Multi-CAST

Sanzhi Dargwa
annotation notes

Diana Forker

Nils Norman Schiborr

July 2019
v1.1
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)

The Multi-CAST collection has been archived at the University of Bamberg, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Sanzhi Dargwa annotation notes  v1.1  last updated 24 July 2019
This document was typeset by NNS with XeLaTeX and the multicast3 class (v3.1.9001).
## Contents

1 **Notes on the GRAID annotations**  .............................................................. 1  
   1.1 Ergativity and the distinction between S and A  ..................................... 1  
      1.1.1 Experiencer verbs  .................................................................. 2  
      1.1.2 Antipassive verbs  .................................................................. 3  
      1.1.3 Verbs of speech  ..................................................................... 3  
   1.2 Complex predicates  ........................................................................... 4  
   1.3 Structurally and pragmatically suppressed arguments  ........................... 6  
      1.3.1 Gapped constituents in relative clauses ...................................... 7  
      1.3.2 Infinitival complement clauses ............................................... 7  
      1.3.3 Subordinate verb forms .......................................................... 8  
      1.3.4 Imperatives  ........................................................................... 9  
   1.4 Predicative expressions of possession .................................................... 10  
   1.5 Third person pronominal forms ............................................................ 11  
   1.6 On the relative order of additional symbols ......................................... 11  

2 **Notes on the RefIND annotations** .............................................................. 12  
   2.1 Cataphoric introductions ................................................................... 12  
   2.2 Referents in clauses otherwise not considered ...................................... 12  

References  ...................................................................................................... 13  

Appendices  ...................................................................................................... 14  
   A List of corpus-specific GRAID symbols .................................................. 14  
   B List of abbreviated morphological glosses .............................................. 15
1 Notes on the GRAID annotations

The following comprises selected notes on the GRAID (Haig & Schnell 2014) and ReflND (Schiborr et al. 2018) annotations of Sanzhi Dargwa. A grammar of Sanzhi (Forker Under revision), compiled on the basis of the material from which the texts in this corpus were taken, is currently in preparation, and will be freely available from the website of the publisher in the future.

This document corresponds to version 1907 of the Sanzhi annotations, published in July 2019. Unless a more recent version of this document exists, it also applies to any later versions of the annotations.

1.1 Ergativity and the distinction between S and A

Sanzhi is a language with ergative alignment, that is, the Patient-like argument (P) of transitive clauses and the single argument (S) of intransitive clauses receive the same marking in the morphology, distinct from that of the Agent-like argument (A) of transitive clauses. In Sanzhi, the A argument is marked with the ergative case, while S and P arguments are in the unmarked absolute case. S and P control gender agreement – where instantiated – while A may trigger person agreement under certain conditions.

As a general rule for the GRAID annotations of A, S, and P in Sanzhi, we follow the language-specific morphology in glossing the ergative-marked argument as ⟨:a⟩, an absolutive-marked object as ⟨:p⟩, and an absolutive-marked subject as ⟨:s⟩. Examples (1) and (2) illustrate this practice:

(1) dul abiʔu nda qərq: a, čak’al dul akaxubda.
\[
\begin{array}{llll}
\text{du-l} & \text{a-b-iʔ-u’n-da} & \text{qərq: a} \\
\text{1SG-erg} & \text{neg-N-steal.PFV-PRET-1} & \text{stone} \\
\end{array}
\]

\[
\begin{array}{llll}
\text{ča-k’al} & \text{du-l} & \text{a-kax-ub-da} \\
\text{who-INDEF} & \text{1SG-erg} & \text{neg-kill.PFV-PRET-1} \\
\end{array}
\]

[mc_sanzhi_barkalla_0059]

‘[He said,] “I did not steal the stone, I did not kill anyone.”’

(2) hinna šuleb kabiš: uncab šajt’an.
\[
\begin{array}{llll}
\text{hin-na} & \text{šule-b} & \text{ka-b-is:−un} & \text{ca-b} & \text{šajt’an} \\
\text{water-gen} & \text{at.side-N} & \text{down-N-sleep.PFV-PRET} & \text{be-N} & \text{devil} \\
\end{array}
\]

\[
\begin{array}{llll}
\text{1n_np:poss} & \text{np:1} & \text{v:pred} & \text{rv_aux np:h:s} \\
\end{array}
\]

[mc_sanzhi_devils_0019]

‘The devil was sleeping at the water’s edge.’

