

History of Polynesian Languages

Yuko Otsuka
University of Hawai‘i

We have been using the term “Polynesian languages” implying that these languages belong in a single group and that they are somehow related. But what does it mean to say that languages are related? What is the basis for the assumption that, say, Hawaiian and Samoan are related? There are two major methods that are relevant to these questions: the comparative method and lexicostatistics. The comparative method is a general tool that can be used to determine linguistic relationship, to reconstruct proto-languages, and to determine the subgrouping relationships within a language family based on evidence of exclusively shared innovations. Lexicostatistics provides an alternative method for subgrouping which is in many ways easier to use, but which suffers from drawbacks which will be discussed later in this chapter. In this chapter, we will discuss these two methods and see how we can reach the conclusion that the languages we refer to as Polynesian languages are indeed related to each other. Studying the relation between these languages is in a sense reconstructing the past of their speakers. Where did they come from? How did they end up scattered across such a wide area? How did they end up speaking different languages? As we will see in this chapter, linguistic evidence can indeed tell us quite a lot about the history of the Polynesian-speaking people.

3.1 *Historical Linguistics and The Comparative Method*

The branch of linguistics that involves studying of the history of languages is called historical linguistics. Historical linguistics is like archaeology and also like detective work. Your goal is to reconstruct the past of a language. All that is available to you is the languages currently spoken in the relevant area, say, Polynesia in our case. You investigate the data available, search for remains of the past, and put together whatever pieces you manage to dig up to reconstruct the past. Archaeologists excavate a site and look for pieces of pottery, bones, and other artifacts. Historical linguists plough through a language and seek linguistic artifacts. What exactly are linguistic artifacts? How do we “dig up” the remains of a language that once existed? Historical linguists use a technique called the comparative method. Let us see how it works.

3.1.1 Linguistic change

First of all, we must agree on one thing: languages are constantly changing. No language remains the same over a period of time. In other words, languages that are spoken now are different from those spoken five hundred years ago in the same areas. Think about the parents complaining that they do not understand the language their kids speak, or teachers who consider it dreadful that their students do not know how to speak “properly”. Young people create new expressions and use them as slang. Some of such innovations may be accepted by a larger community and eventually become a stable part of the language. At any given point of time, a change may occur in a language. The change may be initiated by one speaker or a group, and it may be imitated and adopted by others in the same speech community. The members of a speech community continue

to understand one another despite the accumulating changes in the language, for they all experience the same changes at the same time. But the language, slowly but steadily, is changing its form.

Let us digress a little bit from Polynesia and take a language that is familiar to us, English, as an example. Take a look at the following two versions of the first line of the Lord's Prayer.

Contemporary English: *Our Father, who is in heaven, may your name be kept holy.*
Old English (c. 1000): *Faeder ure thu the eart on heofonum, si thin nama gehalgod.*

The first version is in Contemporary English. You shouldn't have any problem understanding this sentence. The second version, however, is unintelligible to the speakers of Contemporary English. You may be surprised at how different they look from each other. True, we do find some vague similarities between the two. For example, we may recognize the similarity between *Father* and *Faeder*, *our* and *ure*, or *name* and *nama*. These items are similar enough, but that is about all we can find. As a whole, they are very different. Just by looking at the two sentences, one could not know that one of them derived from the other. Why, then, are they both called "English"? Their names suggest that they are considered to be (different versions of) the same language, English. But if they represent the same language, why are they so different? It is very puzzling if we only look at Old English and Contemporary English.

But now, let us look at two other versions of the same line.

Middle English (c. 1400): *Oure fadir that art in heuenes halowid be thi name.*
Early Modern English (c.1600): *Our father which are in heaven, hallowed be thy Name.*

If you compare Middle English and Early Modern English, you will notice the striking similarity between the two. Similarly, comparing Early Modern English and Contemporary English, you would say that they are indeed two different versions of English. Given these intermediate stages, you can now see the gradual change from Old English to Middle English, then to Early Modern English, and finally to Contemporary English. This way, we can trace back the history of English and establish the relation between Old English and Contemporary English: Old English is an ancestor of Contemporary English. In order to claim that two languages are related, we need to show this kind of historical link between them.

3.1.2 When a speech community breaks up ...

As mentioned above, languages constantly change as their speakers start to use new expressions. Certain changes are accepted by the speech community as a whole and become stable part of the language. As long as the relevant speakers stay together in the same community, they keep understanding each other and possibly are unaware of the changes that have occurred. They believe that they speak the same language. No matter how many changes the language undergoes, it really does not matter to the speakers because all of them experience the same changes. But it is not always the case that people stay in the same place all their lives. People move. People move from one place to another for various reasons. Suppose that a village experienced a massive population growth and the little village has not enough room or food for

everybody. A group of people may decide to move to look for another place to live and form a new community.

Suppose there was a speech community, A. One group of people left their homeland and form another community, B. Yet another group of people moved to another place and start a new community, C. Let's call the remainder of the original community, D. The tree diagram below illustrates this situation. The top node A represents the original speech community. This community breaks up as migration takes place.

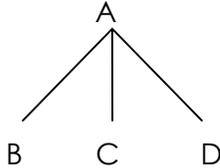


Figure 3.1 Proto language and its daughters

When the new speech communities B and C are formed elsewhere, A no longer exists. Instead, we consider the remainder of the community A a new community, D. Although the location of this community hasn't changed, its constituents have changed. Thus, strictly speaking, the remainder of A is not the same as the original A itself. Obviously, when the community first breaks up, the language spoken in A is spoken in each of the new communities B, C, and D. Once divided, however, these groups no longer communicate with one another regularly. Thus, any new changes in a particular group will not be necessarily spread over to the others. At first, the differences between the languages spoken in the four communities are minor. B, C, and D can be considered variants, or **dialects** of a single language (i.e., Language A). An illustrative situation would be various dialects of English spoken all around the world: American English, British English, Australian English, New Zealand English, etc.

What happens if the isolation continues for a long time? In each community, changes constantly emerge, but these new changes will not be shared by the other communities. As a result, the intelligibility between these languages becomes more and more limited. Eventually, they will reach a point where Languages B, C, and D are taken to be distinct languages rather than dialects of a single language. That is, they are no longer mutually intelligible. People from Community B cannot understand the language spoken by the people of Community C.

Sometimes, however, the boundaries are not so clear as we would like them to be. We find a situation where among the three languages, those spoken in adjoining regions are mutually intelligible, but mutual intelligibility does not extend to the next region. For example, we may have a scenario in which B and C, as well as C and D, are mutually intelligible, but B and D are not. In this case, we may wish to say B and C are dialects of a single language and C and D are also dialects of a single language. But then, we have a problem. B and D are not dialects of the same language. Mutual intelligibility does not extend over the whole range of languages, but only the subparts. On the other hand, there is overlap between these subparts. This curious situation is referred to as a **dialect chain**.

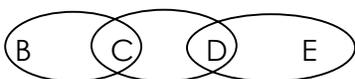


Figure 3.2 Dialect chain

It is possible that dialects of a certain language form a dialect chain before evolving into distinct languages. Such a situation is likely to arise when the neighboring communities are in close contact after the breakup of the original speech community.

By virtue of their historical relationship, Languages B, C, and D are said to constitute a **linguistic family**. They descend from the same ancestor, A. We refer to this common ancestor as a **proto language**. B, C, and D are A's daughter languages. A proto language is an interesting concept in that it is a language that no longer is actually spoken. In the present example, Languages B, C, and D are actually used in Communities B, C, and D, respectively. But at this point we cannot find a speaker of Language A. To get a more concrete picture, let's give specific names to these labels A, B, C, and D in the context of Polynesian languages. The figure below illustrates a grossly simplified relation between three Polynesian languages, Tongan, Samoan, and Māori. All of these languages derive from one ancestral language, Proto Polynesian.

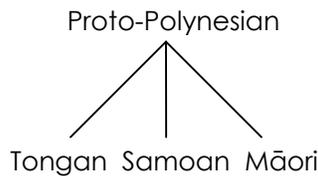


Figure 3.3 Proto Polynesian and its daughters

We know what Tongan, Samoan, and Māori sound like because there are people who speak these languages. But what about Proto Polynesian? Who speaks it? Where is it spoken? The answer is that nobody speaks Proto Polynesian any more. It is a language that was spoken long ago in a place where the ancestors of the speakers of Tongan, Samoan and Māori came from.

So far, we have seen how one speech community breaks up into several subcommunities and how a relation between a proto language and its daughter languages is formed. A similar division may take place within a daughter language. This will give us a more complex picture. Consider Figure 3.4 below.

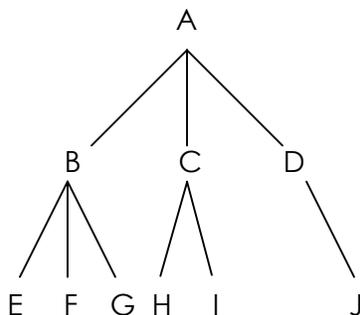


Figure 3.4 A linguistic family with a subgroup

B splits into three subcommunities, giving rise to three daughter languages, E, F, and G. C also divides into two subcommunities, giving rise to two daughter languages, H and I. In contrast, Community D does not break up. However, over time Language D also undergoes some changes, resulting in the daughter language J. In this tree diagram, E, F, G, H, I, and J are languages currently spoken. A, B, C, and D are proto languages, and therefore, we do not find a speaker of these languages. Languages E, F, G, H, I, and J now form a linguistic family. In this

linguistic family, E, F, and G form a **subgroup**, for they descend from the same proto language B. Similarly, H and I form another subgroup, whose proto language is C. If we say that two languages belong to the same subgroup, we imply that they have gone through a period of common descent apart from other languages in the same family.

