

# Multi-CAST

*Nafsan*  
*corpus counts*

---

*Nick Thieberger*

*Timothy Brickell*

*August 2019*

v1.0



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*

Thieberger, Nick & Brickell, Timothy. 2019. Multi-CAST Nafsan corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. ([multicast.aspra.uni-bamberg.de/](http://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/](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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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 Nafsan corpus counts* v1.0 last updated 5 August 2019  
This document was typeset by NNS with  $\text{Xe}_{\text{L}}\text{A}_{\text{T}}\text{E}_{\text{X}}$  and the *multicast3* class (v3.2.0).

## Contents

<b>1</b>	<b>Notes on the GRAID counts</b>	<hr/>	<b>1</b>
<b>2</b>	<b>The Nafsan corpus</b>	<hr/>	<b>2</b>
2.1	<i>kori</i>		3
2.2	<i>lelep</i>		4
2.3	<i>lisau</i>		5
2.4	<i>litog</i>		6
2.5	<i>maal</i>		7
2.6	<i>nmatu</i>		8
2.7	<i>ntwam</i>		9
2.8	<i>taapes</i>		10
2.9	<i>tafra</i>		11



## 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 Nafsan 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 Nafsan corpus

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	46	30	0	1	0	1	0	0	0	0	78
<∅ .2>	44	24	0	2	0	0	0	0	0	0	70
<∅ .h>	140	121	0	11	0	0	0	0	0	0	272
<∅ .d>	98	149	0	24	0	0	0	0	0	0	271
<∅>	30	52	0	72	7	0	0	0	0	0	161
<pro .1>	15	10	0	10	4	1	0	9	0	0	49
<pro .2>	6	11	0	12	4	1	0	5	0	0	39
<pro .h>	7	4	0	15	11	3	0	37	0	0	77
<pro .d>	2	9	0	3	1	0	0	17	0	0	32
<pro>	3	9	0	25	19	1	10	12	2	4	85
<np .h>	22	34	0	48	1	1	0	6	7	1	120
<np .d>	21	42	0	14	2	1	0	5	7	0	92
<np>	7	60	0	180	14	50	38	31	26	52	458
<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	0	0	4	2	0	5	0	11
<i>totals</i>	441	555	0	417	63	63	50	122	47	57	
<##>											778
<#>											234
<i>totals</i>											1012

**Table 1** Summarized GRAID counts for the entire Nafsan corpus.

## 2.1 kori

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	18	17	0	1	0	0	0	0	0	0	36
<∅ .2>	12	7	0	0	0	0	0	0	0	0	19
<∅ .h>	24	38	0	0	0	0	0	0	0	0	62
<∅ .d>	25	65	0	17	0	0	0	0	0	0	107
<∅>	13	10	0	14	0	0	0	0	0	0	37
<pro .1>	2	2	0	3	2	0	0	0	0	0	9
<pro .2>	1	4	0	1	2	0	0	0	0	0	8
<pro .h>	2	0	0	1	1	0	0	17	0	0	21
<pro .d>	0	2	0	0	0	0	0	7	0	0	9
<pro>	0	1	0	0	1	0	6	2	0	0	10
<np .h>	1	15	0	7	0	0	0	1	1	0	25
<np .d>	3	8	0	2	0	0	0	0	0	0	13
<np>	2	9	0	49	5	20	11	4	3	21	124
<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	0	0	2	1	0	1	0	4
<i>totals</i>	103	178	0	95	11	22	18	31	5	21	
<##>											230
<#>											54
<i>totals</i>											284

