

# LIMITATIONS OF LEXICO-STATISTICS FOR SUBGROUPING

## EXAMPLES FROM FOUR EASTERN OCEANIC LANGUAGES

D.S. Walsh  
(University of Sydney)

*The starting-point for this paper is a sample lexico-statistical comparison which uses basic vocabulary from four Eastern Oceanic languages as its raw data. The subgrouping implications of the cognacy percentages produced by this comparison turn out to be largely indeterminate. Some possible reasons for this are suggested.*

### 1. THE LEXICO-STATISTICAL COMPARISON

#### 1.1 THE LANGUAGES AND THEIR DEGREES OF GENETIC RELATIONSHIP

The four languages from which the data are drawn are New Zealand Maori, Tongan, East Fijian (represented by the Bau dialect)<sup>1</sup> and Raga (spoken on Pentecost Island in the North-East New Hebrides).

These languages are widely considered (ref. Pawley 1972, following Biggs 1965) to be members of a higher level subgrouping of East Austronesian languages which is termed Eastern Oceanic (EO). This grouping also includes the other Polynesian (PN) languages, West Fijian, and most languages of the Northern New Hebrides, the Banks and Torres Islands, and the South-East Solomons.<sup>2</sup> Pawley (1972:98) has proposed two highest order subgroups within EO - Southeast Solomonian and North Hebridean-Central Pacific, which latter has as its highest order subgroups North Hebridean (which includes Raga) and Central Pacific (which includes Maori, Tongan and East Fijian). This Central Pacific grouping is essentially compatible with a grouping proposed by Grace (1959 and 1961)<sup>3</sup> which, in addition to the PN and Fijian languages, included Rotuman.<sup>4</sup> There is general agreement that the PN languages form a well-defined lower order grouping within EO. This PN grouping has two highest order subgroups (ref. Walsh and Biggs 1966:iii-iv), one of which includes Maori and the other Tongan.<sup>5</sup>

In terms of these EO subgroupings, then, of the four languages here being considered, Maori and Tongan are more closely related to each other than either of them is to East Fijian or Raga, and Maori, Tongan and East Fijian are more closely related to each other than any of them is to Raga.

## 1.2 DATA AND COGNACY ASSESSMENT

The data (ref. Table 1) consist of equivalents for the semantic categories of the 93-item test list used by Grace (1961) and Walsh (1963). This list is a slightly modified version of the Swadesh 100-item test list of basic vocabulary. The equivalents were obtained from native-speaking informants who considered them to be the most common conversational forms for the categories of the test list.

Table 1: Data and Cognacy Assessment<sup>6</sup>

Notes:-

- (i) Orthographic symbols have conventional phonetic values, except that Tongan ' represents *glottal stop*; East Fijian c represents *voiced apico-dental fricative*; Raga g represents *velar fricative*; and East Fijian and Raga  $\bar{g}$  represent *homorganically prenasalised voiced velar stops*.
- (ii) Raga bw, mw and vw represent unit labio-velar phonemes.
- (iii) Portions of equivalents that have been disregarded in cognacy assessment are enclosed in parentheses.
- (iv) Cognacy has been assessed on the basis of phoneme correspondences which do not depart radically from those proposed by Pawley (1972:27-30). The assessments are indicated as follows:-
  - + definite cognate
  - (+) probable cognate
  - non-cognate
- (v) [+] indicates a positive cognacy rating for an item that has not been included in the total for a given language pair because of duplication of equivalents. This has occurred with the pairings Maori/Tongan, Maori/East Fijian and Tongan/East Fijian, whose equivalents for item 3 duplicate those for item 68. It has occurred again with the pairing Maori/Tongan, whose equivalents for item 58 duplicate those for item 45.

