Monolingual mania: Current trends

in Pacific dictionary making

Robert Early
University of the South Pacific

A paper presented at The Second Asian Lexicography Conference,
Chiangmai, Thailand
24th – 26th May, 2004

1. Introduction

The South Pacific region is an area of widely dispersed human communities living in an environment that is sometimes idyllic, and sometimes threatening. The physical, social and cultural integrity and viability of these communities cannot be taken for granted. Languages in particular may be endangered on a wholesale basis. Of the 1200 or so languages identified in the region (non-Austronesian; Oceanic Austronesian [including Polynesian, Micronesian, and most of the languages of Melanesia], and Australian) the large majority have had very little attention from linguists. While there are some exceptions, most are still unwritten, and of those which have been studied, the materials produced consist mainly of basic descriptive items and missionary translations.

Within the Pacific, there is an enormous regional variation in language situations. Throughout the more recently settled scattered islands and atolls of Polynesia and Micronesia, there is a tendency for one language to be spoken by dispersed speech communities, whose spread often corresponds with modern political entities (e.g. Tonga). On the other hand, the larger islands of Melanesia are known for the high density of languages that they support relative to geographical area and to population (e.g. Vanuatu, with around 100 languages spoken by a population of 200,000; no language greater than about 8000 speakers).

Since the time of significant European contact, beginning about 200 years ago, the best described languages have been those of the Central Pacific, namely Fijian and the Polynesian languages. In most cases these vernacular languages have become established national languages, including being vehicles of a comparatively rich culture of literacy, whereby in addition to their role in the religious domain, they are used as mediums of instruction to higher levels in the education system, and in the production of newspapers and many other kinds of printed material.

The task of documenting these languages was willingly undertaken by many colonial administrators and missionaries, and many of the standard grammars and substantial bilingual dictionaries they produced are still respected as representing the best of the scholarship of the time. A very useful historical survey of lexicographic work in all of the 35 Polynesian languages is given by Sperlich (19??), and as he and others have noted, there are still major gaps remaining in bilingual lexicography for the Polynesian languages, and certainly no monolingual dictionary has appeared in print

for any of them, as well as for the languages of Micronesia and Melanesian. However, in recent decades a new phase of development has led to interest in preparing monolingual dictionaries for some of these languages, particularly Fijian (a Central Pacific language spoken in geographical Melanesia), Māori (New Zealand), Niuean, Samoan, Tongan, and Tuvaluan (all Polynesian languages), and most recently in Kiribati (Micronesian).

This paper describes these projects, which are of special interest because unlike most other linguistic work on these languages, they are characterised by a high level of ownership and participation by motivated native speakers. The paper also discusses the provenance and perceived purpose of these projects, and looks at issues related to the training and skills of compilers, the particular approach to dictionary writing that has been adopted in each case, the use of computers and data management, and project organisation and funding.

The map below from Lynch (1998:24; modified to include Niue and Māori) shows the countries of the Pacific region, and the general geographical boundaries of the commonly understood areas of Polynesia, Micronesia, and Melanesia. In terms of distribution of language groups, it has to be pointed out that there are numerous "outlier" Polynesian languages spoken on the eastern fringes of other countries within Melanesia, and that within Melanesian itself, there are both Oceanic Austronesian languages and non-Austronesian (Papuan) languages which reach as far as the southeastern parts of the Solomon Islands.
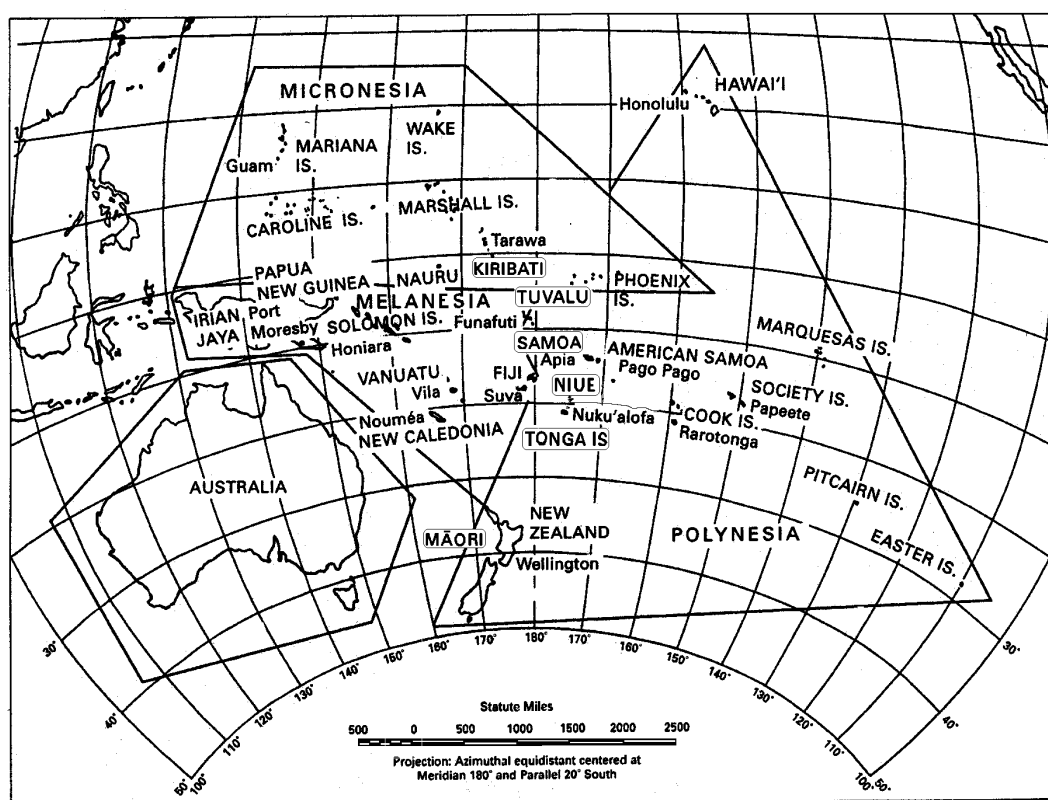
Modern times have also seen other developments in the language situation of the

Pacific with the spread throughout Melanesia of the English-based creole variously called Tok Pisin, Pijin, and Bislama, and the development of a Hindi koine among the Indian community in Fiji, now know as Fiji Hindi or Fiji Baat, as well as the extensive movement within the Pacific Islands and also to Australia, New Zealand, Hawaii and the continental United States of significant groups of speakers of Pacific languages.

