NICK THIEBERGER
SCHOOL OF LANGUAGES AND LINGUISTICS
THE UNIVERSITY OF MELBOURNE
ARC CENTRE OF EXCELLENCE FOR THE DYNAMICS OF LANGUAGE

THE SECOND LIFE OF DICTIONARIES

AUSTRALEX
ANU 2019

I acknowledge and pay respect to the original owners of the land I work on, the Gadigal people of the Eora Nation, the Ngunawal, and the Woiwurrung

WHY A SECOND LIFE?

- Specific needs for Indigenous languages – 'small' languages following Dorian (2014)
- Updated dictionary is needed for revitalisation work
- Updating existing dictionaries, 3 case studies
- *Urgent need to make dictionaries available to the speakers of the languages*

592 Toqabaqita – English dictionary

Frantisek Lichtenberk

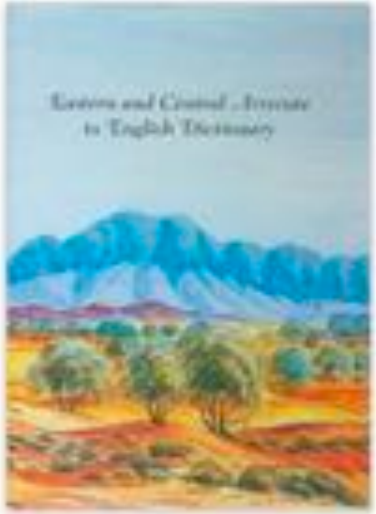
Toqabaqita is an Austronesian, more specifically an Oceanic, language spoken on the island of Malaita in the Solomon Islands. This is the first published dictionary of the language, based on the author's work on the language for over two decades, starting in 1981. The volume contains a Toqabaqita–English dictionary (nearly 7,000 entries) and an English–Toqabaqita finderlist.

Frantisek Lichtenberk is Associate Professor of Linguistics in the Department of Applied Language Studies and Linguistics at the University of Auckland. He has done field work in Papua New Guinea and the Solomon Islands. He is also the author of *A grammar of Manam* and *A grammar of Toqabaqita*.

2008 ISBN 9780858835849 407 pp

Prices: Australia AUD $75.90 (incl. GST) International AUD $69.00

Books › Reference › Dictionaries & Thesauruses



See this image

Eastern and Central Arrernte to English Dictionary Hardcover – January 1, 1994

by John Henderson (Author), Veronica Dobson (Author)

★★★★★ 2 customer reviews

See all formats and editions

Hardcover
—

Eastern and Central Arrernte are strong languages, spoken every day by about 2,000 people in Mparntwe (Alice Springs), Utyerrtye Apurte (Santa Teresa) and other communities in Central Australia. For Arrernte people, their language is important in a way that goes beyond straightforward communication; it links them with their family, with their country and with their dreamings.

Share

Out of Print–Limited Availability.

Deliver to Australia

Find this book on AbeBooks.com

AbeBooks, an Amazon company, offers millions of new, used, and out-of-print books.

Go to AbeBooks.com

Add to List

Have one to sell? Sell on Amazon

SMALL LANGUAGES

- There is no tradition of literacy
- "less-resourced"
- Wordlists may have been made by earliest observers
  - Require conversion and special treatment to decipher content
- There is only one *dictionary*, typically the result of a linguist's fieldwork
  - Ideally, this means there is a structured data file, or at least, a MS Word document that can be archived

PROBLEMS

- Locating dictionaries that were produced in the past
  - Few copies ever produced (duplicated copies, handwritten copies)
    - microfilmed
  - Published versions no longer in print, copies difficult to obtain
  - Obtaining textual versions
    - Typescripts of sufficient quality to OCR
    - Digital files of page layout
    - Digital files of underlying data

MANY HAPPY RETURNS

- Australex 2011 I talked about methods for reusing dictionary files (Thieberger 2011). I concluded then that there was a need:
  - For training in creation of lexical databases from which dictionaries can be derived
  - To preserve digital versions of underlying lexical databases (or MS Word if there is nothing else)
  - For conversion from legacy formats to structured lexical databases
  - For a service that will assist in the above

About us | Aboriginal Studies Press | Get involved | Catalogue search | Cultural sensitivity

AIATSIS
AUSTRALIAN INSTITUTE OF ABORIGINAL AND TORRES STRAIT ISLANDER STUDIES

Explore | Collection | Research | Publications | News and events

News

- Events
- Conferences
- Newsletters
- Blog
- Subscribe

Home › News and events › News

Tweet

The Sydney Language now available to a new generation

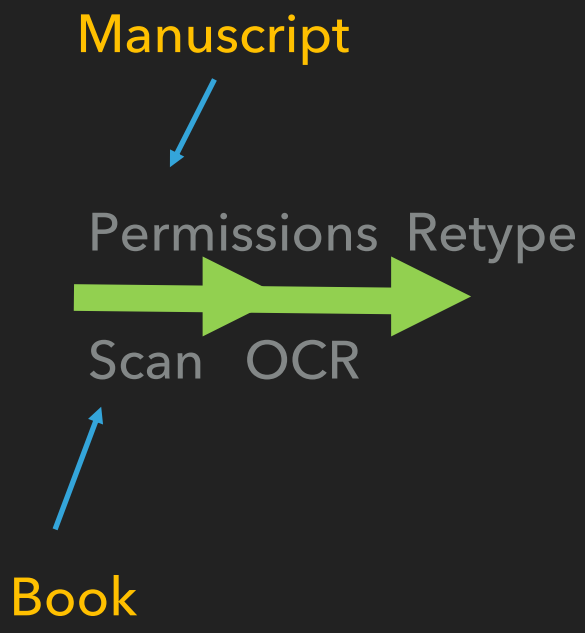
Monday, 22 July 2019 - 2:00pm

The Sydney Language, a long out of print dictionary first published by the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) in 1994, is now available to a new generation of Australians thanks to the fundraising efforts of Australian band The Preatures.

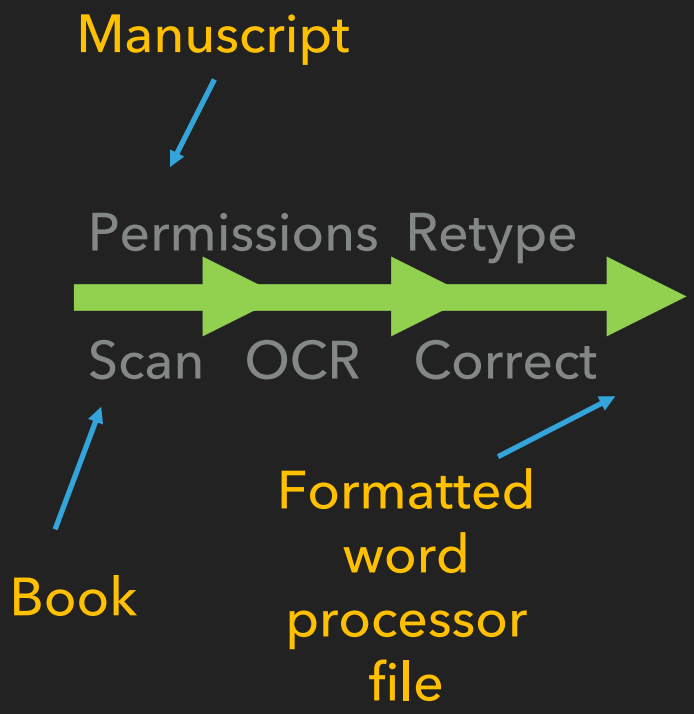
AIATSIS CEO and Co-Chair of the UNESCO Steering Committee for 2019 International Year of Indigenous Languages celebrations, Craig Ritchie, said language is intrinsic to life for Indigenous peoples around the world.

