

Multi-CAST

English *annotation notes*

Nils Norman Schiborr

August 2019
v2.0



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

Schiborr, Nils N. 2019. Multi-CAST English annotation notes. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (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/) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 English annotation notes v2.0 last updated 24 August 2019
This document was typeset by NNS with $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and the *multicast3* class (v3.2.1).

Contents

1	Notes on the GRAID annotations	1
1.1	Structurally and pragmatically suppressed arguments	1
1.1.1	Non-finite clauses	1
1.1.2	Relative clauses	2
1.2	Imperatives	3
1.3	Same-subject deletion	3
1.4	Ellipsis in VP-echo structures	4
1.5	Direct speech	4
1.6	Complex predicates	5
1.7	Possessive pronouns	6
1.8	Generic references	6
2	Notes on the RefIND annotations	6
2.1	Referents in clauses otherwise not considered	6
	References	7
	Appendices	8
A	List of corpus-specific GRAID symbols	8
B	List of abbreviated morphological glosses	9

1 Notes on the GRAID annotations

This document contains notes on the implementation of the GRAID (Haig & Schnell 2014) and RefIND (Schiborr et al. 2018) annotation conventions in the Multi-CAST English corpus. It corresponds to version 1908 of the annotations, published in August 2019. Unless a more recent version of this document exists, it also applies to any later versions of the annotations.

1.1 Structurally and pragmatically suppressed arguments

In standard GRAID, unexpressed clausal referents are annotated, as $\langle \emptyset \rangle$, only when they are

- ◆ licensed by the predicate,
- ◆ specific and retrievable from the discourse context, and, crucially,
- ◆ not in an argument slot that is systematically suppressed by the predicate.

The third criterion assumes that it is possible to distinguish two types of referential null argument: zero that is structurally licensed, but omitted due to context-specific pragmatic factors, and zero that is either systematically suppressed or not licensed due to purely structural factors. As noted above, GRAID captures only the former, as only in this case are speakers understood to exercise any choice of expression. The latter cases remain unannotated.

For the annotation of English we have decided to introduce a form symbol $\langle f\emptyset \rangle$ ‘forced zero’, which aims to capture those categorically suppressed referents that contrastive zero $\langle \emptyset \rangle$ does not. As $\langle f\emptyset \rangle$ is not a kind of $\langle \emptyset \rangle$, the two categories should never be conflated during analysis. Currently, the $\langle f\emptyset \rangle$ symbol is used only in the English and Sanzhi Dargwa corpora in Multi-CAST. It is planned to become part of the standard GRAID specification as an optional gloss in the future. In English, $\langle f\emptyset \rangle$ occurs in non-finite clauses (Section 1.1.1) in one type of relative clause (Section 1.1.2).

1.1.1 Non-finite clauses

Non-finite clauses in English in general do not allow overt expression of their subjects, which accordingly are annotated $\langle f\emptyset \rangle$. The head of the verbal complex receives the $\langle \text{vother} \rangle$ ‘non-canonical verb form’ gloss. The following examples showcase infinitival clauses with and without *to* (1–2), as well as present and past participial clauses (3–4).¹

- (1) *And Father went down to see the agent, ...*

<i>and</i>	<i>father</i>	<i>went</i>	<i>down</i>	
and	father	go.PST	down	
##	other	np.h:s	v:pred	rv

	<i>to</i>	<i>see</i>	<i>the</i>	<i>agent</i>
0_father	to	see.INF	the	agent
##ac	<i>f\emptyset.h:a</i>	lv	vother:pred	ln_det np.h:p

[mc_english_kent02_0428]

¹ For an alternative interpretation of (4) and similar constructions (e.g. *eat the porridge hot*), see the literature on secondary predicates, for example Schultze-Berndt & Himmelmann (2004).

