

Multi-CAST

Tabasaran annotation notes

Natalia Bogomolova

Dmitry Ganenkov

Nils Norman Schiborr

August 2021
v1.1



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

Bogomolova, Natalia & Ganenkov, Dmitry & Schiborr, Nils N. 2021. Multi-CAST Tabasaran annotation notes. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#tabasaran) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Tabasaran annotation notes v1.1 last updated 1 August 2021
This document was typeset by NNS with X_YL^AT_EX and the *multicast3* class (v3.2.4).

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 | Predicative expressions of possession | 2 |
| 1.1.3 | Verbs of speech | 2 |
| 1.2 | Complex predicates | 3 |
| 1.3 | Structurally and pragmatically suppressed arguments | 4 |
| 1.3.1 | Gapped constituents in relative clauses | 5 |
| 1.3.2 | Subordinate verb forms | 5 |
| 1.3.3 | Infinitival complement clauses | 7 |
| 1.3.4 | Imperatives | 7 |
| 1.4 | Verbal cross-indices | 7 |
| 1.5 | Third person pronominal forms | 9 |
| 1.6 | Verbs of speech used as quotatives | 9 |
| 1.7 | On the relative order of additional symbols | 10 |
| 2 | Notes on the RefIND annotations | 11 |
| 2.1 | Referents in clauses otherwise not considered | 11 |
| | References | 11 |
| | Appendices | 13 |
| A | List of corpus-specific GRAID symbols | 13 |
| B | List of abbreviated morphological glosses | 14 |

1 Notes on the GRAID annotations

The following comprises selected notes on the GRAID (Haig & Schnell 2014) and RefIND (Schiborr et al. 2018) annotations of Tabasaran. It corresponds to version 2108 of the annotations, published in August 2021. 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

Tabasaran is a language with ergative-absolutive alignment (Bogomolova 2021): The Patient-like argument (P) of transitive clauses and the single argument (S) of intransitive clauses receive the same (absolutive) case marking in the morphology, distinct from the (ergative) marking of the Agent-like argument (A) of transitive clauses. In Tabasaran, the absolutive case (ABS) is not formally marked.

For the GRAID annotations of the A, S, and P roles in Tabasaran, 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) *hamu bic'i č:uč:u čan čuč:un va'č'.niqinži qipru hadmu vič.*

| | | | | |
|-----------------|--------------|---------------|----------------|----------------|
| <i>ha-mu</i> | <i>bic'i</i> | <i>č:uč:u</i> | <i>ča-n</i> | <i>čuč:u-n</i> |
| EMPH-PROX(ATTR) | small | brother(ERG) | REFL-GEN | brother-GEN |
| ## ln_dem | ln_adj | np.h:a | ln_refl.h:poss | ln_np.h:poss |

| | | | |
|-------------------------|---------------------|----------------|------------|
| <i>va'č'.ni-q-in-ži</i> | <i>q-ip-ru</i> | <i>ha-dmu</i> | <i>vič</i> |
| armpit-POST-LAT-DIR | POST-<NSG>throw-FUT | EMPH-3.P(ATTR) | apple(ABS) |
| np:g | v:pred | ln_dem | np:p |

'The small brother threw the apple into the armpit of his brother.'

[mc_tabasaran_horse_0094]

- (2) *rabadan jivar.ikan sad jiv.an vu'ru mažglar.in vul.az.*

| | | | | | |
|----------------|----------------------|-------------|----------------|---------------|---------------------|
| <i>rabadan</i> | <i>jiv-ar.i-k-an</i> | <i>sa-d</i> | <i>jiv.a-n</i> | <i>vu'-ru</i> | <i>mažgl-ar.i-n</i> |
| Rabadan(ABS) | day-PL-CONT-ELAT | one-NSG | day-GEN | come-FUT | Mezhgul-PL-GEN |
| ## pn_np.h:s | np:other | ln_num | np:other | v:pred | ln_pn_np.h:poss |

vul.a-z
village-DAT
np:g

'One day, Rabadan goes to the village of the Mezhgul people.'

[mc_tabasaran_belt_0030]

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). This approach 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, the morphology does not align with our cross-linguistic definition of S, A, and P, which necessitates adaptations to our annotation conventions, discussed in the sections below.

