

Multi-CAST

Arta

corpus counts

Yukinori Kimoto

September 2021
v1.3



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

Kimoto, Yukinori. 2021. Multi-CAST Arta corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#arta) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Arta corpus counts v1.3 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	<hr/>	1
2	The Arta corpus	<hr/>	2
2.1	<i>alisiya</i>		3
2.2	<i>arsenyo</i>		4
2.3	<i>child</i>		5
2.4	<i>delia</i>		6
2.5	<i>disubu</i>		7
2.6	<i>hapon</i>		8
2.7	<i>husband</i>		9
2.8	<i>marry</i>		10
2.9	<i>swateng</i>		11
2.10	<i>typhoon</i>		12
2.11	<i>udulan</i>		13

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 Arta 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_a⟩	A argument of a transitive clause, actor voice
⟨:p_a⟩	P argument of a transitive clause, actor voice
⟨:a_a⟩	A argument of a transitive clause, undergoer voice
⟨:p_a⟩	P argument of a transitive clause, undergoer voice
⟨:p⟩	direct object
⟨:obl⟩	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 Arta corpus

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	3	8	0	1	3	0	1	0	0	0	0	16
<∅ .2>	1	6	0	1	0	0	0	0	0	0	0	8
<∅ .h>	52	22	2	5	25	1	2	0	0	0	0	109
<∅ .d>	2	1	0	1	1	0	0	0	0	0	0	5
<∅>	24	7	6	0	63	0	1	0	0	0	0	101
<pro .1>	63	45	5	64	10	0	2	0	2	132	2	325
<pro .2>	20	9	0	30	5	1	0	0	3	21	0	89
<pro .h>	58	33	4	113	17	1	0	0	2	92	1	321
<pro .d>	0	0	0	1	0	0	0	0	0	0	0	1
<pro>	29	1	21	3	20	3	11	8	7	9	7	119
<np .h>	77	21	30	19	28	4	11	0	6	30	7	233
<np .d>	0	0	2	0	0	0	0	0	0	0	0	2
<np>	82	3	83	3	47	19	9	20	67	14	68	415
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	2	0	3	0	1	0	0	1	125	0	1062	1194
<i>totals</i>	413	156	156	241	220	29	37	29	212	298	1147	
<##>												690
<#>												340
<i>totals</i>												1030

Table 1 Summarized GRAID counts for the entire Arta corpus.

2.1 *alisiya*

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	1	0	0	1	0	0	0	0	0	0	2
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	3	0	0	1	0	0	0	0	0	0	0	4
<∅ .d>	2	1	0	1	1	0	0	0	0	0	0	5
<∅>	0	1	0	0	0	0	0	0	0	0	0	1
<pro .1>	15	4	1	1	2	0	0	0	0	15	0	38
<pro .2>	0	0	0	1	0	0	0	0	0	0	0	1
<pro .h>	1	0	0	3	0	0	0	0	0	3	0	7
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	1	0	0	0	1	0	0	0	0	2
<np .h>	4	1	1	1	1	0	0	0	0	0	0	8
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	7	0	5	0	3	0	1	1	3	0	3	23
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	0	6	0	34	40
<i>totals</i>	32	8	8	8	8	0	2	1	9	18	37	
<##>												37
<#>												18
<i>totals</i>												55

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

2.2 arsenyo

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	1	0	0	0	0	0	0	1
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	6	1	0	0	1	0	0	0	0	0	0	8
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	1	0	0	0	4	0	0	0	0	0	0	5
<pro .1>	2	5	1	10	1	0	0	0	1	7	1	28
<pro .2>	0	1	0	0	2	0	0	0	0	0	0	3
<pro .h>	2	1	0	6	4	0	0	0	0	1	0	14
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	1	0	2	0	0	1	1	0	1	8
<np .h>	5	0	4	2	3	0	0	0	2	0	3	19
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	5	0	1	0	0	0	0	0	0	0	4	10
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	0	10	0	57	67
<i>totals</i>	23	8	7	18	18	0	0	1	14	8	66	
<##>												39
<#>												26
<i>totals</i>												65

