

The Daakaka Archive

Kilu von Prince

15th June 2016

Contents

1	Introduction	1
2	Annotated Recordings	2
2.1	Overview	2
2.2	Browsing <i>Annotated recordings</i>	5
2.2.1	Keywords	5
2.2.2	Actors	6
3	Contact	7

Acknowledgements

I thank the people of Emyotungan, Sesivi and Baiap, who let me stay with them, and who shared their food, space and their stories, insights and reflections. I hope that they and their children may find this archive accessible and useful. Also, this archive would not exist without the ingenuous efforts by Lena Karvovskaya, who worked as a student assistant in our DoBeS project; and Nicola Klingler, who worked as a student assistant in Manfred Krifka's department at Humboldt-Universität zu Berlin, and who helped me clean up the archive more recently. I also thank Manfred Krifka for allowing Nicola to do this on her job at his department. A heartfelt *thank you* also goes to the support at the MPI archives in Nijmegen, especially Alexander König, who has never failed to patiently address any number of technical catastrophies, which have been our constant companion during the development of this archive.

All annotations were done by myself, always with the help of one or more consultants, who are specified in the metadata. All short-comings and imperfections are my responsibility.

1 Introduction

The data archived here were collected during several field trips between 2009 and 2012 in the context of a DoBeS project led by Manfred Krifka. The language they document is Daakaka, which is spoken

by about one thousand speakers on the western corner of the island of Ambrym, Vanuatu. A detailed account of the language and the conditions of the data collection can be found in [von Prince \(2015\)](#).

This archive is far from perfect. There are many things I would like to add and change over time. However, updating and changing the archive has proven to be tremendously time-consuming and much of my time and energy has evaporated into bug reports. At this point, I feel that the archive can be a useful resource for speakers and for researchers from a variety of backgrounds. And while I still hope to do more work on the archive in the coming years, I decided that I should write up a documentation of its current state, rather than postpone this step until some indefinite time in the future when everything will be perfect.

The Daakaka archive contains the following parts, apart from this documentation:

1. DAA-dictionary.pdf: The illustrated Daakaka dictionary I compiled for local use on Ambrym. This dictionary can also be ordered as a book via Amazon for the production price. For more academic purposes, the dictionary data will be published in the [dictionaria series](#).
2. Annotated Recordings: This is by far the biggest node in the archives and contains individual session bundles with annotated recordings. This part of the archive is described in more detail in [section 2](#).
3. Written Texts: Here is a small collection of written texts that were submitted to a writing competition we organized in Baiap. They include scanned images of the originals and, in most cases, annotations in ELAN.
4. Video recording: This node is further split into the names of the places where the recordings took place: *Emyotungan* (spelled Emiotungan here) and *Sesivi*. Most of these recordings depict sand drawings or various steps of traditional kava production. They do not contain much spoken language, which is the main reason they come without annotations.
5. Deko: Here are a few pictures from Ambrym that are meant to be used for decoration on the [DoBeS portal](#).
6. Template: This is a metadata template that our student assistant Nicola Klingler and I have used.
7. Toolbox collection: For easy download of the annotations, this file contains most of the annotated recordings in the archive, plus a few that I haven't gotten around archiving yet.

2 Annotated Recordings

2.1 Overview

This node contains a total of 100 recording sessions. All bundles minimally contain a sound file and an ELAN file. Of those, 15 bundles are under the node labelled *Bundles with incomplete metadata*, which is quite self-explanatory. Among other things, they contain the sound files of several video recordings of sand drawings, which were done by my colleague Soraya Hosni. I have at this point no access to the video files or to the corresponding metadata.

For the rest of the bundles, there is metadata about the time and place of the recording, the actors involved, a brief description of the nature of the recording and a summary of the content (expand the *Content* window in the metadata view to see those).

Some bundles contain more than one ELAN-file, in which case one file only contains a transcriptoin and translation, while the other file also contains glosses and pos-tags. All bundles except for *Youth Service* contain glosses and pos-tags.

This gives you a rough total of 8:30 h and 50,000 words worth of fully transcribed, translated, glossed

and pos-annotated recordings (refer to table 2.1 for details).

Several bundles also come with a video file. Much to my chagrin, these are typically not played in the *Annex* view of the files. I have often tried to do something about this, but without much success so far. But you should be able to download them and play them in ELAN along with the annotations.

The names of the bundles are usually not very informative. There are several ways to explore the contents properties of the recordings:

- Clicking through bundle nodes and reading the descriptions to get an idea about their contents
- Browsing by metadata (see section 2.2).
- Content searches: You can search for specific words, glosses etc. with the Trova search tool. You may use the dictionary file and my grammar to search for specific Daakaka words or glosses, although the glosses in the archive are not entirely consistent, I'm afraid. Please also consult the archive's [manual](#) for more information on metadata and content searches.
- Table 2.1 gives an overview of all the bundles, with information on the duration, number of words and (main) speaker. For more information on speakers, see table 2. The column *Code* maps the recordings to the codes I am using for reference to the corpus in my published work.
- If you find the information in the archive insufficient for your specific research question, please get in touch, in case I might be able to help you further.

