

Multi-CAST

Vera'a
corpus counts

Stefan Schnell

September 2021
v2.3



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. 2021. 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/#veraa) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 licence can be reviewed online at creativecommons.org/licenses/by/4.0/.

Multi-CAST Vera'a corpus counts v2.3 last updated 13 September 2021
This document was typeset by NNS with X_YL^AT_EX and the *multicast3* class (v3.2.5).

Contents

1	Notes on the GRAID counts	1
2	The Vera'a corpus	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 2108 (from August 2021) of the Multi-CAST Vera'a corpus. Unless a more recent version of this document exists, it also applies to any later versions of the annotations. Note that the tables are intended to offer only 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

⟨0⟩	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

⟨:s⟩	subject of an intransitive clause
⟨:a⟩	subject of a transitive clause
⟨:ncs⟩	non-canonical subject
⟨:p⟩	direct object
⟨:ob1⟩	oblique argument
⟨:g⟩	goal argument
⟨:l⟩	locational argument
⟨:pred⟩	predicate
⟨:poss⟩	possessive
⟨: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.

2 The Vera'a corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	9	3	0	2	0	0	0	0	0	0	14
<∅ .2>	29	13	0	1	0	0	0	0	0	0	43
<∅ .h>	369	181	0	20	10	1	0	0	0	0	581
<∅ .d>	134	86	0	14	1	0	0	0	0	0	235
<∅>	79	17	0	154	14	8	3	0	0	0	275
<pro .1>	231	97	0	38	11	6	0	3	69	1	456
<pro .2>	124	65	0	38	5	8	0	0	41	1	282
<pro .h>	693	345	0	116	11	54	0	0	278	0	1497
<pro .d>	132	43	0	16	0	13	0	0	23	0	227
<pro>	64	9	0	4	0	1	0	5	26	1	110
<np .h>	260	62	0	87	11	45	0	48	53	5	571
<np .d>	110	27	0	24	5	21	0	12	11	0	210
<np>	175	26	0	478	68	245	100	90	4	204	1390
<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>	5	0	0	0	12	59	111	283	0	1922	2392
<i>totals</i>	2414	974	0	992	148	461	214	441	505	2134	
<##>											3201
<#>											407
<i>totals</i>											3608

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

2.1 *anv*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	0	0	0	0	0	0	0	0	0	1
<∅ .2>	1	1	0	0	0	0	0	0	0	0	2
<∅ .h>	28	25	0	3	0	0	0	0	0	0	56
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	9	4	0	9	0	0	0	0	0	0	22
<pro .1>	7	4	0	1	0	0	0	0	1	0	13
<pro .2>	1	1	0	1	0	0	0	0	0	0	3
<pro .h>	28	15	0	15	1	4	0	0	25	0	88
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	2	0	1	0	0	0	0	0	0	4
<np .h>	28	12	0	6	1	4	0	2	6	0	59
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	5	2	0	30	1	10	1	3	0	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>	1	0	0	0	0	9	4	10	0	101	125
<i>totals</i>	110	66	0	66	3	27	5	15	32	102	
<##>											172
<#>											10
<i>totals</i>											182

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

2.2 *as1*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	3	4	0	0	0	0	0	0	0	0	7
<∅ .h>	23	10	0	0	0	0	0	0	0	0	33
<∅ .d>	15	16	0	0	0	0	0	0	0	0	31
<∅>	3	2	0	17	0	0	0	0	0	0	22
<pro .1>	13	1	0	2	1	0	0	0	8	0	25
<pro .2>	6	2	0	1	0	0	0	0	2	0	11
<pro .h>	31	18	0	0	2	6	0	0	10	0	67
<pro .d>	5	2	0	0	0	0	0	0	0	0	7
<pro>	4	1	0	0	0	0	0	0	1	0	6
<np .h>	9	4	0	0	1	4	0	1	1	0	20
<np .d>	9	2	0	1	0	2	0	1	0	0	15
<np>	8	2	0	44	2	11	2	5	0	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	33	0	148	190
<i>totals</i>	129	64	0	65	6	28	6	40	22	150	
<##>											204
<#>											9
<i>totals</i>											213