Some further background is provided for the countries of each of the languages to be discussed in this paper. This information comes mainly from Crocombe 2001:685-709. Māori is not a name for New Zealand (known as Aotearoa in Māori), but for the indigenous language spoken there.

	geography	country size	population
Fiji	many large and small islands	18,272 sq km	810,000 52% Fijian
Kiribati	dispersed atolls	811 sq km 3.5 million sq km of sea	90,000
Māori	temperate New Zealand	comparatively huge	(est.) 350,000
Niue	a single raised atoll	259 sq km	1,600 +20,000 New Zealand
Samoa	two main large islands	2934 sq km	173,000 +66,000 Amer. Samoa
Tonga	smaller islands in three dispersed groups	699 sq km	99,000
Tuvalu	dispersed atolls	26 sq km	10,500

Behind the population figures in the above table hide some complex demographic situations. The Pacific region was settled by means of some of the greatest voyages of exploration and discovery in history, and there continues to be significant population mobility to this day. The colonial era has resulted in minorities of naturalised citizens of especially European and Chinese backgrounds in most countries, but also the early colonial plantation economy resulted in movements of indentured laborers from India to Fiji, of Melanesians to Australia, and of other Pacific Islanders from one country to another (for example, there are still sizeable communities of Kiribati speakers in Fiji and the Solomon Islands today).

These factors make it difficult to determine exactly the numbers of speakers of the various languages. The table below gives estimates of the number of resident L1 speakers for each language within the country of origin, and less certain guesstimates of the number of people who might speak the language who have emigrated away from the country of origin, and of their descendants who still identify with the country of origin of their parents or grandparents. This table shows that population outflow from some particular Pacific island nations is a huge demographic issue which has the potential to impact enormously on the viability of the Pacific island languages spoken by those communities. Crocombe gives figures which show that while there are two million people who would be called Polynesians, only 14% of them live in the politically independent Polynesian countries of Cook Islands, Niue, Samoa, Tonga and Tuvalu. More Polynesians live in the USA, New Zealand and Australia than within Polynesia itself (Crocombe 2001:66-67). Further, while census data from those

countries can provide figures for the absolute numbers, the extent to which their heritage languages are being maintained after three or four generations is another matter entirely, but is something that is beginning to be studied (e.g. Taumoefolau et al. 2002). Certainly, there are numerous active language maintenance activities taking place among Pacific island immigrant communities in Australia, New Zealand, Hawaii and the mainland USA: for example, for Niuean in New Zealand see Falesima and Fuemana-Foa'i 2000, and for Samoan see Fetui and Mālaki-Williams 1996. Many of these activities are modeled on the successful *kohanga reo* 'language nest' model implemented for New Zealand Maori, which may turn out to be the single-most important factor in arresting the decline of that language.

The following table also doesn't take into account non-indigenous residents who have learned the language as L2, but in each case the numbers of such would be negligible. For Fiji, the designation of L1 overlooks the fact that a very large number of Fijian Islanders are first language speakers of some other Fijian language, and the standard language is either a coordinate L1 or an L2.

	resident L1	non-resident L1	total
Fiji	400,000	10,000	410,000
Kiribati	85,000	5,000	90,000
Māori	35,000 (10% of pop.)	few	35,000
Niue	1,400	20,000	21,400
Samoa	170,000 + 60,000	200,000	430,000
Tonga	95,000	40,000	130,000
Tuvalu	10,000	2,000	12,000

2. How the projects began

2.1 Fiji

The first monolingual dictionary project in the Pacific region was the project in Fiji, for the standard dialect of Fijian, which has now developed into the Institute of Fijian Language and Culture. The beginnings of this project can be traced back to 1971. As reported by Geraghty (1996:6): "The American actor Raymond Burr had had long-standing interests in Fiji…and he sponsored a meeting of linguists and interested Fijians … to discuss ways of promoting the Fijian language." One of the founding fathers of modern Pacific linguistics, Prof. Bruce Biggs, is credited with mooting the idea of compiling a monolingual dictionary, and as Geraghty notes, this was a "radical departure, since all previous dictionaries of Pacific dictionaries of Pacific languages had been bilingual" (op. cit.). The first project director, Dr Albert Schütz of the University of Hawai'i lost no time in sending two Fijian schoolteachers to undergo training in Hawai'i, one of whom, Tevita Nawadra, became the first editor of the dictionary and later took over as director of the project. Dr Paul Geraghty, who had been a consultant to the project for several years previously was appointed as acting director in 1986, and along with his substantial academic research and publication on Fijian languages and dialectology, and other promotional activities like a popular weekly TV show on language matters, Geraghty has ushered the Fijian monolingual dictionary through to its final stages.

Geraghty has noted (pers. comm.) that once the dictionary was prepared to a

publishable draft stage in 1996 the then Permanent Secretary of Education decided to refer it to a "review committee". The may be partly in response to the fact that Geraghty, the primary compiler, is not an indigenous Fijian, and so there have been those who regard his involvement as controversial. The fact that progress appeared to stall once the manuscript got to the hands of this committee could be a consequence of the disruptions in national life caused by a series of political coups and upheavals which began in 1987 and which have continued to affect Fiji until the present. Also, Geraghty has subsequently moved to the University of the South Pacific (USP) and those who are now responsible for seeing the dictionary published have not been associated or involved with it from the start and may lack an appreciation of the importance and value that it could have to national goals relating to vernacular language development. Fortunately, Geraghty believes that when it does come time to publish the dictionary, there should not be any difficulty in finding the funds for it to be printed.

2.2 Tuvalu

The Tuvaluan Language Board was set up in 1972, and one of its first decisions was to designate the dialect of the southern islands of the group as an official standard (Siegel 1996). The board also made some decisions about orthography, and one of them was a poorly-informed ruling that the important phonemic contrast between short and long sounds (vowels and consonants) should not be marked (Besnier 1981:xi). After independence in 1979 an official vernacular education policy and system was introduced, with Tuvaluan being the sole language of instruction for the first four years of education. The Language Board also proposed to start writing a (monolingual?) dictionary, but with the board only meeting occasionally and no funds for more serious work, nothing happened on this until 1981, when the idea was reactivated. Since then there has been some funding from Australia in 1988, for helping in the development of educational terminology in Tuvaluan and for dictionary work. More recently, Canadian government funding has allowed for a photocopier and upgraded computer to be purchased.

Another source (Taomia 2000 and pc.) suggested that it was not until around 1992, when Mr Pasoni Tafaaki, a Senior Education Officer, was the chairman of the board, that it was agreed to prepare a monolingual dictionary. It seems that the Prime Minister at the time, Kamuta, and the Minister of Education (Kamuta's wife Maama) were also influential in supporting this proposal. In 1993, a team comprising Pasoni Tafaaki (as chairperson), Iosia Taomia, Vavae Katalake, Lagi Etoma, Siuila, and Mafalu was assembled, and government funding was provided for them to work full time on the monolingual dictionary project.

One factor which motivated the dictionary aspirations of the TLB was widespread negative reaction to the orthographic conventions adopted in a substantial bilingual dictionary which was published by the Peace Corps (Besnier 1981). This dictionary was prepared at a time when computers and printers were unable to render special characters like vowels with macrons, and so the long vowels in the language were written as geminate clusters. Tuvaluan also has long consonants, and these have usually been written with an apostrophe before the consonant (e.g. *fakatauga* 'shopping' in contrast with *faka'tauga* 'opinion'), but the Peace Corps volume also represented these as doubled consonants (*fakatauga* vs. *fakattauga*). The Tuvaluan

community generally was opposed to these orthographic convention, and basically refused to accept them or any publication which used them. In the event, the TLB never actually produced a bilingual dictionary, but subsequently, a privately published Tuvaluan-English dictionary was produced (Jackson 1991), which did consistently marked long segments, but using the orthographic conventions (macron for vowels and appostrophe for consonants) which had been used previously. This volume has been republished only recently with an added English-Tuvaluan section (Jackson 200?).

