

DICTIONARIES VERSUS INFORMANTS: AN ASPECT OF GLOTTOCHRONOLOGY

D. S. Walsh

(University of Auckland)

1. A Brief Outline of the Theory and Method of Glottochronology.

Glottochronology¹ is a field of study which has recently attained some prominence in historical linguistics. It is concerned with the study of the rate of vocabulary change in language, the use of this rate for historical inference (especially for estimating time depths), and the use of such time depths to assist in the establishment of sub-groups within a language family².

The assumptions of the theory of glottochronology are that the rate of morpheme decay in the basic vocabulary of languages is constant, and that therefore a retention rate constant can be derived from historically known languages and applied to languages whose history is unknown. On this basis, pairs of languages can be compared for the percentage of cognacy in basic vocabulary, and the time depth since the separation of those languages can be calculated, using a suitable formula.

Current discussions among linguists who have worked in this field indicate that there is still scope for increased rigour in assessing the rate of decay on which the retention rate is based, and even that the assumption of a constant rate of change for all languages is fallacious.³ In this paper the validity of the theory is assumed, and consideration is given mainly to some of the problems which arise when the theory is put into practice.

In order to compare two languages so that the percentage of shared cognates in basic vocabulary can be determined, the method which has been established is to obtain for each item of a standard English basic word list the single most commonly used equivalents in the languages under consideration, and to assess the resulting pairs of equivalents for shared cognacy.

Equivalents are assessed as cognate if they are derivable one from the other by the use of a systematic set of phoneme correspondences furnished by the application of the traditional comparative method to the language family under consideration.

This method introduces two operations which are of potentially crucial significance to the quality of the final results — a) the procedures used to obtain the equivalents; b) the procedures used in assessing cognacy.

The optimum requirements for the compiling of lists of equivalents for the test list items are that one satisfactory equivalent be obtained for each of as many items as possible, and that the equivalents be those most commonly used at the time the study is made. These requirements would seem best to be met when the lists are elicited from informants by an investigator who has an adequate knowledge of the languages concerned.

The optimum requirements for rigorous and consistent assessment of shared cognacy for pairs of equivalents are that there be one equivalent in each language for each item being compared, and that the comparative phonology of the languages concerned be known.

2. An Application of Glottochronology in Which Dictionary Sources Were Used.

In Grace (1961) a lexicostatistical comparison of all the possible pairings for six eastern Malayo-Polynesian languages — *Tongan*, *Maaori*, *Fijian*, *Rotuman*, *Mota* (Banks Islands, New Hebrides), and *Sa'a* (Malaita, Solomon Islands) — was made. The comprehensive publication of the raw data and its processing for this set of comparisons has made possible the comparative re-study of the same problem area.

1 Gudschinsky, 1956b; Hockett, 1958: 526–35; Hymes, 1960: 3–4.

2 Hymes, 1960: 4.

3 Bergsland and Vogt, 1962; Chretien, 1962; Greenberg, 1957; Gudschinsky, 1956a; Holjer, 1956; Hymes, 1960, 4–15; Swadesh, 1955.

The hypothesis which Grace was testing was that *Fijian*, *Rotuman*, and the Polynesian languages are more closely related to each other than any of them is to either *Mota* or *Sa'a*.

For his test list Grace used a 200 word list based largely on the standard 215 word list with which the retention constant was originally determined⁴. This 200 word list contained 93 of the 100 words of Swadesh's revised list of basic vocabulary items⁵.

Grace's method of obtaining equivalents was as follows:

The equivalents were compiled almost exclusively from published works. This is not an entirely satisfactory procedure: it has certainly resulted in some measure of error. However, it should be pointed out that no method of determining test-list equivalents, short of acquiring an intimate knowledge of each of the languages, appears to be entirely satisfactory. One aspect of the difficulty in determining equivalents was that often it was impossible to choose among two or more alternative forms.... I have regretfully omitted any indication of vowel length for the Polynesian languages because the available sources provided either inadequate or conflicting information⁶.

The sources which Grace used for his equivalents were:— *Tongan*: Churchward, C.M., *Tongan Grammar* (1953); Colomb, A., *Dictionnaire Toga-Francaise et Francais-Toga-Anglais* (1890); and a short *Tongan* vocabulary which Grace himself had collected. *Maaori*: Reed, A.W., *Concise Maori Dictionary* (1951). *Fijian*: Capell, A., *A New Fijian Dictionary* (1941); Churchward, C.M., *A New Fijian Grammar* (1941). *Rotuman*: Churchward, C.M., *Rotuman Grammar and Dictionary* (1940). *Mota*: Codrington, R.H., *The Melanesian Languages* (1885); Codrington and Palmer, J., *A Dictionary of the Language of Mota* (1896). *Sa'a*: Ivens, W.G., *A Dictionary of the Language of Sa'a and Ulawa* (1929).