For the vast majority of predicates in the corpus, adherence to the language-specific morphology yields results that are compatible with our general approach to A and P, as outlined in Andrews (2007: 137). Our adaptation has the advantage of providing clearly defined criteria for the purposes of annotation, and we have maintained it as the default throughout the corpus.

Under certain circumstances, however, the morphology does not align with our cross-linguistic definition of S, A, and P, which necessitated adaptations to our annotation conventions. These cases are discussed in Sections 1.1.1, 1.1.2, 1.1.3, and 1.2 below.
1.1.1 Experiencer verbs

Sanzi has a number of predicates that assign quirky subject case. The most notable of these are experiencer verbs (“affective verbs” in Forker Under revision) such as ‘be cold’, ‘want’, ‘see’, and certain other verbs such as ‘find’, whose Experiencer-argument takes dative case. With bivalent experiencer verbs, the Stimulus is in the absolutive, and with certain verb forms the Experiencer may alternatively be marked with ergative case (see Forker Under revision: 395).

There are numerous ways of interpreting the argument structure of these verbs. As a compromise solution, we have treated the Stimulus as ⟨:p⟩, since it is coded as absolutive, while the Experiencer receives the gloss ⟨:ncs⟩ ‘non-canonical subject’ if it is in dative case and ⟨:a⟩ if it is ergative. If the subject of a bivalent experiencer verb is unexpressed, the glossing defaults to ⟨:ncs⟩. Examples of experiencer verbs with dative and ergative Experiencers are illustrated in (3) and (4), respectively.

(3) a. rucbaj qixbe dik:ul, ...
   ruc-b-a-j   qix-be   d-ik:-ul
   sister-PL-OBL-DAT nut-PL NPL-want.IPFV-ICVB
   #ds_cv np:h:nrs_cv np:p v:pred
   ['She said,] “My sisters want nuts, ...”' [mc_sanzhi_patima_0012]

b. dam ?a’h musa balyad.
   dam    ?a’h    musa    b-al-y-ad
   1SG.GEN good place N-KNOW.IPFV-PRS.1
   #ds pro.1:nrs ln_adj np:p v:pred
   ['The fox said,] “I know a good place.”' [mc_sanzhi_patima_0027]

(4) a. italasuna amzude du:nalla čibažibil ak:u.
   ita-la =buna amzu-dex du-l nalla
   3PL.OBL-GEN =EQ clean-NMLZ 1SG-ERG until.then
   #pc.neg ln_pos.pro.h =ln np:p pro.1:a_pc other
   či-b-až-ib-il   ak:u
   SPR-N-SEE.PFV-PRET-PTCP NEG
   v:pred rv
   ‘A cleanliness like theirs I have seen nowhere else.’ [mc_sanzhi_ramazan_0067]

b. hana ik’ude dul lidil sahadarčibda!
   hana    ik’-ul-de
   now 0.Ismail say.IPFV-ICVB-2SG
   #ds_cv other 0.2:s_ds_cv v:pred
   du-l    lidil    sa-ha-d-arč-ib-da
   1SG-ERG all<PL> ANTE-UPWARDS-NPL-find.PFV-PRET-1
   #ds pro.1:a np:p v:pred
   ['He said,] “And now (you) say I made it all up!”' [mc_sanzhi_barkalla_0069]

1 The conditions that trigger this alternation are as yet improperly understood and bear further investigation.
c. **dul urč’em’cuń urč’em’ra juldaš barčıbdə.**

```
du-l urč’em-c’a-nu urč’em-ra juldaš b-arč-ib-da
1SG-ERG nine-ten-tenth nine-NUM friend IPL-find.PFV-PRET-1
``` #ds pro.1:a ln ln np.h:p v:pred

‘[He said,] “I have found ninety-nine friends.”’

[mc_sanzhi_happy_0013]

### 1.1.2 Antipassive verbs

Sanzhi has an antipassive construction in which the case marking of the A and P arguments of a canonical transitive verb is reversed (Forker Under revision: 398). Notably, there is no formal marking on the verb for this type of construction. It most frequently occurs with the consumption verbs ‘eat’ and ‘drink’, and with the verb *b-irq’-‘do, make* as in (5).

In line with the morphology, antipassive constructions have been annotated as special types of intransitive clauses, that is, the absolutive-marked Agent is ⟨:s⟩ and the ergative-marked Patient ⟨:obl⟩.

