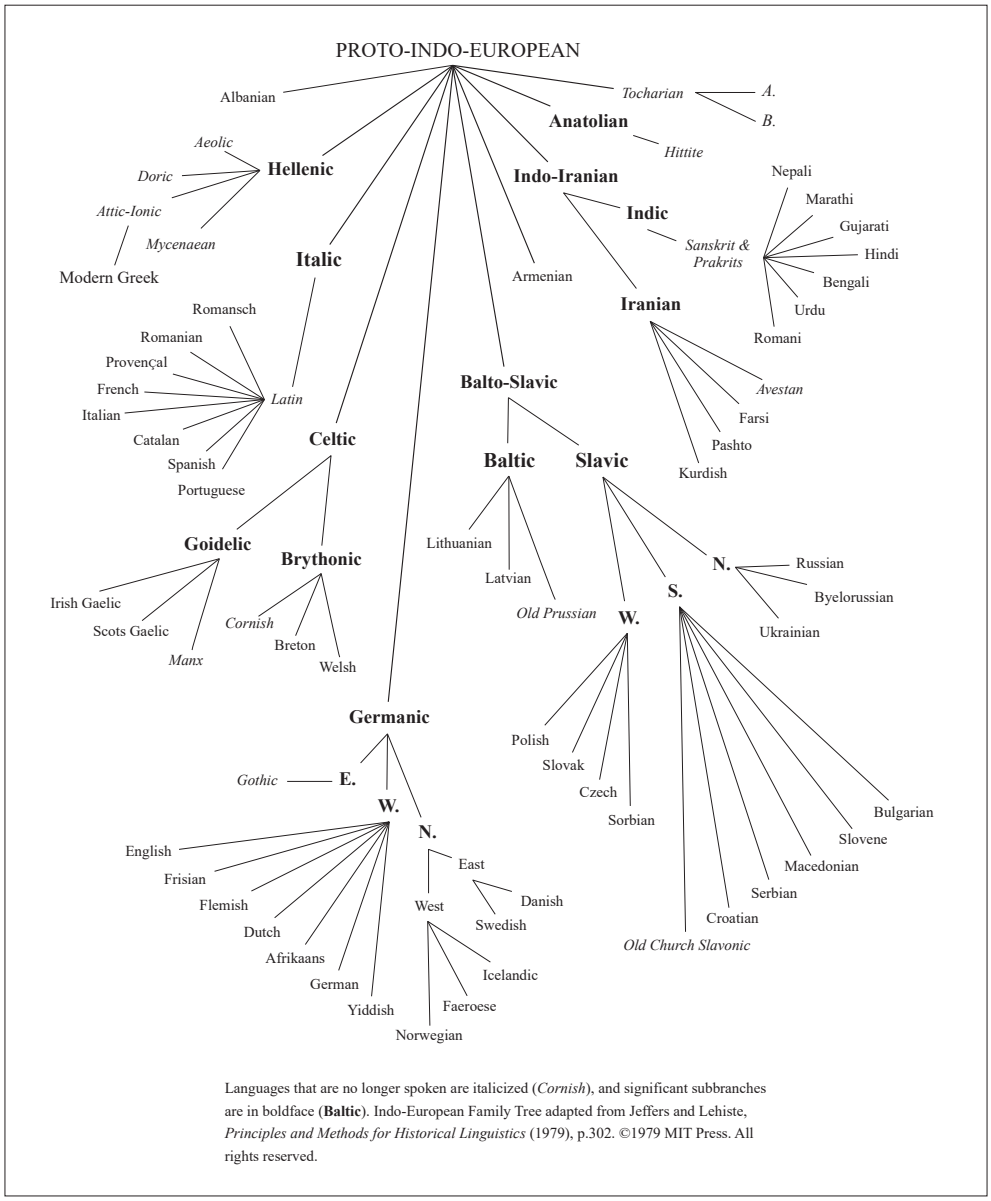
only changes in supposed cognate sets and does not show the effects of forms borrowed from other languages, either genealogically related or not. Another problem is that a family tree model implies complete splitting at some point in time between two groups. This is possible if a group moves to a distant geographical point, and there is little or no back movement between the groups. But such a situation must have been rare in the past. The usual situation would be dialectal difference maintained by mutual contact between the groups, with eventual accumulation of differences that would result in non-intelligibility between them. A third problem occurs when two groups share a body of lexical items with no distinctive sound changes. It is impossible to know whether such forms are inherited from a parent language, or have been borrowed. Another problem has to do with morphology, pronouns, affixes, clitics and related forms, and their functions. When are they the same, and when different? Similar problems have to do with syntactic details, such as word order and other typological features.

