Citation for this document

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)
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 (Schibor et al. 2018) annotation conventions in the Multi-CAST English corpus. It corresponds to version 2002 of the annotations, published in February 2020. 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 \((0)\), 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 \((f0)\) ‘forced zero’, which aims to capture those categorically suppressed referents that contrastive zero \((0)\) does not. As \((f0)\) is not a kind of \((0)\), the two categories should not be conflated during analysis. Currently, the \((f0)\) symbol is used only in a subset of corpora in Multi-CAST. It is planned to become part of the standard GRAID specification as an optional gloss in the future. In English, \((f0)\) 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 \((f0)\). The head of the verbal complex receives the \((vother)\) ‘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) And Father went down to see the agent, ...

   and father went down
   and father go.pst down
   ## other np.h:s v:pred rv

   to see the agent
   0_father to see.inf the agent
   #ac f0.h:a 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.

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

(4) And I couldn’t go see him killed.

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.
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.

```
that =’s the first thing
DIST.SG =be.PST:3SG the first thing
## dem_pro:s =cop ln_det ln_adj np:pred

you put in on a farm
0_peas 2SG put.PST in on a farm
#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, ...

```
if he got one
if 3SG.M get.PST one
#ac adp pro.h:a v:pred num_np:p

that was-n’t genuine
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.

```
have a look at him
0_father have.IMP a look at 3SG.M.OBL
##ds imp_0.2:s v:pred other other:1vc adp pro:obl

try him
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 \( \langle 0 \rangle \), no zero element is added for unexpressed auxiliaries.

(9)  \[ I \text{ used to go up there, and load it, and take it home, pitch it on a stack, and stack it.} \]

\[ I \quad \text{used} \quad \text{to go up there} \]
\[ 1sg \quad \text{used} \quad \text{to go.INF} \quad \text{up there} \]
\[ ## \text{pro.1:s lv_aux 1v v:pred adp other:g} \]
\[ \text{and} \quad \text{load} \quad \text{it} \]
\[ \text{and} \quad 0.1 \quad \text{load.INF} \quad 3SG.N.OBL \]
\[ ## \text{other 0.1:a v:pred pro:p} \]
\[ \text{and} \quad \text{take} \quad \text{it} \quad \text{home} \]
\[ \text{and} \quad \text{take.INF} \quad 3SG.N.OBL \quad \text{home} \]
\[ ## \text{other 0.1:a v:pred pro:p other:g} \]
\[ \text{pitch} \quad \text{it} \quad \text{on a stack} \]
\[ \text{pitch.INF} \quad 3SG.N.OBL \quad \text{on a stack} \]
\[ ## 0.1:a \text{ v:pred pro:p adp ln_deti np:g} \]
\[ \text{and} \quad \text{stack} \quad \text{it} \]
\[ \text{and} \quad \text{stack.INF} \quad 3SG.N.OBL \]
\[ ## \text{other 0.1:a v:pred pro:p} \]

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.  \[ \text{Can’t you build me one with three rows?} \]
\[ \text{can’t} \quad \text{you} \quad \text{build} \quad \text{me} \quad \text{one} \quad \text{with} \quad \text{three} \quad \text{rows} \]
\[ \text{can.NEG} \quad 2SG \quad \text{build.INF} \quad 1SG.OBL \quad \text{one} \quad \text{with} \quad \text{three row-pl} \]
\[ ##\text{ds lv_aux pro.2:a v:pred pro.1:p num_np:p2 rn_adp rn_num rn_np} \]

b.  \[ \text{Yeah, can, if you like.} \]
\[ \text{yeah} \quad \text{can} \quad \text{if} \quad \text{you} \quad \text{like} \]
\[ \text{yeah} \quad 0.\text{manager} \quad \text{can} \quad 0.\text{that} \quad \text{if} \quad 2SG \quad \text{like.PRS} \]
\[ ##\text{ds other 0.1:a lv_aux 0:p ac adp pro.2:s v:pred} \]

As seen here, the substituted phrases may, at least conceptually, include direct objects, for which zero glosses \( \langle 0 \rangle \) 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. \textit{she said nothing}) or ditransitive (e.g. \textit{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.

```
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
```

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

```
s so she tell.PST 3SG.F.POSS groom
## other pro.h:a_ds v:pred ln_pro.h:poss np.h:p

put the horse in the cart #ds imp_0.2:a vother:pred ln_det np:p adp ln_det np:g
```

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, ...

```
if the pony do.PST-NEG take.INF notice
#ac adp ln_det np:s lv_aux v:pred other:lvc
```

(14) So we got hold of the police.

```
s so we get.PST hold of the police
## other pro.1:a v:pred other:lvc rv ln_det np:h:p
```
1.7 Possessive pronouns

English possessive determiners (mine, her) may occur without an explicitly mentioned posses- 
sum, in which case they assume an altered form (mine, hers). Both are annotated as subconstitu-
ents of the possessed NP, as ⟨ln_pro:poss⟩. With the latter, the omitted possessum is inserted as ⟨0⟩.

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

(16) You know, they want theirs.

1.8 Generic references

Constructions invoking generic referents, for example with you or one, are annotated with the speci-
ifier ⟨gen_⟩, such as ⟨gen_pro_2⟩ 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.

2 Notes on the ReflIND 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 ⟨#nc⟩ 
‘not considered’, and all of the elements inside them are glossed ⟨nc⟩.
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?

\[
\begin{array}{l}
\text{did} \quad \text{people} \quad \text{like} \quad \text{the} \quad \text{gypsies} \quad \text{in} \quad \text{those} \quad \text{day-s} \\
\#nc \quad nc \quad nc \quad nc \quad nc \quad nc \quad nc \quad nc \quad nc \\
0022 \quad 0003
\end{array}
\]

b. Oh, we didn’t mind ’em.

\[
\begin{array}{l}
\text{oh} \quad \text{we} \quad \text{did-n’t} \quad \text{mind} \quad =’em \\
\text{oh} \quad 1pl \quad \text{do}.p\text{st-NEG} \quad \text{mind}.\text{INF} \quad =3\text{PL.OBL} \\
\#nc \quad nc \quad nc \quad nc \quad nc \quad nc \quad nc \quad nc \quad nc \\
0023 \quad 0003
\end{array}
\]

References


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 that
(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_) specifier: form with generic reference (e.g. you, one)

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_) specifier: marks form glosses with RefIND indices in segments otherwise not considered (i.e. those marked (#nc))
List of abbreviated morphological glosses

<table>
<thead>
<tr>
<th>Code</th>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>first person</td>
<td>PL</td>
</tr>
<tr>
<td>2</td>
<td>second person</td>
<td>POSS</td>
</tr>
<tr>
<td>3</td>
<td>third person</td>
<td>PROX</td>
</tr>
<tr>
<td>CMPR</td>
<td>comparative</td>
<td>PRS</td>
</tr>
<tr>
<td>DIST</td>
<td>distal</td>
<td>PST</td>
</tr>
<tr>
<td>F</td>
<td>feminine</td>
<td>PTCP</td>
</tr>
<tr>
<td>HORT</td>
<td>hortative</td>
<td>REFL</td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
<td>SG</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
<td>SUBJ</td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
<td>SUPR</td>
</tr>
<tr>
<td>N</td>
<td>neuter</td>
<td>NC</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
<td></td>
</tr>
<tr>
<td>OBL</td>
<td>oblique case</td>
<td></td>
</tr>
</tbody>
</table>
Multilingual Corpus of Annotated Spoken Texts

multicast.aspra.uni-bamberg.de/