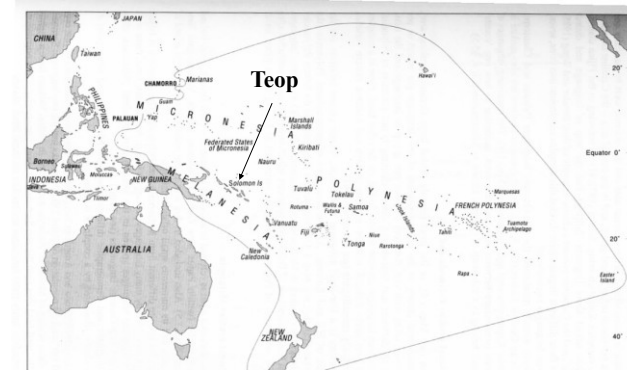


Turning a linguist's lexical data base into a community dictionary
Ulrike Mosel, University Kiel

Saami Winterschool of Language Documentation and Revitalisation
1-12. March 2010

The Teop language documentation project

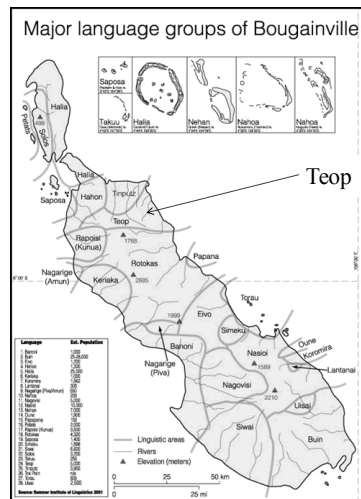


Bougainville
8800 km²
175 000 people

28 Languages

Teop
Austronesian,
Oceanic

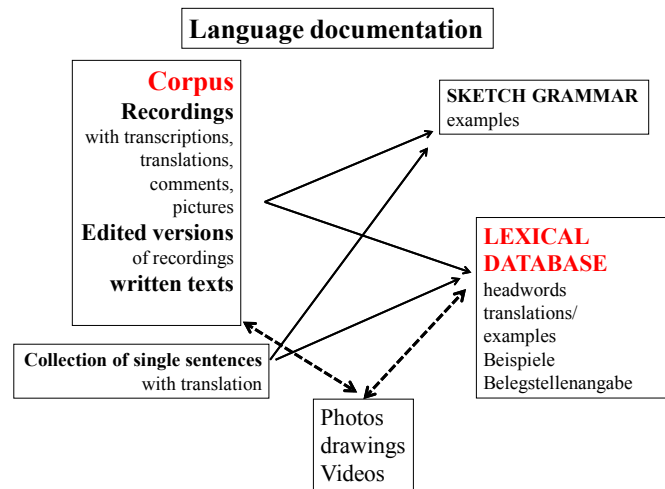
Contact languages:
Tok Pisin
English



Documentation of the Teop Language

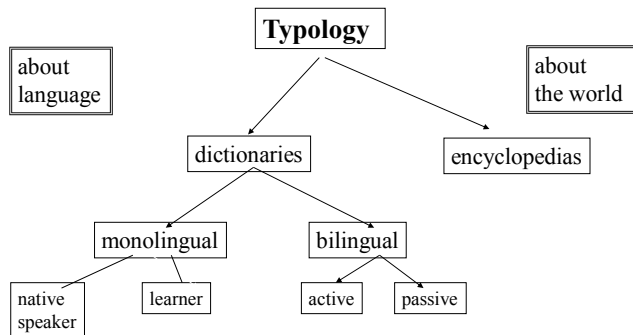
funded by the Dokumentation für Bedrohte Sprachen
program of the Volkswagen Foundation
2000-2007