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

2.3 *child*

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	7	2	0	1	1	0	0	0	0	0	0	11
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	0	3	0	11	0	0	0	0	0	0	20
<pro .1>	2	13	0	15	1	0	0	0	0	45	0	76
<pro .2>	0	0	0	0	0	0	0	0	0	1	0	1
<pro .h>	15	7	3	10	2	1	0	0	0	16	0	54
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	7	0	2	0	3	0	1	0	5	1	1	20
<np .h>	15	2	0	0	1	0	0	0	1	4	1	24
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	16	2	17	0	7	3	3	1	13	3	10	75
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	0	15	0	220	235
<i>totals</i>	68	26	25	26	26	4	4	1	34	70	232	
<##>												91
<#>												72
<i>totals</i>												163

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

2.4 *delia*

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<θ .1>	1	4	0	1	0	0	0	0	0	0	0	6
<θ .2>	0	3	0	0	0	0	0	0	0	0	0	3
<θ .h>	2	4	0	0	0	0	0	0	0	0	0	6
<θ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<θ>	2	1	0	0	5	0	0	0	0	0	0	8
<pro .1>	18	14	1	13	2	0	1	0	0	15	0	64
<pro .2>	1	6	0	5	2	1	0	0	0	4	0	19
<pro .h>	2	1	0	2	3	0	0	0	0	0	0	8
<pro .d>	0	0	0	1	0	0	0	0	0	0	0	1
<pro>	7	1	5	1	2	0	0	1	0	0	0	17
<np .h>	0	1	7	0	1	0	1	0	0	1	0	11
<np .d>	0	0	2	0	0	0	0	0	0	0	0	2
<np>	6	0	18	0	2	2	0	1	9	0	6	44
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	1	0	0	0	0	1	12	0	146	160
<i>totals</i>	39	35	34	23	17	3	2	3	21	20	152	
<##>												76
<#>												44
<i>totals</i>												120

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

2.5 *disubu*

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	1	0	0	0	0	0	0	0	0	0	1
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	2	0	0	0	0	0	0	0	0	0	0	2
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	3	0	1	0	7	0	0	0	0	0	0	11
<pro .1>	2	1	0	5	0	0	0	0	0	17	0	25
<pro .2>	0	0	0	0	0	0	0	0	0	0	0	0
<pro .h>	2	0	0	3	0	0	0	0	0	2	1	8
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	7	0	0	0	3	0	0	1	0	2	0	13
<np .h>	1	0	0	2	0	0	0	0	0	1	0	4
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	8	0	1	0	1	1	2	7	16	0	11	47
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	0	8	0	95	103
<i>totals</i>	25	2	2	10	11	1	2	8	24	22	107	
<##>												48
<#>												11
<i>totals</i>												59

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

2.6 hapon

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	5	2	1	0	2	1	0	0	0	0	0	11
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	0	0	0	7	0	1	0	0	0	0	8
<pro .1>	4	0	0	0	0	0	1	0	0	14	1	20
<pro .2>	3	0	0	0	0	0	0	0	1	2	0	6
<pro .h>	17	14	1	21	1	0	0	0	0	16	0	70
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	1	0	4	0	2	2	0	0	3	14
<np .h>	11	1	2	5	1	1	2	0	0	12	0	35
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	10	0	12	0	8	1	1	2	11	4	8	57
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	2	0	1	0	0	0	0	0	20	0	188	211
<i>totals</i>	54	17	18	26	23	3	7	4	32	48	200	
<##>												99
<#>												33
<i>totals</i>												132

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

2.7 *husband*

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	1	0	0	0	0	0	0	0	0	0	2
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	8	2	0	0	1	0	0	0	0	0	0	11
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	0	0	0	1	0	0	0	0	0	0	1
<pro .1>	6	1	0	2	0	0	0	0	0	3	0	12
<pro .2>	1	0	0	0	0	0	0	0	0	0	0	1
<pro .h>	2	2	0	8	5	0	0	0	0	4	0	21
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	1	0	1	0	0	0	0	0	0	2
<np .h>	3	0	1	0	2	0	0	0	0	1	0	7
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	3	0	3	1	1	2	2	0	4	0	3	19
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	1	0	0	0	0	0	12	0	35	48
<i>totals</i>	24	6	6	11	11	2	2	0	16	8	38	
<##>												35
<#>												12
<i>totals</i>												47