First published 1994, republished 2019

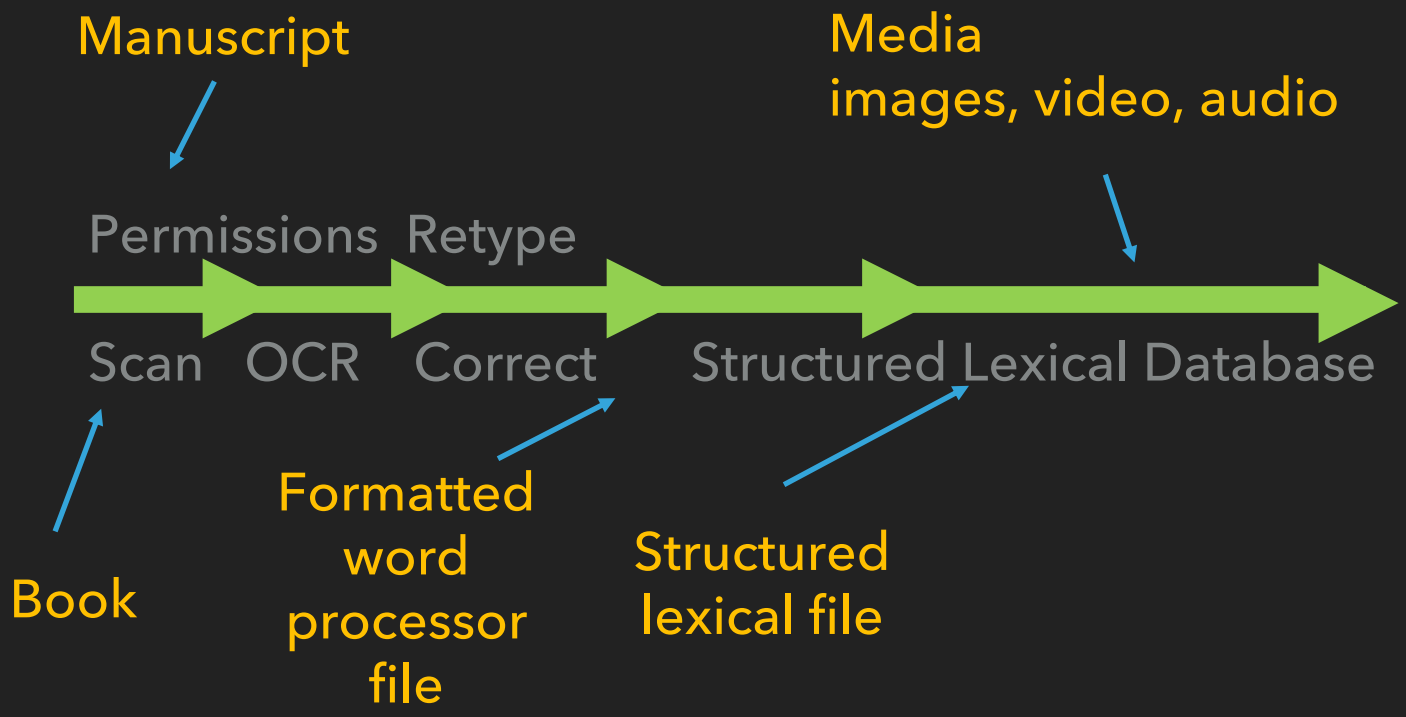
UPDATING EXISTING DICTIONARIES



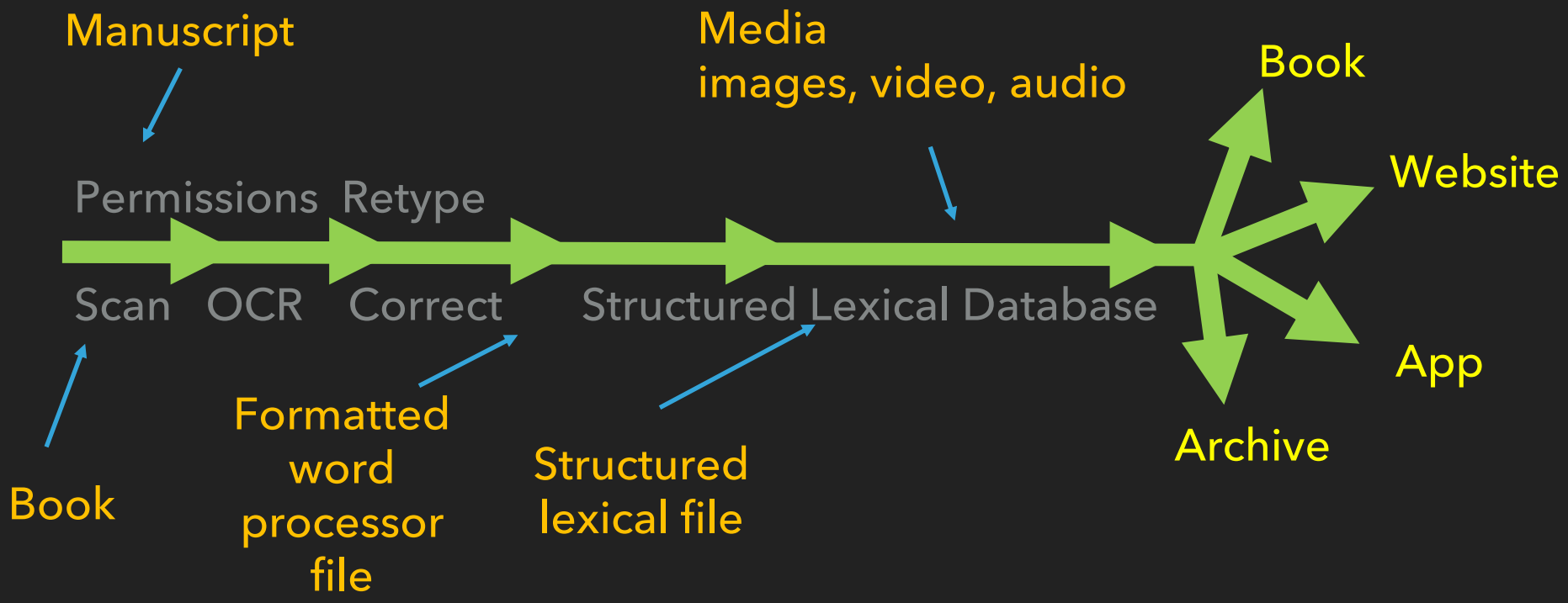
UPDATING EXISTING DICTIONARIES



UPDATING EXISTING DICTIONARIES

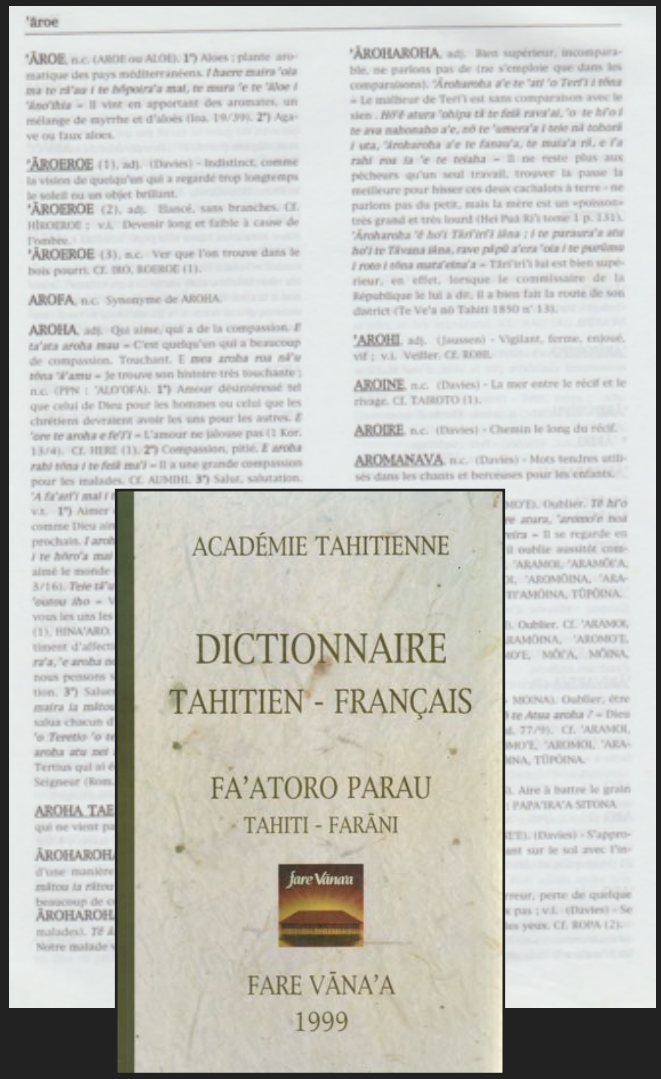


UPDATING EXISTING DICTIONARIES



THE DICTIONARY OF THE TAHITIAN ACADEMY (1999, 574 PP)

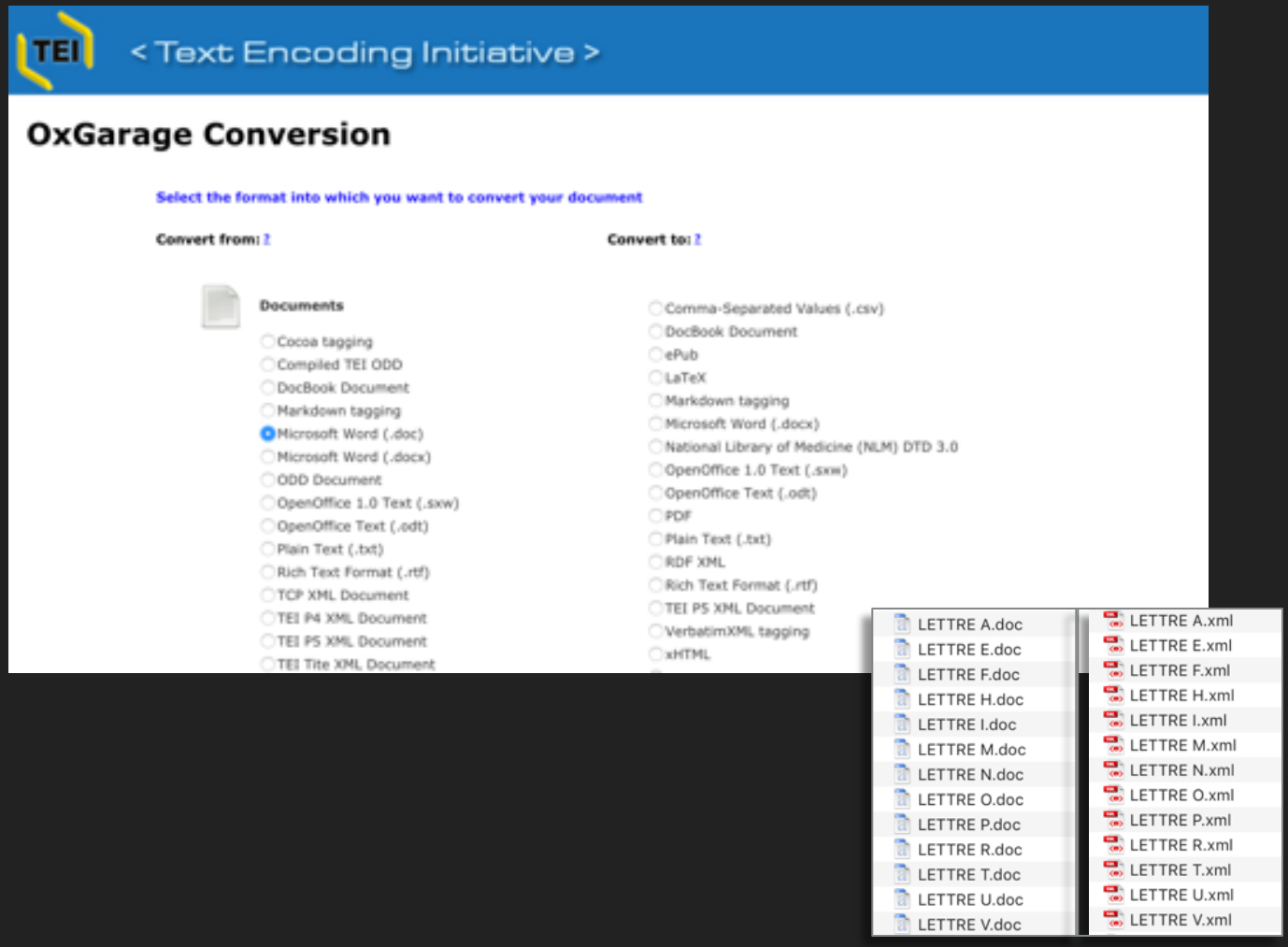
- Aggregates two centuries of lexicographic works, from the dictionary of John Davies (1851) to the contemporary neologisms of the Tahitian Academy
- It contains 13,880 entries
- From Tahitian to French



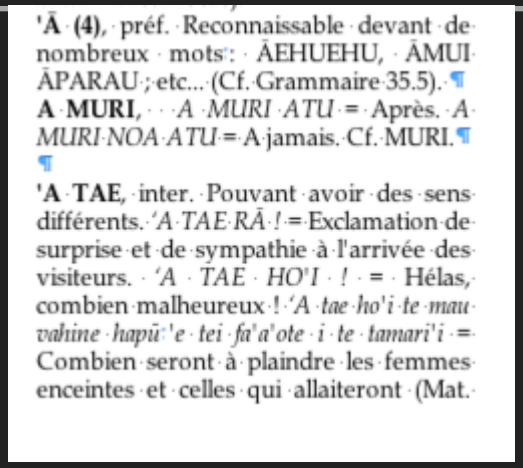
- The Tahitian dictionary ran on a Mac IIfx (that still worked after more than 20 years!) in 4th Dimension (obsolete DBMS).
  The Académie Tahitienne exported a copy to a set of MS Word docs for their printed dictionary

CASE STUDY 1

http://www.tei-c.org/oxgarage



CASE STUDY 1



'Ā (4), préf. Reconnaissable devant de nombreux mots: ĀEHUEHU, ĀMUI ĀPARAU; etc... (Cf. Grammaire 35.5). ¶
A MURI, *A MURI ATU* = Après. *A MURI NOA ATU* = A jamais. Cf. MURI. ¶
¶
'A TAE, inter. Pouvant avoir des sens différents. *'A TAE RĀ !* = Exclamation de surprise et de sympathie à l'arrivée des visiteurs. *'A TAE HO'I !* = Hélas, combien malheureux ! *'A tae ho'i te mau vahine hapū 'e tei fa'a'ote i te tamari'i* = Combien seront à plaindre les femmes enceintes et celles qui allaiteront (Mat.