The quality of these sources varies considerably, and the *Mota* sources are non-contemporary to an extent which could be significant for glottochronology. At the theoretically normal rate of change of 38 items per millenium *Mota* would be expected to change at least 3 items since 1880. This amount of change could be critical in a study in which nearly 1 in 3 of the relevant decisions as to relative closeness of relationship are made on the basis of differences of under 6 % between the percentages of shared cognates of the various language pairs.

When Grace was assessing cognacy, because his sources often left him with two or more equivalents in one language for one test list item, he was forced, since he needed to use those items to achieve a significant quantity of comparisons, to adopt a more than usually complex procedure. In his initial scoring he used six categories:— i) cognate beyond reasonable doubt; ii) non-cognate beyond reasonable doubt; iii) doubtful cognate; iv) doubtful non-cognate; v) partially cognate item (where it was impossible to select a single best equivalent of the test-list item for one or both languages, and where some selections would produce cognate pairs and some would not); vi) partially doubtful cognate (where it was impossible to select a single best equivalent, and where some selections would produce doubtfully cognate pairs)⁷.

Grace's comment on his categories *doubtful cognate* and *doubtful non-cognate* is:

There were a considerable number of cases where the equivalents in the two languages approached agreement according to the known sound correspondences, but showed some irregularity. Such cases might be due to chance or borrowing, on the part of one of the languages, of a cognate form in a related language. On the other hand they might be genuine cognates, in the sense of comparative linguistics, whose status was obscured by the incompleteness of our knowledge of the phonological history of one of the two languages. Such items were scored 'doubtful cognate'. Where the equivalents in the two languages appeared not to be cognate, but there was a significant possibility that they contained cognate morphemes, or there were other suspicious similarities, they were scored 'doubtful non-cognate'⁸.

He further notes:

Doubtful cases, where they encountered in considerable numbers, pose a problem in the application of lexicostatistics. If we simply choose to eliminate them from consideration,

⁴ Swadesh, 1952

⁵ " " "

⁶ Grace, 1961: 2-3

⁷ " " " : 3

⁸ Grace, 1961: 3

we run the risk of introducing serious error into our calculations, since, by definition we do not know what kind of a selection the omitted items represent. However, if we choose to count them as either cognate or non-cognate according to our best guess we again run the risk of introducing an error. The course followed in the present study was simply to make one count which omitted the doubtful items and another which included them⁹.

For the items which came into the category *partially cognate* and *partially non-cognate* in his initial scoring because of lack of single equivalents, Grace adopted a re-scoring procedure of some mathematical complexity which had the net result of producing "the same results as would occur on the average if selection among the alternatives were made by tossing a coin"¹⁰

Grace's comments on his method included the following points:

"The treatment in this study of 'partially cognate' and 'partially doubtful cognate' items does not seem very satisfactory. It is too time-consuming, and the formula concerned is not above reproach"¹¹.

"Alternative equivalents will probably not be frequent enough in most studies to merit serious concern. However, it does seem on the basis of the experience here that the only completely satisfactory solution of the problem they raise is a thorough-going knowledge of the languages being compared"¹².

Thus it can validly be said that several potentially serious problems were created directly for Grace by his use of dictionary sources. The more important of these problems (the existence of most of which has been recognised by Grace) are:—

a) The omission of phonemically significant vowel length could affect cognacy assessments, especially in doubtful cases.

b) The quality of the vocabulary sample may be adversely affected for glottochronological purposes by the use of less common, archaic, or even semantically erroneous equivalents.

c) The necessity to include two or more equivalents for a considerable number of test list items leads directly to an error-producing cognacy scoring procedure.

d) The use of non-contemporary sources for *Mota* (a crucial language for Grace's hypothesis) provides strictly non-comparable data for the comparisons in which this language is involved.

e) Incomplete structural knowledge of one or more of the languages concerned can adversely affect the quality of the test list items.

