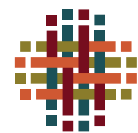


Multi-CAST

Vera'a
corpus counts

Stefan Schnell
University of Bamberg

May 2019
v2.0



ARC CENTRE OF EXCELLENCE FOR
THE DYNAMICS OF LANGUAGE



Australian Government
Australian Research Council



University of Bamberg

DFG

Multi-CAST

*Multilingual Corpus of
Annotated Spoken Texts*

Citation for this document

Schnell, Stefan. 2019. Multi-CAST Vera'a corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/) (date accessed)

Citation for the Multi-CAST collection

Haig, Geoffrey & Schnell, Stefan (eds.). 2015. *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/) (date accessed)

The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.

The entirety of Multi-CAST, including this document, is published under the *Creative Commons Attribution 4.0 International Licence* (CC BY 4.0), unless noted otherwise. The licensing terms can be reviewed online at creativecommons.org/licenses/by/4.0/.

Multi-CAST Vera'a corpus counts v2.0 last updated 8 May 2019
This document was typeset by NNS with X_YL^AT_EX and the *multicast3* class (v3.0.9045).

Contents

1	Notes on the GRAID counts	<hr/>	1
2	The Vera'a corpus	<hr/>	2
2.1	<i>anv</i>		3
2.2	<i>as1</i>		4
2.3	<i>gabg</i>		5
2.4	<i>gaqg</i>		6
2.5	<i>hhak</i>		7
2.6	<i>isam</i>		8
2.7	<i>iswm</i>		9
2.8	<i>jjq</i>		10
2.9	<i>mvbw</i>		11
2.10	<i>pala</i>		12

1 Notes on the GRAID counts

This document collects tables with frequency counts for combinations of selected GRAID symbols in version 1905 (from May 2019) of the Multi-CAST Vera'a corpus. The tables are intended to offer cursory impressions of the relative proportions between different types of referring expression; they do not provide exact summaries of the annotations.

Only a small number of basic GRAID symbols are counted:

Function symbols

⟨∅⟩	zero
⟨pro⟩	definite pronoun
⟨np⟩	full noun phrase
⟨other⟩	form not further specified

Person/Animacy symbols

⟨.1⟩	first person
⟨.2⟩	second person
⟨.h⟩	third person, human
⟨.d⟩	third person, anthropomorphic
∅	third person, non-human

Function symbols

⟨:a⟩	subject of a transitive clause
⟨:s⟩	subject of an intransitive clause
⟨:ncs⟩	non-canonical subject
⟨:p⟩	direct object
⟨:ob1⟩	oblique argument
⟨:g⟩	goal argument
⟨:l⟩	locational argument
⟨:poss⟩	possessive
⟨:pred⟩	predicate
⟨:other⟩	function not further specified

Clause boundary symbols

⟨##⟩	independent clause
⟨#⟩	other clause

Only basic categories are listed; categories represented by complex symbols with additional specifiers (e.g. ⟨dem_pro⟩ ‘demonstrative pronoun’) have been subsumed under the more basic category (e.g. ⟨pro⟩ ‘definite pronoun’). Please refer to the annotation notes for this corpus for information on all annotated categories, including those not listed here.

The tables in this document can be recreated with the `mc_table` function from *multicastR*, the companion package to Multi-CAST for the statistical computing language R. Please refer to the package documentation (`?multicastR`, `?mc_table`) for more information.

