

# Multi-CAST

*English  
corpus counts*

---

*Nils Norman Schiborr*

*January 2020*  
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*

Schiborr, Nils N. 2020. Multi-CAST English 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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at [multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/).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 English corpus counts* v3.1 last updated 4 January 2020  
This document was typeset by NNS with  $\text{\LaTeX}$  and the *multicast3* class (v3.2.2).

## 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 2001 (from January 2020) 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*

⟨: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 English corpus

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	219	137	0	1	0	0	0	0	0	0	357
<∅ .2>	138	50	0	0	0	0	0	0	0	0	188
<∅ .h>	251	110	0	10	0	1	0	0	1	1	374
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	49	163	0	89	10	7	2	0	20	6	346
<pro .1>	780	650	0	107	28	18	1	249	0	4	1837
<pro .2>	287	175	0	50	15	13	1	67	1	1	610
<pro .h>	463	530	0	83	26	11	1	112	0	2	1228
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	128	674	0	586	84	31	18	72	20	27	1640
<np .h>	136	196	0	111	71	46	3	37	104	36	740
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	36	360	0	1206	312	334	344	35	371	668	3666
<other .h>	15	12	0	11	1	1	0	0	1	2	43
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	8	29	0	91	17	137	125	0	876	3	1286
<i>totals</i>	2510	3086	0	2345	564	599	495	572	1394	750	
<##>											3907
<#>											1742
<i>totals</i>											5649

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

2.1 *devon01*

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	28	22	0	0	0	0	0	0	0	0	50
<∅ .2>	9	2	0	0	0	0	0	0	0	0	11
<∅ .h>	22	12	0	0	0	1	0	0	0	0	35
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	20	0	13	2	2	0	0	0	0	43
<pro .1>	92	111	0	17	4	4	0	26	0	0	254
<pro .2>	20	13	0	5	3	0	0	5	0	0	46
<pro .h>	37	40	0	8	0	2	0	8	0	0	95
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	10	62	0	37	6	2	3	1	0	2	123
<np .h>	9	16	0	18	7	8	0	2	10	3	73
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	0	44	0	123	36	42	29	2	39	86	401
<other .h>	2	0	0	1	0	0	0	0	0	0	3
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	4	2	0	5	2	7	18	0	88	0	126
<i>totals</i>	239	344	0	227	60	68	50	44	137	91	
<##>											418
<#>											172
<i>totals</i>											590

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

## 2.2 kent01

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	40	16	0	0	0	0	0	0	0	0	56
<∅ .2>	18	5	0	0	0	0	0	0	0	0	23
<∅ .h>	27	13	0	1	0	0	0	0	0	0	41
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	14	0	10	1	4	1	0	7	0	43
<pro .1>	89	49	0	16	1	1	0	32	0	1	189
<pro .2>	22	12	0	1	0	0	0	6	0	0	41
<pro .h>	68	67	0	18	4	2	0	20	0	0	179
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	7	76	0	79	8	4	1	10	0	4	189
<np .h>	22	18	0	13	6	6	0	5	9	6	85
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	4	37	0	143	37	47	44	3	29	60	404
<other .h>	2	1	0	1	0	0	0	0	0	0	4
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	1	2	0	10	2	17	14	0	95	0	141
<i>totals</i>	306	310	0	292	59	81	60	76	140	71	
<##>											451
<#>											171
<i>totals</i>											622

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



## 2.3 kent02

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	78	30	0	0	0	0	0	0	0	0	108
<∅ .2>	46	14	0	0	0	0	0	0	0	0	60
<∅ .h>	90	30	0	1	0	0	0	0	0	0	121
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	11	38	0	23	1	1	0	0	5	1	80
<pro .1>	289	228	0	36	11	5	1	82	0	1	653
<pro .2>	82	31	0	17	3	4	1	18	0	1	157
<pro .h>	134	179	0	24	7	5	0	23	0	1	373
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	18	157	0	237	24	7	3	21	2	8	477
<np .h>	39	58	0	27	14	19	1	7	24	5	194
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	7	52	0	386	90	71	84	7	89	169	955
<other .h>	5	4	0	0	0	0	0	0	0	0	9
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	3	7	0	29	5	60	26	0	238	0	368
<i>totals</i>	802	828	0	780	155	172	116	158	358	186	
<##>											1243
<#>											394
<i>totals</i>											1637

Table 4 Summarized GRAID counts for the kent02 text.

## 2.4 kent03

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	49	35	0	0	0	0	0	0	0	0	84
<∅ .2>	29	11	0	0	0	0	0	0	0	0	40
<∅ .h>	49	16	0	2	0	0	0	0	0	0	67
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	20	57	0	22	4	0	1	0	4	3	111
<pro .1>	182	129	0	21	5	6	0	40	0	0	383
<pro .2>	72	42	0	14	6	7	0	5	1	0	147
<pro .h>	93	109	0	14	9	1	0	19	0	0	245
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	77	199	0	150	37	13	6	38	2	7	529
<np .h>	18	13	0	17	14	5	1	4	21	6	99
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	14	84	0	272	55	66	75	11	79	153	809
<other .h>	3	4	0	6	1	0	0	0	0	1	15
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	12	0	31	5	34	34	0	206	1	323
<i>totals</i>	606	711	0	549	136	132	117	117	313	171	
<##>											870
<#>											465
<i>totals</i>											1335

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

## 2.5 london01

GRAID	<:a>	<:s>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:poss>	<:pred>	<:other>	<i>totals</i>
<∅ .1>	24	34	0	1	0	0	0	0	0	0	59
<∅ .2>	36	18	0	0	0	0	0	0	0	0	54
<∅ .h>	63	39	0	6	0	0	0	0	1	1	110
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	6	34	0	21	2	0	0	0	4	2	69
<pro .1>	128	133	0	17	7	2	0	69	0	2	358
<pro .2>	91	77	0	13	3	2	0	33	0	0	219
<pro .h>	131	135	0	19	6	1	1	42	0	1	336
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	16	180	0	83	9	5	5	2	16	6	322
<np .h>	48	91	0	36	30	8	1	19	40	16	289
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	11	143	0	282	94	108	112	12	135	200	1097
<other .h>	3	3	0	3	0	1	0	0	1	1	12
<other .d>	0	0	0	0	0	0	0	0	0	0	0
<other>	0	6	0	16	3	19	33	0	249	2	328
<i>totals</i>	557	893	0	497	154	146	152	177	446	231	
<##>											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/)