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

2.8 marry

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	1	3	0	0	0	0	0	0	0	0	0	4
<∅ .h>	0	0	0	0	4	0	0	0	0	0	0	4
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	0	0	0	3	0	0	0	0	0	0	3
<pro .1>	3	0	0	1	1	0	0	0	0	0	0	5
<pro .2>	7	0	0	12	0	0	0	0	2	9	0	30
<pro .h>	0	0	0	2	0	0	0	0	2	0	0	4
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	0	0	0	0	1	0	0	1
<np .h>	0	0	1	0	3	0	0	0	0	1	0	5
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	1	0	2	0	2	1	0	0	1	0	0	7
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	1	0	0	0	7	0	46	54
<i>totals</i>	12	3	3	15	14	1	0	0	13	10	46	
<##>												17
<#>												28
<i>totals</i>												45

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

2.9 *swateng*

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	0	0	0	1	0	0	0	0	0	0	2
<∅ .2>	0	0	0	1	0	0	0	0	0	0	0	1
<∅ .h>	14	4	0	2	16	0	1	0	0	0	0	37
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	2	1	1	0	10	0	0	0	0	0	0	14
<pro .1>	4	4	0	8	3	0	0	0	1	3	0	23
<pro .2>	7	1	0	10	1	0	0	0	0	4	0	23
<pro .h>	12	5	0	43	1	0	0	0	0	38	0	99
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	6	0	5	2	5	0	0	0	2	22
<np .h>	26	8	7	7	14	3	5	0	0	2	2	74
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	11	0	10	2	13	7	0	1	5	0	10	59
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	0	20	0	87	107
<i>totals</i>	79	23	24	73	64	12	11	1	26	47	101	
<##>												141
<#>												49
<i>totals</i>												190

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

2.10 typhoon

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	1	0	0	0	0	1
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	0	1	0	0	0	0	0	0	0	0	0	1
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	8	2	0	0	6	0	0	0	0	0	0	16
<pro .1>	6	2	2	8	0	0	0	0	0	10	0	28
<pro .2>	0	0	0	0	0	0	0	0	0	0	0	0
<pro .h>	1	1	0	3	1	0	0	0	0	5	0	11
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	1	2	0	0	0	1	0	6	0	11
<np .h>	0	0	0	0	0	0	0	0	0	3	0	3
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	15	1	5	0	5	2	0	5	4	7	9	53
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	0	11	0	86	97
<i>totals</i>	31	7	8	13	12	2	1	6	15	31	95	
<##>												47
<#>												25
<i>totals</i>												72

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

2.11 *udulan*

GRAID	<:s>	<:a_a>	<:p_a>	<:a_u>	<:p_u>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	1	0	0	0	0	0	0	0	0	0	1
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	5	6	1	1	0	0	1	0	0	0	0	14
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0	0
<∅>	2	2	1	0	9	0	0	0	0	0	0	14
<pro .1>	1	1	0	1	0	0	0	0	0	3	0	6
<pro .2>	1	1	0	2	0	0	0	0	0	1	0	5
<pro .h>	4	2	0	12	0	0	0	0	0	7	0	25
<pro .d>	0	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	3	0	0	1	2	2	0	0	0	9
<np .h>	12	8	7	2	2	0	3	0	3	5	1	43
<np .d>	0	0	0	0	0	0	0	0	0	0	0	0
<np>	0	0	9	0	5	0	0	2	1	0	4	21
<other .h>	0	0	0	0	0	0	0	0	0	0	0	0
<other .d>	0	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	0	4	0	68	72
<i>totals</i>	26	21	21	18	16	1	6	4	8	16	73	
<##>												60
<#>												22
<i>totals</i>												82

Table 12 Summarized GRAID counts for the *udulan* text.

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/