

Multi-CAST

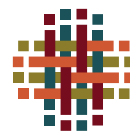
Tabasaran corpus counts

Natalia Bogomolova

Dmitry Ganenkov

Nils Norman Schiborr

September 2021
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

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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.

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 Tabasaran corpus counts v1.1 last updated 13 September 2021
This document was typeset by NNS with X_YL^AT_EX and the *multicast3* class (v3.2.5).

Contents

1	Notes on the GRAID counts	1
2	The Tabasaran corpus	2
2.1	<i>belt</i>	3
2.2	<i>horse</i>	4
2.3	<i>naz</i>	5
2.4	<i>nuradin</i>	6
2.5	<i>work</i>	7

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 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
∅	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 Tabasaran corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	18	66	11	2	1	6	0	0	0	0	104
<∅ .2>	38	67	9	0	2	15	0	0	0	0	131
<∅ .h>	174	268	21	13	9	8	0	1	0	0	494
<∅ .d>	6	18	0	1	0	0	1	0	0	0	26
<∅>	13	3	0	93	1	1	0	1	0	1	113
<pro .1>	13	20	6	4	8	9	0	0	37	0	97
<pro .2>	13	19	7	2	5	5	0	0	28	3	82
<pro .h>	64	65	21	12	15	22	0	1	27	2	229
<pro .d>	3	3	0	1	0	0	0	0	0	0	7
<pro>	5	0	0	5	5	7	12	2	4	30	70
<np .h>	105	67	14	27	26	32	0	17	37	16	341
<np .d>	5	6	0	0	0	0	2	0	0	2	15
<np>	162	1	4	342	36	89	66	14	19	105	838
<other .h>	4	3	2	0	1	3	0	3	3	0	19
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	13	0	0	30	1	4	3	53	0	1042	1146
<i>totals</i>	636	606	95	532	110	201	84	92	155	1201	
<##>											869
<#>											514
<i>totals</i>											1383

Table 1 Summarized GRAID counts for the entire Tabasaran corpus.

2.1 *belt*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	0	1	0	0	0	0	0	0	0	0	1
<∅ .2>	4	8	0	0	0	0	0	0	0	0	12
<∅ .h>	38	30	8	3	0	0	0	1	0	0	80
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	3	0	0	10	0	0	0	0	0	0	13
<pro .1>	1	1	2	1	0	1	0	0	1	0	7
<pro .2>	1	2	1	1	0	0	0	0	3	0	8
<pro .h>	4	7	1	0	0	0	0	0	6	0	18
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	0	2	2	3	0	1	0	0	9
<np .h>	18	11	3	4	3	7	0	4	7	2	59
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	12	0	1	31	2	13	10	8	4	11	92
<other .h>	0	3	2	0	0	1	0	3	2	0	11
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	5	0	0	0	7	0	83	95
<i>totals</i>	82	63	18	57	7	25	10	24	23	96	
<##>											95
<#>											75
<i>totals</i>											170

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

2.2 horse

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	7	8	4	2	1	1	0	0	0	0	23
<∅ .2>	4	18	3	0	0	1	0	0	0	0	26
<∅ .h>	54	65	3	6	8	1	0	0	0	0	137
<∅ .d>	6	18	0	1	0	0	1	0	0	0	26
<∅>	4	1	0	32	0	0	0	1	0	1	39
<pro .1>	7	3	0	2	3	1	0	0	16	0	32
<pro .2>	5	4	2	0	2	2	0	0	6	2	23
<pro .h>	26	25	11	3	7	8	0	0	9	0	89
<pro .d>	3	3	0	1	0	0	0	0	0	0	7
<pro>	2	0	0	1	2	2	2	1	2	0	12
<np .h>	36	30	4	7	8	11	0	4	19	9	128
<np .d>	5	6	0	0	0	0	2	0	0	2	15
<np>	41	0	1	104	13	25	15	2	4	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>	2	0	0	10	0	3	0	22	0	338	375
<i>totals</i>	202	181	28	169	44	55	20	30	56	391	
<##>											267
<#>											155
<i>totals</i>											422

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

2.3 *naz*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	3	6	0	0	0	0	0	0	0	0	9
<∅ .2>	7	10	0	0	0	1	0	0	0	0	18
<∅ .h>	13	20	1	1	0	1	0	0	0	0	36
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	1	0	0	8	0	0	0	0	0	0	9
<pro .1>	1	1	0	1	0	1	0	0	3	0	7
<pro .2>	1	2	2	0	2	0	0	0	2	0	9
<pro .h>	1	4	4	0	0	1	0	0	1	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	4	2	0	41
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	12	0	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>	2	0	0	4	1	1	0	4	0	74	86
<i>totals</i>	49	51	11	49	8	17	12	8	8	85	
<##>											82
<#>											36
<i>totals</i>											118

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

2.4 *nuradin*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	3	17	4	0	0	0	0	0	0	0	24
<∅ .2>	1	2	1	0	1	1	0	0	0	0	6
<∅ .h>	25	34	5	1	0	0	0	0	0	0	65
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	0	0	0	12	0	0	0	0	0	0	12
<pro .1>	0	2	0	0	0	0	0	0	5	0	7
<pro .2>	0	0	0	0	0	0	0	0	0	0	0
<pro .h>	8	16	3	1	2	1	0	0	6	0	37
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	1	1	2	0	1	18	23
<np .h>	5	4	1	5	9	1	0	3	2	4	34
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	13	0	1	52	7	6	14	1	8	22	124
<other .h>	0	0	0	0	1	0	0	0	1	0	2
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	1	0	0	0	0	0	1	1	0	82	85
<i>totals</i>	56	75	15	71	21	10	17	5	23	126	
<##>											77
<#>											73
<i>totals</i>											150

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

2.5 work

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	5	34	3	0	0	5	0	0	0	0	47
<∅ .2>	22	29	5	0	1	12	0	0	0	0	69
<∅ .h>	44	119	4	2	1	6	0	0	0	0	176
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	5	2	0	31	1	1	0	0	0	0	40
<pro .1>	4	13	4	0	5	6	0	0	12	0	44
<pro .2>	6	11	2	1	1	3	0	0	17	1	42
<pro .h>	25	13	2	8	6	12	0	1	5	2	74
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	2	0	0	2	0	1	6	0	1	11	23
<np .h>	38	14	2	7	3	5	0	2	7	1	79
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	84	1	1	124	12	41	17	3	3	23	309
<other .h>	4	0	0	0	0	2	0	0	0	0	6
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	8	0	0	11	0	0	2	19	0	465	505
<i>totals</i>	247	236	23	186	30	94	25	25	45	503	
<##>											348
<#>											175
<i>totals</i>											523

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

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/