TEST LIST	MAORI (M)	TONGAN (T)	EAST FIJIAN (F)	RAGA (R)	MT	MF	MR	TF	TR	FR
1 <i>all</i>	katoa	kaatoa	taucoko	ŋoto	+	-	-	-	-	-
2 <i>ashes</i>	(puna) rehu	efu(efu)	dravu(sa)	(tani) avu	+	+	+	+	+	+
3 <i>bark</i>	kiri	kili('i' akau)	kuli	vinungai	[+]	[+]	-	[+]	-	-
4 <i>belly</i>	puku	kete	kete	sibweni	-	-	-	+	-	-
5 <i>big</i>	nui	lahi	levu	gaivua	-	-	-	-	-	-
6 <i>bird</i>	manu	manu(puna)	manu(manu)	manu	+	+	+	+	+	+
7 <i>bite</i>	kati	u'usi	kati	gasi	-	+	+	-	-	+
8 <i>black</i>	manu	'uli'uli	loaloa	meto	-	-	-	-	-	-
9 <i>blood</i>	toto	toto	dra	daga	+	-	-	-	-	-
10 <i>bone</i>	iwi	hui	sui	hui	+	+	+	+	+	+
11 <i>burn</i>	tahu	vela	kama	tunu	-	-	-	-	-	-
12 <i>cloud</i>	kapua	'ao	oo	tenlani	-	-	-	(+)	-	-
13 <i>cold</i>	makariri	momoko	liliwa	masisi	-	-	-	-	-	-
14 <i>come</i>	haere (mai)	ha'u	lako (mai)	mai	-	-	-	-	-	-
15 <i>die</i>	mate	mate	mate	mate	+	+	+	+	+	+
16 <i>dog</i>	kurii	kulii	kolii	vwiriu	+	+	-	+	-	-
17 <i>drink</i>	inu	inu	ŋunu	mwinu	+	+	+	+	+	+
18 <i>dry</i>	maroke	momoa	mamaca	mamaha	-	-	-	-	-	+
19 <i>ear</i>	tariŋa	telina	daliŋa	bwero	+	+	-	+	-	-
20 <i>earth</i>	fenua	kele(kele)	gele	tano	-	-	-	+	-	-
21 <i>eat</i>	kai	kai	kana	gani	+	+	+	+	+	+
22 <i>egg</i>	heeki	fo'imoa	yaloka	idoli	-	-	-	-	-	-
23 <i>eye</i>	mata	mata	mata	mata	+	+	+	+	+	+
24 <i>feather</i>	huruhuru	fulufulu	vuti	lalau	+	-	-	-	-	-
25 <i>fire</i>	ahi	afi	bukawaŋa	gabi	+	-	-	-	-	-
26 <i>fish</i>	ika	ika	ika	ige	+	+	+	+	+	+
27 <i>fly (v.)</i>	rere	puna	vuka	gaga	-	-	-	-	-	-
28 <i>foot</i>	wae(wae)	va'e	yava	bwalage	+	-	-	-	-	-
29 <i>give</i>	ho(mai)	'o(mai)	solii	lai	-	-	-	-	-	-





62	root	take	aka	vuna	garo	-	-	-	-	-	-	-	-
63	sand	one	'one	nuku	one	+	-	-	-	-	-	-	-
64	say	mea	lea	kaya	veve	-	-	-	-	-	-	-	-
65	see	kite	sio	raica	gita	-	(+)	-	-	-	-	-	-
66	seed	kaakano	teŋa	sorenikau	biri	-	-	-	-	-	-	-	-
67	sit	noho	ta'utu	dabe	togo	-	-	-	-	-	-	-	-
68	skin	kiri	kili	kuli	guli	+	+	+	+	+	+	+	+
69	sleep	moe	mohe	moce	maturu	+	+	-	-	-	-	-	-
70	small	iti	si'isi'i	lailai	tirigi	(+)	-	-	-	-	-	-	-
71	smoke	au(ahi)	'ahu	kubou	ahu	+	-	+	+	+	+	+	+
72	stand	tuu	tu'u	tuu	tu	+	+	+	+	+	+	+	+
73	star	fetuu	fetu'u	kalo	visiu	+	-	(+)	-	-	-	-	-
74	stone	(poo) fatu	maka	vatu	vatu	-	+	+	+	+	+	+	+
75	sun	raa	la'aa	matanisiŋa	alo	+	-	-	-	-	-	-	-
76	swim	kaukau	kakau	galo	gagaru	+	-	(+)	-	-	-	-	-
77	tail	fiore	hiku	bui	gere	-	-	-	-	-	-	-	-
78	that	teenaa	ko ena	o ya	gea	(+)	-	-	-	-	-	-	-
79	this	teenei	ko eni	oŋo	geki	(+)	-	-	-	-	-	-	-
80	tongue	arero	'elelo	yame	mea	+	-	-	-	-	-	-	(+)
81	tooth	niho	nifo	bati	livo	+	-	(+)	-	-	-	-	-
82	tree	raakau	'akau	kau	gai	+	(+)	(+)	(+)	(+)	(+)	(+)	(+)
83	two	rua	ua	rua	rua	+	+	+	+	+	+	+	+
84	walk	haere	'alu	taubale	lago	-	-	-	-	-	-	-	-
85	warm	mahana	mafana	tunutunu	aruaru	+	-	-	-	-	-	-	-
86	water	wai	vai	wai	wai	+	+	+	+	+	+	+	+
87	we (pl. excl.)	maatou	kimautolu	keimami	kamai	(+)	-	-	-	-	-	-	(+)
88	what?	aha	haa	cava	hava(nau)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
89	white	maa	hinehina	vulavula	maita	-	-	-	-	-	-	-	-
90	who?	ko wai	hai	cei	ihei	+	+	+	+	+	+	+	+
91	woman	wahine	fefine	yalewa	vavine	+	-	+	+	+	+	+	+
92	yellow	koofai	eŋeŋa	dromodromoa	aŋoga	-	-	-	-	-	-	-	-
93	you (sg.)	koe	koe	ko	giŋo	+	+	+	+	+	+	+	(+)