and
German Research Foundation
2009-2011

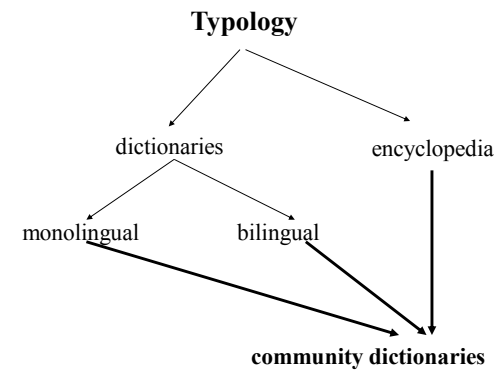


1. Typology of ordinary dictionaries
2. Why dictionaries for endangered languages are special dictionaries
3. Content and structure of the Teop Lexical Database (TLD)
4. Why the print-out of the TLD is not a dictionary
5. Transforming the TLD into mini-dictionaries
6. What linguists can learn from mini-dictionaries

6



7



8

Factors that determine kind and size of dictionaries

intended users:

general public, learners, linguists, specialists ...

purpose

translating, language learning

resources

staff, equipment, money for publication

time frame

(3 years, 10 years, ...?)

9

1. Typology of ordinary dictionaries
2. **Why dictionaries for endangered languages are special dictionaries**
3. Structure and content of the Teop Lexical Database (TLD)
4. Why the print-out of the TLD is not a dictionary
5. Transforming the TLD into mini-dictionaries
6. What linguists can learn from mini-dictionaries

10

Dictionaries for endangered languages are special dictionaries

	ordinary dict.	community dict.
economic basis	commercial	funding agencies
time frame	decades	3-10 years, part time
purpose	translating, L2 learning	language maintenance
users	general public	small community, academics
lexicographers	professionals	linguists, community members
linguistic resources	huge corpora, old dictionaries	language documentation

11

Dictionaries for endangered speech communities: users & purpose

Linguists	Native speakers
understand and analyse texts	Preservation of cultural memory education language maintenance,
passive dictionary for translation	passive & active learner dictionary
linguistic information	linguistic & encyclopedic information

12

1. Typology of ordinary dictionaries
2. Why dictionaries for endangered languages are special dictionaries
- 3. Content and structure of the Teop Lexical Database (TLD)**
4. Why the print-out of the TLD is not a dictionary
5. Transforming the TLD into mini-dictionaries
6. What linguists can learn from mini-dictionaries

13

Content and structure of the TLD

Not an end product,
but a **dynamic tool**, containing information on

-the inherent grammatical features of lexical units
(gender, valency)
- grammatical relations between words
(conversion, derivation, composition)

-semantic features
- semantic relations to other lexical units

-translation equivalents
-extralinguistic reality

15

The Teop Language Corpus

(ca. 170 000 words)

	new genres
Legends	no
conversations on history, social and material culture	no/yes
Personal histories	yes
Procedural texts (How to make an X?)	yes
For the dictionary:	
Descriptions (What is an X?)	yes

14

Structure of the lexical entry of the TLD:

\lx	head word in Teop
\(mn	crossreference to main entry)
(\mr	morphological structure))
\ps	part of speech
\sn	sense
\de	English translation equivalent of lx\ (if possible)
\pc	picture (link to the Teop Picture Dictionary)
\sd	semantic domain
\ev	Teop encyclopedic information
\ee	English translation of ev\
\rf	reference of ev\
\xv	Teop example
\xe	English translation of xv\
\rf	reference of example
\se	subentry

16

1. Typology of ordinary dictionaries
2. Why dictionaries for endangered languages are special dictionaries
3. Structure and content of the Teop Lexical Database (TLD)
- 4. Why the print-out of the TLD is not a dictionary**
5. Transforming the TLD into mini-dictionaries
6. What linguists can learn from mini-dictionaries

17

	Lexical database	Dictionary
orthography	can easily deal with variation	standardization preferable
grammar	citations contain speech errors, interferences from dominant languages	standard forms preferable

19

	Lexical database	Dictionary
Space	unlimited	restricted
Macro-structure	multi-dimensional	linear
Purpose	constantly growing unbiased resource and tool for researchers	user oriented, for the speech community
Content	moderately selective	highly selective
Meanings	text meanings	meaning potentials
examples	citations	illustrative examples

18

Problem:
Text meaning vs. meaning potential

Not the meaning of the lexemes captured, but only the various **senses** the word has in the particular **contexts** of the corpus.