Even though there are obvious problems with using a tree diagram to model the relationship between languages, many introductory linguistic texts provide information about how to do comparative-historical linguistics, based on the regularity of sound change, and how to model the results. Typically, such descriptions cover the formation of family tree diagrams, and the problems associated with them. They typically cover also models of contact relationship, specifically the wave theory, first proposed by Johannes Schmidt in 1872 for Indo-European languages. Figures 1-1 and 1-2 are cited from a major introductory text, *Language Files* (Mihalcek and Wilson 2011), in which tree diagrams for the Indo-European language family (also for the Uralic language family), and a wave diagram of the Indo-European family are provided.

1.2. The Tree Model in Historical Linguistics

This section provides a background about Historical Linguistics, since many of the participants in the symposium were from different disciplines. It is a well-known fact that languages change across generations. This results in language diversification, when separate groups of related peoples change their language in different ways. The techniques of Historical Linguistics, first developed in the later half of the 19th century, are based on the premise that sound changes are systematic and regular, and sound correspondences across languages can then be used to establish the genetic relatedness of languages. Earlier sound systems are reconstructed based on the sound correspondences, and based on this knowledge, earlier lexical items and grammatical forms are also reconstructed. To illustrate this, consider the forms in Table 1-1.[1)]

On the basis of the forms in Table 1-1, certain assumptions can be made about their relatedness. Although there are only 5 words, the correspondence sets in Table 1-2 can be developed, and the sounds from which they developed can be reconstructed (these are the sounds in the top row of Table 1-2, and are preceded by an asterisk, the common device in historical linguistics to indicate a hypothetical, reconstructed sound). Since as noted earlier, sound change is systematic and regular, we can assume that the forms given with the sound changes provided are exemplary of many such forms with the same sound



Languages that are no longer spoken are italicized (*Cornish*), and significant subbranches are in boldface (**Baltic**). Indo-European Family Tree adapted from Jeffers and Lehiste, *Principles and Methods for Historical Linguistics* (1979), p.302.

Figure 1-1 An Indo-European family tree (based on Mihalcek and Wilson 2011: 528)

changes in each of the languages.

Figure 1-3 is a family tree diagram based on the sound changes (reflexes) of the reconstructed sounds shown in the top row of Table 1-2. Only three sets of consonant reflexes are illustrated, those in the boxed section with thick lines in Table 1-2. Those languages which share reflexes are grouped together. For example, there are three

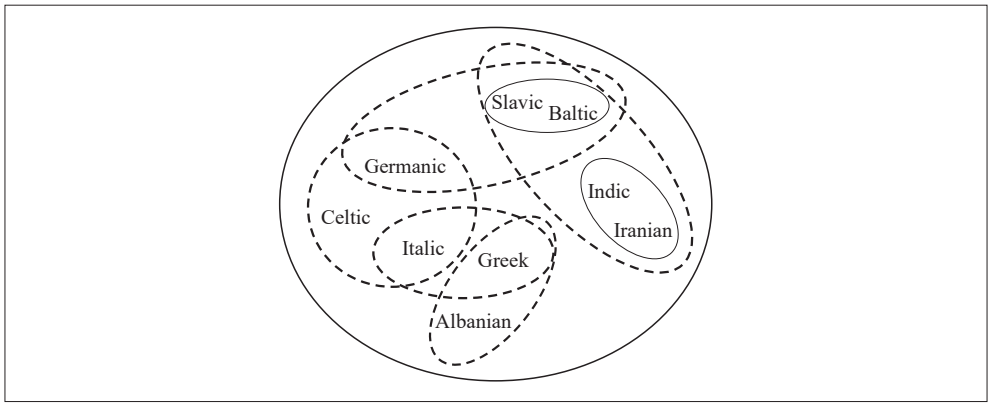


Figure 1-2 An Indo-European wave diagram (based on Mihalcek and Wilson 2011: 531)