### 1.3 THE COGNACY PERCENTAGES AND THEIR SUBGROUPING IMPLICATIONS

Table 2: Raw Cognacy Scores and Cognacy Percentages

Language pair	No. of pairs of equivalentents rated + or (+)	No. of pairs of equivalentents used in calculation, i.e. 93 minus no. of [+] ratings	Cognacy percentages (to nearest whole no.)
Maori/Tongan	56	91	62
Maori/East Fijian	34	92	37
Maori/Raga	37	93	40
Tongan/East Fijian	37	92	40
Tongan/Raga	35	93	38
East Fijian/Raga	35	93	38

Of the two subgroupings outlined above (at the end of section 1.1), the percentages in Table 2 strongly support that of Maori and Tongan as against the other two (ungrouped) languages, but they do not support that of Maori, Tongan and East Fijian as against Raga. Furthermore, they do not support either of the other theoretically possible binary subgroupings of the set Maori-Tongan, East Fijian and Raga, *viz.* Maori, Tongan and Raga as against East Fijian or East Fijian and Raga as against the two PN languages. To put it another way, with the exception of the very clear support they give for the PN subgrouping, these figures do nothing to clarify the subgrouping picture within EO below the level of Pawley's North Hebridean-Central Pacific grouping (ref. section 1.1). Some possible reasons for this indeterminacy will now be considered.

## 2. IN WHAT CIRCUMSTANCES IS COMPARISON OF BASIC VOCABULARY LIKELY TO PRODUCE RESULTS THAT ARE INDETERMINATE FOR SUBGROUPING PURPOSES?

### 2.1 THE PROPORTIONAL INCIDENCE OF "COMMON COGNACY"

"Common Cognacy" is here defined as cognacy that is present for all possible pairings of the members of a given set of languages. In the lexico-statistical comparison of any set of related

languages in terms of basic vocabulary there are inevitably (*because* the languages are related) going to be some test list items the equivalents for which are going to manifest Common Cognacy. Table 3 illustrates some striking differences in the levels of incidence of Common Cognacy between Maori/Tongan and the other language pairs in the above example.

Table 3: Levels of Incidence of Common Cognacy

Note:- Common Cognacy is manifested by the equivalents for 27 of the items in the example detailed in Table 1. All percentages are rounded to the nearest whole number.

Language pair	Incidence of Common Cognacy expressed as a percentage of the total no. of cognates	Percentage incidence of Common Cognacy expressed as a percentage of the total cognacy percentage
Maori/Tongan	48	48
Maori/East Fijian	79	78
Maori/Raga	73	73
Tongan/East Fijian	73	73
Tongan/Raga	77	76
East Fijian/Raga	77	76

This table makes it obvious that where the range of variation between language pairs in the total number of items counted for purposes of cognacy percentage calculation is small - in this case it is only from 91 to 93 - it makes no significant difference whether the levels of incidence of Common Cognacy are calculated on the basis of Common Cognacy as a proportion of raw cognacy scores or of Common Cognacy as a percentage of the total number of items counted for a given language pair considered as a porportion of the cognacy percentage for that pair. However, when this range of variation is large, use of the second method will produce sounder results.

