

Multi-CAST

Northern Kurdish corpus counts

Geoffrey Haig

Maria Vollmer

Hanna Thiele

August 2019
v3.1



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

Haig, Geoffrey & Vollmer, Maria & Thiele, Hanna. 2019. Multi-CAST Northern Kurdish 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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Northern Kurdish corpus counts v3.1 last updated 13 August 2019
This document was typeset by NNS with X_YL^AT_EX and the *multicast3* class (v3.2.0).

Contents

1	Notes on the GRAID counts	1
2	The Northern Kurdish corpus	2
2.1	<i>muserz01</i>	3
2.2	<i>muserz02</i>	4
2.3	<i>muserz03</i>	5

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 Northern Kurdish 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 Northern Kurdish corpus

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	43	33	0	14	0	1	0	0	0	0	91
<∅ .2>	56	32	0	2	0	0	0	0	0	0	90
<∅ .h>	275	364	0	71	0	1	0	0	1	0	712
<∅ .d>	0	7	0	0	0	2	0	0	0	0	9
<∅>	22	63	0	135	0	4	0	0	0	0	224
<pro .1>	149	162	0	32	27	40	0	122	1	1	534
<pro .2>	58	72	0	19	14	27	0	44	2	0	236
<pro .h>	29	52	0	11	28	22	2	73	1	0	218
<pro .d>	0	1	0	0	1	1	0	0	0	0	3
<pro>	1	7	0	2	11	6	8	17	1	18	71
<np .h>	55	115	1	46	32	30	1	95	46	28	449
<np .d>	1	2	0	0	0	0	0	0	0	0	3
<np>	10	186	0	331	94	158	88	62	59	74	1062
<other .h>	1	5	0	0	1	0	0	0	0	0	7
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	2	4	1	22	5	43	15	3	125	0	220
<i>totals</i>	702	1105	2	685	213	335	114	416	236	121	
<##>											980
<#>											861
<i>totals</i>											1841

Table 1 Summarized GRAID counts for the entire Northern Kurdish corpus.

2.1 *muserz01*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	13	7	0	11	0	0	0	0	0	0	31
<∅ .2>	12	9	0	0	0	0	0	0	0	0	21
<∅ .h>	104	127	0	25	0	0	0	0	0	0	256
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	2	10	0	31	0	0	0	0	0	0	43
<pro .1>	44	50	0	8	6	9	0	42	0	1	160
<pro .2>	21	25	0	5	8	3	0	13	0	0	75
<pro .h>	15	22	0	2	18	14	0	32	1	0	104
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	4	5	6	7	1	1	24
<np .h>	23	44	0	18	20	9	0	40	26	22	202
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	2	79	0	136	20	55	40	16	18	25	391
<other .h>	0	0	0	0	1	0	0	0	0	0	1
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	0	0	3	4	17	7	0	48	0	79
<i>totals</i>	236	373	0	239	81	112	53	150	94	49	
<##>											158
<#>											469
<i>totals</i>											627

Table 2 Summarized GRAID counts for the *muserz01* text.

2.2 *muserz02*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	7	1	0	0	0	1	0	0	0	0	9
<∅ .2>	7	3	0	0	0	0	0	0	0	0	10
<∅ .h>	102	110	0	20	0	0	0	0	0	0	232
<∅ .d>	0	2	0	0	0	0	0	0	0	0	2
<∅>	6	33	0	55	0	2	0	0	0	0	96
<pro .1>	26	18	0	3	4	3	0	18	0	0	72
<pro .2>	10	14	0	0	2	6	0	7	0	0	39
<pro .h>	9	15	0	3	1	4	0	25	0	0	57
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	3	0	1	3	1	1	6	0	4	19
<np .h>	17	43	0	10	8	9	1	32	2	4	126
<np .d>	0	1	0	0	0	0	0	0	0	0	1
<np>	3	44	0	72	12	58	16	24	10	22	261
<other .h>	0	5	0	0	0	0	0	0	0	0	5
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	1	0	0	12	0	10	1	2	23	0	49
<i>totals</i>	188	292	0	176	30	94	19	114	35	30	
<##>											121
<#>											361
<i>totals</i>											482

Table 3 Summarized GRAID counts for the *muserz02* text.

2.3 *muserz03*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	23	25	0	3	0	0	0	0	0	0	51
<∅ .2>	37	20	0	2	0	0	0	0	0	0	59
<∅ .h>	69	127	0	26	0	1	0	0	1	0	224
<∅ .d>	0	5	0	0	0	2	0	0	0	0	7
<∅>	14	20	0	49	0	2	0	0	0	0	85
<pro .1>	79	94	0	21	17	28	0	62	1	0	302
<pro .2>	27	33	0	14	4	18	0	24	2	0	122
<pro .h>	5	15	0	6	9	4	2	16	0	0	57
<pro .d>	0	1	0	0	1	1	0	0	0	0	3
<pro>	1	4	0	1	4	0	1	4	0	13	28
<np .h>	15	28	1	18	4	12	0	23	18	2	121
<np .d>	1	1	0	0	0	0	0	0	0	0	2
<np>	5	63	0	123	62	45	32	22	31	27	410
<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>	1	4	1	7	1	16	7	1	54	0	92
<i>totals</i>	278	440	2	270	102	129	42	152	107	42	
<##>											701
<#>											31
<i>totals</i>											732

Table 4 Summarized GRAID counts for the *muserz03* text.

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/