2 The Vera'a corpus

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	3	8	0	2	0	0	0	0	0	0	13
<∅ .2>	13	29	0	1	0	0	0	0	0	0	43
<∅ .h>	181	367	0	20	10	1	0	0	0	0	579
<∅ .d>	86	134	0	14	1	0	0	0	0	0	235
<∅>	17	79	0	155	14	8	3	0	0	0	276
<pro .1>	97	230	0	38	11	6	0	66	3	1	452
<pro .2>	65	124	0	38	5	8	0	41	0	1	282
<pro .h>	350	705	0	117	11	54	0	281	0	0	1518
<pro .d>	38	125	0	15	0	13	0	23	0	0	214
<pro>	9	64	0	3	0	1	0	26	5	1	109
<np .h>	62	261	0	87	11	45	0	53	47	4	570
<np .d>	27	109	0	24	5	21	0	11	12	0	209
<np>	26	175	0	478	68	245	100	4	89	204	1389
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	5	0	0	12	59	111	0	284	0	471
<i>totals</i>	974	2415	0	992	148	461	214	505	440	211	
<##>											3201
<#>											407
<i>totals</i>											3608

Table 1 Summarized GRAID counts for the entire Vera'a corpus.

2.1 *anv*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	1	0	0	0	0	0	0	0	0	1
<∅ .2>	1	1	0	0	0	0	0	0	0	0	2
<∅ .h>	25	28	0	3	0	0	0	0	0	0	56
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	4	9	0	10	0	0	0	0	0	0	23
<pro .1>	4	7	0	1	0	0	0	1	0	0	13
<pro .2>	1	1	0	1	0	0	0	0	0	0	3
<pro .h>	15	28	0	15	1	4	0	25	0	0	88
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	1	0	0	0	0	0	0	0	0	3
<np .h>	12	28	0	6	1	4	0	6	2	0	59
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	2	5	0	30	1	10	1	0	3	1	53
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	1	0	0	0	9	4	0	10	0	24
<i>totals</i>	66	110	0	66	3	27	5	32	15	1	
<##>											172
<#>											10
<i>totals</i>											182

Table 2 Summarized GRAID counts for the *anv* text.

2.2 *as1*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	4	3	0	0	0	0	0	0	0	0	7
<∅ .h>	10	23	0	0	0	0	0	0	0	0	33
<∅ .d>	16	15	0	0	0	0	0	0	0	0	31
<∅>	2	3	0	17	0	0	0	0	0	0	22
<pro .1>	1	13	0	2	1	0	0	8	0	0	25
<pro .2>	2	6	0	1	0	0	0	2	0	0	11
<pro .h>	18	31	0	0	2	6	0	10	0	0	67
<pro .d>	2	5	0	0	0	0	0	0	0	0	7
<pro>	1	4	0	0	0	0	0	1	0	0	6
<np .h>	4	9	0	0	1	4	0	1	1	0	20
<np .d>	2	9	0	1	0	2	0	0	1	0	15
<np>	2	8	0	44	2	11	2	0	5	2	76
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	5	4	0	33	0	42
<i>totals</i>	64	129	0	65	6	28	6	22	40	2	
<##>											204
<#>											9
<i>totals</i>											213

Table 3 Summarized GRAID counts for the *as1* text.

2.3 gabg

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	0	0	0	1	0	0	0	0	0	0	1
<∅ .h>	0	0	0	0	0	0	0	0	0	0	0
<∅ .d>	8	26	0	0	0	0	0	0	0	0	34
<∅>	0	5	0	8	0	0	0	0	0	0	13
<pro .1>	11	22	0	2	0	2	0	3	0	0	40
<pro .2>	4	11	0	3	0	1	0	4	0	0	23
<pro .h>	0	0	0	0	0	0	0	0	0	0	0
<pro .d>	6	28	0	2	0	6	0	2	0	0	44
<pro>	0	1	0	0	0	0	0	0	1	0	2
<np .h>	0	0	0	0	0	0	0	0	0	0	0
<np .d>	2	25	0	3	0	10	0	0	1	0	41
<np>	2	12	0	14	3	16	4	0	10	3	64
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	1	0	0	1	7	8	0	6	0	23
<i>totals</i>	33	131	0	33	4	42	12	9	18	3	
<##>											157
<#>											17
<i>totals</i>											174

Table 4 Summarized GRAID counts for the gabg text.

