

Multi-CAST

*Kalamang
corpus counts*

Eline Visser

September 2021
v1.1



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

Visser, Eline. 2021. Multi-CAST Kalamang corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#kalamang) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Kalamang corpus counts v1.1 last updated 13 September 2021
This document was typeset by NNS with $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and the *multicast3* class (v3.2.5).

Contents

1	Notes on the GRAID counts	1
2	The Kalamang corpus	2
2.1	<i>kasuari</i>	3
2.2	<i>keluer</i>	4
2.3	<i>kuawi</i>	5
2.4	<i>monyet</i>	6
2.5	<i>pitiskiet</i>	7
2.6	<i>yardakdak</i>	8

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 Kalamang 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 Kalamang corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	3	8	0	1	0	0	0	0	0	0	12
<∅ .2>	24	33	0	1	0	0	0	0	0	0	58
<∅ .h>	47	45	0	7	0	0	0	0	0	0	99
<∅ .d>	29	22	0	8	0	0	0	0	0	0	59
<∅>	35	11	0	106	0	0	1	17	0	0	170
<pro .1>	58	45	0	18	1	1	1	1	21	0	146
<pro .2>	31	26	0	9	0	3	2	0	15	0	86
<pro .h>	124	67	0	22	3	8	3	0	62	0	289
<pro .d>	71	30	0	12	0	5	0	0	18	0	136
<pro>	41	10	0	10	4	10	2	25	31	0	133
<np .h>	38	25	0	14	9	3	3	4	0	0	96
<np .d>	57	9	0	11	0	0	0	0	0	0	77
<np>	124	16	0	130	6	44	43	81	0	5	449
<other .h>	6	2	0	0	0	0	1	0	0	1	10
<other .d>	0	0	0	0	0	0	0	0	0	1	1
<other>	1	1	0	1	0	0	0	32	0	1068	1103
<i>totals</i>	689	350	0	350	23	74	56	160	147	1075	
<##>											1019
<#>											32
<i>totals</i>											1051

Table 1 Summarized GRAID counts for the entire Kalamang corpus.

2.1 kasuari

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>	0	0	0	0	0	0	0	0	0	0	0
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	7	4	0	3	0	0	0	0	0	0	14
<pro .1>	0	0	0	0	0	0	0	0	0	0	0
<pro .2>	0	0	0	0	0	0	0	0	0	0	0
<pro .h>	0	0	0	0	0	0	0	0	0	0	0
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	19	9	0	2	0	0	0	0	11	0	41
<np .h>	0	0	0	0	0	0	0	0	0	0	0
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	16	1	0	9	0	7	5	8	0	1	47
<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	0	0	0	0	27	27
<i>totals</i>	42	14	0	14	0	7	5	8	11	28	
<##>											53
<#>											4
<i>totals</i>											57

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

2.2 *keluer*

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	4	0	0	0	0	0	0	0	0	6
<∅ .h>	10	9	0	2	0	0	0	0	0	0	21
<∅ .d>	0	2	0	2	0	0	0	0	0	0	4
<∅>	2	0	0	9	0	0	0	9	0	0	20
<pro .1>	7	10	0	0	0	1	0	0	10	0	28
<pro .2>	0	12	0	0	0	0	0	0	7	0	19
<pro .h>	19	8	0	4	0	6	0	0	20	0	57
<pro .d>	2	1	0	0	0	0	0	0	1	0	4
<pro>	0	0	0	0	2	1	0	2	2	0	7
<np .h>	9	7	0	10	3	2	0	0	0	0	31
<np .d>	3	0	0	6	0	0	0	0	0	0	9
<np>	6	0	0	20	0	3	5	7	0	0	41
<other .h>	4	0	0	0	0	0	0	0	0	0	4
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	4	0	99	103
<i>totals</i>	64	53	0	53	5	13	5	22	40	99	
<##>											116
<#>											2
<i>totals</i>											118