1.1.1 Experiencer verbs

In Tabasaran, a number of predicates assign quirky subject case. The most common of these are experiencer verbs such as ‘see’, ‘hear’, and ‘want’, whose Experiencer-argument takes the dative case. We have annotated such subjects as ⟨: ncs⟩ ‘non-canonical subjects’. With bivalent experiencer verbs, the Stimulus is in the absolutive, and is hence treated as ⟨: p⟩. There are numerous ways of interpreting the argument structure of these verbs; these use of the ⟨: ncs⟩ and ⟨: p⟩ symbols is a compromise solution. Examples of experiencer verbs are illustrated in (3) and (4).

- (3) *jarab allah k’ur, hamu uzuz ra^ˆq:u^ˆrajib nimk’i k’ur, ...*
- | | | | | | |
|--------------|--------------|-------------|----------------|--------------|--|
| <i>jarab</i> | <i>allah</i> | <i>k’ur</i> | <i>ha-mu</i> | <i>uzu-z</i> | <i>ra^ˆq:u^ˆra-ji-b</i> |
| VOC | Allah | CIT | EMPH-PROX(ABS) | I-DAT | <NSG>see-PRS-PTCP-NSG |
| #ds_pc | other | other | other | pro.1:ncs_pc | v:pred |
- nimk’i* *k’ur*
dream(ABS) CIT
np:p other
- ‘[He thought,] Oh God, am I seeing dreams?’ [mc_tabasaran_work_0112]

- (4) *ja muvaz fuk’ara gundar*
- | | | | | |
|-----------|---------------|---------------|---------------|----------------|
| <i>ja</i> | <i>muva-z</i> | <i>fuk’a</i> | <i>=ra</i> | <i>gun-dar</i> |
| or | PROX-DAT | nothing(ABS) | =ADD | want-PRS.NEG |
| ##neg | other | dem_pro.h:ncs | indef_other:p | =other v:pred |
- ‘He did not want anything.’ [mc_tabasaran_naz_0031]

1.1.2 Predicative expressions of possession

In predicative expressions of possession in Tabasaran, the possessor normally receives dative case and the possessee absolutive case. We treat these clauses as transitive clauses with quirky subject case, so that the possessor is annotated ⟨: ncs_poss⟩ ‘non-canonical subject of a possessive expression’, and the possessee as ⟨: p⟩ since it is in the absolutive.

- (5) *muvaz šulu sab liž, malar, žerdar, k’erqar, fji-r.*
- | | | | | | |
|---------------|--------------------|-------------|------------|----------------|----------------|
| <i>muva-z</i> | <i>šul-u</i> | <i>sa-b</i> | <i>liž</i> | <i>mal-ar</i> | <i>žerd-ar</i> |
| PROX-DAT | become-FUT | one-NSG | herd(ABS) | cattle-PL(ABS) | bull-PL(ABS) |
| ## | dem_pro.h:ncs_poss | v:pred | ln_num | np:p | rn_np |
- k’erq-ar* *fji-r*
calf-PL(ABS) what-PL(ABS)
rn_np rn
- ‘He had one herd, cattle, bulls, calves, whatever.’ [mc_tabasaran_horse_0146]

An alternative pattern where the possessor receives one of two spatial cases (APUD or POST) and the possessum the absolutive also occurs with marginal frequency; these cases are annotated the same manner as the more common case noted above.

1.1.3 Verbs of speech

Verbs of speech in Tabasaran are generally transitive, that is, they assign ergative case to the NP expressing the speaker. They are very commonly used to introduce direct speech, and only infre-

quently 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.

In the Tabasaran corpus, we have extended the morphological definition of A arguments to verbs of speech, even when they are not paired with a P argument in the annotations. To denote their special status, the subjects of these verbs have been annotated ⟨:a_ds⟩ ‘subject of a transitive verb of speech’ as in the examples below.

- (6) *hamu jic’ub vižiq’ aq:i k’uru muvu.*

| | | | | | |
|------|----------------|-------|----------------|---------------|----------------|
| | <i>ha-mu</i> | | <i>jic’u-b</i> | <i>vižiq’</i> | <i>aq:-i</i> |
| | EMPH-PROX(ABS) | | ten-NSG | skin(ABS) | (PL)bring(IMP) |
| ##ds | 0.2:a | 0.2:g | 1n_dem | 1n_num | np:p v:pred |

| | | |
|----|--------------|----------------|
| | <i>k’uru</i> | <i>muvu</i> |
| | say.FUT | PROX(ERG) |
| ## | v:pred | dem_pro.h:a_ds |

‘He said: Bring (me) these ten skins.’