2.4 *gaqg*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	5	6	0	0	0	0	0	0	0	0	11
<∅ .d>	10	24	0	7	1	0	0	0	0	0	42
<∅>	2	11	0	4	0	0	0	0	0	0	17
<pro .1>	9	16	0	2	1	0	0	3	0	0	31
<pro .2>	2	7	0	7	0	2	0	1	0	0	19
<pro .h>	13	14	0	0	0	0	0	1	0	0	28
<pro .d>	10	40	0	6	0	5	0	8	0	0	69
<pro>	1	8	0	1	0	0	0	1	1	0	12
<np .h>	1	2	0	1	0	2	0	0	0	0	6
<np .d>	3	19	0	6	0	5	0	2	3	0	38
<np>	1	8	0	28	5	22	9	0	2	8	83
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	1	3	11	0	13	0	28
<i>totals</i>	57	155	0	62	8	39	20	16	19	8	
<##>											186
<#>											40
<i>totals</i>											226

Table 5 Summarized GRAID counts for the *gaqg* text.

2.5 *hhak*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	1	0	0	0	0	0	0	0	0	1
<∅ .2>	0	6	0	0	0	0	0	0	0	0	6
<∅ .h>	35	78	0	1	0	1	0	0	0	0	115
<∅ .d>	4	11	0	6	0	0	0	0	0	0	21
<∅>	0	2	0	10	0	0	0	0	0	0	12
<pro .1>	15	13	0	9	2	1	0	10	0	0	50
<pro .2>	16	8	0	6	1	1	0	7	0	0	39
<pro .h>	26	91	0	7	0	10	0	27	0	0	161
<pro .d>	7	13	0	3	0	1	0	0	0	0	24
<pro>	0	4	0	0	0	0	0	0	0	0	4
<np .h>	6	35	0	13	1	6	0	0	1	0	62
<np .d>	5	6	0	7	2	0	0	3	2	0	25
<np>	4	13	0	58	9	21	11	0	7	9	132
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	1	10	5	0	52	0	68
<i>totals</i>	118	281	0	120	16	51	16	47	62	9	
<##>											417
<#>											15
<i>totals</i>											432

Table 6 Summarized GRAID counts for the *hhak* text.

2.6 *isam*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	0	2	0	0	0	0	0	0	0	0	2
<∅ .h>	2	12	0	1	1	0	0	0	0	0	16
<∅ .d>	1	7	0	0	0	0	0	0	0	0	8
<∅>	1	2	0	11	1	0	0	0	0	0	15
<pro .1>	1	18	0	1	0	0	0	2	0	0	22
<pro .2>	5	23	0	2	1	1	0	2	0	0	34
<pro .h>	31	70	0	9	0	5	0	9	0	0	124
<pro .d>	2	13	0	1	0	0	0	3	0	0	19
<pro>	0	8	0	1	0	0	0	0	0	0	9
<np .h>	0	2	0	1	0	2	0	0	4	0	9
<np .d>	1	10	0	0	0	2	0	0	1	0	14
<np>	1	5	0	19	5	23	5	0	3	46	107
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	1	5	10	0	11	0	27
<i>totals</i>	45	172	0	46	9	38	15	16	19	46	
<##>											212
<#>											26
<i>totals</i>											238

Table 7 Summarized GRAID counts for the *isam* text.

2.7 *iswm*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	2	0	0	2	0	0	0	0	0	0	4
<∅ .2>	1	4	0	0	0	0	0	0	0	0	5
<∅ .h>	26	59	0	7	9	0	0	0	0	0	101
<∅ .d>	1	0	0	0	0	0	0	0	0	0	1
<∅>	0	8	0	25	3	3	0	0	0	0	39
<pro .1>	10	19	0	6	2	1	0	5	0	0	43
<pro .2>	11	6	0	2	0	1	0	8	0	1	29
<pro .h>	105	185	0	42	6	8	0	73	0	0	419
<pro .d>	0	1	0	0	0	0	0	0	0	0	1
<pro>	3	18	0	0	0	0	0	11	3	1	36
<np .h>	16	39	0	27	1	10	0	23	15	3	134
<np .d>	0	1	0	0	0	0	0	0	0	0	1
<np>	4	37	0	73	29	47	26	0	11	66	293
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	1	5	23	0	29	0	58
<i>totals</i>	179	377	0	184	51	75	49	120	58	71	
<##>											488
<#>											88
<i>totals</i>											576

Table 8 Summarized GRAID counts for the *iswm* text.

