

Multi-CAST

*Sumbawa
corpus counts*

Asako Shiohara

November 2022
v1.0



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

Shiohara, Asako. 2022. Multi-CAST Sumbawa corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#sumbawa) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Sumbawa corpus counts v1.0 last updated 8 November 2022
This document was typeset by NNS with $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and the *multicast3* class (v3.2.6).

Contents

1	Notes on the GRAID counts	1
2	The Sumbawa corpus	2
2.1	<i>flood</i>	3
2.2	<i>kerekkure</i>	4
2.3	<i>langlelo</i>	5
2.4	<i>menangis</i>	6
2.5	<i>nuntut</i>	7

1 Notes on the GRAID counts

This document collects tables with frequency counts for combinations of selected GRAID symbols in version 2211 (from November 2022) of the Multi-CAST Sumbawa 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

⟨∅⟩	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 Sumbawa corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	8	15	0	0	0	0	0	0	0	0	23
<∅ .2>	8	10	0	1	0	0	0	0	0	0	19
<∅ .h>	154	128	0	31	1	0	0	0	0	0	314
<∅ .d>	3	2	0	0	0	0	0	0	0	0	5
<∅>	72	2	0	69	1	1	8	0	0	9	162
<pro .1>	33	56	0	13	2	3	0	1	41	1	150
<pro .2>	15	20	0	6	1	1	0	0	14	0	57
<pro .h>	41	42	0	1	3	0	0	1	8	2	98
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	34	1	0	6	2	8	14	10	1	4	80
<np .h>	99	49	0	32	15	5	3	10	28	14	255
<np .d>	8	2	0	0	3	3	0	0	1	0	17
<np>	241	9	0	197	10	34	65	87	10	49	702
<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	8	0	898	906
<i>totals</i>	716	336	0	356	38	55	90	117	103	977	
<##>											0
<#>											1084
<i>totals</i>											1084

Table 1 Summarized GRAID counts for the entire Sumbawa corpus.

2.1 flood

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	5	0	0	0	0	0	0	0	0	5
<∅ .2>	1	2	0	0	0	0	0	0	0	0	3
<∅ .h>	16	6	0	1	0	0	0	0	0	0	23
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	14	0	0	7	0	0	1	0	0	1	23
<pro .1>	1	9	0	5	1	0	0	1	7	1	25
<pro .2>	2	5	0	0	0	0	0	0	8	0	15
<pro .h>	1	0	0	0	0	0	0	0	0	0	1
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	5	0	0	0	0	0	0	1	1	3	10
<np .h>	14	8	0	5	4	0	1	3	10	4	49
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	36	0	0	22	1	1	1	20	5	9	95
<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	1	0	124	125
<i>totals</i>	90	35	0	40	6	1	3	26	31	142	
<##>											0
<#>											134
<i>totals</i>											134

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

2.2 kerekkure

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	6	5	0	0	0	0	0	0	0	0	11
<∅ .2>	7	6	0	1	0	0	0	0	0	0	14
<∅ .h>	90	72	0	17	1	0	0	0	0	0	180
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	29	1	0	35	0	0	4	0	0	5	74
<pro .1>	24	20	0	5	1	0	0	0	24	0	74
<pro .2>	11	8	0	2	1	0	0	0	3	0	25
<pro .h>	27	18	0	1	2	0	0	1	5	2	56
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	17	0	0	4	0	6	10	3	0	0	40
<np .h>	42	24	0	5	6	2	1	3	7	7	97
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	135	2	0	102	3	19	37	47	0	30	375
<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	3	0	436	439
<i>totals</i>	388	156	0	172	14	27	52	57	39	480	
<##>											0
<#>											561
<i>totals</i>											561

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

2.3 *langlelo*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	4	0	0	0	0	0	0	0	0	5
<∅ .2>	0	1	0	0	0	0	0	0	0	0	1
<∅ .h>	8	11	0	3	0	0	0	0	0	0	22
<∅ .d>	3	2	0	0	0	0	0	0	0	0	5
<∅>	22	0	0	15	1	1	0	0	0	3	42
<pro .1>	7	18	0	2	0	2	0	0	2	0	31
<pro .2>	1	4	0	2	0	1	0	0	0	0	8
<pro .h>	0	4	0	0	0	0	0	0	0	0	4
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	10	1	0	0	0	0	0	2	0	0	13
<np .h>	11	2	0	3	0	0	0	1	5	0	22
<np .d>	8	2	0	0	3	3	0	0	1	0	17
<np>	42	0	0	26	1	7	5	17	3	2	103
<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	3	0	141	144
<i>totals</i>	113	49	0	51	5	14	5	23	11	146	
<##>											0
<#>											167
<i>totals</i>											167

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

2.4 menangis

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	1	0	0	0	0	0	0	0	0	2
<∅ .2>	0	1	0	0	0	0	0	0	0	0	1
<∅ .h>	22	12	0	8	0	0	0	0	0	0	42
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	3	0	0	2	0	0	0	0	0	0	5
<pro .1>	1	6	0	1	0	1	0	0	7	0	16
<pro .2>	1	2	0	1	0	0	0	0	3	0	7
<pro .h>	4	8	0	0	1	0	0	0	1	0	14
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	0	0	3	2	0	1	6
<np .h>	27	10	0	17	3	3	0	2	5	3	70
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	16	2	0	8	4	6	9	2	2	3	52
<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	88	88
<i>totals</i>	75	42	0	37	8	10	12	6	18	95	
<##>											0
<#>											117
<i>totals</i>											117

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

2.5 *nuntut*

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>	18	27	0	2	0	0	0	0	0	0	47
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	4	1	0	10	0	0	3	0	0	0	18
<pro .1>	0	3	0	0	0	0	0	0	1	0	4
<pro .2>	0	1	0	1	0	0	0	0	0	0	2
<pro .h>	9	12	0	0	0	0	0	0	2	0	23
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	0	2	2	2	1	2	0	0	11
<np .h>	5	5	0	2	2	0	1	1	1	0	17
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	12	5	0	39	1	1	13	1	0	5	77
<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	1	0	109	110
<i>totals</i>	50	54	0	56	5	3	18	5	4	114	
<##>											0
<#>											105
<i>totals</i>											105

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

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/