[mc_tabasaran_nuradin_0039]

- (7) *sarun gabn.i muvaz dupna a’χirra, ...*

| | | | | | | |
|----|--------------|---------------|---------------|----------------|--------------|-----------------|
| | <i>sarun</i> | <i>gabn.i</i> | <i>muva-z</i> | <i>du-p-na</i> | <i>a’χir</i> | <i>=ra</i> |
| | PRT | cowboy(ERG) | PROX-DAT | PFV-say-RES | finally | =ADD |
| ## | other | np.h:a_ds | dem_pro.h:g | v:pred | other | =other ##ds ... |

‘Finally the cowboy told him, ...’

[mc_tabasaran_horse_0147]

1.2 Complex predicates

Complex predicates (CPs) combine a semantically weak light verb (or “vector verb”) such as *do*, *be*, or *become* with a some kind of non-verbal element, the latter of which supplies most of the semantic content to the expression. Crucially, the non-verbal element syntactically behaves like a regular object if nominal (Bogomolova 2021), but is only interpreted and annotated as such (i.e. ⟨:p⟩) if it is unambiguously referential. In all other cases, including where the complement is not nominal, it instead receives the special function gloss ⟨:lvc⟩ ‘light verb complement’, which marks it out as a special type of expression. In this case, it also invariably receives the form gloss ⟨other⟩ as its lexical category cannot always be ascertained. For ease of identification, the subjects of complex predicates further receive the specifier ⟨_cp⟩.

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 transitive, the subject of the complex predicate is ergative as in (8) and (9); if intransitive, it is absolutive as in (10). Lastly, there are a number of CPs derived from unaccusative intransitive constructions (Bogomolova 2021), in which the dative-marked argument has been re-analyzed as the subject of the clause, as in (11).

- (8) *mu bic'i čuč.:u guč' ap'udar.*

| | | | | |
|------------|--------------|---------------|-------------|------------------|
| <i>mu</i> | <i>bic'i</i> | <i>čuč.:u</i> | <i>guč'</i> | <i>ap'-u-dar</i> |
| PROX(ATTR) | small | brother(ERG) | fear | do-FUT-NEG |
| ##neg | ln_dem | ln_adj | np.h:a_cp | other:lvc v:pred |

‘The little brother was not afraid.’ [mc_tabasaran_horse_0034]

(9) *nüra'li ba<r>q:u xp:ir.i mašar č'ur ap'uru.*

| | | | | |
|----------------|------------------------|------------|-----------|----------|
| <i>nüra'li</i> | <i>ba<r>q:-u</i> | | | |
| Nurali(ABS) | PFV-<HSG>see-PCVB | | | |
| ## | #cv | 0.h:ncs_cv | pn_np.h:p | v:pred % |

| | | | |
|----------------|---------------|-------------|----------------|
| <i>xp:ir.i</i> | <i>maš-ar</i> | <i>č'ur</i> | <i>ap'u-ru</i> |
| wife(ERG) | face-PL(ABS) | bad | do-FUT |
| np.h:a_cp | other:lvc | other | v:pred |

‘When she saw Nurali, she made a grimace.’ [mc_tabasaran_belt_0057]

(10) *a'χir rabadan bav.ri aqru.*

| | | | |
|--------------|----------------|-------------------|------------------|
| <i>a'χir</i> | <i>rabadan</i> | <i>bav.ri</i> | <i>aq-ru</i> |
| finally | Rabadan(ABS) | consciousness(IN) | <HSG>fall-FUT |
| ## | other | pn_np.h:s_cp | other:lvc v:pred |

‘Finally, Rabadan figures it out.’ [mc_tabasaran_belt_0014]

(11) *rabadan.žiz χabar šuldar q'an.az.*

| | | | |
|--------------------|----------------|-----------------|-----------------|
| <i>rabadan.žiz</i> | <i>χabar</i> | <i>šul-dar</i> | <i>q'an.a-z</i> |
| Rabadan-DAT | news(ABS) | become-(PRS)NEG | late-DAT |
| ##neg | pn_np.h:ncs_cp | other:lvc | v:pred other |

‘Rabadan knows nothing for a long time.’ [mc_tabasaran_belt_0012]

1.3 Structurally and pragmatically suppressed arguments

In GRAID, unexpressed clausal referents ⟨ \emptyset ⟩ 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 ⟨ \emptyset ⟩, 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 Tabasaran and Sanzhi Dargwa, 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 a subset of Multi-CAST corpora; it is planned to become an optional gloss in the standard GRAID specification in the future.