- (2)
- Didn't dare let the governor see us.*

```

          did-n't    dare
    0_we do.PST-NEG dare.INF
##neg 0.1:a lv_aux    v:pred

          let        governor
    0_we let.INF    governor
##cc:p f0.1:a vother:pred np.h:p

          see        us
    0_governor see.INF    1PL.OBL
##cc:obl f0.h:a    vother:pred pro.1:p

```

[mc_english_kent02_0810]

- (3)
- I was just in the wood, getting a rabbit.*

```

    I        was    just    in    the    wood
    1SG      be.PST.1SG just    in    the    wood
## pro.1:s cop          other adp ln_det np:pred_l

          getting    a    rabbit
    0_I    get.PTCP.PRS a    rabbit
# f0.1:a vother:pred ln_deti np:p

```

[mc_english_kent02_0221]

- (4)
- And I couldn't go see him killed.*

```

    and    I        could-n't    go    see    him
    and    1SG      could-NEG    go.INF    see.INF    3SG.M.OBL
## other pro.1:a lv_aux    lv_v    v:pred pro:p

          killed
    0_Buller kill-PTCP.PST
##cc:obl f0:s    vother:pred

```

[mc_english_kent02_0059]

1.1.2 Relative clauses

English has two formal types of relative clauses: those that contain an anaphoric relative pronoun such as *who* or *which*, and those that do not. In the former, the relative pronoun is annotated <rel_pro> and carries its respective function:

- (5)
- I spoke to the people who lived near that place.*

```

    I        spoke    to    the    people
    1SG      speak.PST to    the    people
## pro.1:s v:pred    adp ln_det np.h:g

    who        lived    near    that    place
    who        live-PST near    DIST.SG place
##rc rel_pro.h:s v:pred adp ln_dem np:l

```

[mc_english_devon01_0136]

In the latter, the gapped constituent is not overtly expressed. While in a bare relative like in (6) a relative pronoun could conceivably be inserted, if the subordinator *that* is present, as in (7), overt expression of the gapped constituent is systematically blocked. In keeping with the rules above, the gap in a bare relative clause is annotated <rel_0> ‘pragmatically omitted argument of a bare relative clause’, and the forced gap in a relative clause with *that* receives the gloss <rel_f0> ‘structurally suppressed argument of a *that*-relative clause’. The subordinator *that* receives the gloss <other>.

- (6) *That’s the first thing you put in on a farm.*

<i>that</i>	=’s	<i>the</i>	<i>first</i>	<i>thing</i>	
DIST.SG	=be.PST.3SG	the	first	thing	
##	dem_pro:s	=cop	ln_det	ln_adj	np:pred
	<i>you</i>	<i>put</i>	<i>in</i>	<i>on</i>	<i>a</i>
0_peas	2SG	put.PST	in	on	a
#rc	rel_0:p	imp_pro.2:a	v:pred	rv	adp ln_deti np:l

[mc_english_kent02_0562]

- (7) *If he got one that wasn’t genuine, ...*

<i>if</i>	<i>he</i>	<i>got</i>	<i>one</i>	
if	3SG.M	get.PST	one	
#ac	adp	pro.h:a	v:pred	num_np:p
	<i>that</i>	<i>was-n’t</i>	<i>genuine</i>	
	that	0_horse	be.PST.3SG-NEG	genuine
#rc.neg	other	rel_f0.h:s	cop	other:pred

[mc_english_kent01_0016]

1.2 Imperatives

In English, subjects are generally but not categorically omitted in imperatives. If omitted, they are annotated as <0> with the additional specifier <imp_>:

- (8) *Have a look at him, try him.*

	<i>have</i>	<i>a</i>	<i>look</i>	<i>at</i>	<i>him</i>
0_father	have.IMP	a	look	at	3SG.M.OBL
##ds	imp_0.2:s	v:pred	other	other:lvc	adp pro:obl
	<i>try</i>	<i>him</i>			
0_father	try.IMP	3SG.M.OBL			
##ds	imp_0.2:a	v:pred	pro:p		

[mc_english_kent01_0193]

1.3 Same-subject deletion

By far the most frequent occurrence of zero in English is in coordinated clauses with co-referential subjects, which often form long chains like the one in (9). In clauses of this kind, echoed auxiliaries

are omitted alongside the subject. While the subject receives the gloss ⟨0⟩, no zero element is added for unexpressed auxiliaries.