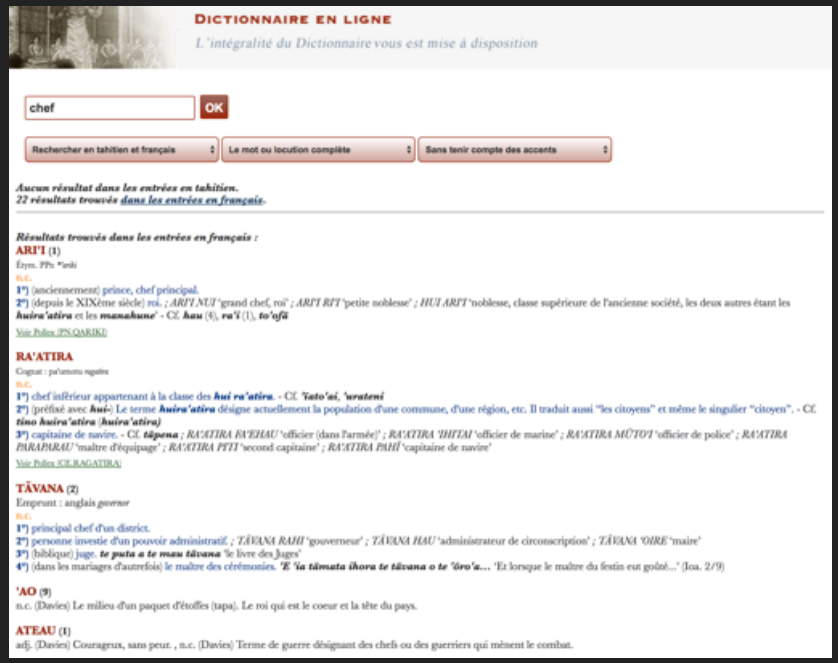
```
        <p rend="justify">¬
            <hi rend="bold">'A TAE</hi>, inter. Pouvant avoir
des sens différents. <hi rend="italic">'A TAE RĀ !</hi> =
Exclamation de surprise et de sympathie à l'arrivée des
visiteurs. <hi rend="italic">'A TAE HO'I !</hi> = Hélas,
combien malheureux ! <hi rend="italic">'A tae ho'i te mau
vahine hapū 'e tei fa'a'ote i te tamari'i</hi> = Combien seront
à plaindre les femmes enceintes et celles qui allaiteront (Mat.
24/19). <hi rend="italic">'A TAE AI HO'I !</hi> = Combien
plus ! <hi rend="italic">'Ia na reira ho'i te Atua i te
fa'a'una'una i te rā'au ri'i o te fenua nei... 'a tae ai ho'i
ia 'outou, 'o 'outou i te fa'aro'o tāpetepete ē !</hi> = Si
Dieu prend ainsi soin de l'herbe des champs... combien plus
(prendra-t-il) soin de vous, hommes de peu de foi ! (Mat.
6/30). </p>¬
        <p rend="justify">¬
```

\lx 'A TAE
\ps inter.
\de Pouvant avoir des sens différents.
\re Pouvant avoir des sens différents
\xv 'A TAE RĀ !
\xe Exclamation de surprise et de sympathie à l'arrivée des visiteurs.
\xv 'A TAE HO'I !
\xe Hélas, combien malheureux !
\xv 'A tae ho'i te mau vahine hapū 'e tei fa'a'ote i te tamari'i
\xe Combien seront à plaindre les femmes enceintes et celles qui allaiteront (Mat. 24/19).
\xv 'A TAE AI HO'I !
\xe Combien plus!
\xv 'Ia na reira ho'i te Atua i te fa'a'una'una i te rā'au ri'i o te fenua nei... 'a tae ai ho'i ia 'outou, 'o 'outou i te fa'aro'o tāpetepete ē!
\xe Si Dieu prend ainsi soin de l'herbe des champs... combien plus (prendra-t-il) soin de vous, hommes de peu de foi ! (Mat. 6/30).

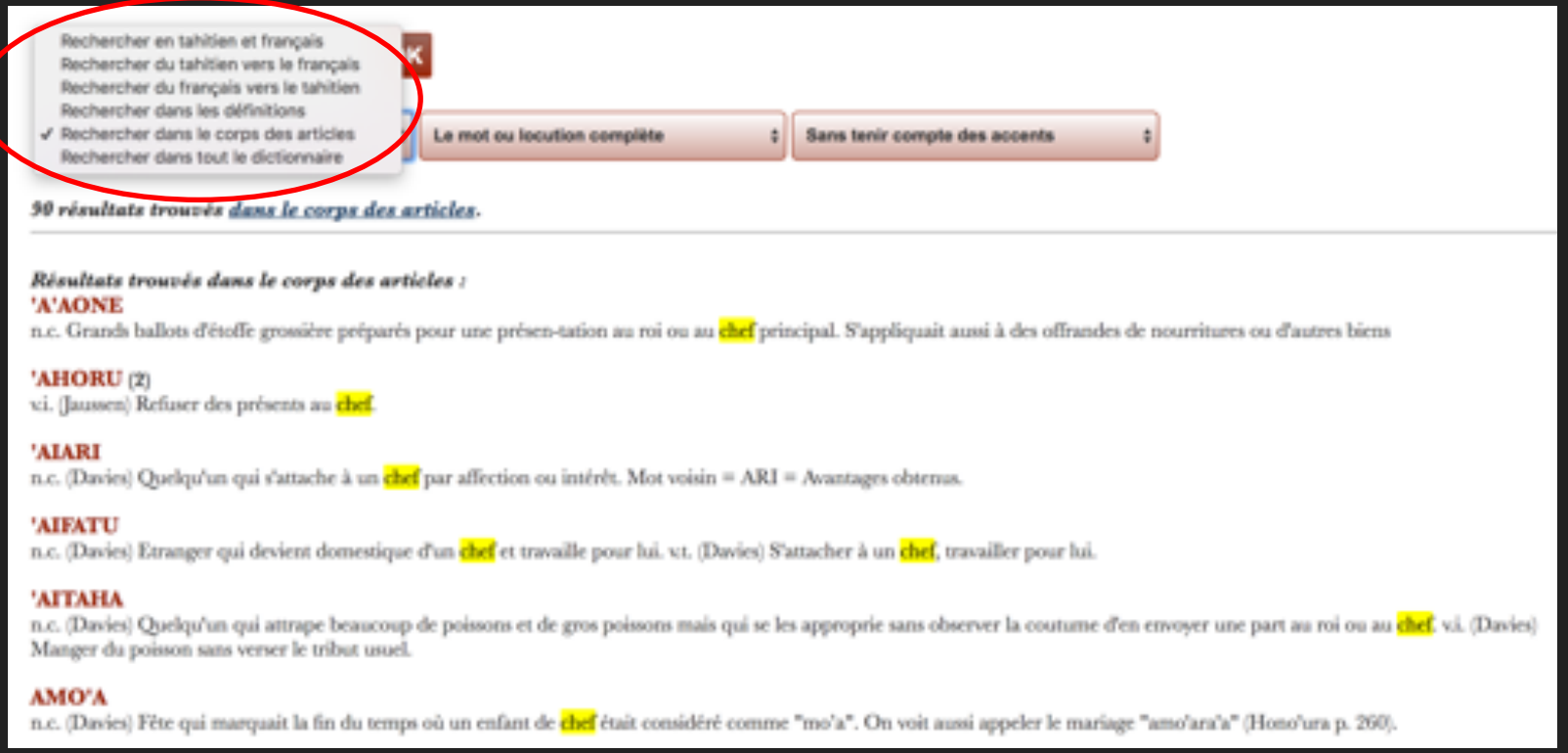
ONLINE DICTIONNARY V2 (2017)