It is also apparent from Table 3 that when the difference between levels of percentage incidence of Common Cognacy for two language pairs is of the order of about 25 to 30, the language pair having the lower level is more likely to emerge as a well-

defined subgroup simply because much more - in this case 52 percent - of the total incidence of cognacy for that pair is potentially significant for subgrouping purposes than is the case with the language pair having the higher level. For the language pairs other than Maori/Tongan in the example under consideration in this paper, the proportion of cognacy for a given language pair that is potentially significant for subgrouping purposes ranges from 22 to 27 percent. In section 2.2 below it will be demonstrated that, in this example at least, the greater part of this 22 to 27 percent of cognacy emerges as non-significant for subgrouping purposes within the set of languages being considered.

## 2.2 THE SIGNIFICANCE OF RETENTION OF A GIVEN FORM-PLUS-MEANING RESEMBLANCE BY SOME BUT NOT ALL OF THE LANGUAGES THAT ARE BEING COMPARED

The Common Cognacy discussed above is eminently attributable to retention, by each of the languages in the set under consideration, of form-plus-meaning resemblances that are ascribable to a period of development common to the entire set. When, however, retention has occurred in some but not all of the languages in the set, this fact will *not* be revealed by the comparison of basic vocabulary. In order to discover the incidence of this kind of retention it is necessary to have recourse to evidence external to the set in question.

Because of the lack of a relevant corpus of reconstructions for Proto-North Hebridean-Central Pacific, the external evidence that is used below consists of the corpus of \*EO and supra-\*EO reconstructions listed in Wurm and Wilson (1975). The presence of a relevant \*EO or supra-\*EO reconstruction in this list is here regarded as *prima facie* evidence of retention by the daughter languages.

Table 4: Incidence of Retention of Form-Plus-Meaning Resemblances in Some But Not All of the Languages in the Set

Language grouping sharing a given resemblance	Items for which a resemblance is shared	Items which manifest reflexes of *EO or supra-*EO reconstructions listed in Wurm and Wilson <sup>7</sup>	No. of items remaining as potentially significant for subgrouping below the level of EO
---	---	--	---

Maori/Tongan/East Fijian	16 19 52 58 69 Total: 5	19 *EO *taliŋa *(n)taliŋa 52 *EO *po'oRu 58 *EO *tamwata 69 *EO *moze Total: 4	1 (item 16)
Maori/Tongan/Raga	56 63 71 73 76 81 91 Total: 7	56 *EO *t[a,i]ka(i) 63 *EO *'one *one(one) 71 *EO *'aðu *qazu 73 *EO *petuqu *pitu'u 81 *EO *(1,n)ivo 91 *EO *vavine Total: 6	1 (item 76)
Maori/East Fijian/ Raga	7 74 Total: 2	7 *EO *kati 74 *EO *vatu Total: 2	-
Tongan/East Fijian/ Raga	-	-	-
Maori/Tongan	1 9 24 25 28 41 43 45 48 50 57 70 75 78 79 80 85 87 Total: 18	9 *OC *toto 25 *EO *api 28 *EO *waqe 45 *OC *taŋmata 78 *EO *ena 79 *EO *eni 85 *EO *mapana(pana) Total: 7	11 (items 1 24 41 43 48 50 57 70 75 80 87)
Maori/East Fijian	-	-	-
Maori/Raga	65 Total: 1	65 *OC *kita Total: 1	-
Tongan/East Fijian	4 12 20 34 49 Total: 5	4 *EO *kete 12 *OC *i(n)ti 20 *OC *ŋkele 34 *OC *qulu 49 *EO *ŋu(s,t)u *ŋu(ts)u Total: 5	-

Tongan/Raga	39	39 *OC *'ilo	-
	Total: 1	Total: 1	
East Fijian/Raga	18 45 50	18 *EO *ma/ma(n)sa	2
	57 80 87	45 *EO *mwane	(items 57 87)
	Total: 6	50 *EO *qansa	
		*'aōa	
		80 *OC *maya	
		*(a)me(a)	
		Total: 4	

Table 4 makes it apparent that for all theoretically possible sub-EO groupings other than Maori/Tongan within the set of languages under consideration there is little or no evidence arising from the comparison of basic vocabulary that is potentially indicative of uniquely shared innovation as the cause of form-plus-meaning resemblance. In this situation it is hardly surprising, since uniquely shared innovation constitutes the crucial evidence for subgrouping, that the cognacy percentages produced by this comparison have virtually no positive significance for subgrouping possibilities within EO other than that of Maori and Tongan.

### 3. CONCLUSION

On the basis of the evidence presented above it is here suggested that, in a lexico-statistical comparison of the kind under consideration, the proportional incidence of *Common Cognacy*, and of *Retention that is shared by some but not all the languages in question*, can provide some explanation for the occurrence of results that lack positive significance for subgrouping purposes.

