

# Print dictionaries for communities and online dictionaries for researchers: Daakaka and Dalkalaen

Kilu von Prince

10.07.2013



## Lexical data on Daakaka and Dalkalaen

- lexical database for Daakaka: ca. 2000 entries
- lexical database for Dalkalaen: ca. 1500 entries



## Lexical data on Daakaka and Dalkalaen

- lexical database for Daakaka: ca. 2000 entries
- lexical database for Dalkalaen: ca. 1500 entries
- published as print dictionaries for local use



## Lexical data on Daakaka and Dalkalaen

- lexical database for Daakaka: ca. 2000 entries
- lexical database for Dalkalaen: ca. 1500 entries
- published as print dictionaries for local use
- possible digital publication formats:
  - The Austronesian Basic Vocabulary database: <http://language.psy.auckland.ac.nz/austronesian/> (Daakaka already included)



## Lexical data on Daakaka and Dalkalaen

- lexical database for Daakaka: ca. 2000 entries
- lexical database for Dalkalaen: ca. 1500 entries
- published as print dictionaries for local use
- possible digital publication formats:
  - The Austronesian Basic Vocabulary database: <http://language.psy.auckland.ac.nz/austronesian/> (Daakaka already included)
  - as Toolbox or XML at The Language Archives



## Lexical data on Daakaka and Dalkalaen

- lexical database for Daakaka: ca. 2000 entries
- lexical database for Dalkalaen: ca. 1500 entries
- published as print dictionaries for local use
- possible digital publication formats:
  - The Austronesian Basic Vocabulary database: <http://language.psy.auckland.ac.nz/austronesian/> (Daakaka already included)
  - as Toolbox or XML at The Language Archives
  - in the Paradisec dictionary database of Vanuatu languages (<http://paradisec.org.au/vandicts/index.htm>)



## Lexical data on Daakaka and Dalkalaen

- lexical database for Daakaka: ca. 2000 entries
- lexical database for Dalkalaen: ca. 1500 entries
- published as print dictionaries for local use
- possible digital publication formats:
  - The Austronesian Basic Vocabulary database: <http://language.psy.auckland.ac.nz/austronesian/> (Daakaka already included)
  - as Toolbox or XML at The Language Archives
  - in the Paradisec dictionary database of Vanuatu languages (<http://paradisec.org.au/vandicts/index.htm>)
  - as a comparative database of Vanuatu languages (not likely to happen any time soon)



# Languages

- For local print dictionaries in Vanuatu:
  - Bislama
  - English
  - **scientific terms**
  - also desirable (but not implemented): French, neighboring vernaculars





# Languages

- For local print dictionaries in Vanuatu:
  - Bislama
  - English
  - **scientific terms**
  - also desirable (but not implemented): French, neighboring vernaculars
- For interlinked digital databases:
  - should all databases have one metalanguage in common?





# Related lexical information

## ① Part of speech



## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses



## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples



## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples
- 4 Variants



## Related lexical information

- ① Part of speech
- ② English definitions in addition to glosses
- ③ Examples
- ④ Variants
- ⑤ Paradigms, inflected forms



## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples
- 4 Variants
- 5 Paradigms, inflected forms
- 6 Derivations (reduplicated forms, nominalizations etc.)



## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples
- 4 Variants
- 5 Paradigms, inflected forms
- 6 Derivations (reduplicated forms, nominalizations etc.)
- 7 Collocations





## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples
- 4 Variants
- 5 Paradigms, inflected forms
- 6 Derivations (reduplicated forms, nominalizations etc.)
- 7 Collocations
- 8 Idioms



## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples
- 4 Variants
- 5 Paradigms, inflected forms
- 6 Derivations (reduplicated forms, nominalizations etc.)
- 7 Collocations
- 8 Idioms
- 9 Subentries



## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples
- 4 Variants
- 5 Paradigms, inflected forms
- 6 Derivations (reduplicated forms, nominalizations etc.)
- 7 Collocations
- 8 Idioms
- 9 Subentries
- 10 Genre, style, dialect, sociolect...