3. A Re-study of the Same Problem Area, Using Informants as Sources.

With these problems in mind, and with an awareness of the considerable shortcomings of Grace's dictionary source for Maaori, the Stage II Maaori Studies Class of 1962 at the University of Auckland made a re-study of the comparison of the six languages in question. For all but one of the languages there was an English-speaking informant available from the ethnically diverse population of Auckland. No speaker of *Sa'a* was available, but there was an informant who spoke *Ulawā*, a closely related language from the island of Ulawa adjacent to the *Sa'a* speaking area of Malaita. According to the *Ulawā* informant there is almost complete mutual intelligibility between the two languages (or dialects?).

Using the same test list as Grace, the class obtained lists of equivalents for each language, with two students working with each informant. Precautions were taken to standardize procedure, and in all cases where equivalents were obtained the single most commonly used word was recorded, often after considerable discussion with the informants. The fact that such discussion was so often found to be necessary and possible when an informant was being used is an indication of the lack of semantic precision which must attend much of the use of dictionary sources for such operations.

The use of informants virtually eliminated multiple equivalents for the test list items, and it was therefore possible to use a much simpler set of categories for cognacy assessment than did Grace. In the re-study three categories were used: i) cognate; ii) non-cognate; iii) doubtful

⁹ Grace, 1961: 3-4
¹⁰ " " : 4

¹¹ Grace, 1961: 7

¹² Grace, 1961: 7

cognate — where the known phoneme correspondences gave partial indication of cognacy, but where *lacunae* in the available knowledge of such correspondences made it impossible to give an assessment of definite cognacy without a significant lowering of standards.

The use of informants in the re-study made it possible to reduce to some extent the areas of problem and difficulty which the use of dictionary sources had imposed on Grace:

a) It was anticipated that the inclusion of phonemic vowel length for all the languages would reduce one potential source of error or confusion in assessing cognacy. That it did not in fact make such a reduction was due to the very incomplete information on the relevant comparative phonologies as far as vowel length was concerned.

b) The obtaining of equivalents which were considered by the informants to be those in most common contemporary use in their speech gave each of the lists of equivalents a more even and high quality for the purposes of glottochronological comparison.

c) It was possible to have a set of categories for cognacy assessment which had fewer error-producing tendencies.

d) The obtaining of contemporary data for all six languages eliminated the source of possible error inherent in Grace's non-contemporary data for *Mota*.

e) In this re-study the possibly adverse effect of incomplete structural knowledge of some of the languages (especially *Mota* and *Ulawa*) was not necessarily greatly reduced by the use of informants.

4. Some Details of the Differences in the Raw Data of the Two Studies.

Table I indicates the extent of the five relevant categories of quantitative difference between the two sets of lists of equivalents. Equivalents were not classed as differing when there was merely variation occasioned by reduplication of morphemes, or by the presence or absence of structural affixes.

T A B L E I

The Quantitative differences between the re-study lists of equivalents and those of Grace.

Language	A	B	C	D	E	TOTAL
Tongan	37	—	2	1	7	47
Maaori	62	8	9	3	5	87
Fijian	20	7	23	—	5	55
Rotuman	18	10	34	—	8	70
Mota	21	15	40	4	4	84
Ulawa	42	14	23	25	7	111

Column: A Grace has a different equivalent.

B Grace has more than one equivalent, all of which differ from the re-study equivalent.

C Grace has more than one equivalent, one of which is the same as the re-study equivalent.

D Only Grace has an equivalent.

E Only the re-study has an equivalent.

In this table phonemic vowel length accounts for 18 of the category "A" differences for *Tongan*, and 17 of those for *Maaori*. Grace's non-inclusion of phonemic vowel length was not one of the factors which were eventually regarded as causing significant difference between his

assessments of cognacy and those of the re-study

The high total of differences for *Sa'a/Ulawa* must be attributable partly to the language difference, but the high proportion of identical equivalents is an indication of the closeness of the relationship between the two languages.

Table 1 also indicates the extent to which multiple equivalents featured in Grace's study - 2 for *Tongan*, 17 for *Maaori*, 30 for *Fijian*, 44 for *Rotuman*, 55 for *Mota*, and 44 (including 7 items for which only Grace had equivalents) for *Sa'a*. The high proportion of multiple equivalents in Grace's lists - almost 1 item in 6 - further emphasises the imprecision attendant on the use of dictionary sources in a study such as this. Grace recognised this, but was unable to eliminate it in the circumstances in which he carried out his study.

Table 2 (see p. 40) gives a detailed indication of the qualitative differences between the re-study list and that of Grace for the language with which those engaged in the re-study were best acquainted, *Maaori*. Of the 87 differences, 17 are due solely to phonemic vowel length, 9 to multiple equivalents of which one was the same as the re-study equivalent, and 8 to either Grace or the re-study not having an equivalent.