Table 1: Items under the node *Annotated recordings*

#	Name	Code	Speaker	Duration	Words
1	Abortion	rep11	FT	05:01	549
2	About a death	con01	DM	11:45	1342
3	Agriculture	exp19	FT	02:14	216
4	Aneityum and Tonga	sto43	DM	04:07	503
5	Animals1	exp50	BV	12:47	1432
6	Ant Story	sto01	FT	03:17	399
7	Ant trail	sto36	FT	03:31	403
8	Baking Chestnuts	sto02	BV	09:16	1006
9	Bamboo1	exp31	FT	00:36	65
10	Bamboo2	exp32	FT	01:14	195
11	Banyan Tree	exp13	FT	04:31	488
12	Bato	exp03	FT	01:52	229
13	Bekeli	sto05	FT	04:38	356
14	Biography	rep01	DM	05:24	304
15	Bush Nut	exp17	FT	06:17	563
16	Coconut	sto09	AK	01:22	201
17	Coconut and Kava	sto16	FT	08:49	853
18	Coconut Palm	exp09	FT	08:09	843
19	Copra	exp18	FT	03:00	392
20	Cottonwood	exp15	FT	05:39	569
21	Crab and Rat	sto10	AN	01:00	286
22	Crab and Swordfish	sto11	AQ	03:27	411
23	Dancing Nettle	exp11	FT	04:08	394
24	Demon	sto12	TB	07:13	821
25	Demon Pig	sto03	RB	04:54	923
26	Digging eggs	rep07	SD	00:17	47
27	Dog swallows cat	-	TS	00:17	103

Table 1: (continued)

#	Name	Code	Speaker	Duration	Words
28	Dove and Lisepsep	sto13	SD	05:52	719
29	Fish species	exp07	JI	21:17	2642
30	Flying Fox and Lorikeet	sto40	AK	01:05	147
31	Girl in Box	rep06	ED	00:53	100
32	Granddad and grandson	sto14	ME	03:33	350
33	Granny and grandson	sto15	BV	03:29	400
34	Imperial pigeon	exp23	FT	03:45	208
35	Insects	exp08	SD	14:32	1576
36	Kaingas	rep09	DM	06:33	579
37	Kingfisher	exp02	SD	14:38	1648
38	Kinship Relations	exp28	FT	03:46	425
39	Laa	exp10	FT	03:57	392
40	Lazarus	sto19	FT	09:31	864
41	Lewaa	exp14	FT	02:04	206
42	Liliwi	exp16	AU	05:04	319
43	Lisepsep at Malver	sto21	JS	09:33	1191
44	Lisepsep encounter	rep10	DM	01:46	191
45	Marriage Lineage	exp27	FT	05:37	634
46	Modern Times	con02	DM	16:08	2173
47	Mystery cave	rep15	BC	05:20	829
48	Name of Sesivi	rep08	ST	05:04	647
49	Names of Yam	-	BV	01:44	170
50	Origin of Kava	sto17	DM	03:35	423
51	Palpalmweli	rep12	FT	05:14	647
52	Pig Hunt	rep13	BT	00:42	110
53	Planting Yam	sto26	BP	02:11	173
54	Preparing Chicken	exp21	RT	00:43	111
55	Rail and Rat	sto31	FT	12:48	1306
56	Rail and Rat II	sto32	DM	05:25	488
57	Ranks	exp25	DM	03:02	315
58	Rat and Cat	sto27	BV	03:50	477
59	Rat and Crab	sto28	AM	01:47	186
60	Rat and Dove	sto29	RT	01:37	252
61	Rat and Kingfisher	sto35	JM	04:57	478
62	Repremalao	sto33	ST	14:58	1831
63	Revenge of the Chickens	sto34	FT	08:08	915
64	Roots of the lemeotir tree	sto22	ME	04:41	419
65	Sesivi and Lamap	rep14	BM	01:27	182
66	Simarongrong	sto38	BV	01:05	333
67	Snake Mother	sto25	FT	21:27	2276
68	Song of the Dancing song	-	FT	04:08	390
69	Song of the Owl	sto20	BV	02:39	174
70	Spear Throwing	exp29	FT	06:54	767
71	Tamadu	sto41	FT	07:07	1033
72	Tamadu Sand Drawing	exp30	FT	03:47	266
73	Taro Laplap	exp51	FT	03:45	242

Table 1: (continued)