http://www.farevanaa.pf/v2/dictionnaire.php

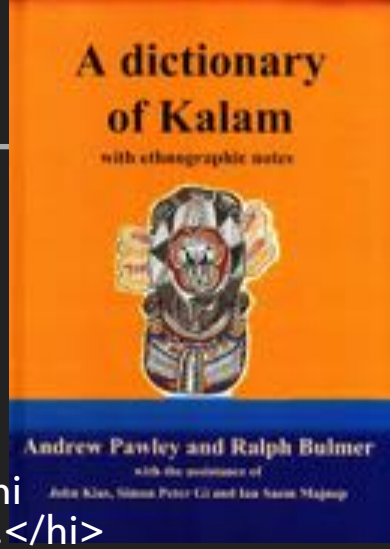
- Tahitian > French
- French > Tahitian



SEARCH OPTIONS



KALAM DICTIONARY – MS WORD DOCS



-E-[1] [-é--], *suffix to medial verbs.* (1) marks the event indicated by the medial verb as prior to that indicated by the next verb of the sentence, and (2) marks the medial verb as having a different subject from the next verb. *Yad den, pkan!* I'll hold it and (then) you hit it. *Dey pkin*! You hold it and I'll hit it! *Apek kaj ñak.* When he came he (different subject) gave him pork. *Ñben, yp tep gp.* It tasted good to me. (lit. When I ate, good happened to me.) Contrasts with other medial verb suffixes **-i, -ig, -ng, -nŋ,** q.v.¶

Conversion using OxGarage to create an archival version of the files

```
<p rend="Kalam Head Para">-<hi rend="bold">E</hi>-<hi rend="superscript">1</hi> [-é·-], <hi rend="italic">suffix to medial verbs.</hi>

(1) marks the event indicated by the medial verb as prior to that indicated by the next verb of the sentence, and (2) marks the medial verb as having a different subject from the next verb.

<hi rend="italic">Yad den, pkan!</hi> I'll hold it and (then) you hit it.  <hi rend="italic">Dey pkin</hi>!

You hold it and I'll hit it!  <hi rend="italic">Apek kaj ñak</hi>.

When he came he (different subject) gave him pork.

<hi rend="italic">Ñben, yp tep gp.</hi> It tasted good to me. (lit. When I ate, good happened to me.) Contrasts with other medial verb suffixes -<hi rend="bold">i</hi>, -<hi rend="bold">ig</hi>, -<hi rend="bold">ng</hi>, -<hi rend="bold">nŋ</hi>, q.v.</p>
```

WUBUY (NUNGGUBUYU) DICTIONARY — PAPER TYPESCRIPT

Original

lhalga NAdj alone, by oneself, isolated. 168.5.2pred.
Ex: n^{g}a=lhalga n^{g}a=ya-n^{g}gi 'I went alone'.
Syn: w$_2$iriwiri (more common term). Cf. w$_1$/w$_2$urag.

lhalgur NcMana a mangrove tree sp. with pointed leaves and edible small green fruits, Avicennia marina. 114.8.1, 118.1.3.
Syn: nambara (uncommon term).
Sapling: wurjulyul (mostly songword).
Fruits: aran^{y}ann^{g}u (common term), n^{g}an^{y}jawalwal.

-lhalma- Pref
Occurs as cpd. initial as a kind of filler with constructions meaning 'very big': see -w$_2$ugag, -yiriwu.
Cf. n^{g}agara- in the same construction type.

-lha:l=ma:jga- VTrA1 (See =ma:jga-).

-lha:l-malan^{g} NAdj (See mamalan^{g}).

-lha:l=ma:ndha- and -lhal=ma:ndha- VTrA2 (See =ma:ndha-).

lhalmar NAdj foreigner, stranger.
HumSg: lhalmar-yun^{g}. Pl: lhalma-lhalmdr. 47.14.6.
May contain lha:l 'country'. Syn: mulguru.

OCR

lhalga NAdj alone, by oneself, isolated. 168.5.2pred.¶
Ex: n^{g}a=lhalga n^{g}a=ya-n^{g}gi 'I went alone'.¶
Syn: w$_2$iriwiri (more common term). Cf. w$_1$/w$_2$urag.¶
lhalgur NcMana a mangrove tree sp. with pointed leaves and edible¶
small green fruits, Avicennia marina. 114.8.1, 118.1.3.¶
Syn: nambara (uncommon term).¶
Sapling: wur'uyul (mostly songword).¶
Fruits: aran^{y}anngT1 (common term), ngan^{y}jawalwal.¶
-lhalma- Pref¶
Occurs as cpd. initial as a kind of filler with constructions
meaning 'very big': see -w ugag, -yiriwu.¶
Cf. n^{g}agara- in the same construction type.¶
-lha:l=ma:jga- VTrA1 (See =ma:jga-).
-lha:l-malan NAdj (See mamalan^{g}).¶
-lha:l=ma:ndha- and -lhal=ma:ndha- VTrA2 (See =ma:ndha-).¶
lhalmar NAdj foreigner, stranger.¶
HumSg: lhalmar-yung. Pl: lhalma-lhalmdr. 47.14.6.
May contain lha:l 'country'. Syn: mulguru.¶

Typed

lhalga NAdj alone, by oneself, isolated. 168.5.2pred.¶
Ex: n^{g}a=lhalga n^{g}a=ya-n^{g}gi 'I went alone'.¶
Syn: w$_2$iriwiri (more common term). Cf. w$_1$/w$_2$urag.¶
lhalgur NcMana a mangrove tree sp. with pointed leaves and edible
small green fruits, Avicennia marina. 114.8.1, 118.1.3.¶
Syn: nambara (uncommon term).¶
Sapling: wurjulyul (mostly songword).¶
Fruits: aran^{y}ann^{g}u (common term), n^{g}an^{y}jawalwal.¶
-lhalma- Pref¶
Occurs as cpd. initial as a kind of filler with constructions
meaning 'very big': see -w$_2$ ugag, -yiriwu.¶
Cf. n^{g}agara- in the same construction type.¶
-lha:l=ma:jga- VTrA1 (See =ma:jga-).¶
-lha:l-malan^{g} NAdj (See mamalan^{g}).¶
-lha:l=ma:ndha- and -lhal=ma:ndha- VTrA2 (See =ma:ndha-).¶
lhalmar NAdj foreigner, stranger.¶
HumSg: lhalmar-yun^{g}. Pl: lhalma-lhalmar. 47.14.6.¶
May contain lha:l 'country'. Syn: mulguru.¶

NATIONAL LIBRARY OF AUSTRALIA: 43 WORDLISTS (MFM)