3.1.3 Reconstruction of a proto language

By now you may be wondering, “If we cannot find a speaker of the proto language, how do we know that it actually existed? How do we know what it looks like? What kind of evidence is there?” Sometimes, we can find some direct evidence for a proto language, particularly when the language has a long tradition of written literature. Languages like Sanskrit and Latin are such languages. These languages are no longer spoken, but we do know what they look like, for we have access to the old manuscripts written in these languages. The situation is different with Polynesian languages, however. As we have seen in the previous chapters, Polynesian languages have a rich oral tradition, but not written literature. Thus, we do not have any record of Proto Polynesian. Well, then, what do we do? Can we still claim that Polynesian languages currently spoken derive from one ancestor language, Proto Polynesian? What evidence can we find to support such a claim?

All that is available to us is languages currently spoken, those located in the bottom layer in the family tree. The task in question is this: based on the data from those languages currently spoken, we put pieces together to reconstruct the proto language. As you can see, this is a step-by-step, bottom-up procedure. Recall that the European seamen and merchants as well as missionaries were all astonished at the striking similarities found across the languages spoken on various islands of Polynesia. In fact, many even assumed these languages to be dialects of a single language, i.e., Polynesian. We now know that it is a serious misconception. Yet, this observation is very important: this very similarity suggests that these languages belong in a single family.

Here is an important question to ask, though. Can we say that two languages are related if they show some similarities? Not necessarily, for superficial similarities may well be a coincidence or could be due to borrowing. When two languages are in contact, it often happens that one language borrows some lexical items from the other language and make them part of its vocabulary. For example, take some Japanese words that are borrowed from English such as *doa* “door”, *juusu* “juice”, and *tenisu* “tennis”. These words sound like the corresponding English words, but that is not because Japanese and English are related. Other than a handful of items, the two languages are completely different. In order to claim that two languages are related, it is not good enough to show that some lexical items in the two languages sound alike. Similarities between two related languages must be systematic and too great to be explained by chance. This is often represented by a set of **regular sound correspondences** in the vocabulary. For instance, it is not enough to show that the word for “eye” sounds very similar in the two languages, e.g., *mata* and *maka*, only differing in one consonant sound, [t] as opposed to [k]. This could be a mere coincidence. Rather, we must show that this t-k correspondence is recurrent. That is, there are a number of examples of [t] in Language A corresponding to [k] in Language B. If that is the case, we have a reason to believe that the two languages are in fact related and that somewhere in its history, [t] was replaced by [k] in Language B (or alternatively, [k] was replaced by [t] in Language A).

To illustrate the point, let us consider a specific example. The following data set contains seven words in three Polynesian languages, Māori, Tahitian, and Hawaiian. At a glance, you notice the striking similarity between the words appearing in the same row. But this alone does not enable us to say that Māori, Tahitian, and Hawaiian are related. We must show that the similarity is systematic.

Māori	Tahitian	Hawaiian	Meaning
ingoa	i'oa	inoa	name
mata	mata	maka	eye
matangi	mata'i	makani	wind
mate	mate	make	dead
ngutu	'utu	nuku	mouth
tangata	ta'ata	kanaka	person
tangi	ta'i	kani	weep

Comparing these words sound by sound, we notice, for example, that whenever Māori has [i], both Tahitian and Hawaiian has [i] in the corresponding position. Similarly, we find correspondence sets such as a-a-a, o-o-o, e-e-e, and u-u-u. Shifting our focus to the consonants, we can extract the following sets: m-m-m, ng-'-n, and t-t-k. Wherever [m] occurs in Māori, both Hawaiian and Tahitian have [m]. Wherever Māori has [ng], we find the glottal stop ['] in Tahitian and [n] in Hawaiian. Similarly, wherever we find [t] in Māori, we also find [t] in Tahitian and [k] in Hawaiian. These correspondences are regular and systematic.

MAO	TAH	HAW
i	i	i
ng	'	n
o	o	o
a	a	a

MAO	TAH	HAW
m	m	m
a	a	a
t	t	k
a	a	a

MAO	TAH	HAW
t	t	k
a	a	a
ng	'	n
i	i	i

Corresponding words in related languages such as *mata* and *maka* in the above example are called **cognates**. These words show systematic similarities, because they all derived from the same proto form. Each of these cognates is a **reflex** of the proto form, from which it is descended. In other words, cognates to historical linguists are like archaeological artifacts to archaeologists in that they tell us something about the history of these languages. They provide evidence for a proto language, giving us clues as to what it looked like.

One must be a little careful, however. Systematic similarities may be shared **retentions** from the proto languages, rather than shared **innovations** since the time of the proto language. That is, two language show systematic similarities because they both kept the relevant properties of the proto language. This can only mean that the two languages belong to the same language family, but nothing more. On the other hand, shared innovations are taken to be evidence to suggest that the two languages have gone through a period of common descent, during which the relevant changes took place. In trying to determine the subgroups within a language family, we must look for systematic similarities that are due to shared innovations. Note also that certain kinds of innovations are likely to be stronger evidence of subgrouping. These are a) particularly unusual changes, or b) sets of quite different types of changes both of which occurring together are unlikely to be due to chance, and c) phonological changes which occur together with grammatical or semantic changes. If two languages have a common sporadic or irregular change, this can be taken as strong evidence for subgrouping, as the same irregular change is unlikely to

take place twice independently. In Section 3.4, we will see some specific examples of shared innovations in Polynesian languages.

3.1.4 Homeland of Proto Polynesian

The comparative method enables us to determine two things: first, whether particular languages belong to the same language family or not; and second, whether there are subgroups within the family. But that's not all. We can learn more about the history of the relevant speech communities by examining their languages using the comparative method. In particular, assuming that in Polynesia, the normal occasion of linguistic split has been the migration of one part of a speech community from one island or archipelago to another, reconstruction of Proto-Polynesian can help us understand the migration pattern and speculate on the location of the homeland of Proto Polynesian.

Reconstruction of proto languages can help us understand migration pattern of the population in question. Two points are relevant here. First, we may assume that greater diversity of daughter languages implies a longer period of separation. This is because the more independent changes two languages undergo during the period of separation, the more different they look from each other. Naturally, more changes are likely to occur in a longer period of time. Second, it is also assumed that the larger the number of shared innovations within a subgroup, the longer the period of separate development before the breakup of the proto language. This is because a large number of shared innovations suggest that the period of common development shared by the relevant languages was long enough for so many changes to take place. Thus, studying the languages spoken today and reconstructing their proto languages can help us understand the course of migration and the relative duration of the stay at each point.

Linguistic evidence can also give us some idea about where the homeland of a language family was. We may safely assume that the homeland of a language family was some part of the territory over which its daughter languages are now spoken. Furthermore, assuming that the earliest migrations from the homeland were nearby areas and that later migrations populated successively more distant areas, the highest order divisions in the family (that is, those located in higher positions in the family tree) will be represented in the area near the homeland. Thus, the homeland of Proto Polynesian is taken to be somewhere around the area in which the languages belonging to higher order subgroups are spoken. As we will see in Section 3.4, the two highest order Polynesian subgroups are Tongic (consisting of Tongan and Niuean) and Nuclear Polynesian (comprising all others). In other words, linguistic evidence implies that the homeland of Proto Polynesian was in Western Polynesia. Pawley (1996) proposes that Proto Polynesian was developed from Pre Polynesian, which was spoken not on a single island, but by a network of several Western Polynesian islands including Tonga, Samoa, 'Uvea and Futuna.

These principles provide an answer to the question we posed earlier regarding the Outlier languages. Recall that some Polynesian languages are spoken in Micronesia and Melanesia, outside the Polynesian triangle. We asked ourselves a) why they are called Polynesian languages when they do not belong to the geographic area of Polynesia and b) how Polynesian languages ended up in Micronesia and Melanesia. We proposed two hypotheses. These languages are either the remainder of the once dominant language, whose speaker continue to move on from these areas to Polynesia, or the result of back-migration from Polynesia. As we will see in Section 3.3, languages of Micronesia and Melanesia are related to Polynesian languages and they are located higher in the family tree than the Polynesian sisters. This suggests that the main stream of

migration was flowing from Micronesia or Melanesia to Polynesia, rather than the other way round. Thus, we may conclude that these Outlier languages are a result of reverse migration at a later stage. In fact, subgrouping facts also provide evidence for the reverse migration hypothesis. The first subgroup branching off from Proto Polynesian is Tongic. The Outlier languages do not belong to this higher order subgroup, suggesting that they developed after the Tongic subgroup branched off from Proto Polynesian.

3.2 *Lexicostatistics and Glottochronology*

Lexicostatistics is another method that is used to subgroup languages. While the comparative method takes a qualitative approach, lexicostatistics is a quantitative method. This technique was first used by Morris Swadesh in early 1950s. He used a list of 200 meanings intended to be, as nearly as possible, universally known and culture independent (e.g., ‘and’, ‘big’, ‘drink’, ‘head’, ‘mother’, ‘skin’, ‘throw’). The rationale of lexicostatistics can be stated as follows. Given two related languages, with a list of the basic words for these meanings in each, we can count the number of cognate items shared by the two languages. Two dialects that have just recently been separated by a migration will share 100 percent cognates. If contact is not reestablished, this percentage will steadily decrease with time as one or the other language loses or replaces an item on the list.

Lexicostatistics is based on some crucial assumptions. First, it is assumed that some parts of the vocabulary of a language are much less subject to borrowing than other parts. Second, this basic ‘core’ of relatively borrowing-resistant vocabulary is taken to be the same for all languages. And third, the actual rate of vocabulary replacement in the core vocabulary is assumed to be more or less stable and is therefore the same for all languages at any period of time. Tested on thirteen languages, there has been average vocabulary retention of 80.5% every thousand years.

Given two languages, we collect the lexical items corresponding to those on Swadesh’s list and calculate the cognate percentage in core vocabulary. Based on this figure, we may determine whether the two languages are dialects of a single language, languages of a single family, or remotely related, if at all. Table 3.1 below provides the general guideline of how we can use lexicostatistical data to determine the relation between two languages.