(5) **nušːa χalq’ ṭaʾči dirq’an χalq’deq’al!**

```
nušːa χalq’
1PL people
``` #ds pro.1:s np.h:appos

```
ṭaʾči-l d-irq’-an
0_people work-ERG 1/2PL-do.PFV-PTCP
``` #rc_pc rel_f0.1:s np:h:obl v:pred %

χalq’-de =q’al
people-PST =MOD
np.h:pred =other

‘[They said,] “We people who are working are people too!”’

[mc_sanzhi_mill_0014]

### 1.1.3 Verbs of speech

Verbs of speech in Sanzhi can be morphologically transitive or intransitive, and hence assign ergative or absolutive case to the NP expressing the speaker. In either case, in the narratives included in this corpus, they are more often than not used to introduce direct speech, and even transitive verbs of speech only infrequently occur with nominal objects. As is standard practice in GRAID, direct speech complements are not formally analyzed as a ⟨:p⟩ arguments of the clauses that introduce them, but merely receive the tag ⟨#ds⟩ ‘direct speech’ (as per Haig & Schnell 2014: 26). As a consequence, in some Multi-CAST corpora (notably Northern Kurdish, among others), the subjects of morphologically transitive verbs of speech have been annotated as cases of S.

For Sanzhi, we have decided to maintain the morphological distinction between A and S for verbs of speech, even when the former is not paired with a P argument in the annotations. To denote their special status, the subjects of these verbs have been glossed either ⟨:a_ds⟩ ‘subject of a transitive verb of speech’ as in (6a) and (6b), or ⟨:s_ds⟩ ‘subject of an intransitive verb of speech’ as in (7).
1.2 Complex predicates

Complex predicates (CPs) combine a semantically weak light verb (or “vector verb”) such as do, take, or be, with a non-verbal element of some kind, the latter of which supplies most of the semantic content to the expression. Crucially, the non-verbal element fails to exhibit many of the properties of regular objects (see, e.g., Berlage 2010), and is hence not identified as such in the annotation. Instead, it is glossed ⟨:other⟩ ‘light verb complement’, which marks it out as a special type of expression. Because the lexical category of the complement cannot always be ascertained, it invariably receives the form gloss ⟨:other⟩.

The complement and light verb contribute jointly to the argument structure of the entire expression (cf. Butt 2010). In most cases, the case marking of the subject is determined by the light verb: if the light verb is agentive (e.g. ‘do’, ‘take’) or experiential (e.g. ‘see’, ‘hear’; see Section 1.1.1), the subject of the complex predicate is, respectively, ergative or dative; if not (e.g. with
'be'), it is absolutive. At the same time, the presence or absence of a direct object in the complex predicate is not strictly controlled by the light verb’s valency. For this reason, it is not uncommon to find complex predicates with two absolutive arguments, as in (9b), or ones with an ergative subject but no discernible (or even implied) object, as in (10a).

A system of extended annotation has been introduced to help capture the complexities of these predicates. The base GRAID function gloss of the subject of a complex predicate is determined by its case marking, usually assigned by the light verb. A subject in the ergative is glossed ⟨:a⟩, a subject in the absolutive ⟨:s⟩, and a subject in the dative ⟨:ncs⟩, as discussed in the previous sections. To this gloss is then added an additional specifier ⟨_cps⟩ or ⟨_cpa⟩, which marks the overall transitivity of the complex predicate, that is, the absence or (at least implied, i.e. ⟨0⟩) presence of a P argument. As such, there are a total of six possible configurations:

1. a. subject is absolutive + CP is intransitive → ⟨:s_cps⟩ e.g. (9a), (13)
   b. subject is absolutive + CP is transitive → ⟨:s_cpa⟩ e.g. (9b)
2. a. subject is ergative + CP is intransitive → ⟨:a_cps⟩ e.g. (10a), (8)
   b. subject is ergative + CP is transitive → ⟨:a_cpa⟩ e.g. (10b)
3. a. subject is dative + CP is intransitive → ⟨:ncs_cps⟩ e.g. (11a)
   b. subject is dative + CP is transitive → ⟨:ncs_cpa⟩ e.g. (11b)

The following examples illustrate each case.