Table 1-1 A sample cognate set of 5 words in 8 Austronesian languages (compiled by the authors)

	'sky'	'road'	'two'	'three'	'five'
Paiwan (Taiwan)	~~*kalevlevan*~~	*djalan*	*ḍusa*	*tjəlu*	*lima*
Bontok (Philippines)	~~*dáya*~~	*dálan*	*duwá*	*tulú*	*limá*
Malagasy (Madagascar)	*lání~~tra~~*	*lálana*	*rúa*	*télu*	*dimi*
Manam (Papua New Guinea)	*laŋ*	*jala*	*rua*	*toil*	*lima*
Tuvaluan (Tuvalu)	*laŋi*	*ala*	-	*tolu*	*lima*
Fijian (Fiji)	~~*lomā*~~-*laŋi*	*sala*	*rua*	*tolu*	*lima*
Niuean (Niue)	*laŋi*	*hala*	*ua*	*tolu*	*lima*
Rapanui (Easter I.)	*raŋi*	*ara*	*rua*	*toru*	*rima*

*Lexical items that are struck-through are identified as non-cognates and excluded from the analysis.

Table 1-2 Example of sound correspondences based on the dataset in Table 1-1 (compiled by the authors)

	*l	*d	*z	*m	*ŋ	*n	*a	*i	*ə	*u
Paiwan (Taiwan)	*l*	*d*	*dj*	*m*	-	*n*	*a*	*i*	*ə*	*u*
Bontok (Philippines)	*l*	*d*	*d*	*m*	-	*n*	*a*	*i*	*u*	*u*
Malagasy (Madagascar)	*l*	*r*	*l*	*m*	*n*	*n*	*a*	*i*	*e*	*u*
Manam (Papua New Guinea)	*l*	*r*	*j*	*m*	*ŋ*	Ø	*a*	*i*	*oi*	*u*
Tuvaluan (Tuvalu)	*l*	-	Ø	*m*	*ŋ*	Ø	*a*	*i*	*o*	*u*
Fijian (Fiji)	*l*	*r*	*s*	*m*	*ŋ*	Ø	*a*	*i*	*o*	*u*
Niuean (Niue)	*l*	*r*	*h*	*m*	*ŋ*	Ø	*a*	*i*	*o*	*u*
Rapanui (Easter Island)	*r*	*r*	Ø	*m*	*ŋ*	Ø	*a*	*i*	*o*	*u*

*The boxed section with thick lines indicates the set of data analyzed in Figure 1-1.