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

1.3.1 Gapped constituents in relative clauses

Relative clauses in Tabasaran are pre-nominal, and generally contain some non-finite (or less finite) form of a verb. The head noun is obligatorily gapped in the relative clause; there are no resumptive pronouns. Where in standard GRAID the gap would remain unannotated, in the Tabasaran corpus (and a number of other corpora) they receive the form gloss ⟨f0⟩ with an additional specifier ⟨rel_⟩ to mark out the context in which they occur. This practice will in the future be adopted into the standard GRAID specification.

- (12) *vadavu hamu jicara qa<v>raji t’ulra, ...*
- | | | | | |
|------------------------|------------------------|---------------------------|-------------------|------------|
| <i>vadavu</i> | <i>hamu</i> | <i>q-a<v>-ra-ji</i> | <i>t’ul</i> | <i>=ra</i> |
| <NSG>take-PCVB | EMPH-PROX(ATTR) | <NSG>POST-lead-PRS-PTCP | stick(ABS) | =ADD |
| #cv 0.h:a_cv v:pred | #rc_pc 0.h:a_pc ln_dem | v:pred | rel_f0:obl % np:p | =other |
| | | | | =a |
| | | | | =ADD |
| | | | | =other |
- ‘Taking the stick which he used to drive the oxen, ...’ [mc_tabasaran_work_0183]

1.3.2 Subordinate verb forms

Tabasaran 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. But 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, with the exception of the subjects of infinitival clauses (Section 1.3.3), omitted arguments are annotated as ⟨0⟩ ‘contrastively suppressed’ rather than ⟨f0⟩ ‘structurally suppressed’, and the head of the verbal complex as regular ⟨v⟩ rather than ⟨vother⟩ ‘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: ⟨_cv⟩ for converb clauses, ⟨_pc⟩ for parti-

cial clauses, and ⟨_in⟩ for infinitival clauses. The masdar is exceedingly rare in the annotated texts, and so is not labelled. 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, (13) for converb clauses, (14) for participle clauses, and (15) for infinitive and subjunctive clauses. See also Section 1.7 for how these extra specifiers are ordered relative to the base GRAID symbols.

- (13) *hamrar.ixā herχri muvu čaz sab χal žibru.*

| | | | | |
|--------|------------------------|----------------|-------------|--------------------------|
| | <i>ha-m-rar.i-x-a</i> | <i>herχ-ri</i> | <i>muvu</i> | <i>ča-z</i> |
| | EMPH-PROX-PL-APUD-ELAT | ask-ICVB | PROX(ERG) | REFL.SG-DAT |
| ## #cv | 0.h:s_cv | dem_pro.h:obl | v:pred | % dem_pro.h:a refl.h:obl |

| | | |
|-------------|------------|---------------|
| <i>sa-b</i> | <i>χal</i> | <i>žib-ru</i> |
| one-NSG | house(ABS) | put-FUT |
| ln_num | np:p | v:pred |

‘Asking them [for help], (Nuradin) built a house for himself.’

[mc_tabasaran_nuradin_0018]

- (14) *bačan hamu la[˘]χn.ika χabar abxir diš šulu va[˘]rac’an.žiz.*

| | | | | |
|-----------|-----------------|----------------|-------------------------------|--------------|
| | <i>bačan</i> | <i>ha-mu</i> | <i>la[˘]χn.i-k-a</i> | <i>χabar</i> |
| | Bachan(ABS) | EMPH-PROX(ABS) | work-CONT-ELAT | news(ABS) |
| ## #ac_pc | pn_np.h:s_cp_pc | ln_dem | np:obl | other:lvc |

| |
|-------------------------|
| <i>a-b-x-i-r</i> |
| PFV-NSG-become-PTCP-HSG |
| v:pred |

| | | |
|------------|--------------|----------------------------------|
| <i>diš</i> | <i>šul-u</i> | <i>va[˘]rac’an.ži-z</i> |
| quick | become-FUT | Qaratsan-DAT |
| % 0.h:s | other | v:pred |
| | | pn_np:g |

‘Having learnt of these happenings, Bachan rushes to Qaratsan.’