And what we consider as distinct senses is influenced by the translation equivalents.

Example: **naovana**

‘and Gaivaa became a **naovana** that we call seagull’

‘then you get two cockatoo feathers, this is a white **naovana**, ...’

naovana ‘bird’ but other **naovana**: flying foxes, insects 20

Lexical database

captures the different aspects of word meaning that are activated in its particular context.

Dictionary

“contains lists of meaning potentials”

(Hanks 2000/2008)

“Word senses are “abstractions over clusters of word usages.”

(Kilgariff 1997/2008)

How can we capture the relevant “meaning potentials” and present them in a sensible way in the dictionary?

21



‘This old woman, Sharphand, stabbed him to death with her hand.’

23

tahi
 v.t.
 1
 spear
 spear, stab s.o./s.th.
 E bubuu paa tahi mau bona iana.
 Grandfather speared the fish.
 v.t.
 2
 pierce, prick
 pierce, prick s.o./s.th.
 O pusu no onoono to paa tahiua bona.
 It was only the edge of the grass skirt that pricked him.
 [Aro. 6.42]
 v.d.t.
 3
 stab s.o. with s.th., spear s.th. with s.th.
 E subuava vai e Kurikakaniva he tahi vamate bona bona kurinae.
 This old woman, Sharphand, stabbed him to death with her hand.
 [Asu. 1.42R]
 Tahi a si babakui bono sikiri tean.
 Spear the small shark with your spear.
 [Su. Sen. 1.20]
 v.d.t.
 4
 butt
 butt s.o. with s.th.
 Meori tahi gunagunaha amamaanam bona vaaporo.
 And they butted us down with shotguns.

“Text meanings arise from combinations, not from any one word individually.”

Problem: citations are not suitable as example sentences.

‘stabbed him ... with her hand?’



the creation of a dictionary is a different job

- revision of sense discrimination
- systematic ordering of senses
- revision of examples

-moderate standardization of orthography and grammar

The same applies to electronic lexica like LEXUS

You need two versions:

- a) database as a tool for researchers and lexicographers
- b) community version

24

Problem: time management

The lexical database is too big
to be transformed into a dictionary within a short time..



no comprehensive dictionary
never promise such a dictionary

25

1. Typology of ordinary dictionaries
2. Why dictionaries for endangered languages are special dictionaries
3. Content and structure of the Teop Lexical Database (TLD)
4. Why the print-out of the TLD is not a dictionary
- 5. Transforming the TLD into mini-dictionaries**
6. What linguists can learn from mini-dictionaries

26

Necessity is the mother of invention

The first monolingual Samoan dictionary

Ministry of Youth, Culture and Sports, Western Samoa
Australian South Pacific Culture Fund

1 year - 10 000 AUD (6000 €)

compiled 1994
published 1997

How much can be done in one year?

Jakob Grimm, Wilhelm Grimm et al. 1852-1960.
Deutsches Wörterbuch (German Dictionary)
16 volumes

1863 (death of Jakob Grimm) : A - Frucht

27

Alphabetical method

A, B, C, D, ...

Thematic method

house building, fishing, ...

How do you collect headwords?

start with A

- filter the database
- interview experts

Time planning? Setting priorities?

difficult in long
projects with
changing staff

Consistency

at least possible for
one theme

Intellectually inspiring?


28

Criteria for selecting a themes:

- which themes do the elders and teachers consider as important?
- which are most endangered?
- which are most suitable for indigenous lexicographers?
- which are easy to understand for the linguist?


