Pacific Languages

Lynch, John

Published by University of Hawai‘i Press

Lynch, John.
Pacific Languages: An Introduction.
University of Hawai‘i Press, 1998.
Project MUSE. muse.jhu.edu/book/67423.

For additional information about this book
https://muse.jhu.edu/book/67423

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2378270
Linguistics: Some Basic Concepts

1.1. The Structure of Language

Linguistics is the systematic study of language, and descriptive linguistics is the branch of linguistics that deals with the analysis and description of languages. Each language is a system with various units and rules for the combination of these units into larger units. These rules are not always formulated in grammar books, but they are there nevertheless—in the brains of speakers of the language.

One simple way of seeing the operation of these rules is through the mistakes children make when they are learning a language. When a four-year-old says *The mans goed away, the sentence is clearly incorrect English. It does, however, follow a pattern. First, the child has deduced that, to make a noun plural in English, you add s to it. She has already produced large numbers of plural nouns, like dogs, cats, cups, bananas, and so on, following this rule. Second, she has also deduced that, to put a verb into the past tense, you add ed to it. Again, she has already produced many English verbs in the past tense this way—laughed, cried, kicked, washed, etc.

In producing the sentence *The mans goed away, the child is not imitating what adults say, since no adult speaker of English would say that sentence. Instead, she is applying two of the many rules she has formulated on the basis of observing how English is spoken.

1. NOUN + s = PLURAL
2. VERB + ed = PAST TENSE

The only problem is that the noun man happens to be an exception to rule (1), and the verb go an exception to rule (2). Looking at this ungrammatical
utterance gives us insight into how the child’s brain is functioning in terms of rules that combine units into larger units.

What are these units I have been talking about? If you asked a non-linguist that question, the answer would probably be sounds, words, and sentences. Unfortunately, the situation is more complex than that.

1.1.1. The Sounds of Language

At the “lowest” level of language we have sounds, which linguists enclose in square brackets [ ] to distinguish them from letters. Individual sounds, like [t], [e], and [n] are meaningless in themselves. Only combinations of sounds provide meaningful utterances: [t] + [e] + [n] = ten, [n] + [e] + [t] = net.

No language uses all the speech sounds human beings can make, and the sound systems of different languages are organized in different ways. The study of sounds is known as phonetics, and the study of the way in which sounds are organized into a system in a language is called phonology (or sometimes phonemics). (A chart of all phonetic symbols used in this book appears in appendix 2.)

Let us take as an example the sounds [p] (represented by the letters p or pp) and [f] (represented by f or ff). These are quite different sounds, but is the difference between them important? In some languages, for example English, it is, as the pairs of words below show.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pull</td>
<td>full</td>
</tr>
<tr>
<td>pig</td>
<td>fig</td>
</tr>
<tr>
<td>supper</td>
<td>suffer</td>
</tr>
<tr>
<td>cup</td>
<td>cuff</td>
</tr>
</tbody>
</table>

The only difference in sound between the words in each pair is the difference between the sounds [p] and [f], but each word has a very different meaning. In English, the sounds [p] and [f] belong to different phonemes; that is, they are different significant units of sound in the language. And linguists write phonemes in slant lines / / to distinguish them from both sounds and letters. Thus English has the phonemes /p/ and /f/.

Compare the same two sounds in the Tok Pisin language of Papua New Guinea:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>paia</td>
<td>faia</td>
</tr>
<tr>
<td>pasim</td>
<td>fasim</td>
</tr>
<tr>
<td>mipela</td>
<td>mifela</td>
</tr>
<tr>
<td>lap</td>
<td>laf</td>
</tr>
</tbody>
</table>
In this language, the difference between [p] and [f] is not significant. You can use either sound without changing the meaning of a word. In Tok Pisin, [p] and [f] belong to the same phoneme, usually written /p/. The same sounds in different languages may therefore have quite different functions in the systems in which they occur, and quite different relationships with each other.

Note that we are dealing with sounds and phonemes here, not with the letters that are used to write them. In the English words we looked at above, the phoneme /f/ is represented by the letter f in full as well as by the combination ff in suffer. The same phoneme /f/ is also represented by ph in phone, by gh in enough, and so on. Our principal concern is with the sound systems of Pacific languages, though we will also look at their orthographies, or writing systems.