(9) a. urk’ uqun il miskin, ...
   urk’    uq-un    il       miskin
   fright   go.PFV-PRET that poor
   other:lvc v:pred  ln_dem np:h:s_cps

   ‘The pauper was frightened,…’ [mc_sanzhi_barkalla_0015]

   b. il šakričibcar beč’ili berk:unce.
   il            šak        r-ič-ib    ca-r
   that          feel        F-occur.PFV-PRET be-F
   dem_pro.h:s_cpa other:lvc v:pred       rv_aux

   beč’-li      b-erko-unc-e
   wolf-ERG  HPL-eat.PFV-PRET-ATTR.SG 0_sisters
   #cc:p np:d:a v:pred        0.h:p

   ‘She suspected that the wolf had eaten (her sisters).’ [mc_sanzhi_patima_0018]

(10) a. dul at kumek birq’an-da.
   du-l        at         kumek     b-irq’-an-da
   1SG-ERG 2SG.DAT help N-do.IPFV-PTCP-1
   #ds_pc pro.1:a_cps_pc pro.2:obl other:lvc v:pred

   ‘[The fox said,] “I will help you.”’ [mc_sanzhi_patima_0022]

   b. ca eča dulira at peškeš birq’id.
   ca         eča       du-li      =ra      at    peškeš     b-irq’-id
   one   nanny.goat 1SG-ERG =and 2SG.DAT gift N-do.IPFV-1.PRS
   #ds ln np:p pro.1:a_cpa =other pro.2:q other:lvc v:pred

   ‘[He said,] “I will give you one goat as a gift.”’ [mc_sanzhi_happy_0016]
(11) a. dam q:ihin d-ič-i-b-le ca-d

[She said,] “[Tell the friends that] I have met with hardship.”

b. ilij han b-ič-i-b, tːura herʁ-an-ne žaˁndar, ...

‘He thought that a Kumyk man would come out,…’

In some cases, the same light verb occurs with different case assignment in different complex predicates. The light verb ‘occur’, for instance, itself governs the absolutive as in (12a), but in (12b), the complex predicate ‘remember occur’ is coded as an experiencer verb with a dative-marked Experiencer, which we have decided to gloss analogously to other bivalent experiencer verbs such as ‘know’ or ‘want’ (see Section 1.1.1 above):

(12) a. waˁw-dex-li-j guči b-ič-i-b ca-b χalq’

‘Because of the shouting, people gathered.’

b. dam han b-ič-i-b na libil, ...

‘I remember it all now,…’

1.3 Structurally and pragmatically suppressed arguments

In GRAID, unexpressed clausal referents (0) are annotated only where they are

1. licensed by the predicate,
2. specific and retrievable from the discourse context, and
3. 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: those that are structurally licensed, but remain empty due to context-specific pragmatic factors, and those that are either systematically suppressed or not licensed due to purely structural factors. As a general rule, GRAID glosses only the former kind of argument with (0), because only in this case do speakers exercise any choice of expression; the latter remains unannotated.

This distinction has proven difficult to maintain in the annotation of a number of languages, including Sanzhi, where it is particularly contentious in the context of certain non-finite verb
forms such as converbs and participles, and with imperatives and certain types of complement clauses. Furthermore, not capturing arguments that fail to meet the third criterion, such as the gapped constituents in relative clauses, leads to conceptual issues regarding the implicitness of discourse.

It is for this reason that we have decided to introduce a form gloss ⟨f0⟩ ‘forced zero’ to capture categorically suppressed referents, as a counterpart to contrastively suppressed zero ⟨0⟩. It should be noted that ⟨f0⟩ is not a type of ⟨0⟩, so the two categories should never be conflated during analysis. In the current version of Multi-CAST, the ⟨f0⟩ symbol is used only in the Sanzhi Dargwa and English corpora; it is planned to become an optional gloss in the standard GRAID specification in the future.

In Sanzhi, the ⟨f0⟩ symbol is primarily applied to gaps in relative clauses, which are discussed in Section 1.3.1, and to the subjects of certain types of complement clause, discussed in Section 1.3.2. Issues with non-finite constructions and related issues are addressed in Section 1.3.3, imperatives in Section 1.3.4.

1.3.1 Gapped constituents in relative clauses

The gapped constituents in relative clauses cannot be overtly expressed. Where in standard GRAID they would remain unannotated, in the Sanzhi corpus they receive the form gloss ⟨f0⟩ with an additional specifier ⟨rel_⟩ to mark out the context in which they occur.

(13) umxu birχʷi ruciceb wac’aceb errirχʷan.

(14) wat’, na ilti il c’ikuri rarc:ibcari, hana xadi karižibil.

