

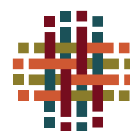
Multi-CAST

Chirag
corpus counts

Dmitry Ganenkov

Nils Norman Schiborr

July 2025
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

Ganenkov, Dmitry & Schiborr, Nils N. 2025. Multi-CAST Chirag corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#chirag) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Chirag corpus counts v1.0 last updated 12 July 2025
This document was typeset by NNS with X_YLA^TE_X and the *multicast3* class (v3.2.6).

Contents

1	Notes on the GRAID counts	1
2	The Chirag corpus	2
2.1	<i>akaccaj</i>	3
2.2	<i>bartishan</i>	4
2.3	<i>bisbisaj</i>	5
2.4	<i>curka</i>	6
2.5	<i>husen</i>	7
2.6	<i>ismail</i>	8
2.7	<i>mannay</i>	9
2.8	<i>sapsap</i>	10
2.9	<i>shamil</i>	11
2.10	<i>urce</i>	12
2.11	<i>zahrat</i>	13

1 Notes on the GRAID counts

This document collects tables with frequency counts for combinations of selected GRAID symbols in version 2507 (from July 2025) of the Multi-CAST Chirag 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
⟨: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 Chirag corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	40	36	12	5	0	9	0	0	0	0	102
<∅ .2>	26	41	6	3	0	2	0	0	0	0	78
<∅ .h>	218	290	23	39	4	10	0	0	0	0	584
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	60	11	1	51	6	0	5	1	0	2	137
<pro .1>	38	18	8	7	17	8	0	0	75	1	172
<pro .2>	8	7	5	3	1	9	0	0	7	0	40
<pro .h>	58	36	17	15	30	9	0	1	37	2	205
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	13	0	0	20	9	10	22	4	2	26	106
<np .h>	117	78	9	41	33	23	0	27	48	17	393
<np .d>	0	0	0	0	0	0	0	0	0	1	1
<np>	131	6	7	198	49	66	83	36	21	120	717
<other .h>	8	5	2	5	3	3	0	1	3	1	31
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	3	0	0	21	4	11	5	107	0	712	863
<i>totals</i>	720	528	90	408	156	160	115	177	193	882	
<##>											0
<#>											1409
<i>totals</i>											1409

Table 1 Summarized GRAID counts for the entire Chirag corpus.

2.1 *akaccaj*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	5	6	0	0	1	0	0	0	0	13
<∅ .2>	1	5	0	0	0	1	0	0	0	0	7
<∅ .h>	30	48	0	0	0	0	0	0	0	0	78
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	19	5	0	13	0	0	0	0	0	1	38
<pro .1>	1	1	2	0	1	1	0	0	7	1	14
<pro .2>	1	2	1	0	1	0	0	0	0	0	5
<pro .h>	5	4	1	0	3	0	0	0	7	0	20
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	5	0	0	7	3	4	2	1	0	2	24
<np .h>	10	6	0	4	1	3	0	6	2	1	33
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	24	3	0	36	5	11	8	4	6	15	112
<other .h>	0	0	0	1	0	1	0	0	0	1	3
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	6	2	2	0	17	0	109	136
<i>totals</i>	97	79	10	67	16	24	10	28	22	130	
<##>											0
<#>											193
<i>totals</i>											193

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

2.2 *bartishan*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	3	0	0	0	0	0	0	0	0	3
<∅ .2>	2	2	0	3	0	0	0	0	0	0	7
<∅ .h>	27	21	3	5	1	0	0	0	0	0	57
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	7	0	0	2	0	0	1	0	0	0	10
<pro .1>	2	1	0	0	0	1	0	0	1	0	5
<pro .2>	1	0	0	1	0	1	0	0	0	0	3
<pro .h>	5	3	1	1	0	0	0	0	7	0	17
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	1	0	0	0	1	1	2	5
<np .h>	4	9	2	1	2	3	0	0	1	1	23
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	8	0	1	5	3	3	5	6	3	5	39
<other .h>	0	2	0	0	0	0	0	0	0	0	2
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	2	0	0	0	7	0	59	68
<i>totals</i>	56	41	7	21	6	8	6	14	13	67	
<##>											0
<#>											105
<i>totals</i>											105