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

2.3 gabg

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<: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>	26	8	0	0	0	0	0	0	0	0	34
<∅>	5	0	0	8	0	0	0	0	0	0	13
<pro .1>	22	11	0	2	0	2	0	0	3	0	40
<pro .2>	11	4	0	3	0	1	0	0	4	0	23
<pro .h>	0	0	0	0	0	0	0	0	0	0	0
<pro .d>	28	6	0	2	0	6	0	0	2	0	44
<pro>	1	0	0	0	0	0	0	1	0	0	2
<np .h>	0	0	0	0	0	0	0	0	0	0	0
<np .d>	25	2	0	3	0	10	0	1	0	0	41
<np>	12	2	0	14	3	16	4	10	0	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>	1	0	0	0	1	7	8	6	0	96	119
<i>totals</i>	131	33	0	33	4	42	12	18	9	99	
<##>											157
<#>											17
<i>totals</i>											174

Table 4 Summarized GRAID counts for the *gabg* text.

2.4 *gaqg*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<: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>	6	5	0	0	0	0	0	0	0	0	11
<∅ .d>	24	10	0	7	1	0	0	0	0	0	42
<∅>	11	2	0	4	0	0	0	0	0	0	17
<pro .1>	16	9	0	2	1	0	0	0	3	0	31
<pro .2>	7	2	0	7	0	2	0	0	1	0	19
<pro .h>	14	13	0	0	0	0	0	0	1	0	28
<pro .d>	40	10	0	6	0	5	0	0	8	0	69
<pro>	8	1	0	1	0	0	0	1	1	0	12
<np .h>	2	1	0	1	0	2	0	0	0	0	6
<np .d>	19	3	0	6	0	5	0	3	2	0	38
<np>	8	1	0	28	5	22	9	2	0	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	13	0	108	136
<i>totals</i>	155	57	0	62	8	39	20	19	16	116	
<##>											186
<#>											40
<i>totals</i>											226

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

2.5 *hhak*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	0	0	0	0	0	0	0	0	0	1
<∅ .2>	6	0	0	0	0	0	0	0	0	0	6
<∅ .h>	78	35	0	1	0	1	0	0	0	0	115
<∅ .d>	11	4	0	6	0	0	0	0	0	0	21
<∅>	2	0	0	10	0	0	0	0	0	0	12
<pro .1>	13	15	0	9	2	1	0	0	10	0	50
<pro .2>	8	16	0	6	1	1	0	0	7	0	39
<pro .h>	91	26	0	7	0	10	0	0	27	0	161
<pro .d>	13	7	0	3	0	1	0	0	0	0	24
<pro>	4	0	0	0	0	0	0	0	0	0	4
<np .h>	35	6	0	13	1	6	0	1	0	0	62
<np .d>	6	5	0	7	2	0	0	2	3	0	25
<np>	13	4	0	58	9	21	11	7	0	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	52	0	156	224
<i>totals</i>	281	118	0	120	16	51	16	62	47	165	
<##>											417
<#>											15
<i>totals</i>											432

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

2.6 *isam*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	2	0	0	0	0	0	0	0	0	0	2
<∅ .h>	14	2	0	1	1	0	0	0	0	0	18
<∅ .d>	7	1	0	0	0	0	0	0	0	0	8
<∅>	2	1	0	11	1	0	0	0	0	0	15
<pro .1>	18	1	0	1	0	0	0	0	2	0	22
<pro .2>	23	5	0	2	1	1	0	0	2	0	34
<pro .h>	68	31	0	9	0	5	0	0	9	0	122
<pro .d>	13	2	0	1	0	0	0	0	3	0	19
<pro>	8	0	0	1	0	0	0	0	0	0	9
<np .h>	2	0	0	1	0	2	0	4	0	0	9
<np .d>	10	1	0	0	0	2	0	1	0	0	14
<np>	5	1	0	19	5	23	5	3	0	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	11	0	123	150
<i>totals</i>	172	45	0	46	9	38	15	19	16	169	
<##>											212
<#>											26
<i>totals</i>											238