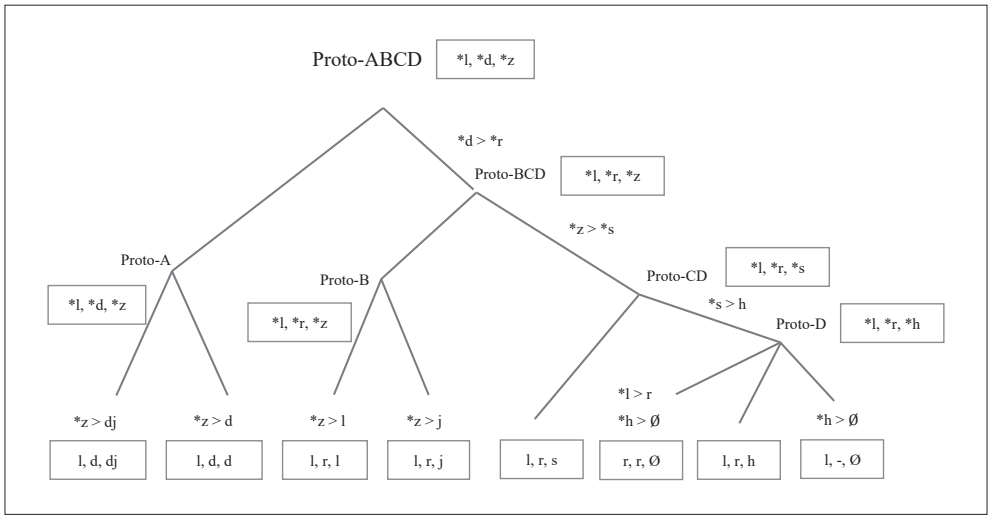


Figure 1-3 A sample tree diagram showing a subgrouping hypothesis based on the data given in Tables 1-1 and 1-2 (compiled by the authors)

reconstructed consonants in the parent language which are used to subgroup the languages, *l, *d, and *z. There are a group of 6 languages, that have changed *d to *r, these are grouped together as Proto-BCD, as against two languages, that keep *d unchanged. The two languages of Proto-A are distinguished by their respective changes of *z. One has changed it to /dj/, and the other to /d/. Two languages of Proto-B are part of the group that changed *d to *r, but they are distinguished from the other languages of this subgroup by their reflexes of *z. One has changed it to /l/ and the other to /j/. The four languages of Proto-CD are united because in their history, they change *z to *s. One language retains *s as /s/, while the others have changed *s to /h/ or have lost it completely. Other changes distinguish each of these three languages of Proto-D.

1.3. In This Volume

The chapters in this volume are organized in the following order.

In chapter 2, Nobuhiro Minaka, a specialist of biostatistics and evolutionary biology, discusses the phylogenetic relationship of different disciplines, such as evolutionary biology, textual stemmatics, and historical linguistics. All search for the best solution among possible alternative phylogenetic hypotheses, which may or may not be true to history. They share not only the basic characteristics of historical sciences but also those of data visualization and information graphics. He discusses how to understand the diversity formed by spatiotemporally changing objects with special reference to the problem of estimating the phylogeny of manuscripts, languages, and organisms. He points out that a common logic has been used by the humanities, social sciences, and natural sciences, despite the wide separation in research methods and data. No matter

what kinds of object were studied, almost the same methodology was independently used to reconstruct genealogical relationships among objects.

In chapter 3, Ryosuke Kimura, a specialist in evolutionary molecular genetics, discusses the development of a wide range of phylogenetic diagrams that have been proposed. He notes that increased computer processing power and the development of generalized programs allow greater access for performing phylogenetic analysis. However, he warns that population history is sometimes too complicated to be represented with a simple diagram. He suggests instead using a phylogenetic analysis in combination with other statistical methods, such as the principle component analysis. This would be more effective for elucidating population history. Remarkable advances have been made in the previous two decades in statistical methods to efficiently and accurately infer the demographic history of and relationships among human populations from genomic variation.

In chapter 4, Søren Wichmann and Taraka Rama, specialists in computational methods in descriptive and historical linguistics, discuss problems related to the Automated Similarity Judgment Program (ASJP) analysis of Austronesian languages. This is a project, they say, dedicated to the diachronic analysis of the world's linguistic diversity, including the specific task of language classification. These result in different classifications of Austronesian languages than those that are proposed by 'experts'. After discussing various proposed algorithms for developing phylogenetic classifications of Austronesian languages, they examine several problematic groups and conclude that future work concerned with methods in quantitative historical linguistics should be directed at the identification of a set of widely accepted gold-standard linguistic phylogenies which may serve as stable points of reference and help towards the improvement of the quality of evaluations.

In chapter 5, Siva Kalyan and Alexandre François discuss a specific case where a family tree diagram is not adequate for displaying the relationships of the languages. These are the 17 Oceanic languages spoken in northern Vanuatu. A tree structure is unsuited for dealing with dialect continua and language families that develop out of them, since the isoglosses of their innovations are not nested, as implied in a tree diagram, but constantly intersect. They propose a new methodology, which they label Historical Glottometry, and outline the procedures necessary for positing intersecting subgroups, and quantifying the strength of the genealogical evidence in favor of each language cluster.