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

2.3 *bisbisaj*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	1	0	0	0	7	0	0	0	0	9
<∅ .2>	0	12	0	0	0	0	0	0	0	0	12
<∅ .h>	14	37	0	0	0	4	0	0	0	0	55
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	3	0	0	6	0	0	0	0	0	0	9
<pro .1>	2	1	0	0	1	1	0	0	6	0	11
<pro .2>	0	0	0	0	0	0	0	0	0	0	0
<pro .h>	4	5	1	0	2	1	0	0	2	0	15
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	0	6	0	0	4	0	0	1	13
<np .h>	5	2	0	6	2	0	0	0	2	1	18
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	9	0	0	28	2	9	2	0	0	1	51
<other .h>	0	0	0	0	1	0	0	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	0	0	62	63
<i>totals</i>	40	58	1	47	8	22	6	0	10	65	
<##>											0
<#>											104
<i>totals</i>											104

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

2.4 curka

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	totals
<∅ .1>	0	0	1	0	0	0	0	0	0	0	1
<∅ .2>	2	0	0	0	0	0	0	0	0	0	2
<∅ .h>	5	20	1	4	0	0	0	0	0	0	30
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	5	0	1	0	0	0	0	0	0	1	7
<pro .1>	0	0	0	0	1	0	0	0	3	0	4
<pro .2>	0	0	1	0	0	1	0	0	1	0	3
<pro .h>	0	0	2	0	1	0	0	0	1	0	4
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	0	0	0	1	2	0	0	5	10
<np .h>	7	14	4	2	2	3	0	0	4	1	37
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	10	0	4	8	6	0	9	4	2	14	57
<other .h>	1	0	0	0	0	0	0	0	0	0	1
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	1	0	0	4	0	0	0	3	0	58	66
totals	33	34	14	18	10	5	11	7	11	79	
<##>											0
<#>											93
totals											93

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

2.5 *husen*

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	0	0	0	0	0	0	0	0	0	1
<∅ .h>	15	35	2	2	1	0	0	0	0	0	55
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	0	0	6	0	0	1	0	0	0	13
<pro .1>	0	0	0	0	0	0	0	0	2	0	2
<pro .2>	1	1	0	0	0	0	0	0	0	0	2
<pro .h>	13	6	2	3	8	1	0	0	2	1	36
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	1	1	1	2	0	0	1	6
<np .h>	9	6	0	5	4	1	0	1	6	1	33
<np .d>	0	0	0	0	0	0	0	0	0	1	1
<np>	16	0	0	24	6	5	14	1	1	9	76
<other .h>	0	0	0	1	0	0	0	0	0	0	1
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	1	1	1	2	10	0	67	82
<i>totals</i>	62	48	4	43	21	9	19	12	11	80	
<##>											0
<#>											119
<i>totals</i>											119

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

2.6 ismail

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	totals
<∅ .1>	12	7	1	1	0	0	0	0	0	0	21
<∅ .2>	4	2	3	0	0	0	0	0	0	0	9
<∅ .h>	30	24	1	4	1	1	0	0	0	0	61
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	1	0	6	0	0	0	0	0	0	13
<pro .1>	17	7	1	1	6	2	0	0	38	0	72
<pro .2>	1	0	0	0	0	0	0	0	1	0	2
<pro .h>	9	6	0	6	1	3	0	1	6	0	32
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	0	0	1	1	4	1	0	4	12
<np .h>	31	8	2	8	10	4	0	11	12	8	94
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	22	3	2	26	12	16	15	4	0	37	137
<other .h>	0	1	0	0	0	0	0	0	0	0	1
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	1	0	0	1	0	0	1	15	0	86	104
totals	134	59	10	53	31	27	20	32	57	135	
<##>											0
<#>											220
totals											220

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

2.7 *mannay*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	4	0	0	0	0	0	0	0	0	0	4
<∅ .2>	5	1	1	0	0	0	0	0	0	0	7
<∅ .h>	20	22	2	4	0	0	0	0	0	0	48
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	1	0	0	0	0	0	0	0	0	0	1
<pro .1>	2	0	0	1	1	0	0	0	4	0	8
<pro .2>	0	2	0	0	0	0	0	0	0	0	2
<pro .h>	5	1	1	0	5	1	0	0	0	1	14
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	0	0	0	1	0	0	1
<np .h>	10	5	1	2	1	2	0	2	2	3	28
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	2	0	0	5	2	1	3	0	2	8	23
<other .h>	3	0	0	2	0	0	0	0	0	0	5
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	2	0	0	0	8	0	45	55
<i>totals</i>	52	31	5	16	9	4	3	11	8	57	
<##>											0
<#>											93
<i>totals</i>											93