[mc_tabasaran_belt_0025]

- (15) *qa murar.in jib živru pač:ih.ži sumčir ap’uz, čan šubar.in.*

| | | | | |
|-----------|-------------------|------------|---------------|------------------|
| <i>qa</i> | <i>mu-rar.i-n</i> | <i>jib</i> | <i>živ-ru</i> | <i>pač:ih.ži</i> |
| then | PROX-PL-GEN | day(ABS) | put-FUT | king(ERG) |
| ## other | ln_dem_pro.h:poss | np:p | v:pred | np.h:a |
| | | | | #ac_in f0.h:a_in |

| | | | |
|---------------|---------------|----------------|------------------|
| <i>sumčir</i> | <i>ap’-uz</i> | <i>ča-n</i> | <i>šubar.i-n</i> |
| wedding(ABS) | do-INF | REFL.SG-GEN | girl+PL-GEN |
| np:p | vother:pred | rn_refl.h:poss | rn_np.h:poss |

‘The king set a day to hold the weddings of his daughters.’

[mc_tabasaran_horse_0120]

A last 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 Tabasaran 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 Tabasaran 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, it should be noted), we believe the benefits of this approach outweigh its disadvantages.

1.3.3 Infinitival complement clauses

The subjects of infinitival clauses are structurally suppressed, and are hence annotated ⟨f0⟩:

- (16) *vallah k'ur, hamc:i k'ur vardaš.*
vallah k'ur ha-m-c:i k'ur vardaš
 INTERJ CIT EMPH-PROX-ADV CIT brother(ABS)
 other other ln other np.h:s #ds_ac_in f0.h:s_in
liχu-z B-uš-nu
 <HSG>work-INF PFV-<HSG>go-AOR
 vother:pred % v:pred
 'He said,] Oh, my brother has gone off to work.' [mc_tabasaran_work_0094]

- (17) *qa a'χu' čuč:uz a'bu'z k:un šulu.*
qa a'χu' čuč:u-z a'bu'z k:un
 then big brother-DAT (HSG)go-INF want
 ## other ln_adj np.h:ncs_cp #cc_in f0.h:s_in vother:pred % other:lvc
šul-u
 become-FUT
 v:pred
 'Then the big brother wanted to go.' [mc_tabasaran_horse_0074]

1.3.4 Imperatives

While the subjects of imperatives are almost always left unexpressed, they are in fact not categorically suppressed. For this reason, like the subjects of subordinate verb forms discussed in Section 1.3.2, they are annotated ⟨0⟩ rather than ⟨f0⟩:

- (18) *uvu fu dap'naš jip k'ur!*
uvu fu d-ap'-na-š jip k'ur
 you.SG(ERG) what PFV-do-RES-COND say(IMP) CIT
 ##ds pro.2:a #cc:p 0.2:a intrg_other:p v:pred % v:pred other
 'Say what you have done!' [mc_tabasaran_belt_0075]

1.4 Verbal cross-indices

In addition to the gender-based agreement system, in which absolutive-marked arguments are generally the controllers, Tabasaran also has a second layer of indexing (Bogomolova 2018), presumably an innovation. In this system, the controllers are first and second person arguments only. We have adopted a conservative (possibly redundant) system of glossing which is sufficiently explicit to allow recognition of the relevant items.

The basic outline of the system is as follows: First or second person core arguments, including the dative-marked subjects discussed in Section 1.1.1 above, are indexed via a clitic on the verb. This applies to both transitive and intransitive predicates, but is limited to finite verb forms in matrix clauses, hence the lack of indexing on the imperative verb in (24). With subjects, the clitics are obligatory; with other arguments they may be optionally present as in (19), (22), and (23).

- (19) *uzu saban jiz uzu pač:ih vap'unza k'ur, qa jiz xp:iri vap'unzu k'ur, qa jiz xp:ir.in daši.ji vap'unzu k'ur.*

uzu saban jiz uzu pač:ih vap'un =za k'ur
 I(ERG) first my I(ABS) king(ABS) PFV-do-PST =1SG:AG CIT
 ##ds pro.1:a other 1n refl.1:p np.h:obl v:pred =rv-pro_1_a other

qa jiz xp:iri vap'un =zu
 then my wife(ERG) PFV-do-PST =1SG:PAT
 ##ds other 1n_pro.1:poss np.h:a 0.1:p 0.h:obl v:pred =rv-pro_1_p

qa jiz xp:ir.i-n daši.ji vap'un
 then my wife-GEN father(ERG) PFV-do-PST
 ##ds other 1n_pro.1:poss 1n_np.h:poss np.h:a 0.1:p 0.h:obl v:pred

=zu k'ur
 =1SG:PAT CIT
 =rv-pro_1_p other

‘[He said,] First, I made myself the king, then my wife made me the king, then my wife’s father made me the king.’ [mc_tabasaran_horse_0203]