After a promising start in 1993 the work only proceeded for a couple of years before funding problems brought it to a halt for some time. In 1997, another Prime Minister, Bikenibeu, reinstituted the project, with most of the previous board members continuing on. When Pasoni Tafaaki died, Iosia Taomia became chairman.

Although work on the dictionary continued, there were a number of language planning issues which needed to be resolved. The two most important ones related firstly to orthography, and secondly, to dialectal variation. There are distinct dialects spoken on each of the eight islands, and a separate common variety, known as *'gana masani* 'common language' is still emerging, and remains uncodified. There was some degree of confusion regarding which dialect or variety should be represented in the monolingual dictionary. Generally, all common variety forms were being included, but any other form from any other dialect which is different from this can be included with a reference to its particular source island. I was asked in 2000 to run a national workshop which brought together a wide selection of people to discuss these issues, and a great deal of unanimity resulted. In particular, it was agreed that the long segments should be represented in the orthography, with macrons for long vowels and apostrophes for long consonants (Early 2000).

Following this workshop, two further training inputs were provided to teach lexicographical skills and good practise, and to introduce the SHOEBOX lexical database into the project (Early 2001, 2003).

2.3 Niue

There has been a history of production of extensive scholarly bilingual dictionaries for Polynesian languages and Niuean had figured with the 1970 dictionary by McEwen. However, this was one of the more modest Polynesian dictionaries, and in 1992 a project began with eleven Niuean speakers recruited to work with linguist editor Dr Wolfgang Sperlich and consultant Prof. Bruce Biggs. The involvement of professional linguists ensured that the final production is a high quality bilingual dictionary, published as Sperlich 1997.

Once this project was a completed, a smaller group was reconstituted as the Niue Language Commission, without a resident expatriate linguist. One of the first activities of this group, along with their work as official translators for the Niue Government, was to begin work on a monolingual dictionary.

This work has been almost entirely drafted, and like its bilingual predecessor, was typed up on computer in the format required by the Lexware software program developed by Dr Robert Shu at the University of Hawaii.

At a fairly late stage, the Niue Government requested assistance from the Pacific Languages Unit of the University of the South Pacific with the final formatting and preparation of the volume. I conducted a workshop in Niue in 2003 when a decision was made to introduce Shoebox to the project, and so all the existing files had to be modified into the record and field formatting required by Shoebox. The macro feature in Microsoft Word made this process relatively straightforward. At this point, it was also decided to restructure the way in which subentries are handled in the dictionary, and this has yet to be handled. Following the bilingual dictionary, all derived forms were listed as subentries within the entry for the headword, as it was considered that speakers of Niuean, for whom the dictionary was intended, would always know which headword to look under for any derived forms. However, as was pointed out by a reviewer of the bilingual dictionary (Hooper 1998:194), this resulted in some idiosyncratic compounds and derived forms resulting from less productive forms of prefixation especially being buried in subentries in contradiction to their alphabetisation, rendering them inaccessible to not only most L2 speakers of the language but potentially many L1 speakers as well.

It is therefore proposed to carry out a fairly major restructuring of the lexical database, by including main entry crossreferences in these cases. However, with partial and full reduplication being very productive processes in the language, and with some very common derivational prefixes having fairly predictable and semantically apparent consequences (e.g. *faka-* 'causative, manner'), there is a real dilemma here, which is probably the single-most critical factor that compilers of Polynesian dictionaries have to consider. If the above principle of cross-referencing is followed exhaustively, it has the potential to multiply the number of total entries for dictionaries of these languages by as much as a factor of three, four, or five times, i.e. from a range of say 5,000-7,500 entries to a range of maybe 15,000 – 40,000 entries. Many of these entries would, even from the point of view of L2 speakers, be potentially redundant, especially, as in the case of reduplications, they will occur in very close proximity to the headword that they are cross referenced to. On the other hand, the level of redundancy can only be reduced if the compilers are prepared to make a judgement call in each individual case as to whether or not they feel that the derived form warrants having its own main entry, perhaps because of alphabetisation considerations, reinforced by a subjective assessment of the extent to which meaning or function of the derivation is not completely predictable from its constituent morphemes.

Another change which is proposed is to include a new semantic domain field in the database, to facilitate the filtering and extraction of sets of related terms, such as all fish names. Some such sets of terms may occur as topical appendices in the final volume.

It had been hoped to have the dictionary printed during this year, ready for dedication at Niue's 30th independence celebrations, but a devastating cyclone recently destroyed much of the countries infrastructure, and non-essential activities like the monolingual dictionary project have been temporarily put on hold.

2.4 Tonga
The origins of the Tongan monolingual dictionary project can be traced back to the

mid 1980s when 'Ana Taufe'ulungaki, a Tongan secondary school teacher, was given a British Council's Regional Postgraduate Scholarship to study for a PhD at Birmingham University (Taufe'ulungaki 1988). This placed her in one of the leading centres for the study of corpus linguistics in Britain, where John Sinclair was laying the foundations of corpus-driven lexicography in the COBUILD project. 'Ana became exposed to the innovative techniques being used there, such as the early use of scanning technology for building a textual database that fed directly into lexicographic applications, and so it was just a small step to ask why the same could not be done for Tongan. There were early discussions with John Sinclair for some assistance to be given to the development of a Tongan monolingual dictionary project. This did not eventuate, but somewhere in late 1987 or early 1988 'Ana did communicate her vision to the then Minister of Education (Dr Langi Kavaliku) back in Tonga. Her proposal to work towards a Tongan monolingual dictionary was significantly motivated by the recognition of the impact that the growing prestige of English as the major official language in the Pacific region was having on the Tongan and other language communities. As she expresses it, the monolingual dictionary should be prepared

> not only as a means of promoting Tongan but I thought it was necessary to define Tongan words/concepts, etc. based on the Tongans' usage of language and their own values, thinking, knowledge systems and world views. I thought it would be great to build a Tongan language corpus (taking advantage of the new technology) from which we could develop a monlingual dictionary and a new grammar of the language, etc - all to be done by the Tongans themselves. (email, 10 May 2004)

In her communication to the education authorities back in Tonga 'Ana also made recommendation of another promising secondary school teacher, Melenaite Taumoefolau, who had just finished her MA in the UK. As a result, Melenaite was subsequently tapped on the shoulder and sent to Auckland University for advanced linguistics training with a view to getting the Tongan monolingual dictionary project off the ground. Melenaite was also atracted by the COBUILD project as an example of a brilliant and novel approach to the description of meaning in a language, and recommended that a similar approach be used for Tongan. Her studies resulted in a PhD thesis dealing with Tongan lexicography (Taumoefolau 1998), but her stay in New Zealand got extended for personal reasons and there was no further progress on the dictionary proposal for some years. It was not until 2001 that the then Director of Education in Tonga (Paula Bloomfield) and Melenaite got together and agreed to try to reactivate the project.

