

# Multi-CAST

*English  
corpus counts*

---

*Nils Norman Schiborr*

*September 2021*  
v3.3



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*

Schiborr, Nils N. 2021. Multi-CAST English corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. ([multicast.aspra.uni-bamberg.de/#english](http://multicast.aspra.uni-bamberg.de/#english)) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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 English corpus counts* v3.3 last updated 13 September 2021  
This document was typeset by NNS with X<sub>Y</sub>L<sup>A</sup>T<sub>E</sub>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 English corpus</b>	2
2.1	<i>devon01</i>	3
2.2	<i>kent01</i>	4
2.3	<i>kent02</i>	5
2.4	<i>kent03</i>	6
2.5	<i>london01</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 English 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 English corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	140	222	0	1	0	0	0	0	0	0	363
<∅ .2>	51	138	0	0	0	0	0	0	0	0	189
<∅ .h>	109	252	0	10	0	1	0	1	0	1	374
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	163	49	0	90	10	8	2	20	0	6	348
<pro .1>	645	770	0	107	28	18	1	0	249	4	1822
<pro .2>	175	285	0	50	15	13	1	1	68	1	609
<pro .h>	530	468	0	83	26	11	1	0	113	2	1234
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	676	128	0	584	84	32	17	20	72	28	1641
<np .h>	194	136	0	111	71	46	3	104	37	36	738
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	361	36	0	1204	312	334	344	370	35	667	3663
<other .h>	12	15	0	11	1	1	0	1	0	5	46
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	28	8	0	91	17	137	125	877	0	5452	6735
<i>totals</i>	3084	2507	0	2342	564	601	494	1394	574	6202	
<##>											3907
<#>											1742
<i>totals</i>											5649

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

2.1 *devon01*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	22	28	0	0	0	0	0	0	0	0	50
<∅ .2>	3	9	0	0	0	0	0	0	0	0	12
<∅ .h>	11	22	0	0	0	1	0	0	0	0	34
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	20	6	0	13	2	2	0	0	0	0	43
<pro .1>	109	87	0	17	4	4	0	0	26	0	247
<pro .2>	13	19	0	5	3	0	0	0	5	0	45
<pro .h>	41	43	0	8	0	2	0	0	8	0	102
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	63	10	0	37	6	2	3	0	1	2	124
<np .h>	16	9	0	18	7	8	0	10	2	3	73
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	44	0	0	123	36	42	29	39	2	85	400
<other .h>	0	2	0	1	0	0	0	0	0	0	3
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	2	4	0	5	2	7	18	88	0	529	655
<i>totals</i>	344	239	0	227	60	68	50	137	44	619	
<##>											417
<#>											173
<i>totals</i>											590

**Table 2** Summarized GRAID counts for the *devon01* text.

## 2.2 kent01

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	17	41	0	0	0	0	0	0	0	0	58
<∅ .2>	5	18	0	0	0	0	0	0	0	0	23
<∅ .h>	13	28	0	1	0	0	0	0	0	0	42
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	14	6	0	10	1	4	1	7	0	0	43
<pro .1>	49	88	0	16	1	1	0	0	31	1	187
<pro .2>	12	22	0	1	0	0	0	0	6	0	41
<pro .h>	66	67	0	18	4	2	0	0	20	0	177
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	76	7	0	79	8	4	1	0	10	4	189
<np .h>	18	22	0	13	6	6	0	9	5	6	85
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	37	4	0	143	37	47	44	29	3	60	404
<other .h>	1	2	0	1	0	0	0	0	0	1	5
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	2	1	0	10	2	17	14	95	0	491	632
<i>totals</i>	310	306	0	292	59	81	60	140	75	563	
<##>											451
<#>											171
<i>totals</i>											622

**Table 3** Summarized GRAID counts for the *kent01* text.



## 2.3 kent02

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	31	79	0	0	0	0	0	0	0	0	110
<∅ .2>	14	46	0	0	0	0	0	0	0	0	60
<∅ .h>	30	90	0	1	0	0	0	0	0	0	121
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	38	11	0	23	1	2	0	5	0	1	81
<pro .1>	227	286	0	36	11	5	1	0	83	1	650
<pro .2>	31	81	0	17	3	4	1	0	19	1	157
<pro .h>	179	135	0	24	7	5	0	0	23	1	374
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	157	18	0	235	24	7	3	2	21	9	476
<np .h>	58	39	0	27	14	19	1	24	7	5	194
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	52	7	0	385	90	71	84	89	7	169	954
<other .h>	4	5	0	0	0	0	0	0	0	2	11
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	7	3	0	29	5	60	26	238	0	1326	1694
<i>totals</i>	828	800	0	777	155	173	116	358	160	1515	
<##>											1243
<#>											394
<i>totals</i>											1637

Table 4 Summarized GRAID counts for the kent02 text.

2.4 *kent03*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	36	50	0	0	0	0	0	0	0	0	86
<∅ .2>	11	29	0	0	0	0	0	0	0	0	40
<∅ .h>	16	49	0	2	0	0	0	0	0	0	67
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	57	20	0	23	4	0	1	4	0	3	112
<pro .1>	127	181	0	21	5	6	0	0	40	0	380
<pro .2>	42	72	0	14	6	7	0	1	5	0	147
<pro .h>	110	93	0	14	9	1	0	0	19	0	246
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	199	77	0	150	37	14	5	2	38	7	529
<np .h>	11	18	0	17	14	5	1	21	4	6	97
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	85	14	0	271	55	66	75	78	11	153	808
<other .h>	4	3	0	6	1	0	0	0	0	1	15
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	11	0	0	31	5	34	34	207	0	1412	1734
<i>totals</i>	709	606	0	549	136	133	116	313	117	1582	
<##>											871
<#>											464
<i>totals</i>											1335

**Table 5** Summarized GRAID counts for the *kent03* text.

## 2.5 london01

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<θ .1>	34	24	0	1	0	0	0	0	0	0	59
<θ .2>	18	36	0	0	0	0	0	0	0	0	54
<θ .h>	39	63	0	6	0	0	0	1	0	1	110
<θ .d>	0	0	0	0	0	0	0	0	0	0	0
<θ>	34	6	0	21	2	0	0	4	0	2	69
<pro .1>	133	128	0	17	7	2	0	0	69	2	358
<pro .2>	77	91	0	13	3	2	0	0	33	0	219
<pro .h>	134	130	0	19	6	1	1	0	43	1	335
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	181	16	0	83	9	5	5	16	2	6	323
<np .h>	91	48	0	36	30	8	1	40	19	16	289
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	143	11	0	282	94	108	112	135	12	200	1097
<other .h>	3	3	0	3	0	1	0	1	0	1	12
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	6	0	0	16	3	19	33	249	0	1694	2020
<i>totals</i>	893	556	0	497	154	146	152	446	178	1923	
<##>											925
<#>											540
<i>totals</i>											1465

Table 6 Summarized GRAID counts for the *london01* text.

# Multi-CAST

*Multilingual Corpus of Annotated Spoken Texts*



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