In chapter 6, Lawrence Reid, a linguist specializing in the description and historical relationships of Philippine languages, discusses issues in the subgrouping of Philippine languages. He notes that traditionally, Philippine languages were considered to constitute a single group within the Austronesian language family, developing from a Proto-Philippines. But evidence points to a linkage or network of related languages including all Philippine linguistic groups and others further south because of the rapid spread of Malayo-Polynesian speaking peoples from their first movement out of what is now Taiwan to the western edges of the Pacific. This is supported by both linguistic and archaeological evidence. The paper also discusses how best to model the language

situation of Negrito peoples, the original occupants of the country, who gave up their original languages in favor of communicating with the Malayo-Polynesian group in their vicinity. Philippine languages have also been heavily affected lexically by Chinese, Spanish and English because of trade and other factors, creating modeling problems.

In chapter 7, Weera Ostapirat, a linguist who specializes in the languages of the East Asian mainland, discusses various macrophyletic issues of the five major language stocks of East Asia: Sino-Tibetan, Austronesian, Austroasiatic, Kra-Dai, and Miao-Yao. The chapter is primarily focused on the various proposed relationships of Chinese, noting that competing hypotheses have resulted in some linguists abandoning the tree diagram in favor of what is called a Falling Leaf model. He asserts that a relationship between Kra-Dai languages and Austronesian is now firmly established, based on evidence presented in the chapter and elsewhere.

In chapter 8, Yutaka Yoshida, a specialist in philological and historical studies of Sogdian and other Middle Iranian materials, presented his paper at the Symposium in Japanese. It is here presented in English translation, and deals with problems in drawing a tree diagram of Middle Iranian languages and dialects, when little is known of some languages that are no longer spoken, and whose only data are from ancient texts. He discusses at length linguistic philology and the stemma that are drawn, noting that although the stemma appears to show the chronological development of a text, that is not the case. It is a classification based on the similarities and differences that are reflected in the texts which are all from the same era, namely the tenth century. In other words, what the stemma shows is a typological and not a historical classification of the manuscripts. Discussion follows on the different tree diagrams of Indo-European languages, and particularly of the Indo-Iranian group. Yoshida concludes that when a linguist tries to attach more historical reality to a family tree than can be scientifically shown based on language data, such work would not only be useless but also detrimental to those who are ignorant of linguistics.

In chapter 9, Ritsuko Kikusawa discusses the inherent nature of tree diagrams, noting that they are useful in representing lineal long-term developments in a language family, but are misleading when viewing short-term developments. She distinguishes macro-comparison which is long-term comparison, and micro-comparison which is short-term comparison. She suggests that contact relationships, horizontal transmission, short-term developments of each language, and areal features are all associated with micro-comparison and belong to a different dimension from those expressed by a tree diagram. For examples, she focuses on Hawaiian which most people now speak as a dialect taught from the University of Hawai‘i, and sign languages which are rarely learned from parents, but are laterally transferred from friends or teachers in schools for the deaf.

Note

1) This information is provided as an example only. There are a wide range of factors that influence the choice of reconstructed sounds, and the ways they are reflected in individual

languages.

References

François, A.

2015 Trees, Waves and Linkages: Models of Language Diversification. In C. Bowern and B. Evans (eds.) *The Routledge Handbook of Historical Linguistics*, pp. 161–189. London and New York: Routledge.

Kikusawa, R.

2015 The Austronesian Language Family. In C. Bowern and B. Evans (eds.) *The Routledge Handbook of Historical Linguistics*, pp. 657–674. London and New York: Routledge.

Mihalcek, V. and C. Wilson (eds.)

2011 *Language Files: Materials for an Introduction to Language and Linguistics,* 11th ed. Columbus: The Ohio State University Press.

Schleicher, A.

1853 Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* 1853: 786–787.