1.1.2. The Composition of Words

Phonemes combine to form larger units. Consider the following English examples:

- act
- acted
- react
- reacted

Each of these consists of a number of phonemes, and each is also a word, since it has meaning by itself and, in the written language, appears with a space before and after. The second and third words, however, can also be divided into two meaningful parts, act ‘carry out’ + ed ‘past tense’ and re ‘back’ + act. The fourth word consists of three meaningful parts: re + act + ed.

These smallest meaningful units are called morphemes. Some single morphemes are words (act, dog, house, desire, for example). Other words (acted, react, reacted, dogs, housewife, desirable, for example), consist of multiple morphemes. The study of morphemes and of the way morphemes combine to form words, is known as morphology, a term also used to refer to the patterns by which morphemes combine to form words in a particular language.

The examples given above show one other feature of morphemes. While act can stand on its own as a word (as a free morpheme), re and ed cannot. Morphemes like re and ed are known as affixes, and they must be attached to another morpheme. There are a number of different kinds of affixes, the most common being prefixes, which, like re, come before the root in a word, and suffixes, which, like ed, come after the root. The convention in linguistics is to write prefixes with a following hyphen (re-) and suffixes with a preceding hyphen (-ed), the hyphen indicating where the join takes place.
Another kind of affix occurs in some Pacific languages, namely, the **infix**, which is placed *within* the root. In Roviana (Solomon Islands), for example, verbs are converted to nouns by inserting the infix *-in-* (note the hyphens both before and after the infix) before the first vowel of the root:

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Meaning</th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>habu</td>
<td>‘to fish’</td>
<td>hínabu</td>
<td>‘a catch of fish’</td>
</tr>
<tr>
<td>kera</td>
<td>‘to sing’</td>
<td>kínera</td>
<td>‘a song’</td>
</tr>
<tr>
<td>moho</td>
<td>‘to be sick’</td>
<td>mínoho</td>
<td>‘sickness, disease’</td>
</tr>
<tr>
<td>toa</td>
<td>‘to be alive’</td>
<td>tínoa</td>
<td>‘life’</td>
</tr>
<tr>
<td>zama</td>
<td>‘to talk’</td>
<td>zínama</td>
<td>‘language’</td>
</tr>
</tbody>
</table>

When morphemes combine to form words, the sounds at the boundaries of these morphemes often change. For example, I said above that the four-year-old had learned to form plurals by adding the suffix *-s*, but this is not strictly true. The regular plural morpheme has two spellings and three or four pronunciations in English. The pronunciation of the letter *s* in plurals like *cats, cups, socks* is indeed the phoneme /s/, but the letter *s* of plurals like *dogs, bugs, homes* is pronounced as the phoneme /z/, not as /s/; and the same letter in plurals like *inches, buses, dishes* is pronounced /iz/ or /əz/, depending on the dialect. I also said that the child had learned to form the past tense by adding *-ed* to verbs. Again, this is not strictly true. The pronunciation of *-ed* is /ɪd/ or /əd/ in words like *banded* and *slotted*, /d/ in *killed* and *conned*, and /t/ in *laughed* and *kissed*.

In these examples, the sound at the end of the noun or verb determines the pronunciation of the plural or past-tense suffix. The study of sound changes that take place when morphemes combine to form words is known as **morphophonemics**.

### 1.1.3. Above the Word Level

Words combine to form **phrases**. A phrase is a group of words that functions as a unit in a sentence. Look at the following English sentence (where / marks the boundary between phrases):

*The young boys / were killing / the cats / on the beach.*

Each of these phrases is a unit. When each is moved to some other position in the sentence, it must be moved as a whole entity. For example, the passive equivalent of the sentence above is

*The cats / were being killed / by the young boys / on the beach.*

and not something like

*The young the cats were being killed by boys on the beach.*
(The asterisk marks the sentence as ungrammatical.) That is, it is not just
the noun boys that moves in this change from active to passive, but the
whole noun phrase the young boys.