As a result, in February 2002 Melenaite travelled to Tonga and conducted a one-week workshop on dictionary-making, to commence the specialist training for a selected group made of representatives from the Cultural Affairs section attached to the Royal Palace, the Education department, religious organisations, and other recognised Tongan language experts, several of whom were retired secondary teachers of the Tongan language curriculum. This first workshop introduced the concepts, exposed participants to the Tongan metalanguage refined in Taumoepeau 1998, and covered planning and logistical issues. A second one-week workshop was held later the same year, where Melenaite spent half-days working with the group looking at topics like

polysemy, kinds of meaning, principles of defining, structure of entries, and critical evaluation of some prepared entries. I was invited, over the other half-days, to give an introduction to the use of Shoebox, and teaching the concept of record and field gave an excellent opportunity to further refine understanding of the kinds of information to be included in the dictionary, and the structure of entries (Early 2002). Some of the participants had no computer experience whatsoever, but there were a few who caught on quickly. As I understand it, the small working group is continuing on this project but on an informal, supernumery and voluntary basis.

2.5 Samoa

At present there is no specific Samoan monolingual dictionary project, but there is more printed lexical material available in Samoan, for Samoans, than for any other language in the Pacific. In particular, a number of very significant monolingual lexical resources have been produced by two sections of the government. The curriculum development unit of the Education department has produced, along with numerous course materials, a simple monolingual dictionary for schools. And secondly, the Culture unit has produced four excellent topical monolingual dictionaries, dealing with 'the house' (Matagaluega o Autalovou, Taaloga ma Aganuu, 1996; 136 entries), 'food and

O LE MAFUAALA O UPU O TAUMAFA

O LE
TUSI FAAMATALA UPU
O LE
GAGANA SAMOA

SAUNIA:
MATAGALUEGA O AUTALAVOU,
TAALOGA MA AGANUU

cooking' (Matagaluega o Autalovou, Taaloga ma Aganuu, 1998; 1506 entries), 'fish and fishing' (Matagaluega o Autalovou, Taaloga ma Aganuu, 2001; 1901 entries), and 'traditional sports' (Matagaluega o Autalovou, Taaloga ma Aganuu, 2002; 728 entries). A fifth item is in preparation, which deals with 'traditional arts' and currently has about 600 entries. There are other sections of government, e.g. in law, which have worked to prepare topical lists of translation equivalents of relevant technical vocabulary, which could feed into activity to prepare further mini monolingual dictionaries. These volumes have been prepared in Word format, and were manually alphabetised and formatted. Macrons for long vowels were added by hand to hard

copy after printout.
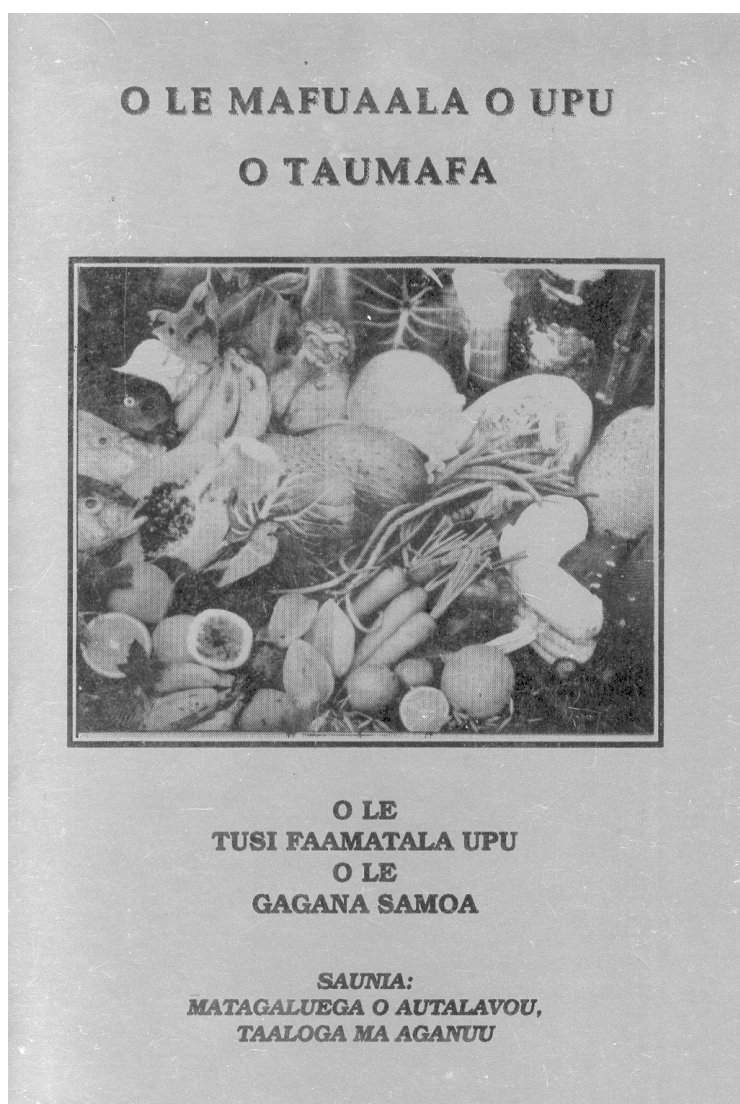
TUAI

tuai n. O le fausaga e fai ona vae e faamau ai se fasi ipu popo poo se uamea ua faatalatala e valu ai popo ia maua ai penu.
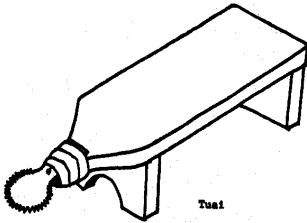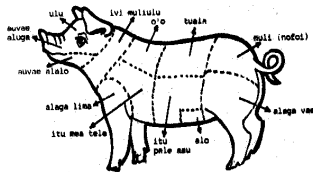
ata 22

tuailetuaniu v. 1. O le lava ma totoe o meaai ae o loo manana'o isi i ni meaai. 2. O le lava o mea tau meaai a se tagata e tua i ai.

tuafaga n. O le igoa o le fue sosolo e faaaoga e lalaga ai le fagai'a poo le enu.

tuala n. O le vaega taua e tipi mai le tua o le puaa ma aofia ai ma puimanava.

ata 23

TUITUI

tualu'ulu'u n. 1. O le ituaiga o pa'a. 2. O le pa fagota ua atiatia.

tuamafa n. O le lupe ua matua ma peti.

tue n. o le ituaiga o ufi vao.

tu'e n. O vaega o le pa'a poo le ula o loo pipii ai vae e ese mai le tua. E lua tu'e o le pa'a poo le ula.

tu'etu'e v. O le sala pe momotu ese o lau, poo fa, e faapei o le talo ua leai ni lau ua ai e manu poo ua pala foi ona fa.

tui n. O lemea e faaaoga pe a fai meaai poo le taimi e aai ai e i ai ona mata ma'ama'ai ma lona 'au e faaaoga e sisi ae ai i luga meaai aemaise meaai vevela.
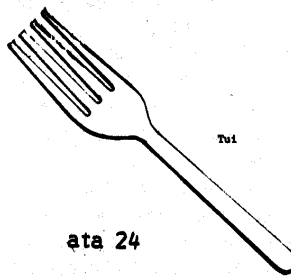
ata 24

tu'ilealelo v. o le upu faalumaluma mo le ai o meaai.

tuitui n. O le figota o le sami e i ai ona tala maamaai mata e faataamilo ai i le tino ae 'aina

There is now a plan to produce a single comprehensive monolingual dictionary based on these previous publications, and I was asked to conduct a workshop in 2002 (Vavao and Early, 2002) to further refine lexicographic skills and expose participants to the possibilities of Shoebox.

