Multi-CAST

Tabasaran

corpus counts

Natalia Bogomolova

Dmitry Ganenkov

Nils Norman Schiborr

January 2021
v1.0

University of Bamberg

DFG

Multi-CAST *Multilingual Corpus of Annotated Spoken Texts*

Citation for this document

Bogomolova, Natalia & Ganenkov, Dmitry & Schiborr, Nils N. 2021. Multi-CAST Tabasaran corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts.* (`multicast.aspra.uni-bamberg.de/#tabasaran`) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at `multicast.aspra.uni-bamberg.de/`.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 Tabasaran corpus counts v1.0 last updated 27 January 2021
This document was typeset by NNS with XeLaTeX and the *multicast3* class (v3.2.3).

Contents

1 Notes on the GRAID counts

This document collects tables with frequency counts for combinations of selected GRAID symbols in version `2101` (from January 2021) of the Multi-CAST Tabasaran 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 Tabasaran corpus

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	66	18	11	2	1	6	0	0	0	0	104
⟨0 .2⟩	67	38	9	0	2	15	0	0	0	0	131
⟨0 .h⟩	268	174	21	13	9	8	0	0	1	0	494
⟨0 .d⟩	18	6	0	1	0	0	1	0	0	0	26
⟨0⟩	3	13	0	93	1	1	0	0	1	1	113
⟨pro .1⟩	20	13	6	4	8	9	0	37	0	0	97
⟨pro .2⟩	19	13	7	2	5	5	0	28	0	3	82
⟨pro .h⟩	65	64	21	12	15	22	0	27	1	2	229
⟨pro .d⟩	3	3	0	1	0	0	0	0	0	0	7
⟨pro⟩	0	5	0	5	5	7	12	4	2	30	70
⟨np .h⟩	67	105	14	27	26	32	0	37	17	16	341
⟨np .d⟩	6	5	0	0	0	0	2	0	0	2	15
⟨np⟩	1	162	4	342	36	89	66	19	14	105	838
⟨other .h⟩	3	4	2	0	1	3	0	3	3	0	19
⟨other .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨other⟩	0	13	0	30	1	4	3	0	53	3	107
totals	606	636	95	532	110	201	84	155	92	162	
⟨##⟩											869
⟨#⟩											514
totals											1383

Table 1 Summarized GRAID counts for the entire Tabasaran corpus.

2.1 *belt*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	1	0	0	0	0	0	0	0	0	0	1
⟨0 .2⟩	8	4	0	0	0	0	0	0	0	0	12
⟨0 .h⟩	30	38	8	3	0	0	0	0	1	0	80
⟨0 .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0⟩	0	3	0	10	0	0	0	0	0	0	13
⟨pro .1⟩	1	1	2	1	0	1	0	1	0	0	7
⟨pro .2⟩	2	1	1	1	0	0	0	3	0	0	8
⟨pro .h⟩	7	4	1	0	0	0	0	6	0	0	18
⟨pro .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro⟩	0	1	0	2	2	3	0	0	1	0	9
⟨np .h⟩	11	18	3	4	3	7	0	7	4	2	59
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	0	12	1	31	2	13	10	4	8	11	92
⟨other .h⟩	3	0	2	0	0	1	0	2	3	0	11
⟨other .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨other⟩	0	0	0	5	0	0	0	0	7	0	12
totals	63	82	18	57	7	25	10	23	24	13	
⟨##⟩											95
⟨#⟩											75
totals											170

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

2.2 horse

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	8	7	4	2	1	1	0	0	0	0	23
⟨0 .2⟩	18	4	3	0	0	1	0	0	0	0	26
⟨0 .h⟩	65	54	3	6	8	1	0	0	0	0	137
⟨0 .d⟩	18	6	0	1	0	0	1	0	0	0	26
⟨0⟩	1	4	0	32	0	0	0	0	1	1	39
⟨pro .1⟩	3	7	0	2	3	1	0	16	0	0	32
⟨pro .2⟩	4	5	2	0	2	2	0	6	0	2	23
⟨pro .h⟩	25	26	11	3	7	8	0	9	0	0	89
⟨pro .d⟩	3	3	0	1	0	0	0	0	0	0	7
⟨pro⟩	0	2	0	1	2	2	2	2	1	0	12
⟨np .h⟩	30	36	4	7	8	11	0	19	4	9	128
⟨np .d⟩	6	5	0	0	0	0	2	0	0	2	15
⟨np⟩	0	41	1	104	13	25	15	4	2	39	244
⟨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	10	0	3	0	0	22	0	37
totals	181	202	28	169	44	55	20	56	30	53	
⟨##⟩											267
⟨#⟩											155
totals											422