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

2.7 *iswm*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	2	0	2	0	0	0	0	0	0	4
<∅ .2>	4	1	0	0	0	0	0	0	0	0	5
<∅ .h>	59	26	0	7	9	0	0	0	0	0	101
<∅ .d>	0	1	0	0	0	0	0	0	0	0	1
<∅>	8	0	0	25	3	3	0	0	0	0	39
<pro .1>	19	10	0	6	2	1	0	0	5	0	43
<pro .2>	6	11	0	2	0	1	0	0	8	1	29
<pro .h>	185	105	0	42	6	8	0	0	73	0	419
<pro .d>	1	0	0	0	0	0	0	0	0	0	1
<pro>	18	3	0	0	0	0	0	3	11	1	36
<np .h>	39	16	0	27	1	10	0	16	23	3	135
<np .d>	1	0	0	0	0	0	0	0	0	0	1
<np>	37	4	0	73	29	47	26	11	0	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	29	0	300	358
<i>totals</i>	377	179	0	184	51	75	49	59	120	371	
<##>											488
<#>											88
<i>totals</i>											576

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

2.8 *jjq*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	7	1	0	0	0	0	0	0	0	0	8
<∅ .2>	9	2	0	0	0	0	0	0	0	0	11
<∅ .h>	63	24	0	3	0	0	0	0	0	0	90
<∅ .d>	38	33	0	0	0	0	0	0	0	0	71
<∅>	31	3	0	42	7	5	3	0	0	0	91
<pro .1>	83	30	0	9	2	0	0	0	24	0	148
<pro .2>	36	15	0	8	2	1	0	0	7	0	69
<pro .h>	143	53	0	26	1	7	0	0	53	0	283
<pro .d>	25	13	0	4	0	1	0	0	6	0	49
<pro>	16	2	0	1	0	1	0	0	11	0	31
<np .h>	71	14	0	8	1	7	0	14	16	2	133
<np .d>	28	8	0	4	2	0	0	1	3	0	46
<np>	62	2	0	98	11	62	33	40	3	44	355
<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>	3	0	0	0	5	13	40	83	0	467	611
<i>totals</i>	615	200	0	203	31	97	76	138	123	513	
<##>											820
<#>											60
<i>totals</i>											880

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

2.9 *mvbw*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	1	0	0	0	0	0	0	0	0	0	1
<∅ .h>	41	33	0	3	0	0	0	0	0	0	77
<∅ .d>	4	0	0	0	0	0	0	0	0	0	4
<∅>	1	1	0	16	0	0	0	0	0	0	18
<pro .1>	24	9	0	3	3	2	0	0	10	1	52
<pro .2>	16	4	0	6	0	0	0	0	7	0	33
<pro .h>	56	34	0	10	1	9	0	0	58	0	168
<pro .d>	2	2	0	0	0	0	0	0	0	0	4
<pro>	3	0	0	0	0	0	0	0	1	0	4
<np .h>	33	3	0	18	3	5	0	6	5	0	73
<np .d>	4	0	0	0	1	1	0	0	0	0	6
<np>	13	6	0	38	2	15	5	6	1	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	28	0	179	214
<i>totals</i>	198	92	0	94	11	33	10	40	82	199	
<##>											189
<#>											118
<i>totals</i>											307

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

2.10 *pala*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	3	5	0	0	0	0	0	0	0	0	8
<∅ .h>	57	21	0	2	0	0	0	0	0	0	80
<∅ .d>	9	13	0	1	0	0	0	0	0	0	23
<∅>	7	4	0	12	3	0	0	0	0	0	26
<pro .1>	16	7	0	3	0	0	0	3	3	0	32
<pro .2>	10	5	0	2	1	1	0	0	3	0	22
<pro .h>	77	50	0	7	0	5	0	0	22	0	161
<pro .d>	5	1	0	0	0	0	0	0	4	0	10
<pro>	1	0	0	0	0	0	0	0	1	0	2
<np .h>	41	6	0	13	3	5	0	4	2	0	74
<np .d>	8	6	0	3	0	1	0	3	3	0	24
<np>	12	2	0	76	1	18	4	3	0	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	18	0	244	265
<i>totals</i>	246	120	0	119	9	31	5	31	38	250	
<##>											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/