It is going to take a lot of work and skill to retrofit the exisiting pub-lications into standard format for Shoebox, and then to merge the databases. There are many overlapping entries between the topical dictionaries and the school dictionary, and compilers who produced these will have to cooperate and collaborate in formulating new single definitions for each entry.

2.6 Māori

I have only just discovered that there is a monolingual dictionary project underway for New Zealand Māori. It has only been in operation for a short period of time, but I

have no other details.

2.7 Kiribati

The Kiribati monolingual dictionary project is still in its earliest stages. There was a Kiribati Language Board which was involved in some attempts to standardise the orthography of the language in the 1980's, but there was a lot of conflict over some of their recommendations to make changes to the established missionary orthography, and the language board has now been defunct for quite a few years. However, on a visit to Kiribati in 2002 I had a meeting with some people from the USP and the Ministry of Education, and there was a great deal of interest in hearing about what is happening in the other Pacific countries.

Attempts to organise a national level planning workshop to initiate a monolingual dictionary project foundered recently when the workshop funding proposal sat unactioned on someone's desk beyond the closing date for a cycle of UNESCO funding, but it is hoped other sources can be found.

3. Purpose

The discussion above has hinted in some cases at the apparent motivations that have led to the current interest in monolingual dictionary production in the Pacific. There is one common theme, which is that the monolingual dictionary is "by the people, for the people", but it is worth specifying some distinct factors that obtain in different countries as well. These factors are largely as I have discerned them in discussions with those involved, as there is little documentary record or evidence of the background thinking that led to the conceptualising and initiating of these projects.

	purpose and aims
Fiji	-repository of the national language and culture -standardisation of the standard Bauan variety -policy commitments to expanding vernacular education
Kiribati	-sense that many unecessary borrowings creeping in to the language -not fully identified, and maybe a sense of "jumping on the bandwagon"
Māori	-no direct information, but undoubtedly a key (but late) component in the struggle by NZ Māori to preserve/maintain/develop their language
Niue	-a key resource for vernacular education programs -documentation of the language in the face of perceived attrition
Samoa	-a key resource for vernacular education programs
Tonga	-intellectualisation of the language -affirming the status of the language
Tuvalu	-an encyclopaedic dictionary to help Tuvaluans understand their own language -making the rest of the world more accessible to Tuvaluans

3. Current status

The following table summarises the current status of the monolingual dictionary projects being discussed:

	entries	further notes

Fiji	25,000+	ready for printing since 1996
Kiribati	0	work yet to start
Māori	?	no information
Niue	5000?	nearly ready for printing, but format revision being considered
Samoa	5000+	several topical dictionaries produced, need to merge into a single database and complete
Tonga	few	project underway?
Tuvalu	40,000	huge amount of work done, but significant reformatting required

4. Methodology and approach

There are some common features that apply across most of the projects:

- there was very little corpus-based research (there are no developed electronic corpora)
- special orthographical characters, especially macrons on long vowels, were a font problem. In Samoa they had adopted the (disputed) Māori solution which was to use the vowels with diaerisis. In Tuvalu, they were either left out, or put in by hand after printing. In Tonga this problem had led to the suggestion that doubled vowels be used.
- alphabetisation is frequently vowels first, then consonants, so English sort order doesn't work
- some kind of semantic domain approach was recognised as useful
- previously prepared bilingual dictionaries were not available in electronic form (except for Niue)

The font problems have been overcome using fonts which I adapted using Fontmonger (now superceded by Unicode), and sort order issues are handled well in Shoebox. MDF does not have an explicit schema for preparing monolingual dictionaries, and some tweaking of the CC (Consistent Change) tables has been required. I will now note various other details for individual projects.

The **Fijian** dictionary began before the personal computer era. Computers were used in the project in more recent years, but the dictionary files are all in word-processor format, and there has been no use of a lexical database package. With professional linguists working in the project from the start, there has been a very rigorous concern for technical quality at all stages, and the project has probably gone too far now for converting the document files to Shoebox format to be considered. However, this would certainly be an extremely useful subsequent step for rendering the extensive information into a format more accessible to sorting and analysis.

The **Kiribati** project is still yet to take any kind of discernible shape. However, this raises the question: in view of what has been seen and learnt in the other monolingual dictionary projects in the Pacific, some nearing completion, are there any recommendations which could be made to assist this or any other new project? Some thoughts along these lines will be presented in section 7.

The **Niuean** monolingual dictionary builds directly on the basis laid by the bilingual dictionary (Sperlich 1997). There was a recognition of the danger of just translating

the bilingual dictionary, but it has nevertheless been a key resource. However, compilers have readily added to this and expanded entries as required, including incorporating more lexical phrases and items like proverbs. This project has had the advantage that the team working on the monolingual dictionary is made up of some of those who went through the whole process of producing the bilingual dictionary in the company of the consultant linguist.

There has already been a lot of monolingual dictionary activity for **Samoan**, under the auspices of a number of different departments. This resulted in some lack of coordination, but a recent restructuring of government ministries has brought the Department of Culture under the Ministry of Education and so there is good prospect for future collaboration. One feature of the data collection approach taken by the Department of Culture was to run one-week community workshops where terminologies were collected and defined.

The way it was described to me for the **Tuvalu** project was that when the team began work, they decided that their first task was to identify all the words that they find for the language. So for the first couple of years (when they were only working part-time), this is essentially what they did: write down all the words they could think of in Tuvaluan, in an ordered way, starting from 'A'. The chairman of the language board at the time is credited with coming up with the idea at some point of using a generative template to produce all possible forms that were phonotactically well-formed, at least for the first three letters of the words, and these lists became a basis for further word identification. They then divided the lists of words up alphabetically, and various compilers were given the task of producing definitions for their assigned lists of words. At some point, individual areas of knowledge were also recognised, so that particular compilers also became responsible for checking the entries for various semantic domains.