In 45 cases out of a possible 200 Grace has obtained a radically different single equivalent from that obtained by investigators who had considerable knowledge of *Maaori*, and who were working with a sophisticated bilingual native speaker as informant. In a further 8 cases Grace obtained multiple equivalents which all differed from the re-study equivalent.

Some of these differing equivalents of Grace's are "old-fashioned" words (e.g., *taru* as opposed to *karaake* for "grass"), and some others are less commonly used words or dialectal variants with a high correspondence to the test list item (e.g., *katau* as opposed to *matau* for "right-hand").

Grace's dictionary source for *Maaori* was certainly the least adequate of his sources, but the example of his *Maaori* list does indicate some of the hazards and shortcomings of dictionary sources at the level of the raw data.

5. Comparison of the Cognacy Percentages for the Two Studies.

Table 3 presents the cognacy percentages for the two studies for each of the four calculations which Grace used. In spite of the high number of differences between the raw data of the studies, the differences in the cognacy percentages are small, ranging from -7 to +4. Differences of this order are statistically non-significant¹³.

T A B L E 3

Differences in cognacy percentages

(For each calculation the restudy % is in the top line and Grace's % in the lower line).

Calculation	TO MA	TO ROT	TO FI	TO MOT	TO SA	MA ROT	MA FI	MA MOT	MA SA	ROT FI	ROT MOT	ROT SA	FI MOT	FI SA	MOT SA
1.	41	25	26	17	15	20	20	16	15	17	13	13	17	15	16
	38	26	24	16	18	20	19	16	15	19	13	11	18	15	17
2.	42	29	29	21	19	24	22	19	17	20	18	18	21	20	24
	41	31	29	24	21	25	22	20	18	22	18	16	23	22	22
3.	54	32	35	19	22	29	30	20	21	28	15	16	20	19	23
	54	33	34	21	23	29	29	21	20	31	17	14	21	20	21
4.	56	37	39	24	28	34	33	24	24	30	24	23	26	27	28
	56	38	36	31	28	33	31	28	24	34	24	19	28	26	26
Calculation.	1. for 200 word list - $\frac{\text{No. of definite cognates}}{200 \text{ minus No. of comparisons assessed as doubtfully cognate.}}$														
	2. for 200 word list - $\frac{\text{No. of definite cognates} + \text{No. of doubtful cognates}}{200}$														
	3. for 93 word list - $\frac{\text{No. of definite cognates}}{93 \text{ minus No. of comparisons assessed as doubtfully cognate.}}$														
	4. for 93 word list - $\frac{\text{No. of definite cognates} + \text{No. of doubtful cognates}}{93}$														

13 Dyen, 1962: 41-42

Why, when there is as much difference as has been indicated between the raw data of the dictionary-based study and the informant-based study, is the extent of difference between the cognacy percentages as low as it is? It is not difficult to see some of the reasons for this:

a) Some of the differences do not affect the cognacy status of the item even when a positive score is involved (e.g., for item 182 the *Maaori* equivalents *tatou* in Grace and *maatou* in the re-study, where *-tou* is the comparable root morpheme).

b) Where the difference consists solely or partly of Grace's having multiple equivalents (columns B and C of *Table I*), his rescoring procedure reduces by 50 % the number of such differences which affect the cognacy totals.

c) Since the percentage of cognates between any two languages in the study is almost always less than 50 (and sometimes less than 15) the chance that any single difference of equivalents will affect a cognacy decision is almost always less than one in two, and sometimes less than one in six.

d) Such alterations in cognacy rating as are made by the differences between the two studies will tend to cancel out.

6. The Effects of the Re-study on Grace's Hypothesis.

Since the relationship between *Sa'a* and *Mota* is not specified by Grace, he is, in effect, postulating two categories — category "A" (*Fijian*, *Rotuman*, *Tongan*, *Maaori*), and category "non-A" (*Sa'a* and *Mota*). If this hypothesis is correct, each language of category "A" should show more relationship with each other language of "A" than it does with either of the languages in the "non-A" category.

Grace's results in general support his hypothesis, except that —

1) In all four calculations, not only does each other language of category "A" have *Tongan* as its closest relative, but also, in some calculations, *Sa'a* (calculations 1, 3, 4) and *Mota* (calculations 2, 4) have *Tongan* as their closest relative.

