

Appendix 1:

Data sources and collation

1 Introduction

Data sources which were consulted in relation to a particular terminology are noted in the chapter on that terminology. However, quite a wide range of sources was consulted in the construction of the data base and we list these here, rather than repeating them in each chapter. Sources are conveniently divided into published and unpublished. In alphabetical sequence of language, published sources are:

Anejom	Lynch (2001)
Araki	François (2002)
'Are'are	Geertz (1970)
Arosi	Fox (1978)
Bareke	Tryon and Hackman (1983)
Bugotu	Ivens (1940)
Carolinian	Jackson and Marck (1991)
Cèmuhî	Rivierre (1994)
Chuukese (= Trukese)	Goodenough and Sugita (1990)
Maringe (= Cheke Holo = Hograno)	White, Kokhonigita and Pulomana (1988)
Bauan Fijian (= Standard Fijian)	Capell (1941)
Boumaa Fijian	Dixon (1988)
Dobu	Grant (1953)
Gapapaiwa	McGuckin and McGuckin (1992)
Gedaged	Mager (1952)
Gela (= Nggela)	Fox (1955)
Ghanongga	Tryon and Hackman (1983)
Hawaiian	Pukui and Elbert (1971)
Kairiru	Wivell (1981)
Katazi	Tryon and Hackman (1983)
Kilivila	Senft (1986)
Kiribatese (= Gilbertese)	Sabatier (1971)
Kosraean (= Kusaiean)	Lee (1976)
Kove	Chowning (1996)
Kwaio	Keesing (1975)
Lau	Fox (1974)

Lenakel	Lynch (1977)
Lewo	Early (1994)
Loniu	Hamel (1994)
Lou	Blust (1998)
Makura	Tryon and Hackman (1983)
Malagheti	Tryon and Hackman (1983)
Maori	Williams (1975)
Marovo	Hviding (1995)
Marshallese	Abo et al. (1976)
Minaveha	Nenegemo and Lovell (1995)
Mokilese	Harrison and Albert (1977)
Mota	Codrington (1896)
Motu	Lister-Turner and Clark (1954) (2nd ed)
Mussau	Blust (1984)
Muyuw	Lithgow and Lithgow (1974)
Nakanai	Chowning (1996)
Niuean	Sperlich (1997)
Nokuku	Tryon (1976)
Paamese	Crowley (1992)
Ponapean	Rehg and Sohl (1979)
Puluwatense	Elbert (1972)
Rennell and Bellona	Elbert (1975)
Roviana	Waterhouse, revised Jones (1949)
Sa'a and Ulawa	Ivens (1918)
Samoan	Milner (1966)
Sengseng	Chowning (1996)
Sesake	Tryon (1976)
Sudest	Anderson (1990)
Tikopia	Firth (1985)
Tolai = Kuanua = Raluana	Rickard (1888), Lanyon-Orgill (1962)
Tolo	Smith Crowley (1986)
Tolomako	Tryon (1976)
Tongan	Churchward (1959)
Ughele	Tryon and Hackman (1983)
Wayan Fijian	Pawley and Sayaba (f.c.)
Wedau	Jennings (1956)
Woleaian	Sohn and Tawerilmang (1976)
Yabem (= Jabêm)	Streicher (1982)

Unpublished sources consisted of manuscript word lists for NE Ambae by Catriona Hyslop, Babatana by Lucy Money, Kiriwina (= Kilivila) by Ralph Lawton, Mekeo by Alan Jones, Molima and Nakanai by Ann Chowning, Mutu by Alice Pomonio, Mwotlap by Alexandre François, Nduke by Ian Scales, Ninigo (= Seimat) by W. Smythe, Tamambo (= Malo) by Dorothy Jauncey, and Zabana (= Kia) by D. Ama and M. Fitzsimons. Computer files were provided by a number of scholars, some of which are themselves based on a variety of primary sources. These files include:

