

# Multi-CAST

*Teop*  
*corpus counts*

---

*Ulrike Mosel*  
University of Kiel

May 2019  
v1.0

*Stefan Schnell*  
University of Bamberg



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*

Mosel, Ulrike & Schnell, Stefan. 2019. Multi-CAST Teop 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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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 licensing terms can be reviewed online at [creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

*Multi-CAST Teop corpus counts* v1.0 last updated 9 May 2019

This document was typeset by NNS with X<sub>Y</sub>L<sup>A</sup>T<sub>E</sub>X and the *multicast3* class (v3.0.9045).

## Contents

<b>1</b>	<b>Notes on the GRAID counts</b>	1
<b>2</b>	<b>The Teop corpus</b>	2
2.1	<i>iar</i>	3
2.2	<i>mat</i>	4
2.3	<i>sii</i>	5
2.4	<i>viv</i>	6



## 1 Notes on the GRAID counts

This document collects tables with frequency counts for combinations of selected GRAID symbols in version 1905 (from May 2019) of the Multi-CAST Teop corpus. The tables are intended to offer 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*

⟨: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.

The tables in this document can be recreated with the `mc_table` function from *multicastR*, the companion package to Multi-CAST for the statistical computing language R. Please refer to the package documentation (`?multicastR`, `?mc_table`) for more information.

## 2 The Teop corpus

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	1	2	0	0	0	0	0	0	0	0	3
<∅ .2>	15	17	0	0	0	0	0	0	0	0	32
<∅ .h>	87	196	0	44	0	0	0	0	0	0	327
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	48	0	54	2	0	0	0	0	0	110
<pro .1>	64	63	0	4	0	0	0	25	0	0	156
<pro .2>	24	24	0	6	1	0	0	16	1	0	72
<pro .h>	161	146	0	55	0	0	0	108	1	0	471
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	3	30	0	38	6	1	2	26	1	4	111
<np .h>	55	122	0	53	5	0	0	21	35	1	292
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	7	65	0	183	17	42	45	37	48	61	505
<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	7	3	0	126	0	136
<i>totals</i>	423	713	0	437	31	50	50	233	212	66	
<##>											1079
<#>											219
<i>totals</i>											1298

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

2.1 *iar*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	1	2	0	0	0	0	0	0	0	0	3
<∅ .2>	6	3	0	0	0	0	0	0	0	0	9
<∅ .h>	15	49	0	14	0	0	0	0	0	0	78
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	5	22	0	7	2	0	0	0	0	0	36
<pro .1>	23	8	0	0	0	0	0	17	0	0	48
<pro .2>	10	5	0	0	1	0	0	3	0	0	19
<pro .h>	44	29	0	19	0	0	0	45	0	0	137
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	16	0	16	4	0	1	2	1	1	41
<np .h>	16	22	0	16	3	0	0	10	11	0	78
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	3	26	0	54	3	15	18	7	22	26	174
<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	2	0	0	33	0	35
<i>totals</i>	123	182	0	126	13	17	19	84	67	27	
<##>											290
<#>											58
<i>totals</i>											348

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

2.2 *mat*

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>	5	5	0	0	0	0	0	0	0	0	10
<∅ .h>	14	44	0	4	0	0	0	0	0	0	62
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	8	0	9	0	0	0	0	0	0	17
<pro .1>	6	6	0	1	0	0	0	4	0	0	17
<pro .2>	0	6	0	0	0	0	0	3	0	0	9
<pro .h>	12	17	0	4	0	0	0	7	0	0	40
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	1	0	1	0	0	0	6	0	0	9
<np .h>	10	26	0	7	1	0	0	5	4	0	53
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	1	11	0	29	0	7	8	8	6	5	75
<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	17	0	17
<i>totals</i>	49	124	0	55	1	7	8	33	27	5	
<##>											174
<#>											33
<i>totals</i>											207

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



2.3 *sii*

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>	4	7	0	0	0	0	0	0	0	0	11
<∅ .h>	33	75	0	8	0	0	0	0	0	0	116
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	1	15	0	38	0	0	0	0	0	0	54
<pro .1>	30	46	0	3	0	0	0	4	0	0	83
<pro .2>	13	11	0	6	0	0	0	10	1	0	41
<pro .h>	82	83	0	23	0	0	0	29	0	0	217
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	12	0	18	2	1	1	13	0	2	51
<np .h>	19	54	0	23	1	0	0	3	14	1	115
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	1	25	0	79	9	11	14	17	19	16	191
<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	5	2	0	62	0	69
<i>totals</i>	185	328	0	198	12	17	17	76	96	19	
<##>											492
<#>											95
<i>totals</i>											587

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

2.4 *viv*

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>	25	28	0	18	0	0	0	0	0	0	71
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	3	0	0	0	0	0	0	0	0	3
<pro .1>	5	3	0	0	0	0	0	0	0	0	8
<pro .2>	1	2	0	0	0	0	0	0	0	0	3
<pro .h>	23	17	0	9	0	0	0	27	1	0	77
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	1	0	3	0	0	0	5	0	1	10
<np .h>	10	20	0	7	0	0	0	3	6	0	46
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	2	3	0	21	5	9	5	5	1	14	65
<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	1	0	14	0	15
<i>totals</i>	66	79	0	58	5	9	6	40	22	15	
<##>											123
<#>											33
<i>totals</i>											156

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



# Multi-CAST

*Multilingual Corpus of Annotated Spoken Texts*



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