2.8 *jjq*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	1	6	0	0	0	0	0	0	0	0	7
<∅ .2>	2	9	0	0	0	0	0	0	0	0	11
<∅ .h>	24	63	0	3	0	0	0	0	0	0	90
<∅ .d>	33	38	0	0	0	0	0	0	0	0	71
<∅>	3	31	0	42	7	5	3	0	0	0	91
<pro .1>	30	83	0	9	2	0	0	21	0	0	145
<pro .2>	15	36	0	8	2	1	0	7	0	0	69
<pro .h>	58	151	0	27	1	7	0	56	0	0	300
<pro .d>	8	18	0	3	0	1	0	6	0	0	36
<pro>	2	16	0	1	0	1	0	11	0	0	31
<np .h>	14	72	0	8	1	7	0	16	14	1	133
<np .d>	8	27	0	4	2	0	0	3	1	0	45
<np>	2	62	0	98	11	62	33	3	39	44	354
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	3	0	0	5	13	40	0	84	0	145
<i>totals</i>	200	615	0	203	31	97	76	123	138	45	
<##>											820
<#>											60
<i>totals</i>											880

Table 9 Summarized GRAID counts for the *jjq* text.

2.9 *mvbw*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	0	1	0	0	0	0	0	0	0	0	1
<∅ .h>	33	41	0	3	0	0	0	0	0	0	77
<∅ .d>	0	4	0	0	0	0	0	0	0	0	4
<∅>	1	1	0	16	0	0	0	0	0	0	18
<pro .1>	9	24	0	3	3	2	0	10	0	1	52
<pro .2>	4	16	0	6	0	0	0	7	0	0	33
<pro .h>	34	56	0	10	1	9	0	58	0	0	168
<pro .d>	2	2	0	0	0	0	0	0	0	0	4
<pro>	0	3	0	0	0	0	0	1	0	0	4
<np .h>	3	33	0	18	3	5	0	5	6	0	73
<np .d>	0	4	0	0	1	1	0	0	0	0	6
<np>	6	13	0	38	2	15	5	1	6	19	105
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	1	1	5	0	28	0	35
<i>totals</i>	92	198	0	94	11	33	10	82	40	20	
<##>											189
<#>											118
<i>totals</i>											307

Table 10 Summarized GRAID counts for the *mvbw* text.

2.10 *pala*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	5	3	0	0	0	0	0	0	0	0	8
<∅ .h>	21	57	0	2	0	0	0	0	0	0	80
<∅ .d>	13	9	0	1	0	0	0	0	0	0	23
<∅>	4	7	0	12	3	0	0	0	0	0	26
<pro .1>	7	15	0	3	0	0	0	3	3	0	31
<pro .2>	5	10	0	2	1	1	0	3	0	0	22
<pro .h>	50	79	0	7	0	5	0	22	0	0	163
<pro .d>	1	5	0	0	0	0	0	4	0	0	10
<pro>	0	1	0	0	0	0	0	1	0	0	2
<np .h>	6	41	0	13	3	5	0	2	4	0	74
<np .d>	6	8	0	3	0	1	0	3	3	0	24
<np>	2	12	0	76	1	18	4	0	3	6	122
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	1	1	1	0	18	0	21
<i>totals</i>	120	247	0	119	9	31	5	38	31	6	
<##>											356
<#>											24
<i>totals</i>											380

Table 11 Summarized GRAID counts for the *pala* text.

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/

