

# Multi-CAST

## *Northern Kurdish corpus counts*

*Geoffrey Haig*

*Maria Vollmer*

*Hanna Thiele*

*September 2021*  
v3.4



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. 2021. 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/#nkurd](http://multicast.aspra.uni-bamberg.de/#nkurd)) (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/](http://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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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/](http://creativecommons.org/licenses/by/4.0/).

*Multi-CAST Northern Kurdish corpus counts* v3.4 last updated 13 September 2021  
This document was typeset by NNS with  $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$  and the *multicast3* class (v3.2.5).

## Contents

<b>1</b>	<b>Notes on the GRAID counts</b>	1
<b>2</b>	<b>The Northern Kurdish corpus</b>	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 2108 (from August 2021) 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*

⟨: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 Northern Kurdish corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	33	51	0	14	0	1	0	0	0	0	99
<∅ .2>	32	57	0	2	0	0	0	0	0	0	91
<∅ .h>	371	283	0	71	0	1	0	1	0	0	727
<∅ .d>	7	0	0	0	0	2	0	0	0	0	9
<∅>	63	22	0	135	0	4	0	0	0	0	224
<pro .1>	164	142	0	32	27	40	0	1	122	1	529
<pro .2>	72	58	0	19	14	27	0	2	44	0	236
<pro .h>	43	19	0	11	28	22	2	1	73	0	199
<pro .d>	1	0	0	0	1	1	0	0	0	0	3
<pro>	7	1	0	2	11	6	8	1	17	18	71
<np .h>	116	55	1	46	32	30	1	46	95	28	450
<np .d>	2	1	0	0	0	0	0	0	0	0	3
<np>	185	10	0	331	94	158	88	59	62	74	1061
<other .h>	5	1	0	0	1	0	0	0	0	0	7
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	4	2	1	22	5	43	15	125	3	794	1014
<i>totals</i>	1105	702	2	685	213	335	114	236	416	915	
<##>											980
<#>											861
<i>totals</i>											1841

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

2.1 *muserz01*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	10	21	0	11	0	0	0	0	0	0	42
<∅ .2>	9	12	0	0	0	0	0	0	0	0	21
<∅ .h>	133	112	0	25	0	0	0	0	0	0	270
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	10	2	0	31	0	0	0	0	0	0	43
<pro .1>	48	36	0	8	6	9	0	0	42	1	150
<pro .2>	25	22	0	5	8	3	0	0	13	0	76
<pro .h>	15	6	0	2	18	14	0	1	32	0	88
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	4	5	6	1	7	1	24
<np .h>	45	23	0	18	20	9	0	26	40	22	203
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	78	2	0	136	20	55	40	18	16	25	390
<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	48	0	286	365
<i>totals</i>	373	236	0	239	81	112	53	94	150	335	
<##>											158
<#>											469
<i>totals</i>											627

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

2.2 *muserz02*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	1	7	0	0	0	1	0	0	0	0	9
<∅ .2>	3	7	0	0	0	0	0	0	0	0	10
<∅ .h>	110	102	0	20	0	0	0	0	0	0	232
<∅ .d>	2	0	0	0	0	0	0	0	0	0	2
<∅>	33	6	0	55	0	2	0	0	0	0	96
<pro .1>	18	26	0	3	4	3	0	0	18	0	72
<pro .2>	14	10	0	0	2	6	0	0	7	0	39
<pro .h>	15	9	0	3	1	4	0	0	25	0	57
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	3	0	0	1	3	1	1	0	6	4	19
<np .h>	43	17	0	10	8	9	1	2	32	4	126
<np .d>	1	0	0	0	0	0	0	0	0	0	1
<np>	44	3	0	72	12	58	16	10	24	22	261
<other .h>	5	0	0	0	0	0	0	0	0	0	5
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	1	0	12	0	10	1	23	2	234	283
<i>totals</i>	292	188	0	176	30	94	19	35	114	264	
<##>											121
<#>											361
<i>totals</i>											482

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



2.3 *muserz03*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	22	23	0	3	0	0	0	0	0	0	48
<∅ .2>	20	38	0	2	0	0	0	0	0	0	60
<∅ .h>	128	69	0	26	0	1	0	1	0	0	225
<∅ .d>	5	0	0	0	0	2	0	0	0	0	7
<∅>	20	14	0	49	0	2	0	0	0	0	85
<pro .1>	98	80	0	21	17	28	0	1	62	0	307
<pro .2>	33	26	0	14	4	18	0	2	24	0	121
<pro .h>	13	4	0	6	9	4	2	0	16	0	54
<pro .d>	1	0	0	0	1	1	0	0	0	0	3
<pro>	4	1	0	1	4	0	1	0	4	13	28
<np .h>	28	15	1	18	4	12	0	18	23	2	121
<np .d>	1	1	0	0	0	0	0	0	0	0	2
<np>	63	5	0	123	62	45	32	31	22	27	410
<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>	4	1	1	7	1	16	7	54	1	274	366
<i>totals</i>	440	278	2	270	102	129	42	107	152	316	
<##>											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/](http://multicast.aspra.uni-bamberg.de/)