

Multi-CAST

*Mandarin
corpus counts*

Maria Vollmer

September 2021
v1.2



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

Vollmer, Maria. 2021. Multi-CAST Mandarin corpus counts. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#mandarin) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Mandarin corpus counts v1.2 last updated 13 September 2021
This document was typeset by NNS with \LaTeX and the *multicast3* class (v3.2.5).

Contents

1	Notes on the GRAID counts	1
2	The Mandarin corpus	2
2.1	<i>hml</i>	3
2.2	<i>jgz</i>	4
2.3	<i>lzh</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 Mandarin 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 Mandarin corpus

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	30	48	0	2	3	0	0	0	0	0	83
<∅ .2>	18	24	0	0	1	0	0	0	0	0	43
<∅ .h>	223	205	0	28	0	2	0	0	0	0	458
<∅ .d>	1	0	0	0	0	0	0	0	0	0	1
<∅>	47	11	0	72	2	0	0	0	0	0	132
<pro .1>	20	29	0	7	5	3	0	0	20	2	86
<pro .2>	20	23	0	6	5	4	0	0	5	0	63
<pro .h>	50	48	0	14	11	0	0	0	37	1	161
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	20	4	0	7	3	1	6	0	0	3	44
<np .h>	146	72	0	84	48	13	0	27	39	1	430
<np .d>	0	0	0	1	0	0	0	0	0	0	1
<np>	102	9	0	227	52	16	86	48	16	156	712
<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>	3	0	0	12	1	3	1	156	0	1643	1819
<i>totals</i>	680	474	0	460	131	42	93	231	117	1806	
<##>											1111
<#>											83
<i>totals</i>											1194

Table 1 Summarized GRAID counts for the entire Mandarin corpus.

2.1 *hml*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	2	7	0	0	0	0	0	0	0	0	9
<∅ .2>	3	0	0	0	0	0	0	0	0	0	3
<∅ .h>	64	78	0	11	0	1	0	0	0	0	154
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	8	2	0	6	0	0	0	0	0	0	16
<pro .1>	5	3	0	0	0	0	0	0	1	0	9
<pro .2>	0	0	0	0	1	2	0	0	0	0	3
<pro .h>	18	22	0	2	0	0	0	0	10	0	52
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	4	0	0	1	1	0	1	0	0	0	7
<np .h>	32	27	0	34	18	6	0	4	20	0	141
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	19	3	0	76	12	4	18	11	7	44	194
<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	2	0	0	0	36	0	412	450
<i>totals</i>	155	142	0	132	32	13	19	51	38	456	
<##>											270
<#>											31
<i>totals</i>											301

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

2.2 *jgz*

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	22	39	0	2	3	0	0	0	0	0	66
<∅ .2>	15	24	0	0	1	0	0	0	0	0	40
<∅ .h>	118	99	0	15	0	1	0	0	0	0	233
<∅ .d>	0	0	0	0	0	0	0	0	0	0	0
<∅>	34	6	0	62	2	0	0	0	0	0	104
<pro .1>	15	24	0	6	4	3	0	0	17	2	71
<pro .2>	16	21	0	6	3	2	0	0	5	0	53
<pro .h>	15	13	0	12	5	0	0	0	17	1	63
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	16	4	0	6	2	1	5	0	0	3	37
<np .h>	93	37	0	28	10	5	0	15	9	1	198
<np .d>	0	0	0	1	0	0	0	0	0	0	1
<np>	58	4	0	125	31	9	55	34	7	80	403
<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>	1	0	0	9	0	3	0	81	0	826	920
<i>totals</i>	403	272	0	272	61	24	60	130	55	913	
<##>											674
<#>											37
<i>totals</i>											711

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

2.3 lzh

GRAID	<:s>	<:a>	<:ncs>	<:p>	<:obl>	<:g>	<:l>	<:pred>	<:poss>	<:other>	<i>totals</i>
<∅ .1>	6	2	0	0	0	0	0	0	0	0	8
<∅ .2>	0	0	0	0	0	0	0	0	0	0	0
<∅ .h>	41	28	0	2	0	0	0	0	0	0	71
<∅ .d>	1	0	0	0	0	0	0	0	0	0	1
<∅>	5	3	0	4	0	0	0	0	0	0	12
<pro .1>	0	2	0	1	1	0	0	0	2	0	6
<pro .2>	4	2	0	0	1	0	0	0	0	0	7
<pro .h>	17	13	0	0	6	0	0	0	10	0	46
<pro .d>	0	0	0	0	0	0	0	0	0	0	0
<pro>	0	0	0	0	0	0	0	0	0	0	0
<np .h>	21	8	0	22	20	2	0	8	10	0	91
<np .d>	0	0	0	0	0	0	0	0	0	0	0
<np>	25	2	0	26	9	3	13	3	2	32	115
<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	1	1	0	1	39	0	405	449
<i>totals</i>	122	60	0	56	38	5	14	50	24	437	
<##>											167
<#>											15
<i>totals</i>											182

Table 4 Summarized GRAID counts for the lzh text.

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/