

Multi-CAST

Cypriot Greek corpus counts

Harris Hadjidas

Maria Vollmer

September 2021
v2.2



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

Hadjidas, Harris & Vollmer, Maria. 2021. Multi-CAST Cypriot Greek corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#cypgreek) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Cypriot Greek corpus counts v2.2 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

1	Notes on the GRAID counts	—————	1
2	The Cypriot Greek corpus	—————	2
2.1	<i>jitros</i>		3
2.2	<i>minaes</i>		4
2.3	<i>psarin</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 Cypriot Greek 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 Cypriot Greek corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	55	116	0	1	0	0	0	0	0	0	172
<∅ .2>	75	116	0	0	0	0	0	0	0	1	192
<∅ .h>	173	182	0	5	0	0	0	0	0	1	361
<∅ .d>	6	7	0	1	0	0	0	0	0	0	14
<∅>	16	13	0	72	0	0	0	2	0	0	103
<pro .1>	9	13	0	10	13	20	0	0	99	3	167
<pro .2>	9	5	0	15	13	21	0	0	45	3	111
<pro .h>	5	2	0	54	20	57	1	1	55	6	201
<pro .d>	2	0	0	0	0	0	0	0	1	0	3
<pro>	2	0	0	59	1	1	0	0	0	1	64
<np .h>	50	28	0	38	17	19	0	11	29	4	196
<np .d>	14	1	0	2	0	0	0	0	0	20	37
<np>	24	9	0	218	43	50	15	18	15	83	475
<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>	8	1	0	23	3	26	19	41	0	767	888
<i>totals</i>	448	493	0	498	110	194	35	73	244	889	
<##>											431
<#>											639
<i>totals</i>											1070

Table 1 Summarized GRAID counts for the entire Cypriot Greek corpus.

2.1 *jitros*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	10	27	0	0	0	0	0	0	0	0	37
<∅ .2>	17	32	0	0	0	0	0	0	0	0	49
<∅ .h>	48	52	0	0	0	0	0	0	0	0	100
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	3	9	0	16	0	0	0	0	0	0	28
<pro .1>	4	5	0	3	4	3	0	0	33	1	53
<pro .2>	3	1	0	6	1	4	0	0	9	0	24
<pro .h>	2	0	0	19	5	8	1	0	25	0	60
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	1	0	0	5	0	0	0	0	0	0	6
<np .h>	16	9	0	13	3	4	0	4	15	1	65
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	2	1	0	61	11	13	4	4	7	27	130
<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	12	1	8	5	16	0	194	238
<i>totals</i>	108	136	0	135	25	40	10	24	89	223	
<##>											104
<#>											167
<i>totals</i>											271

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

2.2 *minaes*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	29	31	0	1	0	0	0	0	0	0	61
<∅ .2>	38	39	0	0	0	0	0	0	0	1	78
<∅ .h>	31	40	0	1	0	0	0	0	0	0	72
<∅ .d>	6	7	0	1	0	0	0	0	0	0	14
<∅>	4	1	0	25	0	0	0	0	0	0	30
<pro .1>	4	3	0	3	7	12	0	0	29	2	60
<pro .2>	4	3	0	1	10	6	0	0	24	1	49
<pro .h>	1	2	0	7	5	13	0	0	19	4	51
<pro .d>	2	0	0	0	0	0	0	0	1	0	3
<pro>	1	0	0	12	1	0	0	0	0	0	14
<np .h>	14	6	0	13	7	10	0	0	9	1	60
<np .d>	14	1	0	2	0	0	0	0	0	20	37
<np>	12	4	0	74	14	15	1	3	3	37	163
<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>	1	1	0	5	0	9	0	6	0	338	360
<i>totals</i>	161	138	0	145	44	65	1	9	85	404	
<##>											161
<#>											198
<i>totals</i>											359

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

2.3 *psarin*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	16	58	0	0	0	0	0	0	0	0	74
<∅ .2>	20	45	0	0	0	0	0	0	0	0	65
<∅ .h>	94	90	0	4	0	0	0	0	0	1	189
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	9	3	0	31	0	0	0	2	0	0	45
<pro .1>	1	5	0	4	2	5	0	0	37	0	54
<pro .2>	2	1	0	8	2	11	0	0	12	2	38
<pro .h>	2	0	0	28	10	36	0	1	11	2	90
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	42	0	1	0	0	0	1	44
<np .h>	20	13	0	12	7	5	0	7	5	2	71
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	10	4	0	83	18	22	10	11	5	19	182
<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>	5	0	0	6	2	9	14	19	0	235	290
<i>totals</i>	179	219	0	218	41	89	24	40	70	262	
<##>											166
<#>											274
<i>totals</i>											440

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

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/