

# Multi-CAST

## *Sanzhi Dargwa corpus counts*

*Diana Forker*

*Nils Norman Schiborr*

*January 2020*

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*

Forker, Diana & Schiborr, Nils N. 2020. Multi-CAST Sanzhi Dargwa corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. ([multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/)) (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/](http://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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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/](http://creativecommons.org/licenses/by/4.0/).

*Multi-CAST Sanzhi Dargwa corpus counts* v1.3 last updated 4 January 2020  
This document was typeset by NNS with  $\text{\LaTeX}$  and the *multicast3* class (v3.2.2).

## Contents

<b>1</b>	<b>Notes on the GRAID counts</b>	_____	<b>1</b>
<b>2</b>	<b>The Sanzhi Dargwa corpus</b>	_____	<b>2</b>
2.1	<i>asabali</i>		3
2.2	<i>bazhuk</i>		4
2.3	<i>dragon</i>		5
2.4	<i>kurban</i>		6
2.5	<i>mill</i>		7
2.6	<i>patima</i>		8
2.7	<i>ramazan</i>		9
2.8	<i>tape</i>		10



## 1 Notes on the GRAID counts

This document collects tables with frequency counts for combinations of selected GRAID symbols in version 2001 (from January 2020) of the Multi-CAST Sanzhi Dargwa 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*

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

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	67	96	13	13	0	4	0	0	0	0	193
<∅ .2>	26	26	1	0	0	0	0	0	0	0	53
<∅ .h>	136	198	13	20	6	10	0	0	0	0	383
<∅ .d>	2	14	1	1	0	0	0	0	0	0	18
<∅>	5	32	0	47	1	2	0	0	0	2	89
<pro .1>	19	33	13	5	10	3	0	25	0	1	109
<pro .2>	12	19	3	2	6	2	0	8	0	1	53
<pro .h>	18	36	2	5	9	9	1	5	0	2	87
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	12	0	16	6	3	4	0	1	7	49
<np .h>	15	100	3	27	16	8	0	10	18	8	205
<np .d>	3	13	1	5	0	2	3	0	0	0	27
<np>	4	90	3	123	39	57	49	36	26	84	511
<other .h>	3	7	0	0	0	0	0	0	1	0	11
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	6	0	18	6	26	22	0	46	0	124
<i>totals</i>	310	682	53	282	99	126	79	84	92	105	
<##>											539
<#>											527
<i>totals</i>											1066

**Table 1** Summarized GRAID counts for the entire Sanzhi Dargwa corpus.

2.1 *asabali*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	15	33	7	8	0	2	0	0	0	0	65
<∅ .2>	1	1	0	0	0	0	0	0	0	0	2
<∅ .h>	13	9	0	1	0	1	0	0	0	0	24
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	1	12	0	1	0	0	0	0	0	0	14
<pro .1>	2	14	4	2	2	0	0	4	0	0	28
<pro .2>	0	2	0	0	0	0	0	0	0	0	2
<pro .h>	3	1	0	0	0	0	0	0	0	0	4
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	3	1	1	2	0	0	0	7
<np .h>	0	11	0	1	1	0	0	0	2	3	18
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	1	10	0	16	5	18	12	7	8	23	100
<other .h>	1	1	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	0	1	3	1	0	8	0	13
<i>totals</i>	37	94	11	32	10	25	15	11	18	26	
<##>											89
<#>											53
<i>totals</i>											142

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

## 2.2 bazhuk

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	1	1	0	0	0	0	0	0	0	0	2
<∅ .2>	1	3	0	0	0	0	0	0	0	0	4
<∅ .h>	19	38	2	4	0	0	0	0	0	0	63
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	2	0	1	0	0	0	0	0	0	3
<pro .1>	2	2	0	0	2	0	0	2	0	0	8
<pro .2>	1	5	0	0	2	0	0	3	0	0	11
<pro .h>	1	3	0	1	0	0	0	2	0	1	8
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	2	0	3	1	0	0	0	0	0	6
<np .h>	2	8	0	0	1	1	0	0	0	0	12
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	0	6	0	10	5	10	6	1	0	10	48
<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	2	0	0	1	0	7	0	10
<i>totals</i>	27	70	2	21	11	11	7	8	7	11	
<##>											47
<#>											52
<i>totals</i>											99

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



## 2.3 dragon

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	0	2	0	0	0	0	0	0	0	0	2
<∅ .h>	27	33	2	1	1	0	0	0	0	0	64
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	1	2	0	4	0	1	0	0	0	0	8
<pro .1>	0	1	0	0	0	0	0	0	0	0	1
<pro .2>	0	3	0	2	0	0	0	3	0	0	8
<pro .h>	2	7	0	0	3	1	0	1	0	0	14
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	1	1	1	0	0	0	0	3
<np .h>	3	21	0	13	5	0	0	1	2	1	46
<np .d>	0	2	0	0	0	0	0	0	0	0	2
<np>	0	6	3	12	6	8	3	0	0	5	43
<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	1	0	2	0	0	2	0	5
<i>totals</i>	33	79	5	34	16	13	3	5	4	6	
<##>											53
<#>											68
<i>totals</i>											121