There are different types of phrases. In this book, I refer to noun phrases,
which are phrases that function like nouns and can be replaced by a single noun
or a pronoun—the young boys and the cats in our sentence above are both noun
phrases (and could be replaced, for example, by they and them). I also refer to
prepositional phrases, which are noun phrases introduced by a preposition:
on the beach and by the young boys in the examples above are prepositional
phrases, introduced by the prepositions on and by. I use the term verb complex
to refer to phrases that function like verbs: were killing and were being killed in the sentences above are both verb complexes.¹

Phrases combine to form clauses. A clause is a group of phrases containing
a subject (the topic being talked about) and a predicate (what is being said
about the topic). A sentence is a group of one or more clauses that can stand
alone. If we return to our example of the cat-killing boys, none of the following is
a sentence, since each requires other phrases to make it complete.²

*The young boys
*Were killing the cats
*The young boys on the beach

English and many other languages usually require each predicate to con-
tain a verb complex, so that a sentence must have at least one verb. Many
languages of the Pacific, however, do not require this, since in these languages
there is no verb equivalent to English be (with its various forms is, are, etc.). So,
for example, English demands the verb be in equational sentences like That man
is a doctor; but many Pacific languages have no verb in equivalent sentences. In
the Lenakel language of Vanuatu, for example, the same sentence would be Wus
aan tokta, literally ‘man that doctor,’ with no verb.

1.2. Common Grammatical Categories and Functions

1.2.1. Subject and Object

The terms subject and object traditionally refer to the performer and re-
ceiver of the action of the verb, respectively. In the sentence The boy is
petting the pig, the performer of the action, the boy, is called the subject,
and the receiver, the pig, is the object. In many languages the verb changes
with a change of subject. In the sentence The boys are petting the pig, the
plurality of the subject, the boys, causes the verb to change from singular (is
petting) to plural (are petting).
This fact is important, because the subject is not always the performer of the action. Look at these sentences:

*The boy likes the pig.*
*The boy was bitten by the pig.*

In these sentences, *the boy* is still the subject, because we can see the same kinds of changes in the verb when *the boy* becomes plural:

*The boys like the pig.*
*The boys were bitten by the pig.*

In the second case, however, *the boy* is not performing the action. The pig is performing the action on the boy.

In other languages, the subject and the object behave in ways different from the way in which English subjects and objects behave, and we cannot give a universal definition of these concepts. But the subject often performs the action, and the object usually receives it.

### 1.2.2. Transitivity and Voice

A sentence that contains no object is **intransitive**, while one that does contain an object is **transitive**. Examples:

**Intransitive:**  
*Mele is eating.*  
*The dogs are sleeping.*

**Transitive:**  
*Mele is eating a banana.*  
*The dogs chased the children away.*

An **active** sentence—a sentence in the **active voice**—is one in which the subject performs the action or where the object has the action performed on it. A **passive** sentence is one in which the action is performed on the subject. For example:

**Active:**  
*Mele ate the banana.*  
*The men cut down the tree.*

**Passive:**  
*The banana was eaten by Mele.*  
*The tree was cut down.*

### 1.2.3. Adjectives and Verbs

Many Pacific languages do not distinguish between adjectives and verbs in the same way English does. The distinction in English is related to the existence of the verb *be*. In English, an **adjective**—like *good*, for example—can either pre-
cede the noun it describes or follow the verb be (or similar verbs like seem or appear), as in A good chief looks after his people and Our chief is/seems good.

In many Pacific languages, however, adjectives belong to a class of stative verbs, verbs that indicate a state rather than an action. In Fijian, for example, a verb is marked as stative by one of a number of markers (e.g., e ‘third person singular subject’). In the first sentence below, the verb is kana ‘eat,’ and the word levu ‘big’ follows the noun it modifies, vuaka ‘pig’:

\[ E \text{kana na vuaka leva oqō}. \]

‘This big pig is eating.’ it eats the pig big this

In the next sentence, the word levu ‘big’ behaves like a verb, that is, just as kana ‘eat’ does in the sentence above.

\[ E \text{levu na vuaka oqō}. \]

‘This pig is big.’ it big the pig this

A stative sentence is an intransitive sentence expressing a state rather than an action. Thus while Mele is eating expresses an action, Mele is fat or Mele is a teacher express a state.

### 1.2.4. Person, Number, and Gender

In English, we are used to distinguishing first, second, and third person pronouns as well as subject, object, and possessive forms. Both nouns and pronouns occur in singular and plural, and in some cases they have masculine, feminine, or neuter gender. The English subject, object, and possessive pronouns illustrate this:

<table>
<thead>
<tr>
<th>Person</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>I, me, my</td>
<td>we, us, our</td>
</tr>
<tr>
<td>Second</td>
<td>you, your</td>
<td>you, your</td>
</tr>
<tr>
<td>Third</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculine</td>
<td>he, him, his</td>
<td>they, them, their</td>
</tr>
<tr>
<td>Feminine</td>
<td>she, her</td>
<td>they, them, their</td>
</tr>
<tr>
<td>Neuter</td>
<td>it, its</td>
<td>they, them, their</td>
</tr>
</tbody>
</table>

Pacific languages differ in a number of ways from the English model.