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

## 2.2 lelep

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	3	1	0	0	0	0	0	0	0	0	4
<∅ .2>	16	11	0	2	0	0	0	0	0	0	29
<∅ .h>	11	4	0	0	0	0	0	0	0	0	15
<∅ .d>	8	5	0	2	0	0	0	0	0	0	15
<∅>	3	22	0	20	1	0	0	0	0	0	46
<pro .1>	2	1	0	0	0	0	0	0	0	0	3
<pro .2>	1	0	0	3	1	0	0	1	0	0	6
<pro .h>	2	2	0	1	0	0	0	0	0	0	5
<pro .d>	1	3	0	0	0	0	0	0	0	0	4
<pro>	2	6	0	4	0	0	2	1	0	1	16
<np .h>	1	1	0	0	0	0	0	1	3	0	6
<np .d>	1	4	0	5	0	1	0	0	5	0	16
<np>	0	18	0	8	0	3	7	5	10	8	59
<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	0	0	0	1	0	1	0	2
<i>totals</i>	51	78	0	45	2	4	10	8	19	9	
<##>											107
<#>											22
<i>totals</i>											129

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



2.3 *lisau*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	1	0	0	0	0	0	0	0	0	0	1
<∅ .2>	2	1	0	0	0	0	0	0	0	0	3
<∅ .h>	18	5	0	2	0	0	0	0	0	0	25
<∅ .d>	6	3	0	3	0	0	0	0	0	0	12
<∅>	0	5	0	3	5	0	0	0	0	0	13
<pro .1>	0	1	0	0	1	0	0	0	0	0	2
<pro .2>	0	0	0	1	1	0	0	0	0	0	2
<pro .h>	1	1	0	2	1	1	0	6	0	0	12
<pro .d>	1	1	0	2	0	0	0	0	0	0	4
<pro>	0	0	0	1	2	0	2	0	0	0	5
<np .h>	4	2	0	9	0	0	0	0	0	0	15
<np .d>	0	3	0	2	0	0	0	0	1	0	6
<np>	0	3	0	8	0	0	1	1	3	0	16
<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	0	0	0	0	0	2	0	2
<i>totals</i>	33	25	0	33	10	1	3	7	6	0	
<##>											40
<#>											18
<i>totals</i>											58

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

2.4 *litog*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	1	1	0	0	0	0	0	0	0	0	2
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	16	21	0	7	0	0	0	0	0	0	44
<∅ .d>	16	16	0	2	0	0	0	0	0	0	34
<∅>	4	3	0	7	0	0	0	0	0	0	14
<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>	0	0	0	0	7	0	0	1	0	0	8
<pro .d>	0	2	0	0	1	0	0	4	0	0	7
<pro>	0	1	0	0	0	0	0	2	0	0	3
<np .h>	1	1	0	11	0	1	0	1	2	0	17
<np .d>	0	3	0	2	0	0	0	0	0	0	5
<np>	0	2	0	10	4	2	2	3	2	4	29
<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	0	0	2	0	0	0	0	2
<i>totals</i>	38	50	0	39	12	5	2	14	4	4	
<##>											77
<#>											9
<i>totals</i>											86

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

2.5 *maal*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	3	2	0	0	0	0	0	0	0	0	5
<∅ .2>	3	1	0	0	0	0	0	0	0	0	4
<∅ .h>	0	0	0	0	0	0	0	0	0	0	0
<∅ .d>	14	11	0	0	0	0	0	0	0	0	25
<∅>	0	0	0	7	1	0	0	0	0	0	8
<pro .1>	1	0	0	0	1	0	0	0	0	0	2
<pro .2>	0	0	0	1	0	0	0	0	0	0	1
<pro .h>	0	0	0	0	0	0	0	0	0	0	0
<pro .d>	0	1	0	0	0	0	0	0	0	0	1
<pro>	0	0	0	1	4	0	0	1	0	0	6
<np .h>	0	0	0	0	0	0	0	0	0	0	0
<np .d>	5	8	0	1	2	0	0	0	0	0	16
<np>	0	2	0	18	1	4	3	2	0	3	33
<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	0	0	0	0	0	0	0	0
<i>totals</i>	26	25	0	28	9	4	3	3	0	3	
<##>											36
<#>											16
<i>totals</i>											52