<u>Level of subgrouping</u>	<u>Cognate percentage in core vocabulary</u>
Dialects of a language	81-100%
Languages of a family	36-81%
Families of a stock	12-36%
Stocks of a microphylum	4-12%
Microphyla of a mesophylum	1-4%

Table 3.1 Classification based on lexicostatistics
(Adopted from Crowley 1987/originally from Wurm 1971)

Swadesh used this method to subgroup Indo-European languages and achieved reasonably accurate results.

3.2.1 Problems with lexicostatistics

Closer to home, several researchers applied lexicostatistics to Polynesian languages (Elbert 1953, Grace 1959, Emory 1963, Walsh 1963, Dyen 1965, Biggs 1978). Table 3.2 shows the cognate percentages among selected Polynesian languages based on Dyen's study.

	NIU	TON	NUK	KAP	REN	PIL	TIK	ECE	SAM	EAS	HAW	MQS	TAH	RAR
Māori	47	45	39	44	43	47	50	49	44	54	51	50	41	56
Rarotongan	56	47	43	47	47	47	54	48	51	57	64	61	54	
Tahitian	38	35	30	32	43	37	42	42	36	44	48	44		
Marquesan	50	44	40	42	45	45	51	50	44	59	57			
Hawaiian	51	44	42	43	49	50	54	54	48	59				
Rapanui	55	50	40	46	51	49	56	56	51					
Samoan	62	61	49	50	57	53	63	63						
Ellicean	63	65	51	53	60	57	67							
Tikopian	60	63	53	54	63	61								
Pileni	54	53	79	47	57									
Rennellese	56	57	53	52										
Kapingamarangi	49	46	53											
Nukuoro	49	46												
Tongan	71													

Table 3.2 Cognate percentages among Polynesian languages (Dyen 1965 cited in Clark 1979)

Based on the lexicostatistical data, Dyen proposed a subgrouping of the Polynesian family, which shows some significant differences from the one based on the comparative method. In Dyen's model, Tongan, Niuean, and Samoan are clustered as the West Polynesian subgroup along with the Outlier languages except for Nukuoro and Kapingamarangi. In addition, Nukuoro, Kapingamarangi and Māori are taken to be isolated, i.e., not belonging to any subgroup. Subgrouping based on the comparative method indicates that Samoan is distinct from Tongan and Niuean at a higher level. Furthermore, reconstruction facts do not support the isolated position of Māori, Kapingamarangi and Nukuoro. In other words, lexicostatistical data do not match up to the comparative data, failing to accurately predict the internal relations among the Polynesian languages.

There are several reasons why lexicostatistics failed to prove to be successful with respect to Polynesian languages. One factor is borrowing. When two languages are in frequent contact, it often happens that one language borrows some lexical items of the other language and incorporates them into its vocabulary. Heavy influence of another language can obscure the true cognate percentage. For instance, the cognate percentage between East Uvea and Tongan is 85%. Thus, according to lexicostatistics, they are considered to be dialects of a single language. However, studies using the comparative method show that Tongan and East Uvea belong to two different subgroups. We have independent evidence to believe that the high cognate percentage between Tongan and East Uvea is due to borrowing from Tongan, rather than reflecting their common ancestry. It is known that Uvea was conquered by Tonga in the 15th century. It is natural that during this period East Uvea borrowed substantial part of vocabulary from Tongan, as it was the ruler's language. Similar arguments can be made for the high cognate percentage between Tongan and Samoan, and that between Pukapukan and other languages spoken in Cook Islands such as Rarotongan.

On the other hand, unexpectedly low cognate percentage can be explained in terms of the custom of name avoidance, which is known to have been practiced in some Polynesian societies. The best known example is Tahiti, where the custom is called *pii*. In Tahiti, it was considered rude to pronounce the king's name. Not only the name itself, but any words that sounded like the name were to be avoided. Hence, it is said that during the reign of Poomare I in the early nineteenth century, words like *poo* "night" and *mare* "cough" were replaced with alternative expressions such as *ru'i* and *hota*, respectively (White 1967). Name avoidance is said to have been practiced also in Samoan (Pratt 1911), Māori (Williams 1915), and Rapa (Stokes 1955). In most cases the avoidance would not extend beyond the chief's lifetime. But in some cases the replacement could have become permanent, obscuring the cognate percentage in the core vocabulary. The fact that Tahitian has abnormally low cognate percentages with all other languages is generally regarded as a result of this practice.

Finally, it has also been pointed out that Swadesh's list of 200 basic meanings is not the most appropriate for Polynesian languages. Swadesh's list includes meanings such as "ice" and "snow", which are (with marginal exceptions) foreign to the areas where Polynesian languages are spoken. In addition, it includes a number of kinship terms such as "mother", "sister", "grandmother", and so on. It is well known that different languages have different ways of classifying kinship terms. For example, in Polynesian languages, the notion of "sister" has two subcategories, "woman's sister" and "man's sister", each represented by a distinctive lexical item: e.g., PPn *taqo-kete and *tua-fafine, respectively (Marck 2000). It should be noted also that a more accurate translation of PPn *taqo-kete is "same-sex sibling". Therefore, the same term is used to refer to "man's brother". Given these factors, it is difficult to obtain accurate cognate percentages among Polynesian languages using Swadesh's list. As a solution, Blust (1981) proposes an alternative list to be used for Polynesian and other Austronesian languages, taking into consideration the cultural factors specific to these languages. (See Appendix 1 for comparison of the original Swadesh list and the alternative version proposed by Blust.)

3.2.2 Glottochronology

What fascinated researchers most about lexicostatistics is the application of this procedure known as glottochronology. Glottochronology attempts to calculate the separation time between two or more languages based on the lexicostatistic data and a special formula. That is, if we know the cognate percentage of two languages, we can also tell how long ago these two languages were one. Naturally, this is an exciting claim. Just as archeologists use radioactive carbon 14 to calculate the date of organic materials, linguists can use linguistic data to calculate glottochronological dates. The magic formula looks like this:

$$t = \frac{\log C}{2 \log r}$$

where C is the percentage of cognates and r is the retention rate (.805/millennium). Roughly, this formula gives us the following figures.

<u>Cognate percentage in core vocabulary</u>	<u>Years of separation</u>
66%	1000 (±250)
44%	2000 (±350)
28%	3000 (±400)
18%	4000 (±500)

For instance, if the cognate percentage between two languages is 66%, we may speculate that they separated some 1000 years ago. Needless to say, the reliability of glottochronological dating depends on the accuracy of the cognate percentages provided by lexicostatistics. As we have seen above, lexicostatistics is not free from problems. For the same reason, therefore, glottochronology has also been subject to criticism.

Having said that, one cannot help but wonder what glottochronology can tell us about the history of Polynesian languages. Pawley (1996), acknowledging the weakness of lexicostatistics and glottochronology, protests that they have been the victims of unfair criticism. Using the figures based on Biggs's (1978) lexicostatistic data, Pawley argues that lexicostatistical data are indeed compatible with the conclusions based on the comparative method, and that glottochronological dates may coincide with radiocarbon dates to some extent. On the other hand, Blust (2000) convincingly argues that there are clear cases where lexicostatistics and glottochronology give false results and points out that although lexicostatistics and glottochronology may sometimes work, the problem is how to distinguish the two cases.

	NIU	TON	EUV	EFU	TUV	TIK	SAM	REN	PIL	NUK	KAP	RAR	MAO	MQA	HAW
Tahitian	40	33	34	42	40	37	33	35	38	35	35	56	41	46	46
Hawaiian	50	44	43	49	48	49	44	49	45	44	43	64	56	56	
Marquesan	46	40	42	45	46	45	38	43	45	39	43	57	51		
Maori	50	45	47	47	52	54	45	46	44	41	49	62			
Rarotongan	53	47	47	50	56	53	47	47	51	45	45				
Kapinga	45	45	43	46	47	58	41	47	50	57					
Nukuoro	49	45	43	50	53	52	53	49	51						
Pileni	56	54	55	59	57	60	57	55							
Rennellese	52	52	49	57	58	56	52								
Samoaan	54	53	53	53	63	57									
Tikopian	57	55	56	58	67										
Tuvaluan	61	61	63	67											
East Futuna	58	67	69												
East Uvea	60	72													
Tongan	64														

Table 3.2 Cognate percentages among Polynesian languages (Biggs 1978 cited in Pawley 1996)

The archaeological evidence for Polynesian migration is multifaceted, but depends in important ways on the presence of a kind of pottery with highly distinctive design features. This pottery is called “Lapita ware”, from the name of the site in New Caledonia where it was first excavated in 1952 (Gifford and Shutler 1956). Lapita sites are scattered across the South Pacific from the Bismarck Archipelago in the west to the islands of Polynesia in the east. According to the radiocarbon dating, the Lapita culture appeared in the St. Matthias archipelago northwest of New Ireland around 1500 B.C. and by around 1300 to 1200 B.C. a very similar culture was present not only in other parts of the Bismarck Archipelago, but also in Santa Cruz, New Caledonia and Vanuatu. By no later than 1200 to 1000 B.C., closely related cultures showed up

in Fiji, Tonga, and Samoa. Archaeological evidence suggests that the Polynesian languages are descended from an ancestral language community that used Lapita pottery within the triangle formed by eastern Fiji, Tonga and Samoa. After more than a millennium of relative isolation in this area and the abandonment of pottery-making they set out on the dramatic migrations which led to the settlement of central and eastern Polynesia.