It is not being suggested here that the subgrouping of the Polynesian and Fijian languages (Pawley's Central Pacific grouping) within EO is necessarily insupportable, but it is being suggested that, for reasons that have been indicated above, the comparison of Maori, Tongan, East Fijian and Raga in terms of basic vocabulary does not support this Central Pacific grouping.

## BIBLIOGRAPHY

- Biggs, B., 1965. "Direct and Indirect Inheritance in Rotuman". *Lingua*, 14: 383-415.
- Dyen, I., 1960. Review of Grace (1959) in *Journal of the Polynesian Society*, 69: 2: 180-4.
- 1965. *A Lexicostatistical Classification of the Austronesian Languages*. Baltimore, Waverly Press, (Indiana University Publications in Anthropology and Linguistics, Memoir 19 of the International Journal of American Linguistics).
- Grace, G.W., 1959. *The Position of the Polynesian Languages Within the Austronesian (Malayo-Polynesian) Language Family*. Baltimore, Waverly Press, (Indiana University Publications in Anthropology and Linguistics, Memoir 16 of the International Journal of American Linguistics).
- 1961. "Lexicostatistical Comparison of Six Eastern Austronesian Languages". *Anthropological Linguistics*, 3: 9: 1-22.
- 1976. Review of Pawley (1972) in *Journal of the Polynesian Society*, 85: 1: 103-12.
- Green, R.C. and M. Kelly (eds.), 1972. *Studies in Oceanic Culture History, Volume 3*. Honolulu, Department of Anthropology, Bernice P. Bishop Museum, (Pacific Anthropological Records, Number 13).
- Pawley, A., 1972. "On the Internal Relationships of Eastern Oceanic Languages". In Green and Kelly (1972: 1-142).
- Pawley, A. and T. Sayaba, 1971. "Fijian Dialect Divisions: Eastern and Western Fijian". *Journal of the Polynesian Society*, 80: 4: 405-36.
- Walsh, D.S., 1963. "Dictionaries Versus Informants: An Aspect of Glottochronology". *Te Reo* 6: 30-38.
- Walsh, D.S. and B. Biggs, 1966. *Proto-Polynesian Word List I*. Auckland, Linguistic Society of New Zealand, (Te Reo Monograph).
- Wurm, S.A. and B. Wilson, 1975. *English Finderlist of Reconstructions in Austronesian Languages (Post-Brandstetter)*. Canberra, Department of Linguistics, Research School of Pacific Studies, Australian National University, (Pacific Linguistics, Series C, No. 33).



## NOTES

- 1 The basis for regarding East Fijian and West Fijian as distinct languages is discussed in Pawley and Sayaba (1971: 427-34).
- 2 The Heonesian Linkage proposed by Dyen (1965: 38-39) includes many of these EO languages.
- 3 Dyen (1960) was less than satisfied with the grounds proposed for this grouping in Grace (1959).
- 4 Pawley (1972) eventually left Rotuman ungrouped within EO.
- 5 Grace (1976: 104-10) provides some further perspective on the sub-groupings mentioned in this paragraph.
- 6 Thanks are tendered to Ross Clark (University of Auckland) for convincing me that ratings of (+) rather than - were warranted in the following cases:- Maori/Tongan, items 56 and 70; Maori/East Fijian, items 53 and 88; Maori/Raga, items 53, 56, 82 and 88; Tongan/East Fijian, items 12, 53 and 88; Tongan/Raga, items 53, 56, 82 and 88; East Fijian/Raga, item 82. Responsibility for the use that has been made of these ratings is, of course, entirely mine. These changes of rating from - to (+) have slightly increased the already strong support given by this sample comparison to the main argument of the paper.
- 7 The orthographic presentation of reconstructions in this column reproduces that of Wurm and Wilson, who followed the various orthographies of their sources. The following orthographic parallels occur in Table 4 in cases where two reconstructions are linked with a given item number:- ' and q, ð and z, (s,t) and (ts), ns and ð. Where a symbol in the reconstructions occurs within parentheses its presence is not conclusively established, and where two symbols occur within parentheses, e.g. 81 \*EO \*(l,n)ivo, or, in one orthography, within square brackets, viz. 56 \*EO \*t[a,i]ka(i) the evidence does not permit firm choice between the alternative reconstructions.