- Anonymous A Wallis Island - French Dictionary.
- Anonymous Lexique de la langue de South West Bay (Malicolo) (New Hebrides)
- Bancarel, C. Dictionaire Français - Sesivi. (New Hebrides)
- Bancarel, P. and others. Lexique Français - Port Sandwich (New Hebrides).
- Bancarel, P. Dictionnaire Sesivi-Français. Ms.
- Bancarel, R.P. Supplement au dictionnaire Sesivi-Français du R. P. Bancarel.
- Bartlett, Rev. H. K. Vocabulary of the Misima-Panaeati Language, Papua. 155pp.
- Bresson, E. A French-Tyamuhi Dictionary. 126pp. Original held by Rt. Rev. E. Bresson, SM, DD, Vicar Apostolic for New Caledonia. SPC mf 2
- Capell, Arthur. A preliminary Maré (Nengone) and English dictionary
- Chauvel, P. Dictionaire Francais - Langue de South West Bay (Malekula, New Hebrides). Ms.
- Chauvel, P. Lexique de la langue de S. W. Bay (Malekula, New Hebrides) 1907.
- Chauvel, P. Lexique: Français-Sesivi. Ms.
- Chauvel, P. Lexique. Langue indigège South West Bay (Malekula, New Hebrides). Ms.
- Courtais, P. Dictionnaire Français-langue de Port-Orly et langue de Port-Olry - Français. Ms.
- Dubois, Marie Joseph . Papers relating to Maré, New Caledonia, including vocabularies in the Nengone Language and an ethnology of Maré.
- Durand, P. Dictionnaire Français-Canaque. Langues du Sud de Pentecôte, 1906-1912. Ms.
- Emprin, R P. Whegnai-French Dictionary. c. 1900.
- Fox, C. E. A Ngella-English Dictionary. 1955.
- Fox, Charles Elliot. A Lau (British Solomon Islands)-English Dictionary, 1950.
- Giblin, Rev. Eric. Mukawa and English Dictionary (with addition of Gapapaiwa) and a grammar and dictionary
- Guiart, M.J. A French-Craig Cove (New Hebrides) Word List. 85pp. Original held by M. J. Guiart. SPC mf 60
- Guiart, M.J. A French-Dubea (New Caledonia) Word List.
- Guiart, M.J. A French-Iai (Loyalty Islands) Word List.
- Guiart, M.J. A French-Olal (New Hebrides) Word List.
- Guiart, M.J. A French-Venan (New Hebrides) Word List.
- Guiart, M.J. A vocabulary of the Kilene dialect as spoken at Sag-Sag, New Britain.
- Holzknecht, K. An Azira (New Guinea), German and English Dictionary
- Kaysser, C. An English-Nauru Dictionary.
- Larson, Gordon F & Mildred O. Moni-Malay-English Dictionary.4
- Loubiere, E. A Dictionary and Grammar of the Melsisi-Namaran Language (New Hebrides).
- Maillard, A. A French-Gilbertese Dictionary, 1917.
- Paton, Frank Hume Lyall. English and Lenakel dictionary from manuscript of F. H. L. Paton. Leichhardt, 1922.
- Paton, J. G.A vocabulary of words in the language of Aniwa, New Hebrides. Ms.,
- Paton, W. F. Dictionary and Grammar of the Language of Dip Point and Northern Ambrym.
- Poole, A. J. An English-New Britain Dictionary, Vunairima, New Guinea, Methodist Mission, 1953.
- Salomon, P. Grammaire de la langue de Wala (Mallicolo, Nouvelles-Hébrides).
- Salomon, P. Lexique Vao-Français (New Hebrides). Ms.
- Tattevin, P. Dictionnaire Français-Melsisi, 1909. Ms.
- Tattevin, P. Dictionnaire Français-Melsisi. Ms.
- Tattevin, P. Grammaire et vocabulaire Oba (New Hebrides) 1935. Ms.,
- Tattevin, P. Vocabulaire Loltong-Français et Français-Loltong. Ms.

LOST AND FOUND – LOCATING DICTIONARY FILES

- Structured source files reduce effort required to rework a dictionary
- Locating these files, responses to request to mailing lists (e.g. Oct 2018)
  - Atong
  - Bilinarra
  - Gamilaraay
  - Gurindji
  - Iatmul
  - Kalamang
  - Kwamera
  - Ngarinyman
  - Nukumanu
  - Rawang
  - Taumako
  - Titan
  - Unua
  - Waima/Roro/Paitana
  - Warlmanpa
  - Warlpiri
  - Xârâcùù

OLAC resources in and about the Nukumanu language

ISO 639-3: nuq

The combined catalog of all OLAC participants contains the following resources that are relevant to this language:

- Primary texts
- Lexical resources
- Language descriptions
- Other resources about the language
- Other resources in the language

Other known names and dialect names: Tasman

Use faceted search to explore resources for Nukumanu language.

Primary texts

1. ONLINE *Languages of the Bougainville District by Jerry Allen and Conrad Hurd.* Jerry Allen; Conrad Hurd. 1964. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai:paradisec.org.au:AC2-NSBCVOC201

Lexical resources

1. ONLINE *Sepik vocabularies, Microfiche. La Jolla, Calif. : University of California, San Diego, Melanesian Studies Resource Center, 1991. 8 microfiches.* Various Various (compiler). 1970. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai:paradisec.org.au:External-sepik
2. ONLINE *Nukumanu dictionary.* Rick Feinberg (compiler); Rick Feinberg (compiler). n.d. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai:paradisec.org.au:RF1-NUKUMANU

Language descriptions

1. ONLINE *Glottolog 4.0 Resources for Nukumanu.* n.a. 2019. Max Planck Institute for the Science of Human History. oai:glottolog.org:nuku1258

Other resources about the language

1. ONLINE *Nukumanu: a language of Papua New Guinea.* n.a. 2018. SIL International. oai:ethnologue.com:nuq
2. ONLINE *LINGUIST List Resources for Nukumanu.* Damir Cavar, Director of Linguist List (editor); Malgorzata E. Cavar, Director of Linguist List (editor). 2017-09-27. The LINGUIST List (www.linguistlist.org). oai:linguistlist.org:lang_nuq

Other resources in the language

1. ONLINE *Counting systems of Papua New Guinea.* Lean, Glendon A. 1991. Lae, Papua New Guinea : Department of Mathematics and Statistics, Papua New Guinea University of Technology. oai:pacific.library.manoa.hawaii.edu:1863773
2. ONLINE *The Polynesian languages in Melanesia. [With notes on the neighbouring Melanesian languages].* Ray, Sidney Herbert, 1858-1939. 1919. Wien. oai:pacific.library.manoa.hawaii.edu:1935814