1.3.2 Infinitival complement clauses

In infinitival complement clauses in which the subject is co-referential with the subject of the matrix clause, the former is structurally suppressed (i.e. complement control, see Forker Under revision: §24.2.6, §24.5), and is hence annotated ⟨f0⟩. In the Sanzhi corpus, infinitival complement clauses most commonly occur with verbs such as ‘want’, ‘decide’, and ‘forget’, and with phasal verbs such as ‘begin’ and ‘end’:
1.3.3 Subordinate verb forms

The concept of finiteness is difficult to apply to Dargwa languages, including to Sanzhi (see, e.g. Kalinina & Sumbatova 2007; see also Forker 2011 and 2013 for a discussions of finiteness in Hinuq, a related Nakh-Daghestanian language). Lack of finiteness, however defined, is often accompanied by an inability to express certain kinds of argument, most notably subjects. Sanzhi has a number of verb forms that lack some of the features exhibited by normal finite verbs, such as the full range of TAM marking and agreement morphology. They may also possess certain nominal properties like case marking. These verb forms include converbs, participles, infinitives, and the masdar. Yet although they may appear morphologically deficient in some ways, distributionally they are often very similar to finite verbs, and appear to govern arguments in an identical manner.

For the GRAID annotations, this means that when one of these “less finite” verb forms governs a referential argument, but that argument is not overtly present, it is difficult to decide whether its absence is caused by the structural inability of the verb to license the argument, or by contextual factors. It is for this reason that we have decided to assume a somewhat non-committal stance in the treatment of these verbs.

Firstly, we annotate any of their omitted arguments as \( \langle 0 \rangle \) ‘contrastively suppressed’ rather than \( \langle f0 \rangle \) ‘structurally suppressed’, and the head of the verbal complex as regular \( \langle v \rangle \) rather than \( \langle vother \rangle \) ‘non-canonical verb form’. Secondly, a series of specifiers are applied to the GRAID function glosses of the subject, allowing these contentious forms to be readily distinguished: \( \langle _{cv} \rangle \) for converb clauses, \( \langle _{pc} \rangle \) for participial clauses, and \( \langle _{in} \rangle \) for infinitival clauses. The masdar is exceedingly rare in the annotated texts, and so is not labelled; the two variants of the infinitive (non-inflecting INF1 and inflecting INF2, called “infinitive” and “subjunctive” in Forker Under revision) are functionally equivalent (Forker Under revision: 351–354), and hence do not receive
Multi-CAST Sanzhi Dargwa annotation notes

separate symbols. Lastly, the same three symbols ⟨cv⟩, ⟨pc⟩, and ⟨in⟩ are added to the clause boundary marker (#). While to a degree redundant with the function specifiers, these tags allow for easier identification of clauses of particular types.

The following examples illustrate the annotation patterns, (17) for converb clauses, (18) for participle clauses, and (19) for infinitive and subjunctive clauses. See also Section 1.6 for how these extra specifiers are ordered relative to the base GRAID symbols.

```
(17) nušːal hinna urχːab birq'ulda.
    nušːa-l  hin-na  urχːab  b-irq'-ul-da
1pl-erg  gen  water-gen  mill  n-do.ipfv-icvb-1
#ds_cv pro.1:a_cv ln_np:poss np:p v:pred

'Theysaid,“Wearebuildingawatermill.”' [mc_sanzhi_mill_0019]
```

```
(18) du hel zamana ala salab arganda.
    du  hel  zamana  ala  sala-b  arg-an-da
1sg  pro.1:s  time  pro.2:l front-n go.ipfv-ptcp-1
#ds_pc pro.1:s_pc ln_dem np:other pro.2:l adp v:pred

'It said,"AndIwillappearbeforeyourightaway.”' [mc_sanzhi_devils_0040]
```

```
(19) tːurabuqij, ca-liiče asːij, ca-ljuldašːe barčːij.
    tːura  b-uq-ij
0.they  outside

#ac_in 0.h:s_in other:g v:pred

    ca-l-li  eč-ne  asː-ij
one-obl-erg  nanny.goat-pl  buy.ipfv-inf1
#ac_in np.h:a_in np:p v:pred

    ca-l-li  juldašːe  b-arčː-ij
one-obl-erg  friend-pl  find.ipfv-inf1
#ac_in np.h:a_in np.h:p v:pred

'(They)left,onetobuynannygoats,onetofindfriends.' [mc_sanzhi_happy_0008]
```

A final point of contention concerns syntactic hierarchization. Especially with converb clauses, which are highly frequent and often appear in long chains, it can be difficult to determine exactly which independent clause, if any, they are subordinated to. The Sanzhi annotations thus implement a slight relaxation of the definition of the two left-edge clause boundary markers in GRAID: while ⟨##⟩ is still defined as the beginning of a fully independent syntactic unit, ⟨#⟩ is not used specifically for identifiably subordinated units, but for all clauses that do not meet the criteria for being glossed ⟨##⟩. In the Sanzhi corpus, then, ⟨#⟩-clauses can freely occur outside of the boundaries of a matrix ⟨##⟩-clause. While this change could cause issues for analyses that rely on the precise syntactic hierarchization of clause units (for which GRAID was not designed in the first place, it should be noted), we believe the benefits of this approach outweigh any conceivable disadvantages.

