Multi-CAST

Tulil

corpus counts

Chenxi Meng

January 2020
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

Meng, Chenxi. 2020. Multi-CAST Tulil 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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at `multicast.aspra.uni-bamberg.de/`.

Multi-CAST Tulil corpus counts v1.1 last updated 4 January 2020
This document was typeset by NNS with XeLaTeX and the *multicast3* class (v3.2.2).

Contents

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 Tulil 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
ø	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
⟨:obl⟩	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 Tulil corpus

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	118	132	0	0	0	0	0	0	0	0	250
⟨0 .2⟩	29	14	0	0	0	0	0	0	0	0	43
⟨0 .h⟩	122	132	0	0	0	0	0	0	0	0	254
⟨0 .d⟩	34	62	0	0	0	0	0	0	0	0	96
⟨0⟩	15	69	0	22	1	0	0	0	0	0	107
⟨pro .1⟩	17	57	0	33	8	0	0	74	1	0	190
⟨pro .2⟩	2	6	0	5	1	0	0	7	0	0	21
⟨pro .h⟩	13	28	0	52	14	7	6	62	1	0	183
⟨pro .d⟩	1	13	0	13	15	1	0	33	1	0	77
⟨pro⟩	11	49	0	98	9	0	5	18	5	4	199
⟨np .h⟩	10	20	0	7	10	1	1	8	11	1	69
⟨np .d⟩	9	27	0	4	5	0	0	6	6	0	57
⟨np⟩	24	119	0	170	49	66	91	12	74	39	644
⟨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	6	0	8	0	49	51	0	169	0	283
totals	405	734	0	412	112	124	154	220	268	44	
⟨##⟩											765
⟨#⟩											499
totals											1264

Table 1 Summarized GRAID counts for the entire Tulil corpus.

2.1 all1

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	1	2	0	0	0	0	0	0	0	0	3
⟨0 .2⟩	0	1	0	0	0	0	0	0	0	0	1
⟨0 .h⟩	12	21	0	0	0	0	0	0	0	0	33
⟨0 .d⟩	6	6	0	0	0	0	0	0	0	0	12
⟨0⟩	2	10	0	0	0	0	0	0	0	0	12
⟨pro .1⟩	0	1	0	1	0	0	0	1	0	0	3
⟨pro .2⟩	1	1	0	1	0	0	0	0	0	0	3
⟨pro .h⟩	0	1	0	3	1	1	0	3	0	0	9
⟨pro .d⟩	0	2	0	1	2	0	0	0	0	0	5
⟨pro⟩	5	1	0	10	0	0	1	7	0	0	24
⟨np .h⟩	1	1	0	0	0	0	0	0	0	0	2
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	6	5	0	18	10	11	20	2	12	5	89
⟨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	1	0	0	0	0	3	0	5	0	9
totals	34	53	0	34	13	12	24	13	17	5	
⟨##⟩											49
⟨#⟩											44
totals											93

Table 2 Summarized GRAID counts for the *all1* text.

2.2 *alrm*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	38	35	0	0	0	0	0	0	0	0	73
⟨0 .2⟩	13	3	0	0	0	0	0	0	0	0	16
⟨0 .h⟩	44	33	0	0	0	0	0	0	0	0	77
⟨0 .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0⟩	10	45	0	10	1	0	0	0	0	0	66
⟨pro .1⟩	10	14	0	19	1	0	0	14	1	0	59
⟨pro .2⟩	0	0	0	0	0	0	0	1	0	0	1
⟨pro .h⟩	7	2	0	12	4	0	0	15	0	0	40
⟨pro .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro⟩	4	35	0	53	2	0	4	9	0	3	110
⟨np .h⟩	4	4	0	2	1	0	0	3	2	1	17
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	8	61	0	49	11	22	25	8	16	19	219
⟨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	1	0	1	0	22	23	0	35	0	82
totals	138	233	0	146	20	44	52	50	54	23	
⟨##⟩											231
⟨#⟩											176
totals											407

