

# Multi-CAST

*Teop*

*corpus counts*

---

*Ulrike Mosel*

*Stefan Schnell*

*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*

Mosel, Ulrike & Schnell, Stefan. 2021. Multi-CAST Teop corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. ([multicast.aspra.uni-bamberg.de/#teop](http://multicast.aspra.uni-bamberg.de/#teop)) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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 Teop corpus counts* v1.3 last updated 13 September 2021  
This document was typeset by NNS with  $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$  and the *multicast3* class (v3.2.5).

## Contents

<b>1</b>	<b>Notes on the GRAID counts</b>	1
<b>2</b>	<b>The 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 2108 (from August 2021) of the Multi-CAST Teop 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 Teop corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	3	1	0	0	0	0	0	0	0	0	4
<∅ .2>	20	15	0	0	0	0	0	0	0	0	35
<∅ .h>	188	90	0	44	0	0	0	0	0	0	322
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	51	6	0	54	2	0	0	0	0	0	113
<pro .1>	61	64	0	4	0	0	0	2	24	0	155
<pro .2>	24	24	0	6	0	0	1	1	17	0	73
<pro .h>	143	163	0	55	0	0	0	1	116	1	479
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	30	3	0	38	4	0	3	1	36	4	119
<np .h>	121	55	0	53	8	1	0	38	21	1	298
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	67	7	0	186	8	44	63	49	40	50	514
<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>	0	0	0	0	6	8	2	130	0	924	1070
<i>totals</i>	709	428	0	440	28	53	69	222	254	980	
<##>											1080
<#>											223
<i>totals</i>											1303

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

2.1 *iar*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	2	1	0	0	0	0	0	0	0	0	3
<∅ .2>	6	6	0	0	0	0	0	0	0	0	12
<∅ .h>	43	18	0	14	0	0	0	0	0	0	75
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	24	5	0	7	2	0	0	0	0	0	38
<pro .1>	8	23	0	0	0	0	0	0	16	0	47
<pro .2>	5	10	0	0	0	0	1	0	3	0	19
<pro .h>	26	47	0	19	0	0	0	0	54	0	146
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	16	0	0	17	2	0	1	1	7	1	45
<np .h>	21	17	0	16	6	1	0	15	11	0	87
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	28	3	0	54	1	14	20	20	10	26	176
<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	1	0	36	0	269	306
<i>totals</i>	179	130	0	127	11	16	22	72	101	296	
<##>											289
<#>											59
<i>totals</i>											348

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

2.2 *mat*

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>	5	5	0	0	0	0	0	0	0	0	10
<∅ .h>	44	14	0	4	0	0	0	0	0	0	62
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	9	0	0	9	0	0	0	0	0	0	18
<pro .1>	6	6	0	1	0	0	0	0	4	0	17
<pro .2>	6	0	0	0	0	0	0	0	4	0	10
<pro .h>	17	12	0	4	0	0	0	0	8	0	41
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	1	0	1	0	0	0	0	6	0	9
<np .h>	26	10	0	7	1	0	0	5	4	0	53
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	10	1	0	29	0	7	8	6	8	5	74
<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	17	0	150	168
<i>totals</i>	124	49	0	55	1	7	9	28	34	155	
<##>											174
<#>											33
<i>totals</i>											207

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



2.3 *sii*

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>	7	4	0	0	0	0	0	0	0	0	11
<∅ .h>	73	33	0	8	0	0	0	0	0	0	114
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	15	1	0	38	0	0	0	0	0	0	54
<pro .1>	44	30	0	3	0	0	0	2	4	0	83
<pro .2>	11	13	0	6	0	0	0	1	10	0	41
<pro .h>	83	82	0	23	0	0	0	0	30	0	218
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	12	2	0	19	2	0	1	0	15	2	53
<np .h>	55	18	0	23	1	0	0	12	3	1	113
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	26	1	0	81	6	13	21	22	18	10	198
<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>	0	0	0	0	0	7	0	63	0	387	457
<i>totals</i>	328	184	0	201	9	20	22	100	80	400	
<##>											494
<#>											96
<i>totals</i>											590

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

2.4 *viv*

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>	2	0	0	0	0	0	0	0	0	0	2
<∅ .h>	28	25	0	18	0	0	0	0	0	0	71
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	3	0	0	0	0	0	0	0	0	0	3
<pro .1>	3	5	0	0	0	0	0	0	0	0	8
<pro .2>	2	1	0	0	0	0	0	0	0	0	3
<pro .h>	17	22	0	9	0	0	0	1	24	1	74
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	0	1	0	0	1	0	8	1	12
<np .h>	19	10	0	7	0	0	0	6	3	0	45
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	3	2	0	22	1	10	14	1	4	9	66
<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	6	0	1	14	0	118	139
<i>totals</i>	78	65	0	57	7	10	16	22	39	129	
<##>											123
<#>											35
<i>totals</i>											158

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/)