Fish-dictionary

aasun *n.a.* tuna, tuna fish.
 A aasun a iana a beera. O iana bari na teitei kooa maari namana, na aniani rori bona vahara iana vaa namana. Vahara aasun me na teitei pete rori namana. A iana bona na aniani ria tavaan. A iana bona a koma kakaavo bara rapanae me na kakaavo nana. A toon nae he o rahirahi paru bara rahirahi kakaavo.
 The tuna is a big fish. These fish only stay in the ocean and eat little ocean fishes. The small tunas also stay in the Ocean. This fish is eaten by the people . This fish has a white belly and its sides are also white, but its back has black and white stripes. A **buu tom keara paa rahi viravira ni mau bona aasun**. The two brothers caught lots of tuna.




aasun

anahi *n.a.* mullet.
 A anahi sa iana beera vira haana,



anahi



anahi

araarata *n.a.* coral fish with green and white stripes.
 A araarata a iana a monamonaa. The araarata is a greasy fish. According to Vaabero [Vaa. Sen. "raarata"], araarata and raarata are the same fish.

aranavi *n.a.* red fish found in the ocean and in big rocks in the deep sea.
 A aranavi na teitei nana namana bara ruene ei taena ruene ae tea boon. A tabaan teori o vahara kehaa vaa tahii ae tea boon bara o vahara katoa me to aniani riori. Aranavi a gogooravi vashum. A iana bona

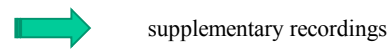
Themes of Teop mini-dictionaries:

house, canoe,
 fishes, fishing, the sea, shellfish,
 trees, other plants, gardening, cooking,
 body, health,
 kinship, ceremonies

Further advantages of mini-dictionaries:

- specialised vocabulary is less frequently used
- more endangered
 - less polysemous

Problem : specialised vocabulary is under-represented in the corpus



Supplementary recordings in Teop for the House-Dictionary:

- How to build a men's house
- How to make bamboo walls
- How to make the floor from the wood of the *kabuu* palm
- How to make the thatch

Further problems of specialised vocabulary

- it is difficult to translate
- linguists have no expertise in ethno-sciences
- indigenous experts lack the necessary proficiency in the target lang.



vernacular encyclopedic descriptions
with translations

33

Sensitive training in writing definitions

marahiri

'The *marahiri* is a fish, that lives in the sea. The *marahiri* has no scales, its body is slippery' (in Teop language)



Colour, size?
Where does it live?
What does it feed on?

What characteristics are essential?
What is a good example? 35

Problem: descriptions are not an indigenous genre

- but show the expressive potential of the language
- reflect the native speaker's metalinguistic knowledge
- can show the native speaker's conceptualisation of extra-linguistic phenomena, e.g. taxonomies
- are useful to understand the meaning of words



training: "Explain what X is to a child."

34

Sensitive training

- avoid patronizing
- encourage them to keep and/or develop their own way of explaining words and things
- avoid style guides

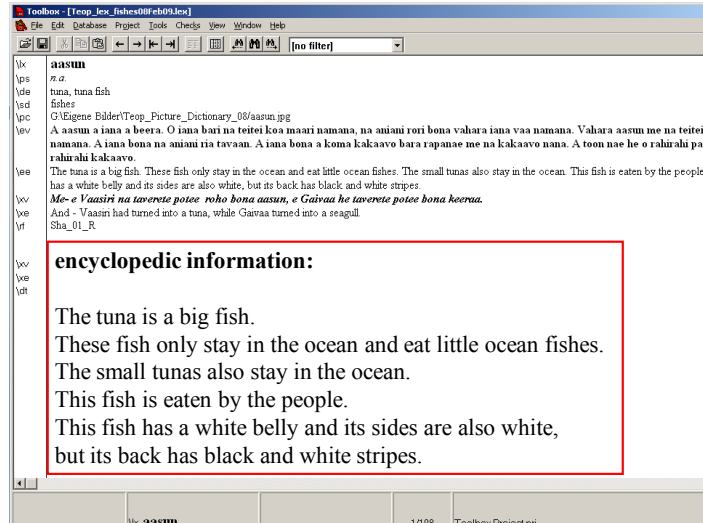
naovana 'bird'

"We eat many birds. Birds are a good food. Only the birds that have a story are the ones that we do not eat. These birds are Pasukokoreo, Topeipei, Toai und Koo."