When I first visited the project in 2000, they were getting to the end of the process of writing definitions for most of the words. There were many thick writing blocks of handwritten work stacked up waiting for the typist to enter them onto the computer in Word format, with random sequences of tabs and spaces providing a hard-coded single-column format. It wasn't until 2001 when the definition writing phase was, in the minds of the compilers, even closer to completion, that they were exposed to some guidelines for writing definitions, and when definitions that had been written by one compiler were presented to the whole group and discussed openly. This exercise showed up that even though the team had been coming along to work together everyday for several years, and had developed a strong team spirit and good working relationships, in fact each compiler had been working quite independently, with little cooperative interaction with the others, so that the workshop was forcing them to do something they had not done previously, which was to critically evaluate each others work. There was a lot of reticence to doing this, but the observation that won the day was that if they didn't learn to criticise each others work, then certainly the public at large would do it for them later on, when the dictionary was published. When presenting their sample work, some of the compilers were apologetic, lacking in confidence, and not willing to strongly defend what they had done. Others presented their sample definitions as if they would defend them to the end. In all cases, there was a lot of unease about having to make comments about another person's work, and

perhaps this was well-founded, because there were cases where some useful criticism was in fact perceived as hurtful by the person whose work was being discussed. This situation highlights the need to be aware of the cultural dynamics of interaction in a group work situation. My view on this is a little hard-nosed: it is true that at all times we all (whether insiders or outsiders) need to understand and appreciate how notions of cultural appropriateness etc. will play out in practise and impose certain constraints on any context. However, in so far as working together to produce a monolingual dictionary of a language is not a context for which there are established patterns of behaviour which can act as precedents, it is necessary to develop a shared commitment to the quality of the end-product which will be strong enough to moderate the niceties of social conventions that might otherwise apply. The team of compilers should ideally show some breadth of representation, and so within their interactions, the younger, or female members, for example, should not feel unable to comment on the work of older, or male members.

Certainly, the Tuvaluan situation shows up some of the difficulties that can arise when native speakers who are recognised for their expertise in their own language are assigned to monolingual dictionary compiling duties, without formal or in-service training, and left to their own devices over a long period of time. It is getting a little too late in the process, when the team considers that its work is almost done, to find out that circular definitions are not acceptable! For example, in Niuean, there is a very productive nominaliser which can operate on most verbs, so this has been employed as a key strategy for giving the meanings of verbs: e.g. *galo* 'to forget' was given the definition *te galoga o se mea* 'the forgetting of something'. This might be well and good, except that the reverse strategy was employed for giving the meaning of nominalisations, whereby *galoga* 'forgetting' was defined as *e galo se mea* 'someone forgets something'. What is one to recommend when this strategy occurs throughout the dictionary, and there is pressure to see the work completed as soon as possible, partly so that the Tuvaluan dictionary just might pip the Fijian dictionary at the post by being the first monolingual dictionary to appear in print for any Pacific language.

Work on the **Tongan** dictionary is still in its infancy, but there are some interesting departures being proposed in how it should be structured. In particular, following the lead of Cobuild, Melenaite Taumoefolau has recommended that the example sentence should come first, straight after the headword (if there is no pronunciation note, which only occurs for some long words). It is intended that definitions should be in the form of explanations rather than following a formula that doesn't work for many lexical items, and that many of them will necessarily be quite encyclopedic, in order to capture relevant cultural information. Following the definition will be a single set of brackets containing any required usage and other information.

Taumoefolau also proposed some orthographical changes for Tongan in her thesis, firstly, that long vowels should be represented by double vowels instead of the macron, and that the definitive accent, which was previously indicated with the acute accent over the final vowel carrying the accent, then make use of the macron. However, I think by the time of the workshop she had changed her mind about these changes, and was no longer suggesting them. However, the question of the extent to which the monolingual dictionary should be used to introduce or impose new orthographical conventions to the speech community is an important one.

5. Kinds of information included

The various monolingual dictionary projects show a great range of approaches regarding what is felt to be appropriate, firstly, with regard to the items which should be catalogued as headword entries, and secondly, with regard to the fields of information that are included for each entry. This variety across the different projects reflects different views regarding the purpose of the monolingual dictionaries, but also suggests that there have been rather ad hoc approaches to this issue in the different projects. This perspective is reinforced when some projects, at a fairly late stage of completion, make some fairly far-reaching decisions about these matters.

Concerning coverage, perhaps the most encyclopaedic approach has been adopted by the compilers of the Tuvaluan dictionary. They have made frequent reference to Pears Cyclopaedia as a source for including entries in the monolingual dictionary for the Tuvaluan forms of names of all countries of the world and their capital cities, important world historical figures and events, as well as places and names found in the Tuvaluan Bible, and so on. Others are more frugal, and the question of the extent to which borrowings should be included is difficult. There are competing motivations: on the one hand, one of the functions of the monolingual dictionary in most cases is to preserve a purer form of the language in the face of the inroads being made on the language by English, and to affirm the status of borrowings by incorporating them in the dictionary would seem counter-productive. On the other hand, if the monolingual dictionary is intended to be an educational resource in the process of standardising and modernising the language, then the place of both established and even newer borrowings in the lexicon of the language should be recognised in the monolingual dictionary. None of the monolingual dictionary projects have set out to undertake intentional language engineering or language development activities towards the deliberate creation of new terminologies for areas of introduced culture (e.g. law) and technology (e.g communications and information technology).

All of the projects have struggled with the issue of how to handle the wide range of slang and vulgar terms which occur in each language. Again, the Tuvaluan commitment to exhaustiveness has seemed to override other sensitivities, and with appropriate usage notes, they seem comfortable about including these terms. In other cases, e.g. Samoa, there is unanimity amongst those involved that these terms should be included, but also a recognition that there also a competing unanimity in the community at large which will totally reject a dictionary containing such words as being in complete contravention of public morality constraints, and the compilers seemed resigned to having to accept this constraint. There is uneasiness in some cases about including terms for private body parts, especially those which can also appear as expletives, and other scatalogical vocabulary.

Monomorphemic words are the most prototyical lexical items, and their status as main entries is unquestioned. However, the lexicality continuum extends either side of forms like this, and with regard to smaller items, none of the projects independently considered giving separate entry status to bound grammatical morphemes. With regard to larger items, compounded forms are usually recognised as words in their own right, but it seems that in each case some kind of external input or training has been required before native speakers understand the concept of lexical phrases, and realise the value of including them in the dictionary. Further, more extended lexical

items like proverbs and other fixed expressions are not usually identified or considered, except where there are bilingual dictionaries for the language which model the incorporation of these.

As noted, there is also a range of approaches adopted towards the kinds of information that the different projects include in entries. Some include etymylogical information including reconstructed Proto-Polynesian proto forms. Others did not really consider including example sentences until the value of this was pointed out at a late stage in workshops. Some have considered that the use of example sentences makes, in many cases at least, formal definitions superfluous. There are few cases where even basic lexical relationships like antonymy have been handled in a principled or extensive way.

Most projects follow the lead of the lexicographic tradition to which they have been exposed, and include word class or part of speech identifications. This raises the whole issue of developing a vernacular metalanguage for doing this, and unfortunately, in a few cases, some metalanguages have been adopted which are simply translations of terms used in English grammar. A whole grammar of Tuvaluan has been written in Tuvaluan by a native speaker which has established terms for a whole range of structural categories (e.g. count vs. mass nouns) which do not occur in Tuvaluan, and which, on the other hand, fails to provide terminology for categories which occur in Tuvaluan (pluralisation by doubling of initial consonant of noun) but which are not found in English.