1. Most Pacific languages do not show gender in pronouns. Raroton-gan (Cook Islands) ia, or Fijian o koya, or Anejo m̃ (Vanuatu) aen all mean ‘he,’ ‘she,’ and ‘it.’

2. A large number of Pacific languages distinguish two types of first person pronouns. **Inclusive** first person pronouns refer to the speaker and the addressee(s). **Exclusive** first person pronouns refer to the speaker and some other person(s), but not the addressee(s). In Bis-
lama, the national language of Vanuatu, for example, yumi is the first person inclusive pronoun (‘I + you’), while mifala is the first person exclusive pronoun (‘I + he/she/it/them [not you]’).

3. Many Pacific languages distinguish more than two numbers, the most common (apart from singular and plural) being the dual number, which refers to two and only two; the trial number, referring to three and only three; and the paucal number, used for a few (three to six or so), or to a small group that is part of a much larger one.

The function of the plural changes depending on how many numbers a language recognizes. In a language with a singular, a dual, a trial or a paucal, and a plural, the role of the plural is much smaller than it is in a language with only a singular and a plural. In Fijian, for example, we have o koya ‘he/she/it’ (singular), o irau ‘they two’ (dual), o ira-tou ‘they (a few)’ (paucal), and o ira ‘they (many)’ (plural).

4. Many Pacific languages have separate object and possessive forms of the pronoun, as English does. But in addition, and unlike English, many also distinguish between an independent pronoun and a subject pronoun. The independent pronoun can be used as an answer to a question, and may be used as a subject, but when it is it is usually emphatic. In Lenakel, for example, in is the third person singular independent pronoun, and r- is the corresponding subject pronoun. The sentences In r-am-apul and R-am-apul both mean ‘He/she is asleep.’ But while the second one is a neutral statement, the first emphasizes that it is he or she, not someone else, who is asleep.

1.2.5. Possessives and Classifiers

In languages like English, there is usually only one kind of possessive construction. No matter what the possessed noun refers to, or what the possessor’s relationship is to that noun, the same construction is used: my hand, my father, my house, my dog are all possessed in the same way, by means of the possessive, my.

Now look at translations of those four phrases in Motu (spoken around Port Moresby in Papua New Guinea), in which the suffix -gu translates ‘my.’ The nouns are ima, tama, ruma, and sisia:

<table>
<thead>
<tr>
<th>Noun</th>
<th>Motu</th>
</tr>
</thead>
<tbody>
<tr>
<td>ima</td>
<td>ima-gu</td>
</tr>
<tr>
<td>tama</td>
<td>tama-gu</td>
</tr>
<tr>
<td>ruma</td>
<td>e-gu ruma</td>
</tr>
<tr>
<td>sisia</td>
<td>e-gu sisia</td>
</tr>
</tbody>
</table>

Here we can see that there are two different constructions: The words for ‘hand’ and ‘father’ attach -gu directly to the noun. I call this type a direct
possessive construction. The words for ‘house’ and ‘dog’ do not attach -gu directly to the noun, but attach it instead to the morpheme e-, and this word (e-gu) precedes the noun. This I call an indirect possessive construction.

In one way or another, most Pacific languages distinguish two types of possessive constructions to which different linguists have given different labels, and which have different semantics. These two types could be classified as follows:

- Close, or subordinate, or inalienable possession is often manifested by direct constructions. This involves the possession of something over which the possessor has no control, and which cannot (normally) be acquired or disposed of. It may be an integral part of the possessor (like a hand), or a relative (we cannot control who our father is).
- Remote, or dominant, or alienable possession is frequently manifested by indirect constructions. This involves the possession of something over which the possessor has control. It can be acquired and disposed of, given away or sold, like a house or a dog.

Some languages are more complex than this, using a system of classifiers, often in both possession and counting, to show what type of thing the noun is, just as in English we normally do not say ten cattle or four breads, but ten head of cattle or four loaves of bread, using head and loaf as kinds of classifiers. Look at the following examples from Ponapean (spoken in Pohnpei, Micronesia):

\[ kene-i-mahi \]
\[ \text{edible:thing-my breadfruit} \]
\[ \text{‘my breadfruit’} \]
\[ nime-i uhpw \]
\[ \text{drinking:thing-my coconut} \]
\[ \text{‘my drinking coconut’} \]
\[ sehuh pah-sop \]
\[ \text{sugarcane four-stalk} \]
\[ \text{‘four stalks of sugarcane’} \]

Ponapean has more than twenty possessive classifiers (like kene- and nime-above), and approximately thirty numeral classifiers (like -sop above).