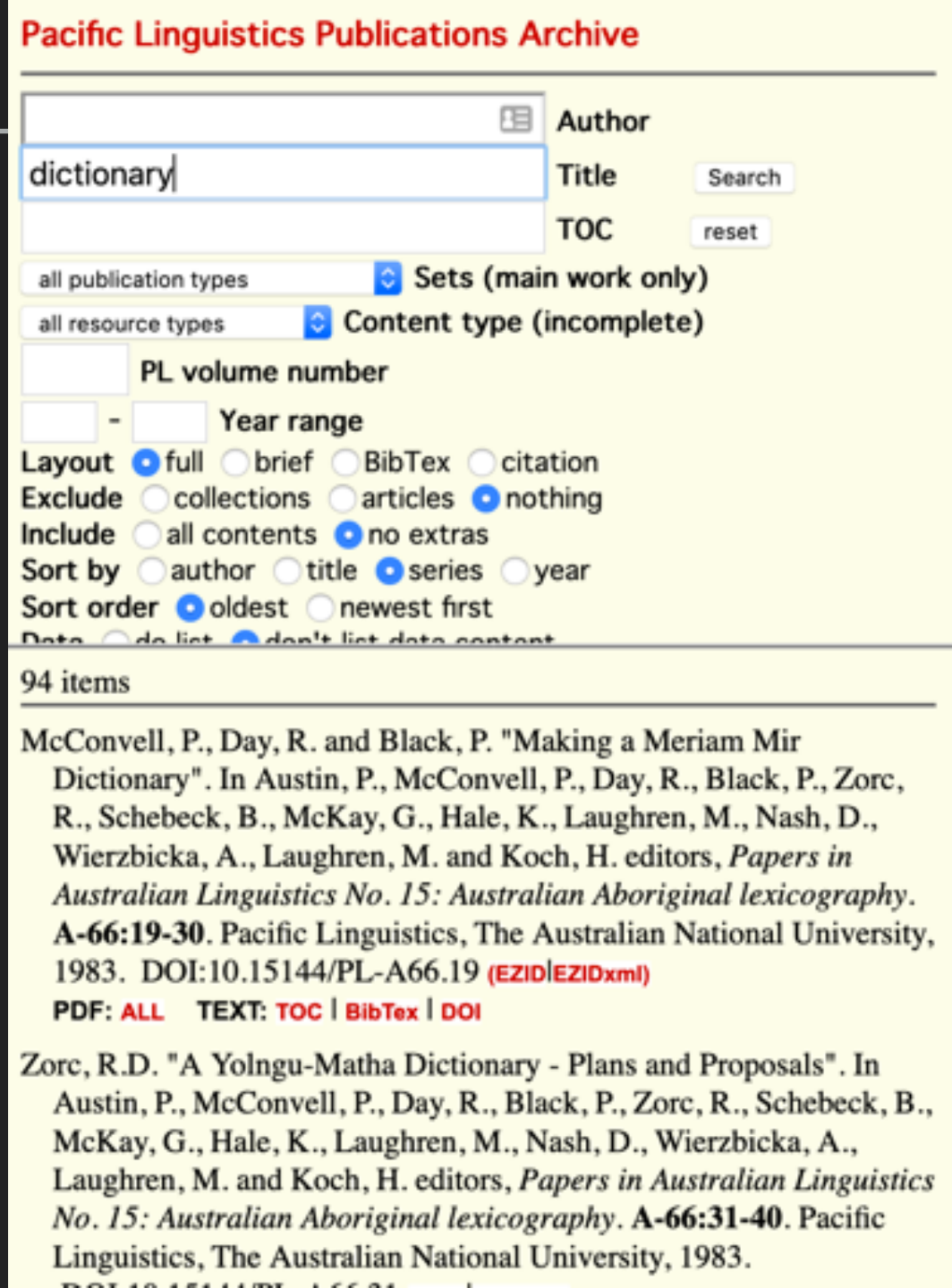
Other known names and dialect names: Tasman

Other search terms: dialect, vernacular, discourse, stories, conversation, dialogue, documentation, lexicon, dictionary, vocabulary, wordlist, phrase book, grammar, syntax, morphology, phonology, orthography

http://www.language-archives.org/language.php/nuq
Up-to-date as of: Tue Aug 6 7:53:16 EDT 2019

FINDING LOST DICTIONARIES

- http://sealang.net/archives/pl/



PACIFIC AND REGIONAL ARCHIVE FOR DIGITAL SOURCES IN ENDANGERED CULTURES (PARADISEC)

- 1,219 languages
- 516 collections
- 25,822 items
- 283,219 files
- 11,000 hours (audio)
- 980 hours (video)
- 66 TB of files

- 37 TB of video files in archive
- 18.4 TB of audio files in archive
- 1.6 TB of image files in archive
- 1.7 GB of text files in archive

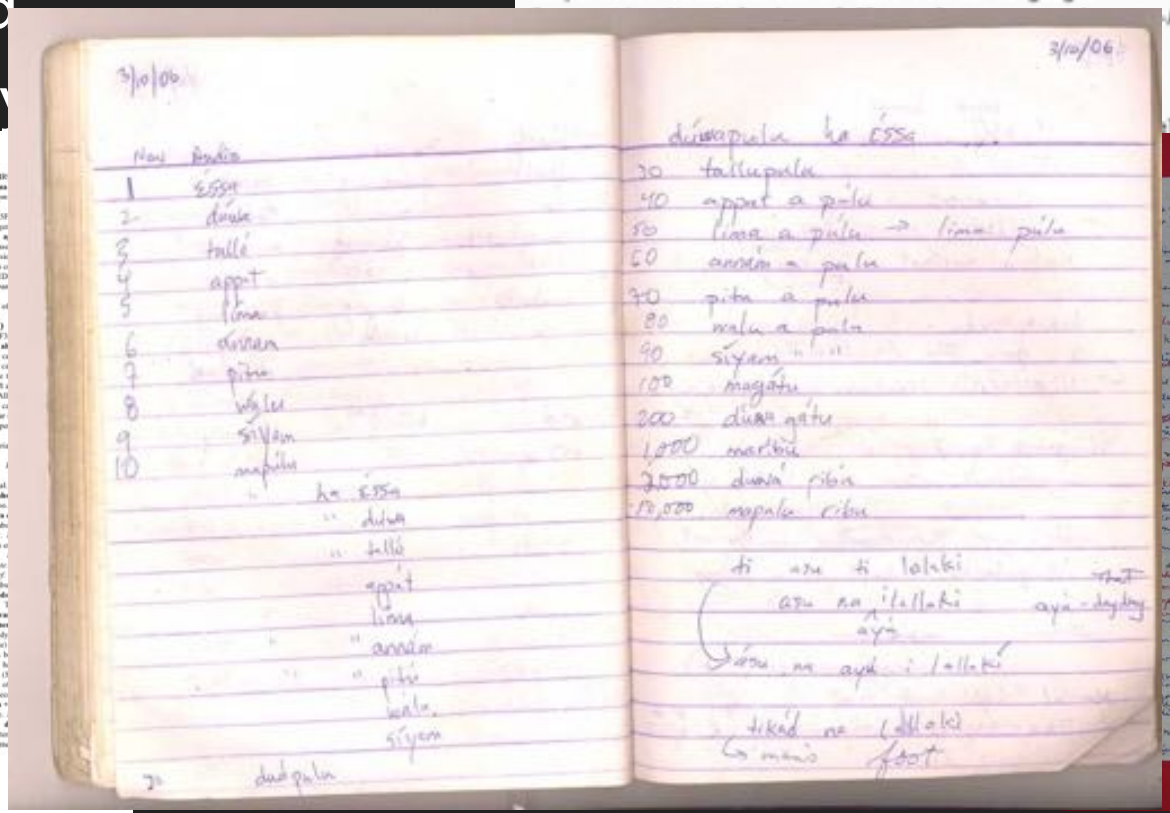
- August 2019

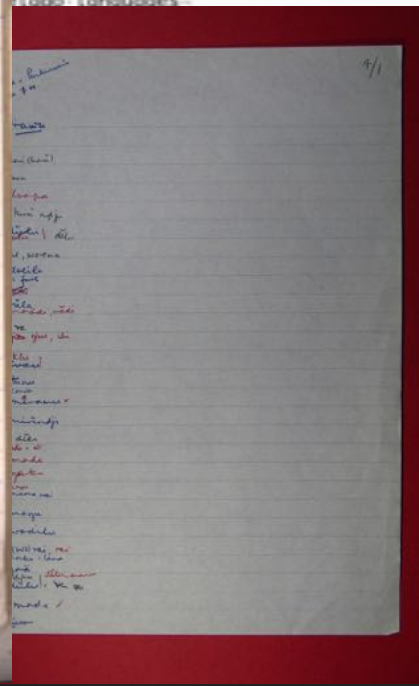
ARC CENTRE OF EXCELLENCE FOR THE DYNAMICS OF LANGUAGE