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

## 2.4 kurban

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	23	6	0	2	0	0	0	0	0	0	31
<∅ .2>	8	1	1	0	0	0	0	0	0	0	10
<∅ .h>	14	32	2	7	1	7	0	0	0	0	63
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	2	0	6	0	0	0	0	0	0	8
<pro .1>	4	3	1	1	0	1	0	10	0	0	20
<pro .2>	7	3	1	0	1	2	0	1	0	0	15
<pro .h>	4	7	1	2	2	4	0	1	0	0	21
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	2	0	1	0	0	0	0	0	1	4
<np .h>	1	27	1	7	5	3	0	4	5	1	54
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	0	4	0	10	4	4	4	7	4	13	50
<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>	0	1	0	6	0	4	5	0	9	0	25
<i>totals</i>	61	89	7	42	13	25	9	23	18	15	
<##>											91
<#>											73
<i>totals</i>											164

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

2.5 *mill*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	7	3	0	0	0	0	0	0	0	0	10
<∅ .2>	2	2	0	0	0	0	0	0	0	0	4
<∅ .h>	24	29	1	1	2	1	0	0	0	0	58
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	3	3	0	14	0	0	0	0	0	0	20
<pro .1>	2	2	0	0	2	2	0	1	0	0	9
<pro .2>	2	0	2	0	0	0	0	1	0	0	5
<pro .h>	5	3	1	1	0	1	0	0	0	0	11
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	3	0	1	0	0	1	0	0	0	5
<np .h>	2	6	1	1	0	1	0	1	1	0	13
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	3	23	0	28	6	1	2	6	0	3	72
<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	2	1	2	4	0	3	0	12
<i>totals</i>	50	74	5	48	11	8	7	9	4	3	
<##>											41
<#>											89
<i>totals</i>											130

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

2.6 *patima*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	2	7	2	0	0	0	0	0	0	0	11
<∅ .2>	0	9	0	0	0	0	0	0	0	0	9
<∅ .h>	13	23	3	1	0	1	0	0	0	0	41
<∅ .d>	2	14	1	1	0	0	0	0	0	0	18
<∅>	0	2	0	3	0	0	0	0	0	0	5
<pro .1>	3	3	4	0	1	0	0	0	0	1	12
<pro .2>	0	3	0	0	1	0	0	0	0	1	5
<pro .h>	0	5	0	1	1	1	0	1	0	1	10
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	1	0	0	0	0	0	2	3
<np .h>	2	9	1	3	0	0	0	1	0	0	16
<np .d>	3	11	1	5	0	2	3	0	0	0	25
<np>	0	2	0	16	7	10	8	1	1	4	49
<other .h>	1	2	0	0	0	0	0	0	0	0	3
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	3	0	1	0	4	0	0	5	0	13
<i>totals</i>	26	93	12	32	10	18	11	3	6	9	
<##>											56
<#>											77
<i>totals</i>											133

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

## 2.7 ramazan

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	16	31	4	3	0	2	0	0	0	0	56
<∅ .2>	10	6	0	0	0	0	0	0	0	0	16
<∅ .h>	24	23	3	5	2	0	0	0	0	0	57
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	5	0	17	1	1	0	0	0	2	26
<pro .1>	5	6	2	2	3	0	0	5	0	0	23
<pro .2>	2	1	0	0	1	0	0	0	0	0	4
<pro .h>	2	4	0	0	2	2	1	0	0	0	11
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	5	0	5	2	1	1	0	0	4	18
<np .h>	4	14	0	2	4	3	0	2	6	3	38
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	0	34	0	26	6	5	12	10	12	22	127
<other .h>	1	1	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	5	3	6	7	0	7	0	28
<i>totals</i>	64	130	9	65	24	20	21	17	25	31	
<##>											119
<#>											90
<i>totals</i>											209

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

2.8 *tape*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	3	15	0	0	0	0	0	0	0	0	18
<∅ .2>	4	2	0	0	0	0	0	0	0	0	6
<∅ .h>	2	11	0	0	0	0	0	0	0	0	13
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	4	0	1	0	0	0	0	0	0	5
<pro .1>	1	2	2	0	0	0	0	3	0	0	8
<pro .2>	0	2	0	0	1	0	0	0	0	0	3
<pro .h>	1	6	0	0	1	0	0	0	0	0	8
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	1	1	0	0	0	1	0	3
<np .h>	1	4	0	0	0	0	0	1	2	0	8
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	0	5	0	5	0	1	2	4	1	4	22
<other .h>	0	0	0	0	0	0	0	0	1	0	1
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	2	0	1	1	5	4	0	5	0	18
<i>totals</i>	12	53	2	8	4	6	6	8	10	4	
<##>											43
<#>											25
<i>totals</i>											68

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



# Multi-CAST

*Multilingual Corpus of Annotated Spoken Texts*



[multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/)