Both archaeological and linguistic data suggest that there were two pauses in the course of this migration. One is known as the West Polynesian pause: a very long period, as long as a millennium or more, between the first settlement of Western Polynesia and the first settlement of any part of Eastern Polynesia. The other pause was between the settlement of some or all the major island groups of Central Polynesia and the settlement of marginal regions such as Hawaii and New Zealand. This one, however, was not as long as the West Polynesian pause, probably about a few centuries long. The archaeological evidence for the West Polynesian pause is as follows. First, there are no archaeological sites in Eastern Polynesia reliably dated to earlier than A.D. 300. This is more than 1000 years after the spread of Lapita across Western Polynesia. The earliest accepted radiocarbon dates for Tonga are around 1100-1000 B.C. while for Samoa, East Futuna and Niuaotupapu, the earliest accepted archaeological dates are around 1000-800 B.C. Second, the transition from classic Lapita to what has been called “ancestral Polynesian” material culture can be traced in the Tonga-Samoa region during the first millennium B.C. (Kirch and Green 2001). This suggests that the first wave of Lapita culture settled in this area long enough to develop its own distinctive cultural characteristics.

Interestingly, linguistic evidence also supports this view. The fact that all the Polynesian languages share many unique innovations points to a long period of unified development between the time of the Polynesian branch split from its nearest relatives and the time the Polynesian branch itself broke up. The initial divergence among contemporary Polynesian languages is generally regarded as being between a Tongic branch and a Nuclear Polynesian branch (Pawley 1966, 1967, Biggs 1978, Clark 1979). If this view is correct, the Tongic-Nuclear Polynesian divergence began in Western Polynesia.

Turning to glottochronology, we get the following figures. First, the glottochronological dates for the divergence of Polynesian from the most closely related sisters, Fijian and Rotuman, range from 1850 to 1000 B.C. Note that this range of glottochronological dates makes some clearly incorrect predictions. Given that the accepted radiocarbon date for the earliest Lapita sites in the Pacific, including the Talepakemalai site in Mussau, is around 1500 B.C. (Kirch 1997), the separation of Proto Polynesian from Fijian and Rotuman could not possibly be as early as 1850 B.C. Second, dates for the break-up of Proto Polynesian can be calculated based on the cognate percentages between Tongic and Eastern Polynesian. These range from 650 B.C. to A.D.400. Again, these dates are earlier than the accepted radiocarbon dates by several centuries. Focusing on the difference between the two sets of glottochronological dates, however, we notice that the length of the Pre-Polynesian period is indicated to be at least 800 years and at most 1850 years, with a mean of 1450 years (Pawley 1996). As noted above, there are some significant discrepancies between the glottochronological dates and radiocarbon dates. Nevertheless, it is worth noting that glottochronological dates also predict that the pause between the divergence of Proto Polynesian and the break-up of Proto Polynesian to be longer than a millennium.

The second pause is also supported by both archaeological and linguistic evidence, although the archaeological evidence in particular is not as conclusive as that for the West Polynesian pause. As we will see below (Section 3.4.2), the Eastern Polynesian subgroup is fairly strongly defined, which suggests that Proto Eastern Polynesian underwent a substantial

amount of time, long enough for a number of shared innovations to develop. The agreements between Eastern Polynesian and Tongic languages range from 33 to 53% with a median of 44-45, giving a most probable separation date of 50 B.C. to A.D.0. Based on the lexicostatistical agreements between the four Eastern languages, the most likely date for the break-up of the four Eastern Polynesian linkage languages is around A.D.450-750, some 400-800 years after the date indicated for the break-up of Proto Polynesian. What about the archaeological evidence? The earliest accepted radiocarbon dates for the settlement of Eastern Polynesia may be taken to represent the latest possible dates for the break-up of Proto Polynesian. According to Spriggs and Anderson (1993), the earliest dates for Eastern Polynesian sites are in the ranges A.D. 300-600 for the Marquesas, A.D. 650-850 for Hawai'i, A.D.700-1000 for the Society Islands and Southern Cooks, A.D. 400-1270 for Easter Island and A.D. 1000-1200 for New Zealand. The earliest radiocarbon dates for the settlement of Hawai'i (A.D. 650-850) and New Zealand (A.D. 1000-1200) may be taken as roughly dating the break-up of the Central Eastern Polynesian subgroup. Thus, although these dates are a few centuries later than the glottochronological dates, the interval between the two break-up points is predicted to be also 400-800 years. Having said that, it should be noted that neither the archaeological nor linguistic facts provide enough information for us to hypothesize on the homeland of Proto Eastern Polynesian.

In sum, we observe that there is a fairly close match between archaeological evidence and linguistic evidence. Furthermore, glottochronology provides reasonably accurate dates for some major reference points in the migration to and within Polynesia, although the glottochronological dates are in each case several centuries earlier than the radiocarbon dates. Despite this discrepancy, it is particularly worth noting that the intervals between the dated events within each chronology are quite similar. This is especially the case in regard to the interval corresponding to the so-called Western Polynesian pause.

3.3 *The Austronesian Family*

The Polynesian language group belongs to a larger language family called Austronesian. It is one of the two largest language families in the world, consisting of more than 900 languages at the most conservative estimate. The areas covered by this large family includes the islands of Polynesia, Micronesia, Melanesia, Indonesia, Madagascar, the Philippines, and Taiwan, as well as continental enclaves in Indochina and the Malay Peninsula. In this section, we will discuss the relation between Proto Polynesian and its relatives in the Austronesian family.

While the fundamental work on the comparative phonology of the Austronesian languages was done by Dempwolff (1934-38), subgrouping within the Austronesian family is a matter of continuing debate. As far as Polynesian is concerned, however, the following subgrouping is generally accepted. In Figure 3.5 below, only the subgrouping that directly involves Polynesian is shown.

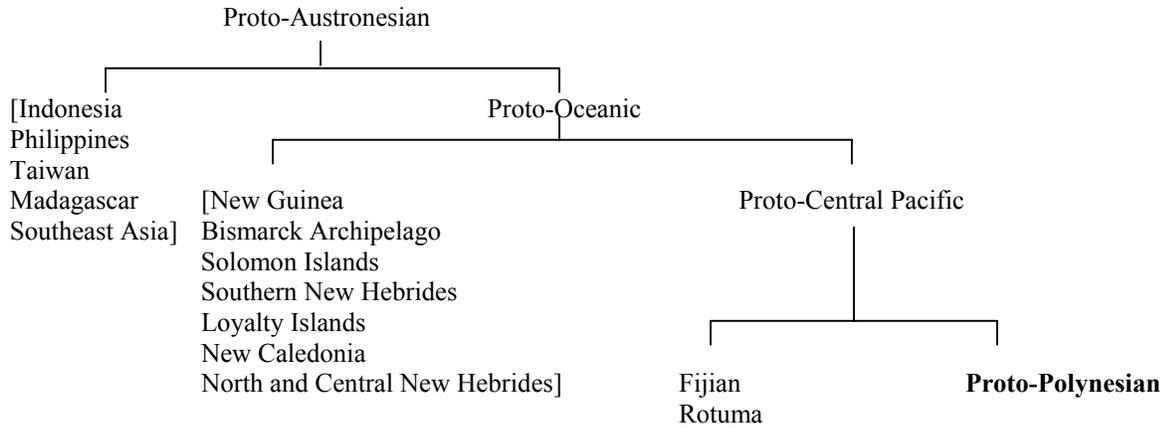


Figure 3.5 Austronesian Family Tree (Adopted from Clark 1979:255)

As you can see, Polynesian languages form one of the lowest order branches on the Austronesian family tree. Its closest sisters are Fijian and Rotuman, which together with Polynesian form a subgroup called Central Pacific. Central Pacific in turn belongs to the Oceanic subgroup, which also includes languages of Melanesia and Micronesia. In addition to all the languages belonging to the Oceanic subgroup, the Austronesian family includes Philippine languages, languages of Indonesia, aboriginal languages of Taiwan (often referred to as Formosan languages), languages of Malaysia, and Malagasy.

These languages, excluding the Formosan languages, divide into five subgroups: Malayo-Polynesian, Western Malayo-Polynesian, Central/Eastern Malayo-Polynesian, central Malayo-Polynesian, and Eastern Malayo-Polynesian. Figure 3.6 illustrates the relations between these subgroups. In short, the Malayo-Polynesian subgroup includes all the Austronesian languages other than Formosan languages. While some have proposed another subgroup consisting of the Formosan languages, Blust (1999) argues that Formosan languages do not form a single subgroup. Rather, Blust's reconstruction shows that Formosan languages divide into nine independent subgroups, all of which are primary branches from Proto Austronesian: Atayalic, East Formosan, Puyuma, Paiwan, Rukai, Tsouic, Bunun, Western Plains, and Northwest Formosan. Applying the reconstruction principles discussed in Section 3.1.4 above, the proposed subgrouping suggests that Proto Austronesian originates in Taiwan.

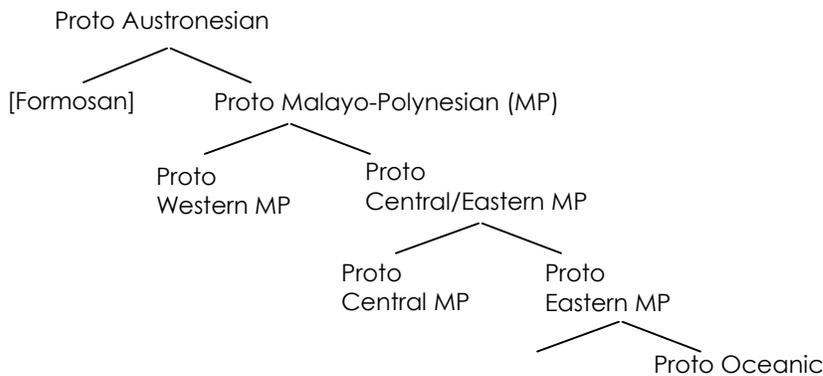


Figure 3.6 Partial Austronesian Family Tree

In the following sections, we will discuss the characteristics that define Oceanic and Central Pacific, the two subgroups that are directly relevant to Proto Polynesian.