1.3. Reconstructing Linguistic History

1.3.1. Genetic Relationship

All languages change. The process of change is gradual, but it is also constant. There are various kinds of evidence for this. For example, earlier written records show a version of the language different from the modern
version, though both are often still recognizable as the “same” language. The two examples given below, of the beginning of the Lord’s Prayer in the English of about 1400 and in modern English, illustrate this principle.

Oure fadir that art in heuenes halowid be thi name, thi kyngdom come to, be thi wille don in erthe es in heuene.

Our Father, who is in heaven, may your name be kept holy. May your kingdom come into being. May your will be followed on earth, just as it is in heaven.

Even if a language does not have written records going back a long time, the fact that people of different generations speak the same language slightly differently shows that languages change. We can even observe changes taking place in a language when we notice competing forms, like the two different pronunciations of a word like either in English (one with an initial vowel sound like that of niece and the other with a vowel like that of nice), or the past tense of the verb dive—dived and dove—in many dialects of American English. Perhaps the most obvious example of language change, however, is the continual introduction of new words into all languages (and, less obvious but also quite frequent, the gradual loss of words that, for one reason or another, have become obsolete).

Imagine now that we have a single speech community speaking a language we will call X. This community splits into four separate groups, A, B, C, and D. Because language change is inevitable and continuous, after a few hundred years these four communities would speak different dialects of the same language. But after a thousand years or more, these four dialects would have changed so much that they had become separate languages, as shown in figure 1. The languages would share many similarities in vocabulary and grammar, since language change is relatively slow. But a speaker of language A would have considerable difficulty in holding a conversation with a speaker of B, C, or D.

Languages A, B, C, and D in figure 1 are all genetically related to each other, because they all descend from language X, which is their common ancestor. Languages A, B, C, and D are often referred to as daughter languages of X, and all four languages belong to the same language family. Figure 1, which represents their relationship, is their family tree.

Where there are historical records of the ancestor language and of the whole period of change, it is easy to establish the relationship between the daughter languages and to see how diversification took place. But in the Pacific, as in many other parts of the world, such records do not go back anywhere near far enough for us to have concrete proof of diversification and relationship. How, then, do linguists establish such languages’ relationship?

Related languages share a number of similarities in vocabulary, pronun-
Linguists look for similarities between various languages, and if the similarities are numerous enough, they assume that the languages involved are related despite the absence of documentary proof and derive from a hypothesized common ancestor, which is referred to as a protolanguage.

But not all similarities between languages can be attributed to genetic relationship. There are two other possible explanations. One is that the similarities are purely accidental. In Motu, Fijian, and many other Pacific languages, the word for ‘eye’ is mata, while in Modern Greek the word for ‘eye’ is mati. This, however, is a purely accidental resemblance, as there are no other connections between Greek and Motu or Fijian. If two or more languages share only a few similarities, these are probably coincidental. It is virtually impossible, however, that languages could accidentally have hundreds of similarities.

The second explanation for similarities between languages is copying or borrowing—that a language has adopted a word (or some other linguistic feature) from some other language. For example, in many Pacific languages the word for ‘radio’ is something like retio or ledio. This word has been copied from English, but this does not mean that these languages are related either to English or to each other.

Copying is a very common phenomenon in all languages (see chapter 9). When new items of technology, new social practices, or new ideas are introduced into a society from outside, often the words for them, modified to fit local pronunciation, will be brought in at the same time. English is full of words copied from other languages: Algebra, boomerang, coup, demonstrator, ghetto, junta, taboo, thug, and yen are just a few examples.

Copying is more likely to take place in certain areas of the lexicon than in others. For example, words like snow, coconut, ice cream, church, team, and television could be easily introduced into a language, since they represent things or concepts that are by no means found in all cultures or environments.
But words like *hand, leg, one, two, black, white, eat, sleep* are much less likely to be taken from another language, since all languages probably have their own words for these concepts, irrespective of the culture of their speakers or the physical environment in which they live. There would be no need for a language to supplement its vocabulary by borrowing them. For similar reasons, certain aspects of grammar (the morphological structure of words, for example) are less likely to be borrowed than others (like word order).