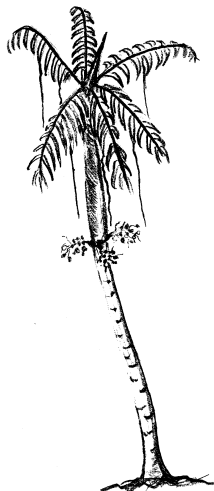
## Related lexical information

- 1 Part of speech
- 2 English definitions in addition to glosses
- 3 Examples
- 4 Variants
- 5 Paradigms, inflected forms
- 6 Derivations (reduplicated forms, nominalizations etc.)
- 7 Collocations
- 8 Idioms
- 9 Subentries
- 10 Genre, style, dialect, sociolect...
- 11 Phonetic information



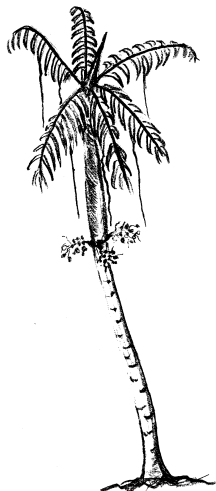
# Multimedia

- **drawings** (local artists, *A guide to the common trees of Vanuatu*, *Topological relations picture series*)



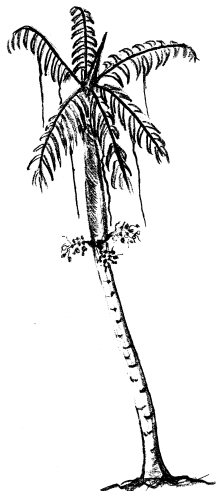
# Multimedia

- **drawings** (local artists, *A guide to the common trees of Vanuatu*, *Topological relations picture series*)
- **photos** (e.g. *Local Language Picture Dictionaries for Vanuatu*)



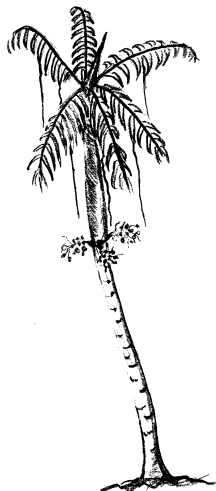
# Multimedia

- **drawings** (local artists, *A guide to the common trees of Vanuatu*, *Topological relations picture series*)
- **photos** (e.g. *Local Language Picture Dictionaries for Vanuatu*)
- **movies** (e.g. *Cut and break clips*)



# Multimedia

- **drawings** (local artists, *A guide to the common trees of Vanuatu*, *Topological relations picture series*)
- **photos** (e.g. *Local Language Picture Dictionaries for Vanuatu*)
- **movies** (e.g. *Cut and break clips*)
- recordings of the lexical entry, examples etc.







# Metadata

Not so relevant for print dictionaries, potentially relevant for databases, especially dynamic ones.

- source (recording, questionnaire, stimulus, text)





# Metadata

Not so relevant for print dictionaries, potentially relevant for databases, especially dynamic ones.

- source (recording, questionnaire, stimulus, text)
- entry author





# Metadata

Not so relevant for print dictionaries, potentially relevant for databases, especially dynamic ones.

- source (recording, questionnaire, stimulus, text)
- entry author
- entry date





# Metadata

Not so relevant for print dictionaries, potentially relevant for databases, especially dynamic ones.

- source (recording, questionnaire, stimulus, text)
- entry author
- entry date
- entry version...



# Censoring and purism

- Certain words were excluded from the print dictionaries at speakers' requests:



# Censoring and purism

- Certain words were excluded from the print dictionaries at speakers' requests:
  - loan words



# Censoring and purism

- Certain words were excluded from the print dictionaries at speakers' requests:
  - loan words
  - some swear words



# Censoring and purism

- Certain words were excluded from the print dictionaries at speakers' requests:
  - loan words
  - some swear words
  - English, French or Bislama words from code-switching contexts





# Censoring and purism

- Certain words were excluded from the print dictionaries at speakers' requests:
  - loan words
  - some swear words
  - English, French or Bislama words from code-switching contexts
- the first two should probably be included in (restricted) versions of a digital database, but what to do with words from code-switching?



