

Vera'a

— metadata sheet —

June 2016

Stefan Schnell
University of Melbourne

cite this document as:

Schnell, Stefan. 2016.
Multi-CAST Vera'a metadata sheet.
In Haig, Geoffrey & Schnell, Stefan (eds.),
Multi-CAST (Multilingual Corpus of Annotated Spoken Texts).
(handle) (date accessed.)

Citing the corpus

The Multi-CAST Vera'a corpus should be cited as follows:

Schnell, Stefan. 2016. Vera'a.
 In Haig, Geoffrey & Schnell, Stefan (eds.),
Multi-CAST (Multilingual Corpus of Annotated Spoken Texts).
 (<https://lac.uni-koeln.de/en/multicast-veraa/>)
 (date accessed.)

Language data

language	Vera'a (ISO 639-3: vra)
affiliation	Austronesian, Malayo-Polynesian, Oceanic, Vanuatu
area spoken	Vanuatu, Banks Islands, Vanua Lava
varieties recorded	Vera'a village
text type	traditional narratives
sources	Schnell 2010, 2011, 2016

Corpus data

file name	speaker	date rec'd	text type	length h:mm:ss	clause units
<i>veraa_ano</i>	VR01	2007	traditional	6:07	185
<i>veraa_as1</i>	VR02	2007	traditional	5:16	213
<i>veraa_gabg</i>	VR03	2007	traditional	8:41	174
<i>veraa_gaqg</i>	VR04	2007	traditional	8:52	226
<i>veraa_hhak</i>	VR05	2007	traditional	12:39	431
<i>veraa_isam</i>	VR06	2007	traditional	7:21	238
<i>veraa_iswm</i>	VR07	2007	traditional	21:43	576
<i>veraa_jjq</i>	VR08	2007	traditional	30:19	879
<i>veraa_mvbw</i>	VR09	2007	traditional	10:07	305
<i>veraa_palaa</i>	VR10	2007	traditional	4:02	140
<i>veraa_palab</i>	VR10	2007	traditional	6:41	239
<i>totals for the Vera'a corpus</i>				2:01:48	3,606

Table 1. List of files in the Vera'a corpus.

Background to the recordings

The Vera'a language has around 450 speakers, 250 of which live in the village of Vera'a, and the other 200 being scattered along the coast line reaching from Vera'a up to the northern shore. The language was initially researched

speaker	gender	age	date born	date rec'd	clause units	texts
VR01	female	~20	~1985	2007	185	<i>veraa_anv</i>
VR02	male	~40	~1965	2007	213	<i>veraa_as1</i>
VR03	male	~40	~1965	2007	174	<i>veraa_gabg</i>
VR04	male	~40	~1965	2007	226	<i>veraa_gaqq</i>
VR05	male	~20	~1985	2007	431	<i>veraa_jhak</i>
VR06	male	~60	~1950	2007	238	<i>veraa_isam</i>
VR07	male	~60	~1950	2007	576	<i>veraa_iswm</i>
VR08	male	~60	~1950	2007	879	<i>veraa_jjq</i>
VR09	male	~30	~1975	2007	305	<i>veraa_mobw</i>
VR10	female	~40	~1965	2007	379	<i>veraa_pala</i>

Table 2. List of Vera'a corpus speakers.

by Catriona Hyslop in the late 1990s and Alexander Francois in the early 2000s (various publications by Francois deal with Vera'a). Since 2007, the language has been extensively documented by Stefan Schnell (SS), first as part of a documentation project funded within the VolkswagenStiftung's DoBeS language documentation programme, and since 2012 as part of Schnell's ASC-funded project on argument realisation in discourse of diverse languages.

To date, several hours of video and audio recordings of speech events have been collected by Schnell and other researchers.¹ A large proportion of the data has been transcribed, hand-written transcriptions being undertaken by native speakers, and later entered into ELAN by SS, together with a translation into English. Some of the recorded narratives have later been edited by a speaker of Vera'a, Makson Vores, and published as a book (Vores et al. 2012).

All speech events recorded took place in the Vera'a community on Vanua Lava. Most of these comprise 'staged events' in the sense of Himmelmann (1998), that is they took place mainly for the sake of being recorded as part of the documentation project. Recording of public events are also included in the Vera'a documentation, and these do not constitute staged events in the strict sense, though speakers were at all times informed about their being recorded. Nevertheless, the Vera'a corpus can be regarded as comprising a large set of fairly natural speech data recorded within the indigenous cultural setting of the speech community. While most recordings were made by SS, Makson Vores collected several narratives in 2012 and 2013.

In addition to narratives and public events, procedural and descriptive texts were recorded. The latter comprise descriptions of plant and fish species. Both of these types of descriptions were recorded in dedicated sessions focussing on ethno-biological aspects of the Vera'a language. In both sessions, speakers were asked to describe the respective plant or fish species, with a

¹ see project website for team members at <http://dobes.mpi.nl/projects/vures-veraa/>

part of the plant or a picture card of the fish in front of them or in their hands.

The Vera'a sub-corpus of Multi-CAST constitutes a relatively small portion of the entire Vera'a corpus. In addition to narratives, some of the plant and fish descriptions have been GRAID'ed and will be added to the corpus in the future, as will edited narratives in order to enable research into medium-related variation in argument expression.

References

- Himmelman, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36(2). 161–195.
- Schnell, Stefan. 2010. *Animacy and referentiality in Vera'a*. Kiel: Kiel University. (Ph.D. dissertation.)
- Schnell, Stefan. 2011. *A grammar of Vera'a*. Kiel: University of Kiel. (Ph.D. dissertation.) (https://www.academia.edu/2317752/Schnell_2011_A_grammar_of_Veraa_an_Oceanic_language_of_North_Vanuatu) (accessed 2016-02-22.)
- Schnell, Stefan. 2016. Multi-CAST Vera'a annotation notes. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST (Multilingual Corpus of Annotated Spoken Texts)*. (<https://lac.uni-koeln.de/multicast-veraa/>) (accessed 2016-05-14.)
- Vores, Makson & Schnell, Stefan & Atkins, George B. & Malau, Armstrong & Malau, Catriona. 2012. *Stories from the village of Vera'a (Vanua Lava, TorBa)*. Port Vila: Spite Design.

Corpus changelog

date	changes	affected texts		
2015-05	added 11 corpus files	<i>veraa_*</i>		
		<i>anv,</i>	<i>as1,</i>	<i>gabg,</i>
		<i>gaqg,</i>	<i>hhak,</i>	<i>isam,</i>
		<i>iswm,</i>	<i>jjq,</i>	<i>mɔbw,</i>
		<i>palaa-b</i>		