- The computer files of lexical data collated during the research leading to the publication of Ross (1988), whose sources are listed in Appendices A and B of that work.
- The computer files from the *Comparative Austronesian Dictionary* project which resulted in Tryon ed. (1995), which lists its own sources.
- The computer files of Robert Blust's *Austronesian Comparative Dictionary* on disk at the University of Hawaii. The version to which we refer dates from 1998.
- The computer files of Biggs and Clark's *POLLEX: Proto Polynesian lexicon* on disk at the University of Auckland. The version to which we refer dates from December 1993.
- Computer files of reconstructions with supporting cognate sets for North/Central Vanuatu (Clark 1994), Southern Vanuatu (Lynch 1996), and Micronesian (Bender et al. 1983).
- Computer files of dictionaries in progress provided by Joel Bradshaw (for Numbami), Deborah Hill (for Longgu) and Malcolm Ross (for Takia).
- Computer files of dictionaries in progress kindly made available by members of the Summer Institute of Linguistics. Languages and those who compiled/supplied the dictionary are as follows: Arop-Lokep (Jeffrey and Lucille D'Jernes), Bariai (Steve Gallagher), Bing (Doug Bennett), Bola (Brent Wiebe), Mapos Buang (Bruce Hooley), Iduna (Joyce Hockett), Dami (George Elliott), Dawawa (Martin and Beate Knauber), Gapapaiwa (Ed and Catherine McGuckin), Gumawana (Clif Olson), Hote (Marguerite Muzzey), East Kara (Perry and Virginia Schlie), Kaulong (Craig Throop), Drehet [= Khehek] (Stephan Beard), Lewo (Robert Early), Lou (Robert and Verna Stutzman), Lukep [= Pono] (Jeff and Sissie D'Jernes), Manam (Stephen and Kim Blewett), Mangseng (Lloyd Milligan), Mangap-Mbula (Robert and Salme Bugenhagen), Mengen (Fred Madden), Misima (Bill Callister), Mumeng [Patep] (Linda Vissering and Karen Wilson), Mussau (John Brownie), Nakanai (Ray Johnston), Nehan (John Glennon), Nochi (Leland and Laurinda Erickson), Patpatar (Ed Condra), Ramoaaina [= Duke of York] (Lisbeth Fritzell and Robyn Davies), Siar (Larry Erdman), Sinaugoro (Gerhard Tauberschmidt), Sio (Stephen and Dawn Clark), Sissano [Arop] (Stephen Whitacre), Sudest (Mike Anderson), Sursurunga (Don Hutchisson), Takia (Salme Bugenhagen, Judy Rehberg, Curtis Thomas), Tawala (Bryan Ezard), Teop (David Snyder), Tinputz (Roman Hostetler), Titan (Keith Lusk).

2 Collation

The collation stage of the project consisted in the first instance of creating a data base of vocabulary materials in a defined set of semantic domains from Oceanic languages for which data were already available (see §1). This data base was kept in text files on Macintosh computers. Files were organised in accordance with a modified version of the Summer Institute of Linguistics' 'standard format' in which fields within each record are labelled with an initial backslash followed by a single letter. In our version of the format, each record was terminated with a carriage return, i.e. each record occupied a single line. Each record contained a single word in a single language with associated information (a

code relating to the language's subgroup, a gloss and any other semantic information, the source, and any other notes the researcher chose to add). The Macintosh allowed non-standard characters to be created and viewed on screen. Records were organised on screen into putative cognate sets. The use of text files rather than files in a proprietary database format meant (i) that it was easy to view them on screen; (ii) that it was easy to manipulate them with a variety of text editors and word processors; (iii) that more complex repetitive processes could be performed by writing small programs in the Icon programming language (Griswold & Griswold 1990); and (iv) that it was relatively easy to import and reformat other people's data sets and to export collated material into publications in preparation.

Although there are accepted or standard orthographies for a number of the languages from which data are cited here, data were transcribed at the collation stage into a standard orthography (see Ross 1988:3–4) to enable us to recognise cognates and to spot regular changes more quickly. This orthography is retained in the citation of data in these volumes.