### 1.3.4 Imperatives

While the subjects of imperatives and prohibitives (i.e. negated imperatives) are almost always left unexpressed, they are in fact not categorically suppressed (Forker Under revision: 460). For this
reason, like the subjects of subordinate verb forms discussed in Section 1.3.3, they are annotated (0) rather than (f0), and the verbal complex (v) rather than (vother):

(20) *Dam kumek barq’aja!*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>help</td>
<td>N-do.PFV-IMP.PL</td>
<td></td>
</tr>
<tr>
<td>0.2:a_cps</td>
<td>pro.1:obl</td>
<td>other:lvc</td>
<td>v:pred</td>
</tr>
</tbody>
</table>

‘[He cried,] “Help me!”’ [mc_sanzhi_barkalla_0006]

(21) *U abdaldew? maˁq’aˁtːa!*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2:s</td>
<td>=w</td>
<td>0.brother</td>
<td>=q</td>
</tr>
<tr>
<td>0.2:s</td>
<td>np.h:pred</td>
<td>=other</td>
<td>=q</td>
</tr>
</tbody>
</table>

‘[They said,] “Are you a fool? Don’t go!”’ [mc_sanzhi_dragon_0006]

Imperatives and prohibitives with non-zero subjects are illustrated in (22a) and (22b).

(22) a. *U sala kabiže*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>2sg</td>
<td>front</td>
<td>down-n-be.PFV-IMP</td>
<td></td>
</tr>
<tr>
<td>2sg</td>
<td>pro.2:s</td>
<td>np.h:pred</td>
<td>v:pred</td>
</tr>
</tbody>
</table>

‘[The fox said,] “You sit down in front.”’ [mc_sanzhi_patima_0031]

b. *Warilla wari u iχ-tːaj erčimahark’utːa!*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>#ds.neg</td>
<td>no way</td>
<td>2sg</td>
<td></td>
</tr>
<tr>
<td>#ds</td>
<td>dem.down-pl-obl-dat</td>
<td>look</td>
<td></td>
</tr>
<tr>
<td>2sg</td>
<td>pro.2:s</td>
<td>dem.pro:obl</td>
<td>other:lvc</td>
</tr>
<tr>
<td>2sg</td>
<td>pro.2:s</td>
<td>np.h:pred</td>
<td>v:pred</td>
</tr>
</tbody>
</table>

‘[He said,] “Whatever happens, do not look at them!”’ [mc_sanzhi_bazhuk_0004]

1.4 Predicative expressions of possession

In Sanzhi, predicative expressions of possession involve locational copulae (Forker Under revision: 641; 467–470), with the possessor in the genitive and the possessee in the absolutive. We annotate these clauses analogously to existential constructions, that is, the copula is glossed (other:predex), the possessor (:obl), and the possessee (:s):

(23) *Dila daršːal juldaš biχ”ar, ...*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg-gen</td>
<td>hundred</td>
<td>friend</td>
<td></td>
</tr>
<tr>
<td>1n</td>
<td>np.h: other:predex</td>
<td>np.h: s</td>
<td></td>
</tr>
</tbody>
</table>

‘[He said,] “If I had a hundred friends, ...”’ [mc_sanzhi_happy_0006]
1.5 Third person pronominal forms

Sanzhi does not have a separate paradigm of third person personal pronouns. Their role is filled by an extensive set of demonstrative pronouns expressing proximity, elevation, and cardinal direction, in addition to case and number. See Forker (Under revision: Ch. 4) for a full listing of pronominal forms in Sanzhi.

Only first and second person pronouns receive the form gloss ⟨pro⟩; third person pronouns and other demonstrative forms are glossed ⟨dem_pro⟩ instead.

\[(24)\] c’ili ilt:ali kaq:ible k’aq:ib-le q’u’q‘,…

c’ili ilt:ali 3pl.erg k-aq:ib-le dem_pro.h: a_cv v:pred

‘Then they brought a hammer,...'  [mc_sanzhi_mill_0031]

\[(25)\] il mar xabar cab, ace dul bursulda.

il mar xabar 2sg-in ca-b that story be-n

a-c:ev du-l bv:pred

2sg-in 1sg-erg n-tell: icvb-1

‘This is a true story, I tell you.'  [mc_sanzhi_kurban_0053]

1.6 On the relative order of additional symbols

The GRAID annotations for Sanzhi Dargwa make use of a number of additional symbols for complex predicates (Section 1.2) and subordinate verb forms (Section 1.3.3) that attach to function glosses and clause boundary markers. While they are not by themselves particularly numerous, some complexity arises from their combination with other symbols that occupy the same space.