If similarities between two languages are only in areas where we might expect to see copying, they do not constitute evidence of genetic relationship. If, however, the similarities are in areas of vocabulary and grammar where borrowing is much less likely to take place, we can reasonably conclude that these are not due to chance or borrowing, but to **genetic inheritance**. The words and structures were present in some form in an ancestor language and have been retained, usually in a modified form, in the daughter languages. This then leads to the conclusion that the languages sharing these similarities are related, belong to the same language family, and derive from the same protolanguage.

### 1.3.2. Reconstructing a Protolanguage

In addition to being able to show, with reasonable confidence, that a set of languages are related and derive from the same common ancestor, historical-comparative linguists can reconstruct what many of the sounds, words, and grammatical structures in the protolanguage were probably like.

An important principle in reconstruction, especially in dealing with similarities in vocabulary, is that of the **regularity of sound correspondences**. Look at the following examples from the Aroma, Hula, and Sinagoro languages spoken on the coast east of Port Moresby in Papua New Guinea:

<table>
<thead>
<tr>
<th></th>
<th>Aroma</th>
<th>Hula</th>
<th>Sinagoro</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘father’</td>
<td>ama</td>
<td>ama</td>
<td>tama</td>
</tr>
<tr>
<td>‘milk’</td>
<td>laa</td>
<td>laa</td>
<td>lata</td>
</tr>
<tr>
<td>‘sew’</td>
<td>uli</td>
<td>uli</td>
<td>tuli</td>
</tr>
<tr>
<td>‘grandparent’</td>
<td>upu</td>
<td>upu</td>
<td>tubu</td>
</tr>
<tr>
<td>‘sago’</td>
<td>lapia</td>
<td>lapia</td>
<td>labia</td>
</tr>
<tr>
<td>‘pigeon’</td>
<td>pune</td>
<td>pune</td>
<td>pune</td>
</tr>
<tr>
<td>‘skin’</td>
<td>opi</td>
<td>kopi</td>
<td>kopi</td>
</tr>
<tr>
<td>‘bird’</td>
<td>manu</td>
<td>manu</td>
<td>manu</td>
</tr>
<tr>
<td>‘mosquito’</td>
<td>nemo</td>
<td>nemo</td>
<td>nemo</td>
</tr>
</tbody>
</table>

There are a number of correspondences between identical phonemes. Aroma *m* corresponds to Hula *m* and Sinagoro *m*. This correspondence is
abbreviated as \textit{m:m:m}. We can also see all the vowels (\textit{a:a:a}, \textit{i:i:i}, and so on). But there are also some correspondences between different phonemes: First, although we have the set \textit{p:p:p} (as in \textit{pune} : \textit{pune} : \textit{pune} ‘pigeon’), we also have another set \textit{p:p:b} (as in \textit{lapia} : \textit{lapia} : \textit{labia} ‘sago’). Then, we also have the set \textit{θ:θ:t} (where \textit{θ} represents the absence of a sound), as in \textit{uli} : \textit{uli} : \textit{tuli} ‘sew.’ The important thing about both types of correspondence sets is that they are regular. They are not random, but occur again and again in many words. Even in the short list above, you can see a number of examples of each.

In the case of correspondence sets of the type \textit{m:m:m}, the original language almost certainly had \textit{m}, and the daughter languages have not altered it. The protolanguage, then, had a phoneme \textit{*m}, where the asterisk denotes a reconstructed form.

In the case of correspondence sets of the type \textit{p:p:p} and \textit{p:p:b}, however, one or more daughter languages has changed. The logical assumption here is that the set \textit{p:p:p} reflects an original \textit{*p}, while the set \textit{p:p:b} represents an original \textit{*b}, which Aroma and Hula have changed to \textit{p}. The \textbf{merger} of phonetically similar phonemes is a very common phenomenon, and this is what seems to have happened: The distinction between the two phonemes \textit{p} and \textit{b} has been lost in these two languages (in the same way as the distinction between the voiced \textit{w} in \textit{witch} and the voiceless \textit{w} in \textit{which} is being lost in most varieties of English). Similarly, the set \textit{θ:θ:t} probably represents an earlier \textit{*t}, which has been lost in Aroma and Hula; again, loss of a phoneme is far more common and natural than the addition of a phoneme.