- (20) *sarun, uvuz uvu a'jdazuz, uvuzra uzu a'jdaruz, fs'o.*

sarun uzu-z uvu a'j-da =zu-z
 PRT I-DAT you know-NEG(PRS) =1SG-DAT
 ##ds.neg other pro.1:ncs pro.2:p v:pred =rv-pro_1_ncs ##ds.neg

uvu-z =ra uzu a'j-dar =u-z fs'o
 2SG-DAT =ADD I know-NEG(PRS) =2SG-DAT that's_it
 pro.2:ncs =other pro.1 p v:pred =rv-pro_2_ncs other

‘Fine, you don’t know me, I don’t know you, that’s it.’ [mc_tabasaran_belt_0028]

- (21) *uvu uzuxna baχnijiš k'ur, uzu χujza k'ur.*

uvu uzu-x-na baχ-ni-ji-š k'ur
 you.SG(ABS) I-APUD-LAT PFV-<NSG>bring-PFV-PST-COND CIT
 ##ds #ds_ac pro.2:a pro.1:g 0:p v:pred other %

uzu χu-j =za k'ur
 I(ABS) bring-PST-COND =1SG:AG CIT
 pro.1:a 0:p 0.2:g v:pred =rv-pro_1_a other

‘[She said,] If you had brought something for me, I would bring out (something for you).’

[mc_tabasaran_naz_0047]

- (22) *qa la^sχin uzu tuvurzavuz k'ur, ...*

qa la^sχin uzu tuv-ur =za =vu-z k'ur
 then work(ABS) I do-FUT =1SG:AG =2SG-DAT CIT
 ##ds other np:p pro.1:a 0.2:g v:pred =rv-pro_1_a =rv-pro_2_g other

‘[He said,] I will give you a job, ...’

[mc_tabasaran_work_0033]

(23) *sarun fu k'uzavuz, ...*

| | | | | |
|--------------|-----------|-------------|---------------|--------------------------------|
| <i>sarun</i> | <i>fu</i> | <i>k'-u</i> | <i>=za</i> | <i>=vu-z</i> |
| PRT | what(ABS) | say-FUT | =1SG:AG | =2SG-DAT |
| ## other | 0.1:a | 0.2:g | intrg_other:p | v:pred =rv-pro_1_a =rv-pro_2_g |

'What else should I tell you, ...' [mc_tabasaran_horse_0081]

(24) *teet uzu k'ur.*

| | | |
|-------------|------------|-------------|
| <i>teet</i> | <i>uzu</i> | <i>k'ur</i> |
| leave(IMP) | I(ABS) | CIT |
| ##ds | 0.2:a | v:pred |
| | pro.1:p | other |

'[It said,] Let me go.' [mc_tabasaran_horse_0057]

1.5 Third person pronominal forms

Like Sanzhi Dargwa, Tabasaran does not have a separate paradigm of third person personal pronouns; their role is filled by an extensive set of demonstrative pronouns. Only first and second person pronouns receive the form gloss ⟨pro⟩; third person pronouns and other demonstrative forms are glossed ⟨dem_pro⟩ instead.

(25) *ap'uru muvu čaz sab užub χalla.*

| | | | | | | |
|----------------|-------------|--------------|-------------|--------------|------------|-------------|
| <i>ap'-uru</i> | <i>muvu</i> | <i>ča-z</i> | <i>sa-b</i> | <i>užu-b</i> | <i>χal</i> | <i>=la</i> |
| do-FUT | PROX(ERG) | REFL(SG)-DAT | one-NSG | good-NSG | house(ABS) | =ADD |
| ## | v:pred | dem_pro.h:a | refl.h:obl | ln_num | ln_adj | np:p =other |

'He built a good house for himself.' [mc_tabasaran_nuradin_0015]

(26) *uzu uxuz la^χin agurza k'ur.*

| | | | | | |
|------------|--------------|-------------------------|-----------------------|------------|-------------------|
| <i>uzu</i> | <i>uxu-z</i> | <i>la^χin</i> | <i>ag-ur</i> | <i>=za</i> | <i>k'ur</i> |
| I | we(INCL)-DAT | work(ABS) | <NSG>search-FUT | =1SG:AG | CIT |
| ##ds | pro.1:a | pro.1:obl | np:p | v:pred | =rv-pro_1_a other |

'[He said,] I will find a job for us.' [mc_tabasaran_work_0021]

1.6 Verbs of speech used as quotatives

In Tabasaran, the verb *k'ur* 'say.FUT' is very frequently used in lieu of a quotative, often interspersed throughout quoted material as in (27). In this use, it always occurs without an overt subject and any clausal modifiers. So as to not inflate the number of embedded clauses in the corpus, it has been glossed it as CIT 'quotative' and annotated simply as ⟨other⟩ rather than as the predicate of a clause where it occurs on its own alongside direct speech.