In the Sanzhi texts, the subject of a complex predicate of speech in convorb form, for instance, might receive the function gloss ⟨:_ds_cps_cv⟩, and in the direct speech that might follow, a negated complement clause with an infinitival predicate in P role would have ⟨:_ds_cc_in.neg:p⟩ as its clause boundary marker. While these are, thankfully, the worst case scenarios, they are far from uncommon occurrences.

In order to avoid confusion, these symbols combine in accordance with a strictly defined pattern. As a general rule, symbols that are not part of the base GRAID inventory always attach after (or outside of) those that are; in the Sanzhi Dargwa annotations, additional specifiers on function glosses (of subjects, mostly) are always added in the following order:

1. base function symbol: e.g. ⟨:_s⟩, ⟨:_a⟩, ⟨:_ncs⟩
2. subject of direct speech: ⟨:_ds⟩
3. complex predicate transitivity: ⟨:_cps⟩, ⟨:_cpa⟩
4. clause type (converb, participle, infinitive): ⟨:_cv⟩, ⟨:_pc⟩, ⟨:_in⟩

The various clause boundary tags likewise combine as follows (an extension of Haig & Schnell 2014: 25, Tab. 6); the first element to follow after the boundary marker ⟨#⟩ has no delimiter ⟨(_⟩ or ⟨_⟩):
2 Notes on the ReflND annotations

2.1 Cataphoric introductions

Occasionally, the first ‘mention’ of a new referent will be in a subordinate clause that precedes the formal introduction in the matrix clause. Mentions are always annotated as they occur in linear order, with the RefLex gloss being applied to the first occurrence of a referent index, irrespective of the form (or lack thereof) of the referring expression. These cases are readily identifiable in the analysis, which allows cataphoric distance and other aspects of the referential relationship to be determined.

(26) helka satːi argan kːurtːa-j či-raž-ible hel.

'sAs soon as she (sat) down, a fox that was passing by saw her.'

2.2 Referents in clauses otherwise not considered

Segments that have not been annotated for whatever reason, be that because they are incomplete or not syntactically well-formed, or because they are taken out of the normal flow of narration (e.g. because they address the listener, directly reply to the interviewer’s questions, or are not produced by the primary speaker), are marked as ⟨#nc⟩ ‘not considered’, and all of the elements they contain are glossed ⟨nc⟩.

However, these segments may still contain identifiable discourse references, which are presumably registered by the listener even in cases where the clause in question is abandoned partway through. So as to preserve the genuine sequence of references in the annotations, mentions in ⟨#nc⟩ segments are indexed with ReflND, even though they do not receive meaningful GRAID annotations. This is true of all Multi-CAST corpora with ReflND.

For the Sanzhi corpus, however, we have attempted to go one step further by adding form and person/animacy glosses back on those ⟨nc⟩’d elements that have referent indices. The glosses
are added as specifiers to the righthand side of the \langle nc \rangle symbol, yielding, for instance, \langle nc_np \rangle or \langle nc_pro.h \rangle. Grammatical functions are not glossed.

(27) a. caw qumuqlandew?
   \begin{align*}
   & ca-w & qumuqla&-\text{del} =w \\
   & \text{REFL-M} & \text{Kumyk-PST} =Q \\
   & \#nc & \text{nc_refl.h} & \text{nc} & \text{nc} \\
   & 0000
   \end{align*}

   [Someone in the audience asks,] 'Was he himself Kumyk?'

b. aʔa, dark:"an-de.
   \begin{align*}
   & aʔa & \text{no} & \text{0_journalist} & \text{Dargwa-PST} \\
   & \#nc & \text{nc_0.h} & \text{nc} & 0000
   \end{align*}

   'No, he was Dargi.'

Why not simply gloss these elements normally? Doing it in this (admittedly) roundabout way makes it clear that while some information can be gleaned from these elements, one cannot (and should not) rely on retrieving any information from the rest of the \langle #nc \rangle segment. For most types of analysis, the \langle nc_ \rangle glosses should not be conflated with related GRAID symbols.