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

2.8 sapsap

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	18	12	0	2	0	0	0	0	0	0	32
<∅ .2>	2	5	0	0	0	0	0	0	0	0	7
<∅ .h>	6	16	0	0	0	0	0	0	0	0	22
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	3	0	3	1	0	0	0	0	0	13
<pro .1>	10	3	1	4	2	0	0	0	4	0	24
<pro .2>	0	0	0	2	0	0	0	0	2	0	4
<pro .h>	4	0	1	0	2	0	0	0	3	0	10
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	1	2	1	0	0	2	6
<np .h>	16	3	0	2	4	4	0	2	0	0	31
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	7	0	0	20	4	7	4	6	0	7	55
<other .h>	2	0	0	0	0	0	0	0	0	0	2
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	1	0	4	1	19	0	89	114
<i>totals</i>	71	42	2	34	14	17	6	27	9	98	
<##>											0
<#>											118
<i>totals</i>											118

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

2.9 *shamil*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	1	0	0	0	0	0	0	1
<∅ .2>	0	2	0	0	0	0	0	0	0	0	2
<∅ .h>	19	26	1	9	0	1	0	0	0	0	56
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	4	0	0	3	0	0	1	1	0	0	9
<pro .1>	1	1	0	1	2	0	0	0	3	0	8
<pro .2>	1	0	0	0	0	0	0	0	0	0	1
<pro .h>	6	4	0	2	3	2	0	0	2	0	19
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	0	0	1	0	3	0	0	4	9
<np .h>	10	2	0	2	1	2	0	1	10	0	28
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	7	0	0	9	6	4	12	1	2	8	49
<other .h>	0	0	1	0	1	0	0	0	0	0	2
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	0	0	0	0	10	0	29	39
<i>totals</i>	49	35	2	27	14	9	16	13	17	41	
<##>											0
<#>											90
<i>totals</i>											90

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

2.10 *urce*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	2	7	3	1	0	0	0	0	0	0	13
<∅ .2>	3	12	0	0	0	1	0	0	0	0	16
<∅ .h>	15	16	5	6	0	1	0	0	0	0	43
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	1	2	0	5	0	0	1	0	0	0	9
<pro .1>	1	3	4	0	1	1	0	0	6	0	16
<pro .2>	1	0	1	0	0	5	0	0	2	0	9
<pro .h>	3	1	3	1	0	1	0	0	5	0	14
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	0	5	0	1	3	0	1	1	13
<np .h>	11	14	0	3	2	0	0	3	6	0	39
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	9	0	0	27	1	6	6	8	2	14	73
<other .h>	2	2	1	1	0	0	0	0	3	0	9
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	1	0	1	1	6	0	47	56
<i>totals</i>	50	57	17	50	4	17	11	17	25	62	
<##>											0
<#>											129
<i>totals</i>											129

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

2.11 *zahrat*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	1	1	0	0	1	0	0	0	0	4
<∅ .2>	6	0	2	0	0	0	0	0	0	0	8
<∅ .h>	37	25	8	5	1	3	0	0	0	0	79
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	2	0	0	7	5	0	1	0	0	0	15
<pro .1>	2	1	0	0	2	2	0	0	1	0	8
<pro .2>	2	2	2	0	0	2	0	0	1	0	9
<pro .h>	4	6	5	2	5	0	0	0	2	0	24
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	2	0	1	0	0	4	7
<np .h>	4	9	0	6	4	1	0	1	3	1	29
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	17	0	0	10	2	4	5	2	3	2	45
<other .h>	0	0	0	0	1	2	0	1	0	0	4
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	1	0	0	2	1	3	0	12	0	61	80
<i>totals</i>	76	44	18	32	23	18	7	16	10	68	
<##>											0
<#>											145
<i>totals</i>											145

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

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/