- (27) *ja žihil k'ur, fu χabar u k'ur, jaman fikrar ap'urava a'χir k'ur, fujav derdi k'ur.*
- | | | | | | | |
|-----------|--------------|-------------|-----------|--------------|----------|----------------|
| <i>ja</i> | <i>žihil</i> | <i>k'ur</i> | <i>fu</i> | <i>χabar</i> | <i>u</i> | <i>k'ur</i> |
| VOC | young(ABS) | CIT | what | news(ABS) | COP | CIT |
| ##ds | ln | np.h:voc | other | ln | np:s | v:predex other |
-
- | | | | | | |
|--------------|-----------------|----------------|------------|--------------|-------------------------|
| <i>jaman</i> | <i>fikr-ar</i> | <i>ap'-ura</i> | <i>=va</i> | <i>a'χir</i> | <i>k'ur</i> |
| bad | thought-PL(ABS) | do-PRS | =2SG:AG | finally | CIT |
| ##ds | 0.2:a_cp | ln_adj | other:lvc | v:pred | =rv-pro_1_a other other |
-
- | | | | |
|-----------|------------------|----------------|-------------|
| <i>fu</i> | <i>=jav</i> | <i>derdi</i> | <i>k'ur</i> |
| what+COP | =2SG:POSS | trouble(ABS) | CIT |
| ##ds | intrg_other:pred | =ln_pro.2:poss | np:s other |
- 'Ah, boy, (he said,) what's up, (he said,) you are having bad thoughts, (he said,) what's bothering you, (he said)?' [mc_tabasaran_work_0031]

1.7 On the relative order of additional symbols

The GRAID annotations for Tabasaran make use of a number of additional symbols for complex predicates (Section 1.2) and subordinate verb forms (Section 1.3.2) 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.

A maximally complex example from the Tabasaran texts might be the subject of a complex predicate of speech in converb form, which would receive the function gloss ⟨:s_ds_cp_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 not uncommon occurrences.

In order to avoid confusion, these symbols combine in a strictly defined order. As a general rule, symbols that are not part of the base GRAID inventory always attach after (or rather, 'outside of') those that are; in the Tabasaran 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⟩, or
subject of verbal expression of possession: ⟨_poss⟩
3. subject of complex predicate: ⟨_cp⟩
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 ⟨.⟩):

1. boundary marker: ⟨#⟩ or ⟨##⟩
2. direct speech clause: ⟨ds⟩
3. clause type (complement, adverbial, relative): ⟨_cc⟩, ⟨_ac⟩, ⟨_rc⟩
4. clause type (converb, participle, infinitive): ⟨_cv⟩, ⟨_pc⟩, ⟨_in⟩
5. negated: ⟨.neg⟩
6. function: e.g. ⟨:s⟩, ⟨:a⟩, ⟨:p⟩, etc.

2 Notes on the RefIND annotations

2.1 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 part-way through. So as to preserve the genuine sequence of references in the annotations, mentions in <#nc> segments are indexed with RefIND, even though they do not receive meaningful GRAID annotations. This is true of all Multi-CAST corpora with RefIND.

For the Tabasaran corpus, however, we have attempted to go one step further by adding form and person/animate glosses back onto those <nc> elements that have referent indices. The glosses are added as specifiers to the righthand side of the <nc> symbol, yielding, for instance, <nc_np> or <nc_pro.h>. Grammatical functions are not glossed.

- (28) a. *hamus šaʃʃban udučʰiš...*
 hamus šaʃʃban udučʰ-iš
 now Shaban(ABS) <HSG>go_out-COND
 #nc nc nc_pn_np.h nc
 0008

‘If Shaban goes out...’

- b. *hamu čan χanuk.rin?*
 ha-mu ča-n χanuk.ri-n
 EMPH-PROX(ABS) REFL-GEN friend-GEN
 #nc nc_dem_pro.h nc_refl.h nc_np.h

‘[Someone in the audience asks,] To his friend?’