#	Name	Code	Speaker	Duration	Words
74	Ten Fish	exp24	DM	01:32	132
75	The talking dove	sto42	FT	04:39	412
76	Three Birds	sto04	FT	05:00	562
77	Tired of Baiap	rep02	BP	02:44	289
78	Tonga and Aneityum	sto44	DM	04:31	487
79	Treefern	exp12	FT	06:15	755
80	Twins	exp42	BB	02:59	175
81	Two Friends	rep03	FT	06:48	768
82	Volcano	sto24	FT	14:03	1569
83	Wano Stone	sto48	FT	05:49	244
84	Youth Service	–	SD	06:25	671
Bundles with incomplete metadata					
85	Duan myató	exp06	DM	05:08	434
86	Love conquers death	sto47	ME	07:27	1024
87	Why chickens look up while drinking	sto45	RT	01:11	180
88	Tis ane aatawa			01:39	108
89	Tis ane apyal’o san longlong			01:55	39
90	Tis ane awes			04:39	59
91	Tis ane dom			01:46	36
92	Tis ane Lewaa ne vis			05:15	192
93	Tis ane Liliwi san eya			03:00	72
94	Tis ane Meap			02:59	172
95	Tis ane Syewuoswuos			01:36	22
96	Tis ane Tamadu			02:55	128
97	Tis ane Tan me nok			05:39	113
98	Tis ane tangtanglii			05:35	68
99	Tis ane vevyos myane yewotopkee			02:36	83
100	Tis ne gaakaa			03:32	55
Total				8:33:30	51,279

2.2 Browsing *Annotated recordings*

You can search the Metadata in this archive to find the files most relevant to your research. The problem is that you have to know what to look for in the first place. This section is meant to give you an overview of the categories and labels that I think may be the most relevant for your search.

Those categories mostly fall under the domains of *Keywords* and properties of *Actors*. They are described in the following two sections. There are also links to example searches.

2.2.1 Keywords

I have supplied some of the recordings with one or more keywords in order to help users find relevant content and browse the archive. These labels are very coarse, not very consistent and pre-theoretical.

They are meant as simple tools for helping users, not a result of careful analysis. Also, the same values may apply to different key names. This is not a purposefully designed feature, but an artefact of the difficulties working with Arbil (the metadata editor) and Lamus (the archiving platform) plus a lack of time to clean things up.

For example, the keyword *Myth* is a value both of the key *Genre* and of the key *Standard Template Key*. In the category *Genre*, it matches 6 files, and in the category *Standard Template Key*, it matches 25 different files.

As a result, I advise the user to use keywords only in simple searches, where they do not have to specify a key name. A simple search for the term *Myth* returns 32 hits, as you can see at [this link](#).

A list of keywords used in this archive is given below, with the number of matching bundles in brackets:

1. *Animals* (27): Recordings (mostly fictional stories), in which animals feature prominently
2. *History* (13): Relating biographical information, or historical explanations for contemporary facts. See also the corresponding value in the *Genre* key above.
3. *Instruction* (4): Instructions for practical procedures, such as preparing chicken.
4. *Kastom* (23): Relating to traditional rituals.
5. *Lisepep* (7): Stories involving the magical, dwarf-like creatures called *lisepep*.
6. *Magic* (11): About supernatural phenomena.
7. *Myth* (32): Relating content which is either not thought to be literally true, relates to mythical times, or involves extraordinary encounters with the supernatural (see also *Genre:myth*)
8. *Nature* (18): Relating to plants, animals, or geological formations.
9. *Plants* (14): Relating to flora.
10. *Sand drawing* (27): Descriptions of or stories about sand drawings
11. *Song* (14): Recordings of songs, or of stories containing songs.
12. *Tradition* (7): About certain traditions, see also *Kastom*.

2.2.2 Actors

Local actors who acted as speakers, writers and/ or consultants have anonymized Codes. I have not been able to determine their exact age in all cases, but there is an age range that I supplied for each actor at the time of the recording in table 2. The metadata in the archive also information about their gender and place of origin. This means that if you are interested only in recordings from female speakers, older speakers, or speakers from a certain region, you can specify this in your metadata search.

An overview of local actors and their properties can be seen in table 2. An example search for recordings involving female actors can be seen [here](#).¹

References

von Prince, Kilu. 2015. *A grammar of Daakaka*. Berlin, Boston: De Gruyter Mouton.

¹The reasons for the relatively low number of female speakers are explained in [von Prince \(2015\)](#).

Table 2: Anonymous actors who acted as speakers and/or consultants

Code	Sex	Age	Home
AD	F	<15	Baiap
AK	M	45-60	Sesivi
AN	F	15-25	Emyotungan
AQ	F	35-45	Sesivi
AU	M	25-35	Sesivi
BP	M	>60	Baiap
BT	M	>15	Emyotungan
BV	M	35-45	Emyotungan
DB	M	25-35	Emyotungan
DM	M	>60	Sesivi
ED	M	25-35	Leelee
FT	M	45-60	Emyotungan
JM	M	15-25	Emyotungan
JS	M	15-25	Malver
GR	M	15-25	Emyotungan
HM	M	15-25	Sesivi
ME	M	>60	Baiap
NO	M	>60	Sesivi
RT	F	15-25	Emyotungan
SD	M	45-60	Emyotungan
ST	M	>60	Sesivi
TB	M	35-45	Sesivi
TS	M	15-25	Emyotungan

3 Contact

You can contact me about all things Daakaka and West Ambrym with the following information:

Office:

Kilu von Prince
Dorotheenstr. 24
Raum 3.302
+49 (0)30 2093-9755

Postal addresss:

Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
kilu.von.prince@hu-berlin.de