- (9) *I used to go up there, and load it, and take it home, pitch it on a stack, and stack it.*

```

I      used  to go  up  there
1SG   used  to go.INF up there
## pro.1:s lv_aux lv v:pred adp other:g

and    load  it
and    0_l   load.INF 3SG.N.OBL
## other 0.1:a v:pred pro:p

and    take  it      home
and    take.INF 3SG.N.OBL home
## other 0.1:a v:pred pro:p other:g

pitch  it      on a      stack
pitch.INF 3SG.N.OBL on a      stack
## 0.1:a v:pred pro:p adp ln_deti np:g

and    stack  it
and    stack.INF 3SG.N.OBL
## other 0.1:a v:pred pro:p

```

[mc_english_kent02_0129]

1.4 Ellipsis in VP-echo structures

Spoken English makes extensive use of various expletive auxiliaries when identical VPs are echoed. This strategy is particularly common in responses to questions:

- (10) a. *Can't you build me one with three rows?*

```

can't  you  build  me  one  with  three  rows
can.NEG 2SG  build.INF 1SG.OBL one with three row-PL
##ds lv_aux pro.2:a v:pred pro.1:p num_np:p2 rn_adp rn_num rn_np

```

- b. *Yeah, can, if you like.*

```

yeah          can          if  you  like
yeah 0_manager can  0_that  if  2SG  like.PRS
##ds other 0.1:a lv_aux 0:p #ac adp pro.2:s v:pred

```

[mc_english_kent02_0550]

As seen here, the substituted phrases may, at least conceptually, include direct objects, for which zero glosses ⟨0⟩ are inserted.

1.5 Direct speech

Direct speech, as it is syntactically independent and may span several clauses, is not annotated as the object of the clauses that introduce or conclude it. In contrast to verbs of speech with NP objects, which may be transitive (e.g. *she said nothing*) or ditransitive (e.g. *she told us a story*), verbs of speech bookending direct speech have been annotated as either intransitive or, if

a direct object addressee is present, as transitive. In order to note their special status, the subjects of these verbs have the additional specifier <_ds> ‘subject of a verb of speech’ attached to their respective function glosses.

- (11) *And Father says, He’ll do.*

	<i>and</i>	<i>father</i>	<i>says</i>		<i>he</i>	<i>=’ll</i>	<i>do</i>
	and	father	say.PRS.3SG		3SG.M	=will	do.INF
##	other	np.h:s_ds	v:pred	##ds	pro.h:s	=lv_aux	v:pred

[mc_english_kent01_0256]

- (12) *So she told her groom, Put the horse in the cart!*

	<i>so</i>	<i>she</i>	<i>told</i>	<i>her</i>	<i>groom</i>		
	so	3SG	tell.PST	3SG.F.POSS	groom		
##	other	pro.h:a_ds	v:pred	ln_pro.h:poss	np.h:p		

		<i>put</i>	<i>the</i>	<i>horse</i>	<i>in</i>	<i>the</i>	<i>cart</i>
	0_groom	put.IMP	the	horse	in	the	cart
##ds	imp_0.2:a	vother:pred	ln_det	np:p	adp	ln_det	np:g

[mc_english_kent02_0047]

1.6 Complex predicates

Complex predicates combine a semantically weak light verb (also called vector verb) such as *do*, *take*, or *be*, with a non-verbal element of some kind. The latter supplies most of the semantic content of the expression, but does not exhibit many of the properties of regular objects (see Berlage 2010), and is hence not identified as such in the GRAID annotations. Instead, the non-verbal element is glossed <:lvc> ‘light verb complement’, marking it out as a special kind of expression. It invariably receives the form gloss <other>, irrespective of its lexical category.