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

2.3 *naz*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	6	3	0	0	0	0	0	0	0	0	9
⟨0 .2⟩	10	7	0	0	0	1	0	0	0	0	18
⟨0 .h⟩	20	13	1	1	0	1	0	0	0	0	36
⟨0 .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0⟩	0	1	0	8	0	0	0	0	0	0	9
⟨pro .1⟩	1	1	0	1	0	1	0	3	0	0	7
⟨pro .2⟩	2	1	2	0	2	0	0	2	0	0	9
⟨pro .h⟩	4	1	4	0	0	1	0	1	0	0	11
⟨pro .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro⟩	0	0	0	0	0	0	2	0	0	1	3
⟨np .h⟩	8	8	4	4	3	8	0	2	4	0	41
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	0	12	0	31	2	4	10	0	0	10	69
⟨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	4	1	1	0	0	4	1	13
totals	51	49	11	49	8	17	12	8	8	12	
⟨##⟩											82
⟨#⟩											36
totals											118

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

2.4 *nuradin*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	17	3	4	0	0	0	0	0	0	0	24
⟨0 .2⟩	2	1	1	0	1	1	0	0	0	0	6
⟨0 .h⟩	34	25	5	1	0	0	0	0	0	0	65
⟨0 .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0⟩	0	0	0	12	0	0	0	0	0	0	12
⟨pro .1⟩	2	0	0	0	0	0	0	5	0	0	7
⟨pro .2⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro .h⟩	16	8	3	1	2	1	0	6	0	0	37
⟨pro .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro⟩	0	0	0	0	1	1	2	1	0	18	23
⟨np .h⟩	4	5	1	5	9	1	0	2	3	4	34
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	0	13	1	52	7	6	14	8	1	22	124
⟨other .h⟩	0	0	0	0	1	0	0	1	0	0	2
⟨other .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨other⟩	0	1	0	0	0	0	1	0	1	1	4
totals	75	56	15	71	21	10	17	23	5	45	
⟨##⟩											77
⟨#⟩											73
totals											150

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

2.5 *work*

GRAID	⟨:a⟩	⟨:s⟩	⟨:ncs⟩	⟨:p⟩	⟨:obl⟩	⟨:g⟩	⟨:l⟩	⟨:poss⟩	⟨:pred⟩	⟨:other⟩	*totals*
⟨0 .1⟩	34	5	3	0	0	5	0	0	0	0	47
⟨0 .2⟩	29	22	5	0	1	12	0	0	0	0	69
⟨0 .h⟩	119	44	4	2	1	6	0	0	0	0	176
⟨0 .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨0⟩	2	5	0	31	1	1	0	0	0	0	40
⟨pro .1⟩	13	4	4	0	5	6	0	12	0	0	44
⟨pro .2⟩	11	6	2	1	1	3	0	17	0	1	42
⟨pro .h⟩	13	25	2	8	6	12	0	5	1	2	74
⟨pro .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨pro⟩	0	2	0	2	0	1	6	1	0	11	23
⟨np .h⟩	14	38	2	7	3	5	0	7	2	1	79
⟨np .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨np⟩	1	84	1	124	12	41	17	3	3	23	309
⟨other .h⟩	0	4	0	0	0	2	0	0	0	0	6
⟨other .d⟩	0	0	0	0	0	0	0	0	0	0	0
⟨other⟩	0	8	0	11	0	0	2	0	19	1	41
totals	236	247	23	186	30	94	25	45	25	39	
⟨##⟩											348
⟨#⟩											175
totals											523

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

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts

multicast.aspra.uni-bamberg.de/