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

2.3 *kuawi*

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>	4	1	0	0	0	0	0	0	0	0	5
<∅ .h>	8	8	0	0	0	0	0	0	0	0	16
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	10	4	0	14	0	0	1	0	0	0	29
<pro .1>	13	8	0	4	0	0	0	0	4	0	29
<pro .2>	11	0	0	1	0	0	0	0	0	0	12
<pro .h>	29	13	0	0	0	0	0	0	8	0	50
<pro .d>	4	4	0	0	0	0	0	0	0	0	8
<pro>	14	1	0	4	0	0	0	11	4	0	34
<np .h>	12	1	0	0	1	0	0	0	0	0	14
<np .d>	1	2	0	1	0	0	0	0	0	0	4
<np>	29	0	0	21	2	8	11	12	0	1	84
<other .h>	2	2	0	0	0	0	0	0	0	1	5
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	1	1	0	0	0	0	0	5	0	222	229
<i>totals</i>	139	45	0	45	3	8	12	28	16	224	
<##>											183
<#>											8
<i>totals</i>											191

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

2.4 monyet

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	2	7	0	1	0	0	0	0	0	0	10
<∅ .2>	11	22	0	1	0	0	0	0	0	0	34
<∅ .h>	2	1	0	0	0	0	0	0	0	0	3
<∅ .d>	28	20	0	1	0	0	0	0	0	0	49
<∅>	11	0	0	40	0	0	0	0	0	0	51
<pro .1>	25	12	0	4	1	0	0	1	5	0	48
<pro .2>	14	1	0	3	0	0	0	0	6	0	24
<pro .h>	13	4	0	1	0	0	0	0	0	0	18
<pro .d>	61	24	0	12	0	0	0	0	16	0	113
<pro>	8	0	0	0	0	8	2	7	1	0	26
<np .h>	4	1	0	0	0	0	0	0	0	0	5
<np .d>	50	7	0	3	0	0	0	0	0	0	60
<np>	49	2	0	36	1	16	12	40	0	1	157
<other .h>	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	1	1
<other>	0	0	0	0	0	0	0	12	0	443	455
<i>totals</i>	278	101	0	102	2	24	14	60	28	445	
<##>											368
<#>											12
<i>totals</i>											380

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

2.5 *pitiskiet*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	1	0	0	0	0	0	0	0	0	1
<∅ .2>	7	6	0	0	0	0	0	0	0	0	13
<∅ .h>	24	21	0	1	0	0	0	0	0	0	46
<∅ .d>	1	0	0	5	0	0	0	0	0	0	6
<∅>	0	3	0	37	0	0	0	8	0	0	48
<pro .1>	13	15	0	10	0	0	1	0	2	0	41
<pro .2>	6	13	0	5	0	3	2	0	2	0	31
<pro .h>	52	31	0	12	2	2	3	0	30	0	132
<pro .d>	4	1	0	0	0	5	0	0	1	0	11
<pro>	0	0	0	4	2	1	0	5	4	0	16
<np .h>	11	16	0	4	4	1	3	4	0	0	43
<np .d>	3	0	0	1	0	0	0	0	0	0	4
<np>	13	13	0	39	2	6	2	9	0	1	85
<other .h>	0	0	0	0	0	0	1	0	0	0	1
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	1	0	0	0	4	0	230	235
<i>totals</i>	134	120	0	119	10	18	12	30	39	231	
<##>											250
<#>											5
<i>totals</i>											255

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

2.6 *yardakdak*

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>	3	6	0	4	0	0	0	0	0	0	13
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	5	0	0	3	0	0	0	0	0	0	8
<pro .1>	0	0	0	0	0	0	0	0	0	0	0
<pro .2>	0	0	0	0	0	0	0	0	0	0	0
<pro .h>	11	11	0	5	1	0	0	0	4	0	32
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	0	0	0	0	9	0	9
<np .h>	2	0	0	0	1	0	0	0	0	0	3
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	11	0	0	5	1	4	8	5	0	1	35
<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	0	0	7	0	47	54
<i>totals</i>	32	17	0	17	3	4	8	12	13	48	
<##>											49
<#>											1
<i>totals</i>											50

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

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/