Using this principle of regularity of correspondence, and also making use of what linguists know generally about language change, it is possible to \textbf{reconstruct} elements of a protolanguage—to make an educated guess about what the phonemes, words, and grammar of the ancestor language might have been. Given that Aroma \textit{nemo}, Hula \textit{nemo}, and Sinagoro \textit{nemo} all mean ‘mosquito,’ for example, and that the correspondences \textit{n:n:n}, \textit{e:e:e}, \textit{m:m:m}, and \textit{o:o:o} are regular, linguists would reconstruct the word \textit{*nemo} ‘mosquito’ in the language ancestral to these three languages. The full set of protoforms for the words given above would be:

\begin{align*}
*\textit{tama} & \quad \text{‘father’} \\
*\textit{lata} & \quad \text{‘milk’} \\
*\textit{tuli} & \quad \text{‘sew’} \\
*\textit{tubu} & \quad \text{‘grandparent’} \\
*\textit{labia} & \quad \text{‘sago’} \\
*\textit{pune} & \quad \text{‘pigeon’}
\end{align*}
1.3.3. Families and Subgroups

The original split of a community may be followed by later splits. Similarly, the original split of a protolanguage may be followed by subsequent splits in intermediate ancestral languages, sometimes called **interstage languages**. Look at the family tree in figure 2, which represents the following historical sequence of events.

First, the original ancestral language, X, initially split into three daughter languages, P, Q, and R. Some time later, (1) language P suffered sufficient divisions to result in the modern languages A and B; (2) language Q split into Z and the modern language C; (3) language Z itself underwent a further split, into the modern languages D and E; and (4) language R split, giving rise to the modern languages F, G, and H.

All of these languages are related, since they all derive from a common ancestor, X. There are, differing however, degrees of relationship in this family tree. For example, languages A and B are more closely related to each other than either is to any other modern member of the family because they share a period of common development that the other languages do not—the period when language P was separated from the others. Similarly, languages F, G, and H are more closely related to each other than to any other modern member of the family. Languages C, D, and E can also be grouped together.
but within the group, D and E are more closely related to each other than ei-
ther is to language C.

Linguists generally use the term **subgroup** to refer to two or more lan-
guages within a family that are more closely related to each other than to
the rest of the family. In figure 2, A and B form one subgroup and F, G, and H
another. C, D, and E make up a third subgroup within which exists a further,
lower-level, subgroup (sometimes called a subsubgroup), D and E.

When the history of a language family is known through written records,
the subgrouping of languages within that family can also usually be estab-
lished by examining those records. But how do we determine subgroups of
a language family in an area like the Pacific, where written records of lan-
guages either do not exist at all or date only from recent times?

One technique for doing this is known as **lexicostatistics**. This involves
the comparison of the basic vocabulary of the languages we are interested
in (using a standard one-hundred-or two-hundred-word list), and expressing
the degree of relationship between any two languages in the sample as a
percentage, which represents the **cognates** (similar vocabulary items pre-
sumed to derive from the same original word in the protolanguage) shared
by each pair of languages. A higher percentage corresponds to a closer rela-
tionship, and members of subgroups should show the highest percentages.

Lexicostatistics has the advantage of allowing quick formulation and quanti-
fication of the internal relationships of a language family, but it also has many
problems. Some of these are theoretical or methodological and need not concern
us here. One obvious problem, however, is that a list of even two hundred words
represents only an extremely small part of a whole language, and the figures ob-
tained from comparing such lists may not accurately represent the relationship
between two languages. Today, most linguists do not rely heavily on lexicostatis-
tics as a method for subgrouping languages, although they might use it to get a
preliminary indication of the possible subgrouping.

The chief method linguists use to establish subgroups is examination of
**shared innovations**. If you go back to the Aroma, Hula, and Sinagoro exam-
ples in the last section, you will see that two changes, or innovations, have
taken place: (1) original *t has been lost in both Aroma and Hula (but not in
Sinagoro); and (2) the distinction between original *b and *p has been re-
tained in Sinagoro, but it has been lost in both Aroma and Hula, where these
two phonemes merge as the single phoneme p.

Aroma and Hula share two innovations that Sinagoro does not, which
would suggest that the two languages are more closely related to each other
than either is to Sinagoro. The family tree in figure 3 shows how these three
descendants of Proto East-Central Papuan are related.

Rather than suggesting that Aroma and Hula both quite independently
made the changes *t > θ, *b > p inferred from a comparison of cognates, it seems logical to assume that the changes happened only once, in the inter-stage language, Proto Aroma-Hula. In this way Aroma and Hula came to share two innovations missing in Sinagoro, which suggests that they belong together in a subgroup.