## Direction of translation

- our print dictionaries are mono-directional, from Daakaka/Dalkalaen to Bislama to English



## Direction of translation

- our print dictionaries are mono-directional, from Daakaka/Dalkalaen to Bislama to English
- reverse lookup, especially from Bislama to Daakaka/Dalkalaen would be very desirable, but has not been implemented because there are no one-word glosses for most entries.



## Direction of translation

- our print dictionaries are mono-directional, from Daakaka/Dalkalaen to Bislama to English
- reverse lookup, especially from Bislama to Daakaka/Dalkalaen would be very desirable, but has not been implemented because there are no one-word glosses for most entries.
- this issue would also limit the usefulness of browsing and searching Bislama entries in a digital database.



# General ordering principles

- Alphabetic: depends on script/writing system

# General ordering principles

- Alphabetic: depends on script/writing system
- Thematic (only possible if entries are categorized thematically)



# General ordering principles

- Alphabetic: depends on script/writing system
- Thematic (only possible if entries are categorized thematically)
- Frequency (only possible in connection with a text database)



# General ordering principles

- Alphabetic: depends on script/writing system
- Thematic (only possible if entries are categorized thematically)
- Frequency (only possible in connection with a text database)
- ...





# Cross-references relevant for local print dictionaries

- semantic relations between lexemes of the same language:  
synonyms, antonyms

# Cross-references relevant for local print dictionaries

- semantic relations between lexemes of the same language:  
synonyms, antonyms
- co-occurring terms

# Cross-references relevant for local print dictionaries

- semantic relations between lexemes of the same language:  
synonyms, antonyms
- co-occurring terms
- morphologically related terms



# Cross-references relevant for local print dictionaries

- semantic relations between lexemes of the same language:  
synonyms, antonyms
- co-occurring terms
- morphologically related terms
- etymological relations to terms in neighboring vernaculars



# Cross-references relevant for interlinked databases

- all of the above

# Cross-references relevant for interlinked databases

- all of the above
- common languages

## Cross-references relevant for interlinked databases

- all of the above
- common languages
- illustrations/ video clips used as stimuli (only automatic if everyone uses the same reference); copyright?



## Cross-references relevant for interlinked databases

- all of the above
- common languages
- illustrations/ video clips used as stimuli (only automatic if everyone uses the same reference); copyright?
- commonly used questionnaires





## Cross-references relevant for interlinked databases

- all of the above
- common languages
- illustrations/ video clips used as stimuli (only automatic if everyone uses the same reference); copyright?
- commonly used questionnaires
- links to concordances



## Cross-references relevant for interlinked databases

- all of the above
- common languages
- illustrations/ video clips used as stimuli (only automatic if everyone uses the same reference); copyright?
- commonly used questionnaires
- links to concordances
- scientific names



## Cross-references relevant for interlinked databases

- all of the above
- common languages
- illustrations/ video clips used as stimuli (only automatic if everyone uses the same reference); copyright?
- commonly used questionnaires
- links to concordances
- scientific names
- etymological relations to terms in related languages (need to be added explicitly)



## Weighting information/ basis of identification

For greater reliability/recoverability, the basis of identification especially of species terms, color terms etc. should be made explicit.  
Example: zebra morey eel vs. banded sea snake



## References

- Bohnmeyer, Jürgen, Melissa Bowerman & Penelope Brown. 2001. Cut and break clips. In Stephen C. Levinson & N.J. Enfield (eds.), *Manual for the field season 2001*, 90-96. Nijmegen: Max Planck Institute for Psycholinguistics. <http://fieldmanuals.mpi.nl/volumes/2001/cut-and-break-clips/>
- Bowerman, Melissa & Eric Pederson. 1992. Topological relations picture series. In Stephen C. Levinson (ed.), *Space stimuli kit 1.2: November 1992*, 51. Nijmegen: Max Planck Institute for Psycholinguistics  
<http://fieldmanuals.mpi.nl/volumes/1992/bowped/>

