

Multi-CAST

Teop
corpus counts

Ulrike Mosel

Stefan Schnell

August 2019
v1.1



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/) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Teop corpus counts v1.1 last updated 28 August 2019
This document was typeset by NNS with $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and the *multicast3* class (v3.2.1).

Contents

1	Notes on the GRAID counts	1
2	The Teop corpus	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 1908 (from August 2019) 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

⟨: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 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	19	0	0	0	0	0	0	0	0	34
<∅ .h>	90	191	0	44	0	0	0	0	0	0	325
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	49	0	54	2	0	0	0	0	0	111
<pro .1>	64	61	0	4	0	0	0	24	2	0	155
<pro .2>	24	24	0	6	0	0	1	17	1	0	73
<pro .h>	163	143	0	55	0	0	0	116	1	1	479
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	3	30	0	38	4	0	3	36	1	4	119
<np .h>	55	121	0	53	8	1	0	21	39	0	298
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	7	68	0	187	8	44	63	40	49	50	516
<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	0	0	0	6	8	2	0	132	0	148
<i>totals</i>	428	709	0	441	28	53	69	254	225	55	
<##>											1079
<#>											223
<i>totals</i>											1302

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	5	0	0	0	0	0	0	0	0	11
<∅ .h>	18	45	0	14	0	0	0	0	0	0	77
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	5	23	0	7	2	0	0	0	0	0	37
<pro .1>	23	8	0	0	0	0	0	16	0	0	47
<pro .2>	10	5	0	0	0	0	1	3	0	0	19
<pro .h>	47	26	0	19	0	0	0	54	0	0	146
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	16	0	17	2	0	1	7	1	1	45
<np .h>	17	21	0	16	6	1	0	11	15	0	87
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	3	28	0	54	1	14	20	10	20	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	0	36	0	37
<i>totals</i>	130	179	0	127	11	16	22	101	72	27	
<##>											289
<#>											59
<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	4	0	0	10
<pro .h>	12	17	0	4	0	0	0	8	0	0	41
<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	4	5	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	1	0	17	0	18
<i>totals</i>	49	124	0	55	1	7	9	34	28	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	74	0	8	0	0	0	0	0	0	115
<∅ .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	44	0	3	0	0	0	4	2	0	83
<pro .2>	13	11	0	6	0	0	0	10	1	0	41
<pro .h>	82	83	0	23	0	0	0	30	0	0	218
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	12	0	19	2	0	1	15	0	2	53
<np .h>	18	55	0	23	1	0	0	3	13	0	113
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	1	26	0	82	6	13	21	18	22	10	199
<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	0	0	0	0	7	0	0	65	0	72
<i>totals</i>	184	328	0	202	9	20	22	80	103	12	
<##>											493
<#>											96
<i>totals</i>											589

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>	22	17	0	9	0	0	0	24	1	1	74
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	1	0	1	0	0	1	8	0	1	12
<np .h>	10	19	0	7	0	0	0	3	6	0	45
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	2	3	0	22	1	10	14	4	1	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	0	14	0	21
<i>totals</i>	65	78	0	57	7	10	16	39	22	11	
<##>											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/