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 MultiCAST Sanzhi Dargwa 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) gapped argument of a relative clause
dem_pro demonstrative pronoun
(pos_pro) possessive pronoun
(pn_np) proper name
(intro_other) interrogative pronoun
(indef_other) indefinite pronoun

Function symbols and specifiers

(:lvc) non-verbal complement of a complex predicate
(_ds) specifier: subject of a verb of speech; attaches to (:s), (:a), and (:ncs)
(_cps) specifier: subject of an intransitive complex predicate
(_cpa) specifier: subject of a transitive complex predicate
(_cv) specifier: subject of a converb clause
(_pc) specifier: subject of a participial clause
(_in) specifier: subject of an infinitival clause

Clause boundary symbols

(ds), (_ds) tag: direct speech clause
(cv), (_cv) tag: converb clause
(pc), (_pc) tag: participial clause
(in), (_in) tag: infinitival clause

Subconstituent symbols

(_adj) adjectival modifier; attaches to (ln) and (rn)
(_dem) demonstrative determiner; attaches to (ln) and (rn)
(_deti) indefinite determiner; attaches to (ln) and (rn)
(_num) numeral modifier; 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))
## B  List of abbreviated morphological glosses

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>first person</td>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>1/2</td>
<td>first/second person</td>
<td>M</td>
<td>human masculine singular</td>
</tr>
<tr>
<td>2</td>
<td>second person</td>
<td>MOD</td>
<td>modal</td>
</tr>
<tr>
<td>3</td>
<td>third person</td>
<td>MODQ</td>
<td>modal interrogative</td>
</tr>
<tr>
<td>ABL</td>
<td>ablative</td>
<td>MSD</td>
<td>masdar</td>
</tr>
<tr>
<td>ADD</td>
<td>additive</td>
<td>N</td>
<td>neuter singular</td>
</tr>
<tr>
<td>ADVZ</td>
<td>adverbializer</td>
<td>NEG</td>
<td>negative</td>
</tr>
<tr>
<td>ANTE</td>
<td>spatial case ‘before’</td>
<td>NMLZ</td>
<td>nominalizer</td>
</tr>
<tr>
<td>ATTR</td>
<td>attributive</td>
<td>NPL</td>
<td>neuter plural</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
<td>NUM</td>
<td>numeral</td>
</tr>
<tr>
<td>COMIT</td>
<td>comitative</td>
<td>OBL</td>
<td>oblique stem</td>
</tr>
<tr>
<td>COMP</td>
<td>comparative</td>
<td>OPT</td>
<td>optative</td>
</tr>
<tr>
<td>CONC</td>
<td>concessive</td>
<td>ORD</td>
<td>ordinal</td>
</tr>
<tr>
<td>COND</td>
<td>conditional</td>
<td>POST</td>
<td>spatial case ‘behind’</td>
</tr>
<tr>
<td>CVB</td>
<td>converb</td>
<td>PRET</td>
<td>preterite</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
<td>PRS</td>
<td>present</td>
</tr>
<tr>
<td>DEM</td>
<td>demonstrative</td>
<td>PROH</td>
<td>prohibitive</td>
</tr>
<tr>
<td>DIR</td>
<td>directional</td>
<td>PTCP</td>
<td>participle</td>
</tr>
<tr>
<td>ERG</td>
<td>ergative</td>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>EQ</td>
<td>equative enclitic</td>
<td>PRT</td>
<td>particle</td>
</tr>
<tr>
<td>F</td>
<td>human feminine singular</td>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
<td>PFV</td>
<td>perfective</td>
</tr>
<tr>
<td>HAB</td>
<td>habitual</td>
<td>Q</td>
<td>question</td>
</tr>
<tr>
<td>HPL</td>
<td>human plural</td>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>ICVB</td>
<td>imperfective converb</td>
<td>SPR</td>
<td>spatial case ‘on’</td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
<td>SUB</td>
<td>spatial case ‘under’</td>
</tr>
<tr>
<td>IN</td>
<td>spatial case ‘in’</td>
<td>NC</td>
<td>not classified</td>
</tr>
<tr>
<td>INF1</td>
<td>non-inflecting infinitive</td>
<td>[I]</td>
<td>code switching to Icari</td>
</tr>
<tr>
<td>INF2</td>
<td>inflecting infinitive</td>
<td>[R]</td>
<td>code switching to Russian</td>
</tr>
<tr>
<td>INDEF</td>
<td>indefinite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDQ</td>
<td>embedded question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPFV</td>
<td>imperfective</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>