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

2.6 *nmatu*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	11	1	0	0	0	0	0	0	0	0	12
<∅ .2>	5	1	0	0	0	0	0	0	0	0	6
<∅ .h>	25	15	0	2	0	0	0	0	0	0	42
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	7	4	0	10	0	0	0	0	0	0	21
<pro .1>	2	0	0	1	0	1	0	4	0	0	8
<pro .2>	3	0	0	4	0	1	0	4	0	0	12
<pro .h>	0	0	0	3	0	0	0	5	0	0	8
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	6	2	0	0	1	0	1	10
<np .h>	6	5	0	15	0	0	0	2	0	0	28
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	1	2	0	19	1	4	0	5	2	5	39
<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	0	0	0	0	0	0	0	0
<i>totals</i>	60	28	0	60	3	6	0	21	2	6	
<##>											65
<#>											23
<i>totals</i>											88

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

2.7 *ntwam*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	5	6	0	0	0	0	0	0	0	0	11
<∅ .2>	2	3	0	0	0	0	0	0	0	0	5
<∅ .h>	40	34	0	0	0	0	0	0	0	0	74
<∅ .d>	13	9	0	0	0	0	0	0	0	0	22
<∅>	2	4	0	5	0	0	0	0	0	0	11
<pro .1>	5	3	0	3	0	0	0	0	0	0	11
<pro .2>	1	5	0	2	0	0	0	0	0	0	8
<pro .h>	1	1	0	3	0	1	0	6	0	0	12
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	0	10	5	1	0	1	0	2	20
<np .h>	6	9	0	5	0	0	0	0	1	0	21
<np .d>	5	9	0	1	0	0	0	0	1	0	16
<np>	1	10	0	42	2	9	7	1	5	5	82
<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	0	0	0	0	0	0	0	0
<i>totals</i>	82	93	0	71	7	11	7	8	7	7	
<##>											129
<#>											57
<i>totals</i>											186

**Table 8** Summarized GRAID counts for the *ntwam* text.

2.8 *taapes*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	0	0	0	0	0	0	0	0	0	0	0
<∅ .2>	1	0	0	0	0	0	0	0	0	0	1
<∅ .h>	0	0	0	0	0	0	0	0	0	0	0
<∅ .d>	14	31	0	0	0	0	0	0	0	0	45
<∅>	0	1	0	4	0	0	0	0	0	0	5
<pro .1>	0	1	0	0	0	0	0	0	0	0	1
<pro .2>	0	1	0	0	0	0	0	0	0	0	1
<pro .h>	0	0	0	0	0	0	0	0	0	0	0
<pro .d>	0	0	0	1	0	0	0	6	0	0	7
<pro>	0	1	0	2	1	0	0	3	2	0	9
<np .h>	0	0	0	0	0	0	0	0	0	0	0
<np .d>	7	2	0	1	0	0	0	5	0	0	15
<np>	1	7	0	11	0	8	5	7	1	3	43
<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	0	0	0	0	0	0	0	0
<i>totals</i>	23	44	0	19	1	8	5	21	3	3	
<##>											49
<#>											18
<i>totals</i>											67

Table 9 Summarized GRAID counts for the *taapes* text.

## 2.9 tafra

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	4	2	0	0	0	1	0	0	0	0	7
<∅ .2>	3	0	0	0	0	0	0	0	0	0	3
<∅ .h>	6	4	0	0	0	0	0	0	0	0	10
<∅ .d>	2	9	0	0	0	0	0	0	0	0	11
<∅>	1	3	0	2	0	0	0	0	0	0	6
<pro .1>	3	2	0	3	0	0	0	2	0	0	10
<pro .2>	0	1	0	0	0	0	0	0	0	0	1
<pro .h>	1	0	0	5	2	1	0	2	0	0	11
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	1	4	0	0	1	0	0	6
<np .h>	3	1	0	1	1	0	0	1	0	1	8
<np .d>	0	5	0	0	0	0	0	0	0	0	5
<np>	2	7	0	15	1	0	2	3	0	3	33
<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	0	0	0	0	0	1	0	1
<i>totals</i>	25	34	0	27	8	2	2	9	1	4	
<##>											45
<#>											17
<i>totals</i>											62

Table 10 Summarized GRAID counts for the *tafra* text.

# Multi-CAST

*Multilingual Corpus of Annotated Spoken Texts*



[multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/)