3.3.1 Proto Oceanic

The Oceanic subgroup includes almost all the Austronesian languages of the Polynesian, Melanesian, and Micronesian islands, except for some languages at the far western end of New Guinea, and Palauan and Chamorro in Micronesia. Its closest relatives are located around Cenderawasih Bay at the western New Guinea and in South Halmahera (Blust 1978). Its next closest relatives appear to be in other parts of East Indonesia (Blust 1984, 1993). The most economical interpretation of the subgrouping of Austronesian is that the common ancestor of Oceanic probably entered Melanesia from eastern Indonesia via the north coast of New Guinea.

As we discussed earlier, we look for shared innovations as evidence for a subgroup. The best-known innovation shared by the Oceanic languages is a series of simplifications of the Proto Austronesian sound system. For example, Proto Austronesian *p and *b both correspond to *p in Proto Oceanic. (The asterisk * beside a letter indicates that it is a reconstructed sound.) The relevant changes are shown below.

Proto-Austronesian	→	Proto-Oceanic
*p, *b	→	*p
*s, *z, *c, *j	→	*s
*k, *g	→	*k

To illustrate, let us compare some cognates in Proto Austronesian and Proto Oceanic. The word for “mouth” in Proto Austronesian is *baqbaq. The *b sound, however, is replaced with *p in its cognate in Proto Oceanic *papaq. Similarly, Proto Austronesian *z is replaced by *s in Proto Oceanic, as illustrated by the pair *zalan and *salan.

<u>Proto</u>	<u>Proto Oceanic</u>	
<u>Austronesian</u>		
*baqbaq	*papaq	“mouth”
*nipen	*nipon	“tooth”
*zalan	*salan	“path/road”
*susu	*susu	“breast”
*katkat	*kakat	“bite”
*gemgem	*kokom	“hold in hand”

(Blust, personal communication)

Shared innovations are not restricted to sound change. There are some changes in the lexicon, too. That is, some lexical items are replaced by completely new items. Compare the Proto Malayo-Polynesian words and Proto Oceanic words listed below. Each pair represents the same meaning, but their phonological forms are totally different. It is impossible to derive the Proto Oceanic form from the Proto Malayo-Polynesian form. In other words, they cannot be cognates.

<u>Proto Malayo-Polynesian</u>	<u>Proto Oceanic</u>	
*hulaR	*m ^w ata	“snake”
*dilaq	*maya	“tongue”
*tuqelan	*suRi	“bone”
*ma-kunij	*ano	“yellow”
	(Blust, personal communication)	

Of these four, *m^wata is taken to be the strongest evidence, for this item is not found anywhere outside Oceanic. Strictly speaking, POc *suRi is a cognate of PMP *zuRi, which specifically refers to “fish bone”. The POc innovation concerning this item is semantic extension from a specific meaning to a general meaning.

3.3.2 Proto Central Pacific

The Central Pacific subgroup consists of Fijian, Rotuman, and the Polynesian languages (Geraghty 1983, 1986, Pawley 1972). The close relation between Polynesian and Fijian has long been recognized (Grace 1959, Pawley 1970, 1972, 1979). As for Rotuman, there has been a debate concerning its position relative to Central Pacific. While Grace (1959) proposed that Rotuman be included in Central Pacific, Pawley (1972) left it unclassified for lack of conclusive evidence. On the other hand, Biggs (1965) showed that the striking similarity of the Rotuman lexicon with Polynesian languages is in part due to massive borrowing and in part due to “direct inheritance”. It is now generally accepted that Rotuman is closely related to Polynesian and belongs in the Central Pacific subgroup (Pawley 1979, Geraghty 1986).

In any case, it should be noted that Central Pacific is only weakly defined. That is, there is not much evidence for this subgroup. According to Geraghty, probable innovations of Proto Central Pacific are as follows. First, the development of the “focus particle” *ko. Second, the development of the prefix *fia- (“want to”). Third, a change from subject initial to verb initial word order; and forth, the numeral four is *faa unlike other Eastern Oceanic languages, which use *fati. These properties are common in the Central Pacific languages, but not found in other parts in the Austronesian family.

The fact that the Central Pacific subgroup is only weakly defined implies that the period of common development must have been brief. Pawley (1996) argues that there is evidence that Proto Central Pacific was not a homogeneous language, but a dialect chain, located in Fiji and possibly extending to Western Polynesia.

3.3.3 Proto-Polynesian

Finally, we have reached the point where Proto Polynesian branches off from Proto Central Pacific. As we will see shortly, the Polynesian family is very strongly defined. This means that the Proto Polynesian speech community underwent a long period of isolated development after its separation from Proto Central Pacific. Recall that this observation is also supported by archaeological evidence, as discussed in Section 3.2.2. Let us look at some of the innovations that define the Polynesian family.

Pawley’s (1966, 1967) subgrouping is based not only on the sound changes, but also on shared innovations in morphology. We will look at some phonological evidence first. The Proto Polynesian sound system is discussed in detail by Biggs (1978). Phonological evidence can be

divided into two kinds: regular sound changes and sporadic sound changes. The most notable regular sound change in Proto Polynesian is the loss of prenasalized stops: while Fijian has prenasalized stops, these sounds correspond to plain voiceless stops in Proto Polynesian. Compare the following cognate pairs.

Fijian	Proto Polynesian	Meaning
^m beka	*peka	“bat”
ⁿ dalo	*talo	“taro”
^ŋ gele	*kele	“earth”

The initial consonant in these words is a prenasalized stop (i.e., ^mb, ⁿd, or ^ŋg) in Fijian, while it is replaced by a plain voiceless stop (*p, *t, or *k) in Proto Polynesian. Other regular changes are summarized in the table below. (Proto Central Pacific *v probably was a voiced bilabial fricative [β]. Proto Central Pacific *c probably was a voiced interdental fricative [ð]).

Proto Central Pacific	Proto Polynesian
*t, *d, *j	*t
*g, *k, *ŋgw, *kw	*k
*n, *ñ	*n
*ŋ, *ŋw	*ŋ
*r, *dr	*r
*y	∅
*v	*f
*c	*h
*k	*ʻ, *k

Table 3.3 Phonological innovations in Proto Polynesian

In addition to these regular sound changes, Proto Polynesian shows some sporadic sound changes, which constitutes strong evidence for this subgroup. Pawley (1996) provides the following four examples. First, Proto Central Pacific *m is lost before *u in Proto Polynesian. Thus, PCP *malumu “soft” becomes *maluu in Proto Polynesian. Note that this is not a regular change, as evidenced by the following examples: POc *qumun “earth oven” > PPn *qumu, POc *lumut “seaweed” > PPn *limu, and POc *muri “behind” > PPn muri. Second, metathesis of *-nia to *-ina occurs in verb endings. Metathesis is a process where two sounds switch their positions. Thus, PCP *kania “eat, eaten” is realized as *kaina “eaten” in PPn. Third, *n is sometimes lost in the sequence *ani: PCP *kani “eat” > PPn *kai, PCP *ʻanitu “spirit” > PPn *ʻaitu. Fourth, there is also the irregular loss of second syllable -ŋwa in some lexical items: e.g., PCP *taŋwaʻane “male” > PPn *taʻane.

With respect to morphology, there are a number of innovations defining Polynesian. Let us take a look at three examples. One is development of articles. In Central Pacific, a common noun article *na* is used. This article disappears in Proto Polynesian and is replaced by contrasting definite and indefinite articles, *te and *sa/ha, respectively. It should be noted that the question has been raised in the literature whether *te is pre-Proto Polynesian, as it is also found in Kiribati (possibly through borrowing) and Marshallese. As far as Marshallese is concerned, borrowing is

less likely. Second, there is an innovation concerning pronouns. Central Pacific distinguishes paucal (“a few”) and unlimited plural, as illustrated in Fijian examples below.

		Fijian	Tongan
Singular	he/she/it	i kwaya	ia
Dual	they two	i kirau	naua
Paucal	they (a few)	i kiratou	nautolu
Plural	they (many)	i kira	

This contrast is lost in Proto Polynesian. The Proto Polynesian pronominal system shows a three-way number distinction: singular, dual and plural, as seen in Tongan.

Another innovation has to do with the expression of possession. In Proto Central Pacific, nouns divide into four categories: inalienable, edible, drinkable, and neutral. When used in a possessive form such as “John’s book”, each noun category selects a particular possessive morpheme. In other words, there are four different morphemes to express possession. For example, Fijian has four possessive morphemes: *-na*, *ke-*, *me-*, and *no-*. In Proto Polynesian, these four categories are reduced to two: dominant and subordinate. The possessive morpheme marking dominant possession is *a and the one marking subordinate possession is *o. The two classes are also referred to as alienable and inalienable in the literature. The actual terms do not really matter, however. What is important to note is that Proto Polynesian has a two-way system unlike Proto Central Pacific, which has a four-way system. Observe the contrast between the two systems by comparing Fijian and Maori.

<u>Meaning</u>	<u>Fijian</u>		<u>Maori</u>	
his head	na ulu- na	inalienable	t- oo -na uru	subordinate
his sweet potato	na ke -na kumala	edible	t- aa -na kuumara	dominant
his water	na me -na wai	drinkable	t- oo -na wai	subordinate
his basket	na no -na kato	neutral	t- aa -na kete	dominant

Finally, there are some lexical (semantic) innovations, too. Pawley (1996) notes that Proto Polynesian has several affixes that have no functional equivalents in Proto Central Pacific. For example, *toko- or *toka-, which is prefixed to numbers to mark reference to a person, as well as a pair of prepositions *ma‘a “for (dominant relation)” and *mo‘o “for (subordinate relation)”. In addition, Clark (1979) notes that the following items are restricted to Polynesian.

*la‘aa	“sun”
*kata	“laugh”
*moana	“sea”
*ma‘uŋa	“mountain”
*kulii	“dog”
*moa	“fowl”

Those and other innovations are taken to be evidence for the Polynesian family, which consists of some thirty-four languages. In the next section, we discuss evidence for subgroups within the Polynesian family.