Content of encyclopedic information:

cf. Coward & Grimes 2000: 138-153)
consult anthropologists, ethnobotanists, etc.

36



Linguistic observations

conventionalized constructions for descriptions of things, properties and events

systematic patterns of polysemy and word formation

Example for systematic heterosemy/conversion

Noun 'part of the house'	Verb 'add this part (e.g. wall) onto the house'
'rafter'	'put up the rafters'
'bamboo wall'	'put up the bamboo wall'
'fence'	'put up the fence'

Conventional and idiosyncratic language use in indigenous lexicography

Let different people work on items

that are presumably described in a similar way; e.g.

house & canoe or **fishes & birds**

similarities of house and canoe descriptions:

- topic = thing
- predicate = put s.th. somewhere

similarities of fishes and birds descriptions:

- topic = animal
- predicate = habit
- = properties like size, colour, shape

Conclusions

A lexical database and a dictionary are two very different things

a dictionary is not a by-product

a dictionary requires hard work and

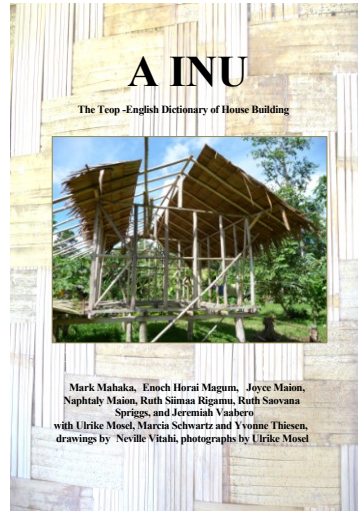
active involvement of the speech community



Start with a mini-dictionary.

Use vernacular encyclopedic descriptions.

Be rewarded by a culturally and linguistically interesting, and **completed** little dictionary!



A-u-a

• ariono¹ 1st fish line used for three special types of fishing and for building houses. Amas ariono: maa: taika: evahie: saka mataa haana. Our ariono dried up, it is not good. 

• atovo² 1st sago palm. E: sumde be tagata bono atovo: mas mas moon pas hosi komma bar. When the old man chopped down the sago palm, the women cut the leaves for haka aravo. 

• 1st thatch of sago palm leaves. O: atovo: eve to lahi vaobe mas: riori toa no ma vuvurara. Na: moori riori: ore pas hosi bono pakanne eori: repas: siri lahi: bona: paka: bono siki:ri: maa: vatame: los: bona no: vuri: na: te: haka. Va: haka: riori: eori: paka to: ipino: eori: repas: luh: bari: te: maa: tataraha. A: bebehu use? Na: tei: pe: hoka: na: te: bebehu: no: i: ni. O: atovo: no: tei: va: bebehu: maa: saka: antee: haana: te: hosa: va: laka: hu. A: kio: va: te: atovo: va: ruru: ruru: laka: maa: hosa: maa: maha: taha: vai to: antee: raara: te: haka: pona: i: ni. The sago palm thatch is put on top of the rafters. They cut the sago palm and then they cut the leaves and remove the midrib from the leaf so that it becomes flexible for sewing the thatch. They break it after they have done the measurement, then they sew it on to the tataraha (a bamboo strip). How long is it? That just depends on the length of the house. The sago palm thatch stays for a long time, it won't rot.

References

Atkins, B.T. Sue & Rundell, Michael. 2008. **The Oxford Guide to Practical Lexicography.** OUP.

Coward, David F. & Grimes, Charles E. 2000. **Making Dictionaries,** SIL International Waxhaw. North Carolina.

Hanks, Patrick. 2008. Do word meanings exist? In Thierry Fotenelle. **Practical Lexicography. A Reader.** OUP, pp. 125-134

Kilgarriff, Adam. 2008. "I don't believe in word senses!" In Thierry Fotenelle. **Practical Lexicography. A Reader.** OUP, pp. 135-151