

# Multi-CAST

*Bora*

*corpus counts*

---

*Frank Seifart*

*Tai Hong*

*July 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*

Seifart, Frank & Hong, Tai. 2022. Multi-CAST Bora corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. ([multicast.aspra.uni-bamberg.de/#bora](http://multicast.aspra.uni-bamberg.de/#bora)) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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 Bora corpus counts* v1.0 last updated 10 July 2022

This document was typeset by NNS with X<sub>q</sub>L<sup>A</sup>T<sub>E</sub>X and the *multicast3* class (v3.2.5).

## Contents

<b>1</b>	<b>Notes on the GRAID counts</b>	1
<b>2</b>	<b>The Bora corpus</b>	2
2.1	<i>ajyuwa</i>	3
2.2	<i>meenujkatsi</i>	4



## 1 Notes on the GRAID counts

This document collects tables with frequency counts for combinations of selected GRAID symbols in version 2207 (from July 2022) of the Multi-CAST Bora 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 Bora corpus

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>	6	3	0	1	0	0	0	0	0	0	10
<∅ .h>	8	2	0	35	0	0	0	0	0	0	45
<∅ .d>	2	2	0	2	0	0	0	0	0	0	6
<∅>	4	2	0	37	0	0	1	0	1	0	45
<pro .1>	40	32	0	28	4	0	0	0	37	3	144
<pro .2>	79	40	0	15	1	0	0	0	13	1	149
<pro .h>	442	196	1	42	7	2	0	0	120	6	816
<pro .d>	44	31	0	6	2	0	0	0	4	1	88
<pro>	25	6	1	18	20	8	25	5	5	9	122
<np .h>	38	23	0	38	4	12	0	5	0	18	138
<np .d>	13	5	0	8	0	0	0	2	0	3	31
<np>	52	11	0	102	85	30	24	33	0	125	462
<other .h>	0	0	0	0	1	0	0	0	0	0	1
<other .d>	0	0	0	0	1	0	0	0	0	0	1
<other>	9	0	0	5	5	0	0	29	0	1123	1171
<i>totals</i>	762	353	2	337	130	52	50	74	180	1289	
<##>											695
<#>											548
<i>totals</i>											1243

**Table 1** Summarized GRAID counts for the entire Bora corpus.

2.1 *ajyuwa*

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>	6	0	0	1	0	0	0	0	0	0	7
<∅ .h>	7	1	0	31	0	0	0	0	0	0	39
<∅ .d>	2	2	0	1	0	0	0	0	0	0	5
<∅>	3	1	0	21	0	0	0	0	1	0	26
<pro .1>	34	24	0	12	4	0	0	0	28	3	105
<pro .2>	74	31	0	12	1	0	0	0	10	1	129
<pro .h>	333	149	1	32	7	2	0	0	88	5	617
<pro .d>	41	26	0	6	2	0	0	0	4	1	80
<pro>	12	4	1	14	18	8	25	4	0	9	95
<np .h>	28	16	0	22	1	12	0	3	0	8	90
<np .d>	12	4	0	6	0	0	0	2	0	2	26
<np>	39	3	0	72	51	19	23	20	0	92	319
<other .h>	0	0	0	0	1	0	0	0	0	0	1
<other .d>	0	0	0	0	1	0	0	0	0	0	1
<other>	8	0	0	5	4	0	0	24	0	801	842
<i>totals</i>	599	261	2	235	90	41	48	53	131	922	
<##>											553
<#>											404
<i>totals</i>											957

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

## 2.2 meenujkatsi

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	3	0	0	0	0	0	0	0	0	3
<∅ .h>	1	1	0	4	0	0	0	0	0	0	6
<∅ .d>	0	0	0	1	0	0	0	0	0	0	1
<∅>	1	1	0	16	0	0	1	0	0	0	19
<pro .1>	6	8	0	16	0	0	0	0	9	0	39
<pro .2>	5	9	0	3	0	0	0	0	3	0	20
<pro .h>	109	47	0	10	0	0	0	0	32	1	199
<pro .d>	3	5	0	0	0	0	0	0	0	0	8
<pro>	13	2	0	4	2	0	0	1	5	0	27
<np .h>	10	7	0	16	3	0	0	2	0	10	48
<np .d>	1	1	0	2	0	0	0	0	0	1	5
<np>	13	8	0	30	34	11	1	13	0	33	143
<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>	1	0	0	0	1	0	0	5	0	322	329
<i>totals</i>	163	92	0	102	40	11	2	21	49	367	
<##>											142
<#>											144
<i>totals</i>											286

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





# Multi-CAST

*Multilingual Corpus of Annotated Spoken Texts*



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