3.4 *Subgroups within the Polynesian family*

As we have seen, Polynesian is part of a subgroup of Austronesian called Oceanic. Generally, the Polynesian family is taken to consist of five major subgroups: Tongic, Nuclear Polynesian, Samoic-Outlier, Eastern Polynesian, and Central Eastern Polynesian. Although the Polynesian family is strongly defined, there still remain several issues concerning the subgrouping within Polynesian.

3.4.1 Shared innovations

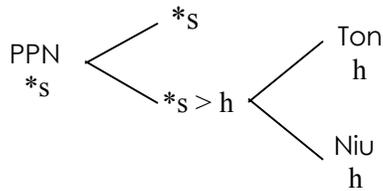
To repeat, linguistic properties that can be used as evidence for subgrouping are shared innovations. Languages that belong to a particular subgroup share some characteristics that resulted from a change. Take a look at the following set of cognates in five Polynesian languages. Proto Polynesian form for each word is *songi “smell” and *sae “to tear”.

	“smell”	“to tear”
Proto Polynesian	*songi	*sae
Tongan	hongi	hae
Niuean	hongi	hee
Samoan	sogi	sae
East Futuna	sogi	sae
Luangiua	songi	sae

Focusing on the initial consonant *s, we find the following sound correspondence set.

Proto Polynesian	*s
Tongan	h
Niuean	h
Samoan	s
East Futuna	s
Luangiua	s

What we observe here is that the reflex of Proto Polynesian *s is [h] in Tongan and Niuean, while it is retained in Samoan, East Futuna, and Luangiua. What is the implication of this observation? We may hypothesize that at some point in the history, there was a phonological innovation, i.e., change from Proto Polynesian *s to [h]. This change took place in Tongan and Niuean, but not in Samoan, East Futuna, and Luangiua. This suggests that Tongan and Niuean shared a period of common development during which the relevant change occurred. In other words, Tongan and Niuean are descendants of a common proto language. The observed change PPN *s > h is a shared innovation defining a subgroup in which Tongan and Niuean belong. The situation can be schematized as follows.



The subgroup Tongan and Niuean form is called Tongic. As we have just observed, one of the shared innovations that define the Tongic subgroup is the merger of *s and *h.

Phonological innovations can be used as strong or weak evidence. Regular changes affect all or nearly all occurrences of a sound under phonologically specifiable conditions and provide only weak evidence. Sporadic changes that affect only one or two forms are considered to be stronger evidence for subgrouping. For example, the reflex of Proto Polynesian *komo “suck” is *omo in Proto Eastern Polynesian. On the other hand, we observe that Proto Polynesian *k is generally retained in Proto Eastern Polynesian as shown in the pair PPN *kiato and PEP *kiato “outrigger boom”. Therefore, the loss of *k in PEP *omo is a sporadic change and provides strong evidence for this subgroup.

As we have seen in previous sections, shared innovations are not necessarily phonological. Pawley (1966) proposes a number of morphological innovations that define the subgroups within Polynesian. One shared innovation defining Proto Nuclear Polynesian is the development of a new demonstrative system. Two of the Proto Polynesian demonstrative pronouns are *eni “this, these” and *ena “that, those”. As you can tell from the English translation, these forms are used to refer to both singular and plural. In contrast, Proto Nuclear Polynesian distinguishes singular and plural, using four different forms, as shown below.

<u>PPn</u>		<u>PNP</u>	
*eni	“this, these”	*teenei	“this”
		*(ee)nei	“these”
*ena	“that, those”	*teenaa	“that”
		*(ee)naa	“those”

Another example of morphological innovation comes from Proto Eastern Polynesian, which introduces a new set of tense/aspect markers. Proto Polynesian past tense marker, *neqe/naqa is replaced by *i. A marker for habitual or progressive, *e-qana is also an innovation in Proto Eastern Polynesian.

Let us look at some semantic innovations, too. Here are some examples from Proto Eastern Polynesian. Proto Eastern Polynesian *fofonu means “deep,” which is derived from Proto Polynesian *fonu “full.” A similar semantic extension is observed with Proto Polynesian *lepo “cover with leaves,” whose reflex in Proto Eastern Polynesian *repo means “dirt, earth.” In addition to these cases of semantic shift, some completely new items are introduced in Proto Eastern Polynesian. For example, *maitaki “good, pleasant” is taken to be an innovation, for it cannot be found outside the Eastern Polynesian subgroup.

3.4.2 Standard model of subgrouping within Polynesian

The standard model of Polynesian subgroups is based on the proposals by Pawley (1966) and Green (1966). Pawley (1966) proposed four subgroups based on shared morphological innovations: Tongic, Nuclear-Polynesia, Samoic-Outlier, and Eastern Polynesian. Note that recently, this model has been revised by Marck (2000), as we will see shortly.

The Tongic subgroup consists of only two languages, Tongan and Niuean. One of the shared morphological innovations defining this subgroup is an intrusive *u* in three plural pronoun markers.

	PPn	PTo	Tongan	Niuean
1.pl.incl.	*kitato(l)u	*kitautolu	kitautolu	tautolu
1.pl.excl.	*kimato(l)u	*kimautolu	kimautolu	mautolu
3.pl.	*kilato(l)u	*kilautolu	kinautolu	lautolu

Phonological innovations that define this subgroup are as follows: 1) loss of PPn **r* as in PPn **rua* “two” > PTo **ua* and PPn **huru* “enter” > PTo **huu*; 2) merger of PPn **s* and **h* > PTo **h* as in PPn **solo* “wipe” > PTo **holo* and PPn **siku* “tail” > PTo **hiku*; and 3) PPn **w* > PTo **v* as in PPn **hiwa* “nine” > PTo **hiva*.

The rest of the Polynesian languages form the Nuclear Polynesian subgroup. Note that this subgroup is not defined by default as those which do not belong to the Tongic subgroup. The fact that certain languages do not belong to a particular subgroup does not necessarily mean that they form a single subgroup. In order to conclude that these languages form a subgroup, we need evidence. We have seen above a morphological innovation in PNP, i.e., number distinction in demonstratives: PPn **eni* > PNP **teenei*, *(*ee*)*nei* and PPn **ena* > PNP **teena*, *(*ee*)*na*. Phonological innovations defining the Nuclear Polynesian subgroup include 1) merger of PPn **l* and **r* > PNP **l* as in PPn **rua* “two” > PNP **lua* and PPn **langi* “sky” > PNP **langi*; 2) loss of PPn **h* as in PPn **hingo* “name” > PNP **ingo* and PPn **hiwa* “nine” > PNP **iwa*.

Nuclear Polynesian divides into two subgroups: Samoic-Outlier and Eastern Polynesian. The Eastern Polynesian subgroup includes Rapanui, Hawaiian, Marquesan, Mangarevan, Tahitian, Maori, Moriori, Tuamotuan, Rarotongan and various other Cook Islands Maori languages. The rest belong to the Samoic-Outlier subgroup. While Eastern Polynesian is strongly defined, evidence for Samoic-Outlier is rather weak. In fact, it appears that this subgroup is indeed defined as those Nuclear Polynesian languages which do not fall under Eastern Polynesian. Pawley (1966) proposed the following morphological shared innovations defining Samoic-Outlier. First, PPn **toko-* “human number prefix” changes into PSO **toka-*. Second, some new items are introduced such as **ngaa-*, a prefix to certain locative bases and **soko-*, a prefix meaning “each, any.” As for Eastern Polynesian, we have seen earlier that the past tense marker **i* and the habitual or progressive aspect marker **e-qana* are PEP innovations. There is also the negative marker **kore*, which is a shared innovation defining this subgroup. As for phonological innovations, we find the following sporadic sound changes: *PPn **nga* > PEP **n* as in PPn **manga-wai* “tributary water course” > PEP **mana-wai* and also PPn **a* > PEP **e* as in PPn **salu* “scrape, comb” > PEP **selu* (Marck 2000).

In addition to the four subgroups discussed above, Green (1966) proposed another subgroup Central Eastern Polynesian, which consists of all the Eastern Polynesian languages but Rapanui. Shared innovations defining Central Eastern Polynesian includes 1) loss of PPn **q* as in PPn **qara* “wake up” > PCE **ara*; 2) loss of PPn **h* as in PPn **quha* “rain” > PCE **ua* and PPn

*qaho “day” > PCE *ao; 3) merger of PPn *l and *r > PCE *r as in PPn *laqaa “sun” > PCE *raa and PPn *maaqoli “true, genuine: > PCE *maoari; and 4) PPn *f > PCE *w word initially before *a(a)h as in PPn *fafine “woman” > PCE *wahine, and PPn *f > PCE *h word medially and before round vowels as in PPn *faho “outside” > PCE *waho. The last one is known as (Central) Eastern Polynesian labial dissimilation.

Central Eastern Polynesian further breaks up into two subgroups, Marquesic and Tahitic. Biggs (1978) argues that Proto Tahitic and Proto Marquesic are not marked by any regular sound changes from Proto Central Eastern Polynesian. According to Marck (2000), some sporadic sound changes defining these subgroups are as follows. Proto Marquesic shows 1) low vowel assimilation PCE *ai > PMq *ei as in PCE *taiti “young male child” > PMq *teiti and 2) loss of the first vowel in a vowel sequence as in PEP *taokete “same-sex sibling-in-law” > PMq *tokete and PEP *haere “to go, walk” > PMq *here. For Proto Tahitic, we find, among others, 1) unexpected change of the first vowel in *PPn *mutie “grass” > *PTa *matie; 2) lowering of the final vowel in *PPn nonu “a kind of tree (Morina citrifolia)” > *PTa “nono; and 3) irregular lowering of the first vowel in PPn *hulufe “a fern” > PTa *aruhe.