The light verb and its complement jointly contribute to the argument structure of the whole expression (cf. Butt 2010). As such, since the complement is not treated as an object, the function gloss of the subject is determined by the absence or (at least implied) presence of (another) object in the clause. The following examples illustrate the annotation schema, (13) for an intransitive and (14) for a transitive predicate.

- (13) *If the pony didn’t take notice, ...*

	<i>if</i>	<i>the</i>	<i>pony</i>	<i>did-n’t</i>	<i>take</i>	<i>notice</i>
	if	the	pony	do.PST-NEG	take.INF	notice
#ac	adp	ln_det	np:s	lv_aux	v:pred	other:lvc

[mc_english_kent01_0202]

- (14) *So we got hold of the police.*

	<i>so</i>	<i>we</i>	<i>got</i>	<i>hold</i>	<i>of</i>	<i>the</i>	<i>police</i>
	so	1PL	get.PST	hold	of	the	police
##	other	pro.1:a	v:pred	other:lvc	rv	ln_det	np.h:p

[mc_english_kent02_0619]

1.7 Possessive pronouns

English possessive determiners (*mine, her*) may occur without an explicitly mentioned possessum, in which case they assume an altered form (*mine, hers*). Both are annotated as subconstituents of the possessed NP, as $\langle \text{ln_pro: poss} \rangle$. With the latter, the omitted possessum is inserted as $\langle \emptyset \rangle$.

- (15) *All the young calves coming in and knew their mothers.*

```

all the young calves coming in
all the young calf.PL come.PTCP.PRS in
# ln ln_det ln_adj np:s vother:pred rv

and knew their mother-s
and 0_calves know.PST 3PL.POSS mother-PL
## and 0:a v:pred ln_pro:poss np:p

```

[mc_english_devon01_0046]

- (16) *You know, they want theirs.*

```

you know they want theirs
2SG know.PRS 3PL want.PRS theirs 0_wurzel
## other other pro:a v:pred ln_pro:poss 0:p

```

[mc_english_kent02_0599]

1.8 Generic references

Constructions invoking generic referents, for example with *you* or *one*, are annotated with the specifier $\langle \text{gen_} \rangle$, such as $\langle \text{gen_pro. 2} \rangle$ in (17). In general, they should not be combined with other forms during analysis. Generic referents do not receive referent indices.

- (17) *When you talk of Churston, you've got to bring in Galmpton as well.*

```

when you talk of Churston
when 2SG talk.PRS of Churston
## #ac adp gen_pro.2:a v:pred rv pn_np:p %

you =ve got to bring in Galmpton as well
2SG =have.PRS get.PTCP.PST to bring.INF in Galmpton as well
gen_pro.2:a =lv_aux lv_aux lv v:pred rv pn_np:p other other

```

[mc_english_devon01_0009]

2 Notes on the RefIND annotations

2.1 Referents in clauses otherwise not considered

Where segments have not been annotated because they are incomplete or not syntactically well-formed, or because they stand outside of the normal flow of narration, they are marked as $\langle \#nc \rangle$ ‘not considered’, and all of the elements inside them are glossed $\langle nc \rangle$.

Some of these segments, however, do contain identifiable discourse references. These are, presumably, still registered by the interlocutors even in cases where the clause they reside in is abandoned partway through. In order to preserve the genuine sequence of reference in the annotations, mentions of referents inside segments that otherwise not considered are nevertheless assigned referent indices. In the English corpus, these elements further receive GRAID form and person/animacy glosses with the ⟨nc_⟩ specifier, noting their extraneous status. Grammatical functions are not glossed.