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

2.3 *jkpp*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	67	78	0	0	0	0	0	0	0	0	145
⟨0 .2⟩	4	2	0	0	0	0	0	0	0	0	6
⟨0 .h⟩	29	46	0	0	0	0	0	0	0	0	75
⟨0 .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0⟩	3	10	0	7	0	0	0	0	0	0	20
⟨pro .1⟩	3	36	0	11	4	0	0	41	0	0	95
⟨pro .2⟩	0	0	0	2	0	0	0	0	0	0	2
⟨pro .h⟩	4	11	0	26	5	1	0	12	1	0	60
⟨pro .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro⟩	2	8	0	20	3	0	0	2	3	0	38
⟨np .h⟩	4	11	0	4	4	0	0	1	8	0	32
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	9	31	0	51	12	13	20	2	27	14	179
⟨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	2	0	2	0	20	12	0	72	0	108
totals	125	235	0	123	28	34	32	58	111	14	

⟨##⟩	270
⟨#⟩	144
totals	414

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

2.4 *lnsl*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	1	1	0	0	0	0	0	0	0	0	2
⟨0 .2⟩	2	1	0	0	0	0	0	0	0	0	3
⟨0 .h⟩	4	3	0	0	0	0	0	0	0	0	7
⟨0 .d⟩	8	19	0	0	0	0	0	0	0	0	27
⟨0⟩	0	2	0	0	0	0	0	0	0	0	2
⟨pro .1⟩	1	3	0	0	0	0	0	6	0	0	10
⟨pro .2⟩	0	4	0	0	0	0	0	0	0	0	4
⟨pro .h⟩	0	0	0	1	1	0	0	0	0	0	2
⟨pro .d⟩	1	5	0	10	10	1	0	13	0	0	40
⟨pro⟩	0	0	0	2	0	0	0	0	2	1	5
⟨np .h⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np .d⟩	4	7	0	1	1	0	0	3	5	0	21
⟨np⟩	0	10	0	5	0	3	9	0	5	0	32
⟨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	2	0	1	0	3	9	0	23	0	38
totals	21	57	0	20	12	7	18	22	35	1	
⟨##⟩											63
⟨#⟩											29
totals											92

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

2.5 *lrdw*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	8	16	0	0	0	0	0	0	0	0	24
⟨0 .2⟩	9	7	0	0	0	0	0	0	0	0	16
⟨0 .h⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0 .d⟩	20	37	0	0	0	0	0	0	0	0	57
⟨0⟩	0	2	0	2	0	0	0	0	0	0	4
⟨pro .1⟩	3	3	0	2	3	0	0	9	0	0	20
⟨pro .2⟩	1	1	0	2	1	0	0	6	0	0	11
⟨pro .h⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro .d⟩	0	6	0	2	3	0	0	20	1	0	32
⟨pro⟩	0	3	0	8	0	0	0	0	0	0	11
⟨np .h⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np .d⟩	5	20	0	3	4	0	0	3	1	0	36
⟨np⟩	1	9	0	28	9	11	10	0	7	0	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	2	0	4	4	0	23	0	33
totals	47	104	0	49	20	15	14	38	32	0	

⟨##⟩	87
⟨#⟩	70
totals	157

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

2.6 *sves*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	3	0	0	0	0	0	0	0	0	0	3
⟨0 .2⟩	1	0	0	0	0	0	0	0	0	0	1
⟨0 .h⟩	33	29	0	0	0	0	0	0	0	0	62
⟨0 .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0⟩	0	0	0	3	0	0	0	0	0	0	3
⟨pro .1⟩	0	0	0	0	0	0	0	3	0	0	3
⟨pro .2⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro .h⟩	2	14	0	10	3	5	6	32	0	0	72
⟨pro .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro⟩	0	2	0	5	4	0	0	0	0	0	11
⟨np .h⟩	1	4	0	1	5	1	1	4	1	0	18
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	0	3	0	19	7	6	7	0	7	1	50
⟨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	0	0	11	0	13
totals	40	52	0	40	19	12	14	39	19	1	
⟨##⟩											65
⟨#⟩											36
totals											101

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

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts

multicast.aspra.uni-bamberg.de/