1,114 "LEXICON" ITEMS IN PARADISEC

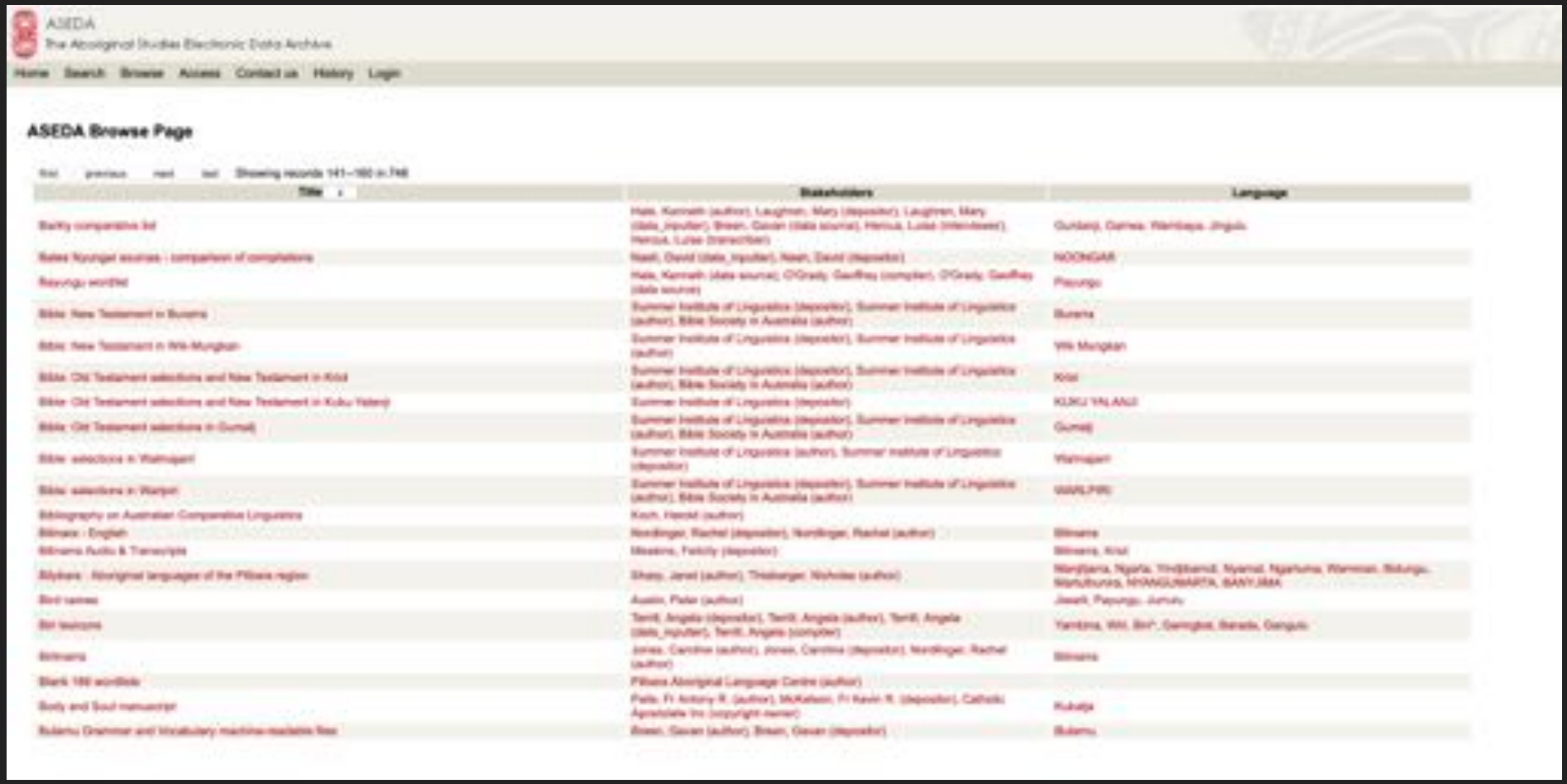
- Scanned images
  - handwritten
  - typed
- Page lay

Pohnpeian
Languages of Tanna Island, Vanuatu
Taiora - Fore comparative vocabulary
Kamano grammatical notes, wordlist and texts
Godzub (Gadsup) vocabulary and phrases
Yamafi - Tjuave (Chuave) comparative vocabulary
Lagaria wordlist and phrases
Ekagi vocabulary
Bokondini vocabulary
Correspondence with Myron Bromley about Abui (incl. vocabulary)
Short Alor - Timor vocabulary
Alor wordlist and phrases
A comparative of Sulawesi Languages
Notes on Spraakunst van le Banggai
Watubela comparative vocabulary
Indonesian Vocabularies
Indonesian Vocabularies II
Comparative vocabularies of various MN and AN languages
Comparative vocabularies of Ngad'a and Manggarai
Comparative Vocabularies of East Flores languages
Austronesian





ABORIGINAL STUDIES ELECTRONIC DATA ARCHIVE (ASEDA)



http://aseda.aiatsis.gov.au/aseda_main.php

EFFORT OF CONVERSION OF EARLIER FILES

- Could be avoided if lexical databases were archived
- Book-based dictionaries have great values but they:
  - do not last long in village/community/camp settings
  - cost too much for speakers to buy
  - go out of print and become unavailable
- Speakers rightly have an expectation that they can access a dictionary of their own language
- We can make an effort to create digital versions of dictionaries in open repositories.

References

Académie tahitienne. 1999. *Dictionnaire tahitien - français*. Papeete: STP Multipress.

Dorian, N. C. 2014. *Small-Language Fates and Prospects: Lessons of Persistence and Change from Endangered Languages: Collected Essays.* Brill's Studies in Language, Cognition and Culture, volume 6. Leiden: Brill.

Heath, Jeffrey. 1982. *Nunggubuyu Dictionary*. Canberra: Australian Institute of Aboriginal Studies.

Pawley, Andrew and Ralph Bulmer ; with the assistance of John Kias, Simon Peter Gi and Ian Saem Majnep. 2011. *A dictionary of Kalam with ethnographic notes*. Canberra: Pacific Linguistics.

Thieberger, Nicholas. 2011. Building a lexical database with multiple outputs: examples from legacy data and from multimodal fieldwork. *International Journal of Lexicography*, 24(3) 463-472; doi:10.1093/ijl/ecr027

Australian Research Council – LIEF grants LE110100142, LE0560711, LE0453247

ARC DP0450342, DP0984419, FT140100214, CE140100041

ARC CENTRE OF EXCELLENCE FOR THE DYNAMICS OF LANGUAGE