- (18) a. [Interviewer:] *Did people like the gypsies in those days?*

	<i>did</i>	<i>people</i>	<i>like</i>	<i>the</i>	<i>gypsies</i>	<i>in</i>	<i>those</i>	<i>day-s</i>
	do.PST	people	like.INF	the	gypsy.PL	in	DIST.PL	day-PL
#nc	nc	nc_np.h	nc	nc	nc_np.h	nc	nc	nc
		0022			0003			

- b. *Oh, we didn't mind'em.*

	<i>oh</i>	<i>we</i>	<i>did-n't</i>	<i>mind</i>	<i>= 'em</i>
	oh	1PL	do.PST-NEG	mind.INF	=3PL.OBL
#nc	nc	nc_pro.1	nc	nc	nc_pro.h
		0023			0003

[mc_english_kent02_0022-0023]

References

- Berlage, Eva. 2010. Lexicalisation of predicative complements in English. *Transactions of the Philological Society* 108(1). 53–67.
- Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Amberber, Mengistu & Harvey, Mark & Baker, Brett (eds.), *Complex predicates in cross-linguistic perspective*, 48–78. Cambridge: Cambridge University Press.
- Haig, Geoffrey & Schnell, Stefan. 2014. *Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators (version 7.0)*. (<https://multicast.aspra.uni-bamberg.de/#annotations>) (Accessed 2019-03-08).
- Haig, Geoffrey & Schnell, Stefan (eds.). 2016. *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. (<https://multicast.aspra.uni-bamberg.de/>) (Accessed 2019-03-08).
- Huddleston, Rodney D. & Pullum, Geoffrey K. (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Schiborr, Nils N. 2015. Multi-CAST English. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (<https://multicast.aspra.uni-bamberg.de/#english>) (Accessed 2016-02-28).
- Schiborr, Nils N. & Schnell, Stefan & Thiele, Hanna. 2018. *RefIND — Referent Indexing in Natural-language Discourse: Annotation guidelines (v1.1)*. Bamberg / Melbourne: University of Bamberg. (<https://multicast.aspra.uni-bamberg.de/#annotations>) (Accessed 2019-03-08).
- Schultze-Berndt, Eva & Himmelmann, Nikolaus P. 2004. Depictive secondary predicates in crosslinguistic perspective. *Linguistic Typology* 8. 59–131.

Appendices

A List of corpus-specific GRAID symbols

The following is a list of the non-standard GRAID symbols used in the annotation of the Multi-CAST English corpus. Please refer to the *GRAID manual* (Haig & Schnell 2014: 54–55) for an inventory of basic GRAID symbols.

Form symbols and specifiers

⟨f0⟩	structurally suppressed argument slot of a predicate
⟨rel_f0⟩	structural gap in a relative clause with <i>that</i>
⟨imp_0⟩	omitted subject of an imperative verb
⟨rel_0⟩	gap in a bare relative clause
⟨dem_pro⟩	demonstrative pronoun
⟨rel_pro⟩	relative pronoun
⟨num_np⟩	numeral
⟨pn_np⟩	proper name
⟨indef_other⟩	indefinite pronoun
⟨intrg_other⟩	interrogative pronoun
⟨gen_⟩	<i>specifier</i> : form with generic reference (e.g. <i>you, one</i>)

Function symbols and specifiers

⟨:lvc⟩	non-verbal complement of a complex predicate
⟨:s_ds⟩	subject of a verb of speech, intransitive
⟨:a_ds⟩	subject of a verb of speech, transitive

Subconstituent symbols

⟨_adj⟩	attributive adjective; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_dem⟩	demonstrative determiner; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_det⟩	definite article; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_deti⟩	indefinite article; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_detq⟩	quantifier; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_num⟩	attributive numeral; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_aux⟩	auxiliary; attaches to ⟨lv⟩ and ⟨rv⟩

Other symbols

⟨nc_⟩	<i>specifier</i> : marks form glosses with RefIND indices in segments otherwise not considered (i.e. those marked ⟨#nc⟩)
-------	--

B List of abbreviated morphological glosses

1	first person	PL	plural
2	second person	POSS	possessive
3	third person	PROX	proximal
CMPR	comparative	PRS	present
DIST	distal	PST	past
F	feminine	PTCP	participle
HORT	hortative	REFL	reflexive
IMP	imperative	SG	singular
INF	infinitive	SUBJ	subjunctive
M	masculine	SUPR	superlative
N	neuter		
NEG	negation		
OBL	oblique case	NC	not classifiable

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/