2) In calculation 2, *Fijian* is more closely related to *Mota* than it is to *Maaori* or *Rotuman*.

3) In calculation 2, *Fijian* is as closely related to *Sa'a* as it is to *Maaori* or *Rotuman*.

Grace considers the first of these exceptions to indicate that *Tongan* has been slightly more conservative than the other languages and has changed its basic vocabulary at a slightly lower rate. He associates the second exception with the high proportion of doubtful cognates in the comparisons involving *Mota*¹⁴.

The results of the re-study also in general support Grace's hypothesis, with a pattern of exceptions that in most respects resembles that of Grace's results.

1) As with Grace's results, all four calculations show each other language of category "A" as having *Tongan* as its closest relative. *Sa'a*, however, has *Mota* as its closest relative in calculations 1, 2, and 3, while in calculation 4 it has both *Mota* and *Tongan* as its closest relatives. *Mota* has *Sa'a* as its closest relative in calculations 2, 3, and 4, while in calculation 1 it has both *Tongan* and *Fijian* as its closest relatives.

2) As with Grace's results again, the re-study results for calculation 2 show a closer relationship between *Fijian* and *Mota* than between *Fijian* and *Rotuman*. In calculation 1 *Fijian* is equally related to both *Mota* and *Rotuman*.

3) In calculation 2, *Fijian* is equally related to both *Sa'a* and *Rotuman*.

Thus the first pattern of exceptions in the results of the re-study conflicts less with Grace's hypothesis than does the comparable pattern in his results.

The patterns of non-congruence with Grace's hypothesis in the second and third groups of exceptions are broadly similar for both studies. That these exceptions are confined to calculation 2 in both Grace's study and (with one exception) in the re-study lends support to the association of these patterns with the high proportion of doubtful cognates in comparisons using the 200 word list and involving *Mota* (and to a slightly lesser extent *Sa'a*).

¹⁴ Grace, 1961: 8-9

Certainly the overall differences between the re-study results and those of Grace are slight, and in spite of the considerable differences between the raw data obtained from dictionaries and that obtained from informants, no significant alteration to Grace's general picture is obtained.

At this stage it seems pertinent to raise the question — does the general support given to Grace's hypothesis by the re-study in any way serve to validate the use of glottochronology for determining sub-groups within a language family, or do both Grace's work and the re-study attribute too great a precision to glottochronology as an instrument for determining such sub-groups?

Dyen maintains that, to be significant in comparisons such as those involved in these studies, differences need to be of the order of at least 10 %.¹⁵ Of the 160 comparisons made in each study to test Grace's hypothesis¹⁶, 101 of Grace's comparisons produced differences of under 10 % (and of these 49 were under 6 %), while 90 of the re-study comparisons produced differences of under 10 % (including 45 of under 6 %).

Bibliography

- Bergsland, K., and Vogt, H., 1962, "On The Validity of Glottochronology", *Current Anthropology*, 3: 115-153.
- Chrétien, C.D., 1962, "The Mathematical Models of Glottochronology", *Language*, 38: 11-37.
- Dyen, I., 1962, "The Lexicostatistical Classification of the Malayopolynesian Languages", *Language*, 38: 38-46.
- Dyen, I., 1963, "Lexicostatistically Determined Borrowing and Taboo", *Language*, 39: 60-66.
- Grace, G.W., 1959, *The Position of the Polynesian Languages Within the Austronesian (Malayo-Polynesian) Language Family*, Honolulu, Bernice P. Bishop Museum (*Special Publication*, 46).
- , ---, 1961, "Lexicostatistical Comparison of Six Eastern Austronesian Languages", *Anthropological Linguistics*, 3: 9: 1-220.
- Greenberg, J.H., 1957, *Essays in Linguistics*, Chicago, University of Chicago Press.
- Gudschinsky, S., 1956a, "Three Disturbing Questions Concerning Lexicostatistics", *IJAL*, 21: 138-49.
- , ---, 1956b, "The ABC of Lexicostatistics (Glottochronology)", *Word*, 12: 175-210.
- Hockett, C.F., 1958, *A Course in Modern Linguistics*. New York, Macmillan.
- Holjer, H., 1956, "Lexicostatistics: A Critique", *Language*, 32: 49-60.
- Hymes, D.H., 1960, "Lexicostatistics So Far", *Current Anthropology*, 1: 3-44.
- Lees, R.B., 1953, "The Basis of Glottochronology", *Language*, 29: 113-127.
- Swadesh, M., 1952, "Lexicostatistic Dating of Prehistoric Ethnic Contacts", *Proceedings of the American Philosophical Society*, 96: 452-63.
- , ---, 1955, "Toward Greater Accuracy in Lexicostatistic Dating", *IJAL*, 21: 121-37.