Once a Shoebox format has been adopted, the possibility of including helpful non-print fields like name of compiler, draft-stage, notes to check, date-stamp and so on has been introduced, but compilers have not usually seen any need to make much use of these. The value of the semantic domain field has been more recognised, but has yet to be rigourously adopted in any of the projects.

6. Strengths and weaknesses

Having had the privelege of observing and being involved in a number of the Pacific monolingual dictionary projects, some particular strengths and weaknesses can be noted, recognising that this is necessarily a somewhat subjective and individual perspective. I could relate these specifically to particular projects, but it will still be of as much interest and value if they are mentioned in more general terms.

Some of the strengths of some of the projects are:
- strong base of long-term funding to initiate and maintain the project for many years, especially in relationship to staff salaries
- sourcing ongoing funding from aid donors for equipment
- inclusion of linguistic expertise from the start and on an ongoing basis
- commitment to selecting qualified native speakers and supporting them to obtain the highest level of training and formal qualifications required
- incorporated as an official government activity or with strong government support through relevant ministries/departments
- involvement of key local language experts and recognised authorities
- high level of motivation
- wide involvement of stakeholders in initial planning and training sessions

- keen to adopt computer database approach
- good experience and public awareness gained through series of topical dictionaries
- monolingual project builds right on the back of other succesful activities, like a bilingual dictionary project, and utilises the skill of experienced staff
- project fits into formal activity plans relating to vernacular language development

Some weaknesses that could be noted are:
- a long gestation process without apparent result can make the public and government somewhat cynical about outcomes
- use of tried-and-true manual methods and document processing rather than embracing computational solutions
- introduction of required level of linguistic and computer database expertise into the project came at too late a stage
- government funding has been inconsistent, or no firm base of financial support established
- a lot of talk over many years, but is it really happening yet?

7. Recommendations

1. It is critical to include ongoing input from a professional linguist/lexicographer from the earliest stages of the project. The availability of such qualified people varies. Tonga, reputed to have the highest ratio of PhDs to population of any country in the world, has a number of people who have completed PhDs in linguistics. Samoa also has some people with postgraduate qualifications. However, this is not the case in Kiribati, and while various political sensitivities and local aspirations are not favourable to the involvement of an outsider in such a culturally significant activity, a monolingual dictionary project should not even be commenced without the input of the expertise of a linguist. The more familiarity the linguist has with the language or other languages of the region the better, but the outside linguist must also understand local sensibilities which require that rather than doing all the work themselves, their input will be most suitably expressed in a skills transfer mode, with an emphasis on training local staff. In some cases, indigenous language rights have become a sensitive political issue, and monolingual dictionaries, perhaps moreso that any other item of literature are regarded as being exclusively geared for the speakers of these languages, and this perception can be extended to the view that only speakers of the language should be involved in their preparation, so that the input of outsiders, not matter how much their well-meaning professional expertise could assist, would simply not be welcomed.

2. Computer technology and the use of a lexical database like Shoebox should be embraced from the start. This means that the secretarial and other staff associated with the project need to have adequate computer skills to learn how the database operates and how to carry out vital functions like backup. In the Tuvalu project, the one and only computer crashed, and while the data files were in some stage of backup the staff were not able to reinstall the database program files on the repaired computer. Trying to get the computer set up properly again from several thousand kilometres away by

phone and fax proved to be a costly, and ultimately unsuccessful enterprise. It was not until I was able to be in Tuvalu again, many months later, that things could get on track again. In the case of the Niue project, at a very late stage of editing and checking, the project staff closed all the data files in Shoebox and then when they reopened Shoebox and none of the data files opened automatically, they gave up on Shoebox and so just started editing in the formatted Word files that were exported from Shoebox. This meant that the Word files had more up-to-date information than the database files, and it is very difficult to repair this situation if the database files are no longer the primary current data source. In another project, the person looking after the computer files has, on more than one occasion, lost significant amounts of data and had to reenter a lot of information as a result of being confused about the roles of the various files used by the lexical database, and not understand the relationship between the Shoebox program files (stored in C:\Windows\Programs\Shoebox), the database settings files (stored in C:\My Shoebox settings), the actual data files (some in *.db database format, some in *.rtf Rich Text Format, and some as *.doc formatted Word documents in C:\My Documents\Dictionary), and the desktop shortcut icon pointing to the *.prj Project file.

3. The government or some other national agency should adopt the project as one of special significance for the whole country or language community. This recognition should translate into efforts being made to develop a broad base of awareness and support for the project, as well as specific budgetary support, realising that it is a long term activity. However, this can be done quite economically in some cases by redeploying staff from other activities and by housing the dictionary project in existing facilities. The project should not be the part-time hobby of a few dedicated enthusiasts, but should be embraced as an activity of national significance.

4. It should be a team effort. Special training and orientation may be required to help members build an acceptable method of working together which reflects socially acceptable patterns of interaction, but also allows for shared learning and skills development, mutual accountability, and commitment to quality.
It also means that the participation of all those involved should be recognised, but this can be difficult. For example, in the Niue bilingual dictionary (Sperlich, 1977) the overseas linguist contracted to work with the project is designated as the editor of the volume, and his name alone appears in citations of the dictionary, whereas so-called "associate editors" and other members of the dictionary writing panel who worked on the project both before, during and after the tenure of the expatriate linguist, are not recognised in the same way. If an outside linguist is involved in providing substantial training, editing, facilitation etc. throughout the project, in association with the L1 compilers, to what extent should that be recognised?

5. Local aspirations for control and participation at all stages need to be recognised. This theme is clearly evident in the narrative above of how the Tongan monolingual dictionary project has developed. There could be many overseas linguists with a knowledge of Tongan who might have welcomed the opportunity to be involved in this project from the late 1980s on, but there has been a very deliberate approach by the Tongan authorities that instead of inviting overseas experts to do the job, they would rather do what is necessary to develop their home-grown experts. This might have meant considerable delay in getting the project underway, but it responds to the

strong sense of responsibility and ownership that speakers of Tonga feel for their language. This is to be applauded, but has to be balanced by an understanding that an expert knowledge of the language is not the only kind of expertise that is required. The Tongan project showed this awareness by inviting me to conduct some early training in Shoebox, and they made sure that some of the participants at this training were computer literate, and some had an IT background, so that after an initial exposure to the features of the lexical database they would be able to pick it up and work on their own without further external input.

5. The use of the topical dictionary approach has been a key factor in maintaining and building interest in the Samoa project, and is being considered for emulation elsewhere, e.g. Tonga. Some key benefits of this approach is that it provides for some amount of trial and error; it allows for the significant satisfaction of seeing something produced and in print along the way; and it gives the project credibility.