To summarize, the following tree diagram shows the relations between the subgroups within Polynesian as proposed by Pawley (1966) and Green (1966).

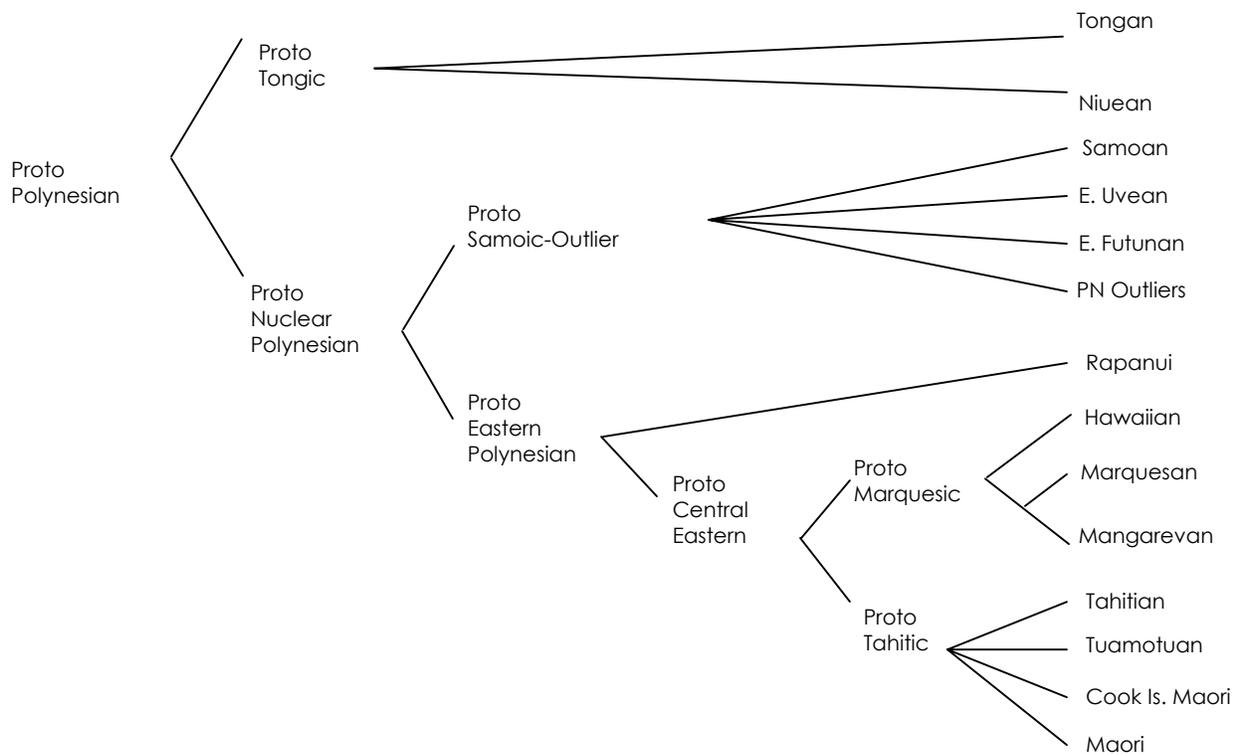


Figure 3.7 The standard subgrouping of Polynesian languages (based on Pawley 1966, 1967, Green 1966)

3.4.3 The Revised Model

As mentioned above, one of the subgroups proposed in the standard model, Samoic-Outlier is only weakly supported. On the other hand, Wilson (1985) proposed another subgroup Ellicean

within Samoic-Outlier, which includes Samoan, Ellicean Outlier and Eastern Polynesian, based on uniquely shared changes in pronominal systems. Ellicean Outlier includes Nukuoro, Kapingamarangi, Nukuria, Takuu, Nukumanu, Luangiua, Sikaiana, Tokelau, and Tuvalu. Marck (1999, 2000) provides further evidence for this subgroup. This can be found in sporadic changes in the following items: PNP *mafu “to heal” > PEc *mafo and PNP *kiu “a kind of bird” > PEc *kiwi, PNP *fuanga “whetstone” > PEc *foanga. As for the other languages formerly classified as Samoic-Outlier, Marck argues that they should be taken to be directly branching off from Proto Nuclear Polynesian, for there is no evidence to suggest a subgroup encompassing all of these languages. Marck hence proposes the revised model of subgrouping of Polynesian languages, illustrated in Figure 3.8 below. In this new model, the Nuclear Polynesian subgroup is reorganized into Proto Ellicean and a bunch of isolated branches. Eastern Polynesian is now included in the Ellicean subgroup rather than directly branching off from Proto Nuclear Polynesian.

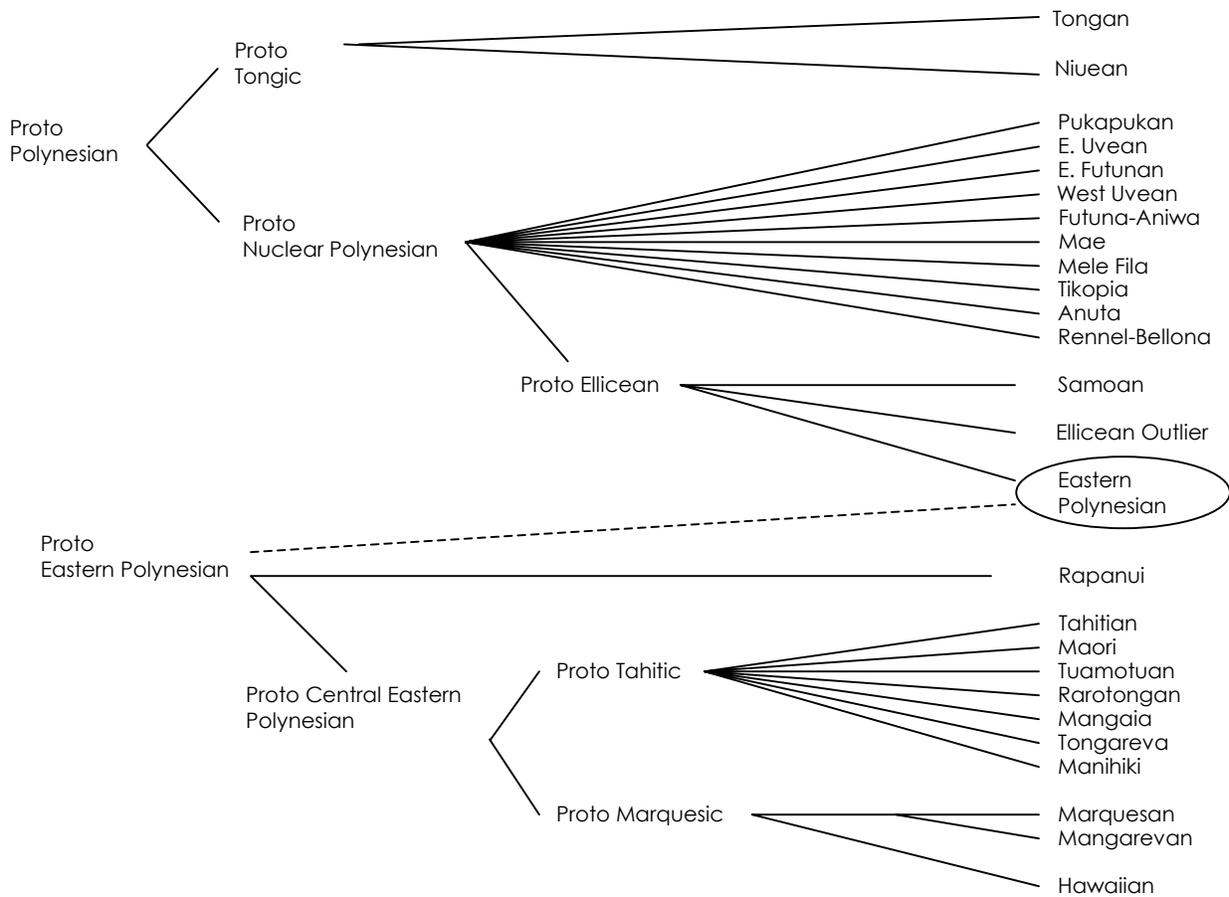


Figure 3.8 Revised subgrouping of Polynesian languages (From Marck 2000)

3.4.4 Subgrouping and the problem of borrowing

In searching for shared innovations to define subgroups within Polynesian, we must take into account the problem of borrowing. As many of the languages in this family were in close contact with one another, some lexical items have been borrowed from another language. As a result, some lexical items do not conform to the reconstructed sound correspondences. Such exceptions

can complicate the process of reconstruction. Let us consider three cases that exemplify the scenario in question.

The first example is found in East Uvean. Recall that East Uvean belongs to the Nuclear Polynesian subgroup. One of the shared innovations defining this subgroup is loss of PPn *h. Now, if we look at the reflex of PPn *mahuku “grass,” we encounter a slight problem. Its reflex in Tongan is *mohuku*. This is no surprise, for we know that PPn *h is retained in Proto Tongic. What is puzzling is that its reflex in East Uvean is also *mohuku*, with [h] retained. If we exclusively focus on this item, we might conclude that East Uvean belongs to the Tongic subgroup rather than Nuclear Polynesian. However, this apparent retention of PPn *h in East Uvean can be explained in terms of borrowing. It is known that East Uvea was under Tongan rule for some time. Considering the strong influence of Tongan, it is natural that some lexical items, including *mohuku*, were borrowed from Tongan into East Uvean.