[mc_tabasaran_belt_0032-0033]

This approach makes it clear that while some information can be gleaned from these elements, one should not rely on being able to retrieve full information from the rest of the <#nc> segment. For most types of analysis, the <nc_> glosses should not be conflated with related GRAID symbols.

References

- Andrews, Avery. 2007. The major functions of the noun phrase. In Shopen, Timothy (ed.), *Language typology and syntactic description*, vol. 1: Clause structure, 132–223. Cambridge: Cambridge University Press.
- Bogomolova, Natalia. 2018. The rise of person agreement in East Lezgi: Accessing the role of frequency. *Linguistics* 56(4). 819–844.
- Bogomolova, Natalia. 2021. Ergativity in Tabasaran: A reply to Woolford (2015). *Linguistic Inquiry*. (https://doi.org/10.1162/ling_a_00420).
- Bogomolova, Natalia & Ganenkov, Dmitry & Schiborr, Nils N. 2021. Multi-CAST Tabasaran. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. (<https://multicast.aspra.uni-bamberg.de/#tabasaran>) (Accessed 2021-01-27).
- 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).
- Schiborr, Nils N. & Schnell, Stefan & Thiele, Hanna. 2018. *RefIND – Referent Indexing in Natural-language Discourse: Annotation guidelines (v1.1)*. University of Bamberg Unpublished manuscript. (<https://multicast.aspra.uni-bamberg.de/#annotations>) (Accessed 2019-03-08).

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 Tabasaran 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 |
| <pn_np> | proper name |
| <intrg_other> | interrogative pronoun |
| <indef_other> | indefinite pronoun |

Function symbols and specifiers

| | |
|---------|--|
| <:lvc> | non-verbal complement of a complex predicate |
| <_ds> | <i>specifier:</i> subject of a verb of speech; attaches to <:s>, <:a>, and <:ncs> |
| <_cp> | <i>specifier:</i> subject of a complex predicate |
| <_cv> | <i>specifier:</i> subject of a converb clause |
| <_pc> | <i>specifier:</i> subject of a participial clause |
| <_in> | <i>specifier:</i> subject of an infinitival clause |
| <_poss> | <i>specifier:</i> subject of a verbal expression of possession; attaches to <:ncs> |

Clause boundary symbols

| | |
|-------------|--------------------------------|
| <cv>, <_cv> | <i>tag:</i> converb clause |
| <pc>, <_pc> | <i>tag:</i> participial clause |
| <in>, <_in> | <i>tag:</i> infinitival clause |

Subconstituent symbols

| | |
|--------|---|
| <_adj> | adjectival modifier; attaches to <ln> and <rn> |
| <_dem> | demonstrative determiner; attaches to <ln> and <rn> |
| <_num> | numeral modifier; 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 | INCL | inclusive |
| 2 | second person | INDEF | indefinite |
| 3 | third person | INF | infinitive |
| ABS | absolutive | INTER | spatial case ‘between’ |
| ABSTR | abstract | INTERJ | interjection |
| ADD | additive | IPFV | imperfective |
| ADV | adverbial | JUSS | jussive |
| AG | agent | LAT | lative |
| AOR | aorist | LOC | locative |
| APUD | spatial case ‘by, near’ | MSD | masdar |
| ATTR | attributive | NEG | negative |
| CIT | verb <i>k’ur</i> ‘say.FUT’ used as a quotative | NMLZ | nominalization |
| COM | comitative | NSG | neuter singular |
| COMP | comparative particle | ORD | ordinal number |
| COND | conditional | PAT | patient |
| CONT | continuative | PCVB | perfective converb |
| CONTR | contrastive | PFV | perfective |
| COP | copula | PL | plural |
| DAT | dative | POSS | possessive |
| DEF | definite | POST | spatial case ‘behind, after’ |
| DIR | spatial case ‘to’ | PROH | prohibitive |
| DIST | distal | PROX | proximal |
| DOWN | prefixal marker in demonstrative pronoun ‘downwards’ | PRS | present |
| ELAT | elative | PRT | particle |
| EMPH | emphasis | PST | past |
| ERG | ergative | PTCP | participle |
| EXCL | exclusive | Q | question particle |
| FOC | focal particle | REFL | reflexive |
| FUT | future | RES | resultative |
| GEN | genitive | SG | singular |
| HSG | human singular | SUB | spatial case ‘under’ |
| ICVB | imperfective converb | SUPER | spatial case ‘on’ |
| IMP | imperative | TEMP | temporal marker ‘when’ |
| IN | spatial case ‘in’ | VOC | vocative |
| | | NC | not classified |

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/