There are various kinds of innovations which, if shared by two or more languages exclusive of others in the family, can be solid evidence for assigning those languages to the same subgroup. Phonological innovations (like the example above) and innovations in morphology are fairly strong evidence; innovations in vocabulary and syntax (sentence structure) are less strong, since changes take place in vocabulary much more easily and rapidly than in phonology or morphology. Quantity is also a factor. Generally speaking, if languages share more innovations (of the stronger kind) the hypothesis that they form a subgroup is more secure.

1.3.4. Reconstructing Linguistic and Cultural History

What use can linguists and others can make of the conclusions reached about the relationships between languages, the subgroups of a language family, and the reconstructed protolanguage?

The branch of linguistics I have been discussing is known as comparative linguistics or historical-comparative linguistics. It involves comparing languages in order to find out something about their history. This branch of linguistics is one of the disciplines contributing to the study of prehistory, the time preceding the existence of written records. (Other such disciplines include archaeology, social anthropology, the study of oral literature and oral traditions, and so on.) So, what can comparative linguistics tell us about prehistory?
First, the fact that languages are related implies that they have a common origin. This often (though not always) implies that the people who speak those languages have a common origin as well, telling us something about the origins of and historical connections between the peoples of a region.

Second, information about subgroupings can give us an idea of the chronology of language divisions (and presumably also divisions in a community), as well as providing indications about the directions in which people migrated. As an example of this, let us consider just the following Pacific languages: Fijian, Tongan, Pukapuka (spoken in the Cook Islands), Tahitian, and Rapanui (Easter Island). A simple family tree for just these five languages would look like the one in figure 4.

The most recent split in this family (which includes hundreds of other languages) is that between Tahitian and Rapanui, with the next most recent that between Pukapuka and the ancestor of Tahitian and Rapanui. Somewhat earlier Tongan and “Proto Pukapuka-Tahitian-Rapanui” divided, and the first split was between Fijian and all the other languages. As you can see by looking at map 1, the splits proceeded from west to east.

On the basis of this subgrouping, most linguists would assume (1) that the original homeland of this group of people was probably somewhere around the Fiji-Tonga area; and (2) that the general direction of migration of these peoples was probably from west to east, as shown in map 1. Note that I have used the terms “assume,” “probably,” and “somewhere.” These conclusions are merely the best educated guesses we can make from the data. We would still want to find supporting evidence from other disciplines—archaeological dates, oral traditions, or the like—before adopting these conclusions firmly.

Third, comparative linguistics can tell us something about the culture of the people who spoke the protolanguage, and about the changes that have taken place in that culture. If a set of words can be reconstructed for

![Figure 4. Establishing Migration Patterns]
a protolanguage, the items or concepts they refer to were probably also present in the protoculture. For example, if we could reconstruct for a protolanguage words for *taro*, *yam*, *coconut*, and *breadfruit*, then we could presume that these items were in the original culture of the people who spoke that language. And if the daughter languages have quite unrelated words for *peanut*, *rice*, *coffee*, and *sweet potato*, then we could assume that these items were not in the original culture, but represent later innovations. The identification of copied words can also tell us quite a bit about another aspect of linguistic and social history—cultural contact between groups of people speaking (related or unrelated) languages.

**1.3.5. Time Depths**

Finally, a word of warning. The principles and techniques of comparative linguistics allow linguists to trace relationships between languages going back perhaps eight or ten thousand years, and to make associated conclusions regarding migrations, cultures, and so on. If, however, the initial breakup of a language family took place longer ago than about ten thousand years, linguists often cannot find sufficient evidence to prove that the languages involved are related. The changes that have taken place in each language over the millennia are usually so great that very few similarities can be distinguished or reconstructed.

The hypothetical family tree in figure 5 helps illustrate this point. The similarities currently existing between the modern languages P through Y would probably lead comparative linguists to divide them into four unrelated families:
1. the C family, with members P and Q;
2. the D family, with members R, S, and T;
3. the E family, with members U and V; and
4. the F family, with members W, X, and Y.

The true historical picture is presented in the diagram, which shows how all these languages are related, deriving from a common ancestor X. Because of the length of time involved, however, the changes have been so great that most similarities between, say, languages P and Y have been lost, which is why linguists would treat these languages as belonging to four distinct families.

The study of prehistory relies heavily on comparative linguistics for many different kinds of information. But it is important also to realize that—at least with the techniques currently at our disposal—comparative linguistics has limitations.