Another instance concerns the reflex of Proto Polynesian *tufunga “expert, craftsman”. Its reflex in Proto Tahitic is reconstructed as *tahunga “priest.” Notice the change in the first vowel from PPn *u to *a in PTa. Now, if we look at Hawaiian, its reflex is *kahuna* “priest”. In Marquesan, its reflex is *tuhuna/tuhuka* “expert, artisan”. At a first glance, it appears that Hawaiian patterns with Tahitian rather than Marquesan in that the first vowel is [a], not [u]. Thus, Hawaiian can be mistakenly taken as Tahitic rather than Marquesic, based on this seeming shared innovation. Recall, however, that in the model(s) discussed above, Hawaiian belongs to the Marquesic subgroup. This is because we have other evidence that supports Hawaiian’s membership in the Marquesic subgroup. But then, how can we account for the fact that the reflex of PPn *tufunga is *kahuna* rather than *kuhuna* in Hawaiian? Marck (2000) argues that this irregular match between Hawaiian and Tahitic can be understood as a consequence of borrowing. The initial vowel is [a], not because it retains the Proto Tahitic *a, but because this lexical item is borrowed from Tahitian into Hawaiian.

Finally, we find a similar case in Rapanui and its relation to Tahitian. Rapanui belongs in the Eastern Polynesian subgroup, but crucially not part of the Central Eastern subgroup. We have also seen that PPn *k is generally retained in PEP. Thus, we expect the reflex of PPn *k to be [k] in Rapanui. For example, PPn *kamu “chew, munch” is expected to be *kamu* in Rapanui. Contrary to what we predict, its reflex in Rapanui is *‘amu* “eat” with a glottal stop instead of [k]. Furthermore, it is the same form as Tahitian *‘amu* “eat.” Based on this particular set, we may mistakenly conclude that Rapanui belongs in the same subgroup as Tahitian, i.e., Central Eastern Polynesian. Again, given the geographical and social proximity of Rapanui and Tahiti, this particular instance can be seen as a result of borrowing. Rapanui *‘amu* does not arise due to a shared innovation, but is merely a result of borrowing from Tahitian.

These are just a few examples to illustrate the problems one encounters in an attempt to subgroup Polynesian languages. We have learned that in order to determine a subgroup, we need to look for shared innovations, sporadic sound changes in particular, as evidence. The above examples show that one must also take into account the possible borrowing between two languages that do not form a single subgroup. Having said that, it should be noted also that the analyses discussed above raise some concerns, too. Robert Blust (p.c.) points out the following as potential problems. With respect to the borrowing between Tahitian and Hawaiian (e.g., *kahuna* “priest” in Hawaiian), a question arises as to how languages that are 2,500 miles apart are able to participate in a borrowing relationship. Similarly, in the case of Rapanui *‘amu* “eat”, if speakers borrow a word like “eat”, it would be expected that they would borrow many others

as well. Thus, it is not always straightforward to account for the counterexamples to the reconstructed sound correspondences in terms of borrowing.

3.5 *Summary*

In this chapter, we discussed how studying the Polynesian languages helps us understand the history of Polynesia. Where do the Polynesian people come from? How did they reach Polynesia? We have seen that reconstruction of proto languages using the comparative method provides answers to these questions.

Given the position of Proto Polynesian in relation to the other Austronesian languages, it is concluded that the ancestor of Proto Polynesian speakers came from Taiwan through the Philippines, Melanesia, and Fiji. Subgrouping within the Polynesian family suggests that the homeland of Proto Polynesian was located in the Western Polynesia, somewhere around the eastern Fiji-Tonga-Samoa region and that the population continued to migrate eastward. The data from the Outlier languages suggest that there was also a reverse migration, going back to Micronesia and Melanesia.

In summary, examining the languages spoken in Polynesia today does reveal quite a lot about the history of the area. It should be also noted that linguistic evidence is generally compatible with the archaeological evidence. We may say that linguistics makes a significant contribution to providing a complete picture of the history of Polynesia.

References

- Biggs, Bruce. 1965. Direct and indirect inheritance in Rotuman. *Lingua* 14: 383-415.
- Biggs, Bruce. 1978. The history of Polynesian phonology. In S.A. Wurm and Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings. Fascicle 2, Eastern Austronesian*. Pacific Linguistics C-61: 691-716. Canberra: Australian National University.
- Blust, Robert. 1981. Variation in retention rate in Austronesian languages. Paper presented at the Third International Conference on Austronesian Linguistics, Den Pasar, Bali, January, 1981.
- Blust, Robert. 1984. More on the position of the languages in Eastern Indonesia. *Oceanic Linguistics* 22-23: 1-28.
- Blust, Robert. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32: 241-93.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-kuei Li (eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics: Symposium Series of the Institute of Linguistics (Preparatory Office) Academia Sinica, Number 1*. Taipei.
- Blust, Robert. 2000. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages. In Colin Renfrew, April McMahon and Larry Trask (eds.), *Time depth in historical linguistics*, pp.311-331. Cambridge: McDonald Institute for Archaeological Research
- Clark, Ross. 1979. Language. In J.D. Jennings (ed.), *The prehistory of Polynesia*, pp.249-70. Cambridge, Mass.: Harvard University Press.
- Crowley, Terry. 1987. *An Introduction to historical Linguistics*. 1st ed. Papua New Guinea: University of Papua New Guinea Press; Suva, Fiji: Institute of Pacific Studies, University of the South Pacific.
- Dempwolff, Otto. 1934-38. *Vergleichende Lautlehre des austronesischen Wortschatzes*. 3 vols. Zeitschrift für Eingeborenen-Sprachen, Beiheft 15 (1934), Beiheft 17 (1937), Beiheft 19 (1938). Berlin: Dietrich Reimer.
- Elbert, Samuel H. 1953. Internal relationships of Polynesian languages and dialects. *Southwestern Journal of Anthropology* 9: 147-173.
- Emory, Kenneth P. 1963. East Polynesian relationships: Settlement pattern and time involved as indicated by vocabulary agreements. *Journal of the Polynesian Society* 72: 78-100.
- Geraghty, Paul. 1983. *The history of the Fijian languages*. Oceanic Linguistics Special Publication 19. Honolulu: University of Hawaii Press.
- Geraghty, Paul. 1986. The sound system of Proto-Central-Pacific. In P. Geraghty, L. Carrington and S.A. Wurm (eds.), *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*. Pacific Linguistics C-94: 289-312. Canberra: Australian National University.
- Gifford, E.W., and R.J. Shutler. 1956. Archaeological excavations in New Caledonia. *Anthropological Records* 18(1). Berkeley and Los Angeles: University of California Press.

- Grace, George. 1959. The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) language family. *International Journal of American Linguistics Memoir no.16*. Baltimore.
- Green, Roger. 1966. Linguistic subgrouping within Polynesia: The implications for prehistoric settlement. *Journal of the Polynesian Society* 75: 6-38.
- Kirch, Patrick V. 1997. *The Lapita peoples: Ancesters of the Oceanic World*. Oxford: Blackwell.
- Kirch, Patrick V. and Roger C. Green. 2001. *Hawaiki, Ancestral Polynesia: An essay in historical anthropology*. Cambridge: Cambridge University Press.
- Marck, Jeff. 1999. Revising Polynesian linguistic subgrouping and its culture history implications. In Roger Blench and Matthew Spriggs (eds.), *Archaeology and language: Language change and cultural transformation* 4: 95-122. New York: Routledge.
- Marck, Jeff. 2000. Topics in Polynesian language and culture history. *Pacific Linguistics* 504. Canberra: Australian National University.
- Pawley, Andrew. 1966. Polynesian languages: A subgrouping based on shared innovations in morphology. *Journal of the Polynesian Society* 75: 39-64.
- Pawley, Andrew. 1967. The relationships of Polynesian Outlier languages. *Journal of the Polynesian Society* 76: 259-96.
- Pawley, Andrew. 1970. Grammatical reconstruction and change in Polynesia and Fiji. In S.A. Wurm and D.C. Laycock (eds.), *Pacific Linguistic Studies in Honour of Arthur Capell*. *Pacific Linguistics* C-13:301-67. Canberra: Australian National University.
- Pawley, Andrew. 1972. On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly (eds.), *Studies in Oceanic Culture History Volume 3*. *Pacific Anthropological Records* 13: 1-142. Honolulu: Department of Anthropology, Bernice P. Bishop Museum.
- Pawley, Andrew. 1979. New evidence on the position of Rotuman. *Working Papers in Anthropology, Archaeology, Linguistics and Maori Studies* 56. Department of Anthropology, University of Auckland.
- Pawley, Andrew. 1996. On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis. In Janet Davidson, Geoffrey Irwin, Foss Leach, Andrew Pawley and Dorothy Brown (eds.), *Oceanic culture history: Essays in honour of Roger Green*, pp. 387-410. New Zealand Journal of Archaeology Special Publication.
- Pratt, George. 1911. *Grammar and dictionary of the Samoan language*, 4th ed. Malua, Western Samoa: London Missionary Society.
- Spriggs, Matthew and Anderson, Atholl. 1993. Late colonization of East Polynesia. *Antiquity* 67: 200-17.
- Stokes, John F.G. 1955. Language in Rapa. *Journal of the Polynesian Society* 64: 315-340.
- Walsh, D. S. 1963. Dictionaries versus informants: An aspect of glottochronology. *Te Reo* 6: 30-38.
- White, Ralph Gardner. 1967. Onomastically induced word replacement in Tahitian. In Genevieve A. Highland, Roland W. Force, Alan Howard, Marion Kelly, and Yosihiko H. Sinoto (eds.), *Polynesian culture history*, pp. 323-338. Honolulu: Bishop Museum Press.
- Williams, William. 1915. A dictionary of the New Zealand language, 4th ed. Wellington: Whitcombe and Tombs.
- Wilson, William H. 1985. Evidence for an Outlier source for the Proto Eastern Polynesian pronominal system. *Oceanic Linguistics* 24: 85-133.

Wurm, Stephen A. 1971. The Papuan linguistics situation. In Thomas A. Sebeok (ed.), *Current trends in linguistics, Vol. 8: Linguistics in Oceania*, pp. 541-657. The Hague: Mouton.