(Information received from Dr. Grace while this paper was being printed indicates that the re-study test list differed from his for item 90 *live* — re-study *dwelt*, Grace *be alive*; and for item 182 *we* — re-study *1st. p. pl. excl.*; Grace *1st. p. pl. incl.* While these differences will affect the detail of the comparison to some extent, they should not affect greatly the general pattern of the results).

15 Dyen, 1962: 41-42
16 Grace, 1961: 8-9

TABLE 2

Detail of the differences between the re-study list and Grace's list for Maaori.
 Items in italics are part of the 93 word revised list used for calculations 3 and 4.

<u>Item:</u>	<u>English</u>	<u>Re-study Equivalent</u>	<u>Grace's Equivalent(s)</u>
2	and	me	ma
6	back	tuaraa	angahate, tua
9	because	noo te mea	hoki
10	<i>belly</i>	puku	kopu
13	<i>bite</i>	ngau	u
14	<i>black</i>	mangu	& pango
18	breathe	haa	ta
20	child	tamaiti	& potiki
21	<i>cloud</i>	kapua	& ao
25	cut	tapahi	motu
26	day	rangi	& ra
30	<i>dog</i>	kurii	kuri
33	dull	puuhuki	puhuki
36	<i>earth</i>	whenua	oneone
38	<i>egg</i>	heeki	hua
39	<i>eye</i>	mata	& kanohi
41	far	tawhiti	taitua
45	<i>feather</i>	huruhuru	raukura
46	few	iti	ouou
51	float	maanuu	tere
53	flower	putiputi	pua
57	four	whaa	wha
60	<i>give</i>	hoatu	homai, whakawhiwhi
62	grass	karaahe	taru
63	<i>green</i>	kaakaariki	kakariki
64	guts	puku	manawa
65	<i>hair</i>	makawe	& huru, mahunga
75	how	peehea	pehea
76	hunt	kimi	whaiwhai
77	husband	taane	tane, hoa
78	<i>I</i>	au	ahau
79	ice	hukaapapa	--
80	if	me	--
81	in	roto	--
83	<i>know</i>	moohio	mohio
84	lake	roto	& moana
90	live (dwell)	noho	ora
94	<i>man</i>	tangata	tane
95	<i>many</i>	maha	wene
99	<i>mouth</i>	waha	& mangai
103	<i>neck</i>	kakii	ua
104	<i>new</i>	houu	hou
105	<i>night</i>	poo	po
107	<i>not</i>	kiihai	kohore, kaore, kore
110	other	kee	atu

111	<i>person</i>	--	tangata
112	play	porohianga	takaro
113	pull	too	kume
114	push	pana	tute
118	right (side)	matatau	katau
121	<i>root</i>	take	pakiaka
125	salt	tote	mataitai
127	<i>say</i>	mea	korero
128	scratch	rakuraku	natu
131	<i>seed</i>	kaakano	kano
135	sing	waiata	tau
140	<i>small</i>	iti	& riki
141	smell (v.)	whakakakara	rongo
143	smooth	--	maeneene
146	some	eetahi	--
148	split	waahi	titore
149	squeeze	kuti	roromi
150	stab	wero	pumuka
151	<i>stand</i>	tuu	tu, mekari
152	<i>star</i>	whetuu	whetu
153	stick	raakau	tokotoko
154	<i>stone</i>	pohatu	kowhatu
155	straight	tika	tohitu
156	suck	ngote	momi
157	<i>sun</i>	raa	ra
160	<i>tail</i>	whiore	hiku
161	<i>that</i>	teenaa	tena, tera
163	they	raatou	ratou
164	thick	maatotoru	matotoru
165	thin	piirahirahi	kohoi, maiaka
167	<i>this</i>	teenei	tenei
170	throw	kuru	maka
174	<i>tree</i>	raakau	rakau
182	<i>we</i> (1st p.pl. excl.)	maatou	tatou
183	wet	maakuu	maku
187	<i>white</i>	maa	ma
189	wide	whaanui	whanui
190	wife	hoa wahine	wahine, hoa
193	wipe	ukui	miri
194	with	me	--
196	woods	ururaakau	ngahere
200	<i>yellow</i>	--	pungapunga