6. Efforts to identify publishers need to begin at an early stage, or better, funding for final publication needs to be included as part of the overall cost of the project. For example, while the Fiji monolingual dictionary has been ready to print for a number of years, no publisher has been found. The main publishing houses for Pacific language materials (University of Hawaii Press, Pacific Linguistics, and the University of the South Pacific) do not consider the printing and publication of the Niuean monolingual dictionary to be a commercially viable project, and would require substantial grants from the government or some donor agency to underwrite this. Clearly a monolingual dictionary is going to have a much smaller market than a bilingual one, which have a much wider international audience. There are other concerns as well. One publishing house expressed concern at taking on the publishing of a monolingual dictionary because it would lack the in-house resources (i.e. a trained copy-editor who knows the language) to adequately edit the volume and be confident that quality is ensured. There can also be copyright issues. For example, the Niue Language Commission, along with the Government of Niue, adopt a strong view on intellectual copyright issues. The language "belongs" to them, and to release the copyright of the monolingual dictionary of their language to some international commercial publishing house or even some other facilitating agency is tantamount to selling their birthright. The model adopted for the Niue bilingual dictionary is acceptable, whereby the volume was "Printed in the United States of America" by "University of Hawai'i Press", which also distributes it, and was "Funded by Niue Government and United Nations Development Programme (UNDP), with additional funding from UNESCO", but is also "© 1997, Government of Niue". There is no clear statement as to who the actual publisher is, although the main title page in the volume, following the name of the editor, simply states "Government of Niue, in association with Department of Linguistics, University of Hawaii, Honolulu". Somehow each of these five entities must have known what part they were playing in the process. It is also understood that in order for the UHPress to undertake the project, there had to be confirmed forward orders of 1000 volumes to be paid for the Government of Niue.

7. Metalanguages should be developed at an early stage for word class labels and field names in particular, but these must reflect the structure and requirements of the language.

Bibliography

Besnier, Nico. 1981. *Tuvaluan Lexicon*. United States Peace Corps.

Crocombe, Ron. 2001. *The South Pacific*. Suva: University of the South Pacific.

Crowley, Terry. 1999. 'The socially responsible lexicographer in Oceania'. *Journal of Multilingual and Multicultural Development* 20(1): 1–12.

Early, Robert. 2000. 'Standardisation and orthography of the Tuvalu language'. USPSolutions and AusAID. Course materials.

Early, Robert. 2001. 'Dictionary-making for the Tuvalu language'. USPSolutions and AusAID. Course materials.

Early, Robert. 2002. 'Monolingual Tonga dictionary project: Introduction to computer database'. USPSolutions and NZAID. Course materials.

Early, Robert. 2003a. 'Advanced dictionary-making for Tuvaluan'. USPSolutions and AusAID. Course materials.

Early, Robert. 2003b. 'Data processing and formatting for the Niuean dictionary'. UNESCO and USP. Workshop notes.

Falesima, Taanefoou Ogotau, and Lisa Fuemana-Foa'i. 2000. 'Niue language: Vagahau Niue'. In Fischer and Sperlich, pp. 84-86.

Fetui, Vāvāo, and Afamasaga Mālia Mālaki-Williams. 1996. In France Mugler and John Lynch, *Pacific Languages in Education*. Suva: Institute of Pacific Studies.

Fetui, Vāvāo, and Robert Early. 2002. 'Indigenous language development skills training workshop'. USPSolutions and AusAID. Course materials.

Fischer, Stephen Roger, and Wolfgang B. Sperlich, *Leo Pasifika: Proceedings of the Fourth International Conference on Oceanic Linguistics*. Auckland: Institute of Polynesian Languages and Literatures.

Geraghty, Paul. 1996. "The Fijian dictionary experience". Unpublished typescript, 17pp.

Head, Bertha. 'Keeping the Niue Leo (Language) Alive'. In Fischer and Sperlich, pp. 142-155.

Hooper, Robin. 1998. Review of Sperlich 1997. In *Oceanic Linguistics* 37:193-200.

Jackson, Geoffrey W. 1991. *Te Tikisonale o te 'gana Tuvalu: A Tuvaluan-English dictionary*. Apia, Samoa: Geoffrey W. Jackson.

Jackson, Geoffrey W. 200?

Lynch, John. 1998. *Pacific languages: an introduction*. Honolulu: University of Hawaii Press.

Matagaluega o Autalovou, Taaloga ma Aganuu (Ministry of Youth, Sports and Culture). 1996. *O le mafuaala o upu o le fale: O le tusi faamatala upu o le gagana Samoa* (A topical dictionary in Samoan of words relating to the house).

Matagaluega o Autalovou, Taaloga ma Aganuu (Ministry of Youth, Sports and Culture). 1998. *O le mafuaala o upu o taumafa: O le tusi faamatala upu o le gagana Samoa* (A topical dictionary in Samoan of words relating to food and cooking).

Matagaluega o Autalovou, Taaloga ma Aganuu (Ministry of Youth, Sports and Culture). 2001. *O le mafuaala o upu o le faigafaiva: O le tusi faamatala upu o le gagana Samoa* (A topical dictionary in Samoan of words relating to fish and fishing).

Matagaluega o Autalovou, Taaloga ma Aganuu (Ministry of Youth, Sports and Culture). 2002. *O le mafuaala o upu o taaloga faasamoa: O le tusi faamatala upu o le gagana Samoa* (A topical dictionary in Samoan of words relating to traditional sports).

Matagaluega o Autalovou, Taaloga ma Aganuu (Ministry of Youth, Sports and Culture). in prep. *O le mafuaala o upu o faiva alofilima: O le tusi faamatala upu o le gagana Samoa* (A topical dictionary in Samoan of words relating to traditional arts).

Siegel, Jeff. 1996. 'Vernacular education in the South Pacific: A report to AusAID'.

Sperlich, Wolfgang B. 19??. 'A review of advances in Polynesian lexicography'. *Rongorongo*.

Sperlich, Wolfgang B. 1997. (ed.) *Tohi Vagahau Niue: Niue Language Dictionary*. Government of Niue and University of Hawai'i.

Sperlich, Wolfgang B. 2000. "Gahua he Tohi Vagahau Niue: Niue dictionary project: orthography and vowel quality." In Bill Palmer and Paul Geraghty, *SICOL: Proceedings of the Second International Conference on Oceanic Linguistics: Vol. 2. Historical and descriptive studies*. Pacific Linguistics 505.

Taomia, Iosia. 2000. 'Tala toetoe i te galuega a te Komiti o te 'Gana Tuvalu' (Description of the work of the Tuvalu Language Board). Typescript, 5 pp.

Taufe'ulungaki, 'A. M. 1988. 'Implications of Language Attitudes to Language Planning in a Bilingual Situation. University of Birmingham. Ph.D. thesis.

Taumoefolau, Melenaite. 1998. 'Problems in Tongan Lexicography'. University of Auckland. PhD thesis.

Taumoefolau, Melenaite, Donna Starks, Karen Davis, and Allan Bell. 2002 'Linguists and Language Maintenance: Pasifika Languages in Manukau, New Zealand'. In *Oceanic Linguistics* 41:1:15–27.