Multi-CaST
Northern Kurdish annotation notes

Geoffrey Haig
Maria Vollmer
Hanna Thiele

January 2021
v2.1
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)
Contents

1 Notes on the GRAID annotations ........................................ 1
  1.1 Transitive and intransitive verbs: Verbs of speech 1
  1.2 Light verb constructions ............................................ 1
  1.3 Adpositions ............................................................ 3
  1.4 Verb serialization and the verb ‘bring’ ............................ 3

References ................................................................. 3

Appendices ................................................................. 4
  A List of corpus-specific GRAID symbols ...................... 4
  B List of abbreviated morphological glosses ................... 5
1 Notes on the GRAID annotations

This document contains notes on the implementation of the GRAID annotation conventions (Haig & Schnell 2014) in the Multi-CAST Northern Kurdish corpus. It corresponds to version 2101 of the annotations, published in January 2021. Unless a more recent version of this document exists, it also applies to any later versions of the annotations.

1.1 Transitive and intransitive verbs: Verbs of speech

Transitive and intransitive verbs can be distinguished on morphological grounds: transitive verbs trigger the ergative construction in the past tenses, while intransitives do not. In general, this was taken as the main diagnostic for distinguishing transitive and intransitive predicates, with one main exception: verbs of speech, in particularly gotin ‘say’. Morphologically, it definitely belongs in the class of transitive verbs. However, it is very rarely used with a nominal object; rather, it generally introduces direct speech, which may continue for several clauses (and hence is difficult to classify as an “object”).

We have therefore decided to treat this verb as intransitive, but attach a special specifier to the form gloss of the verb, ⟨ds_v⟩, in order to facilitate the identification of these (highly frequent) verb forms. If the complement of the verb is direct speech, as it is in the overwhelming majority of cases, then the complement clause receives the clause boundary gloss ⟨#ds⟩, and the subject of the verb of speech is treated as a special case of intransitive subject, namely ⟨:s_ds⟩. In those few cases where the complement is a nominal object and hence glossed as ⟨:p⟩, the subject of the verb of speech is glossed ⟨:a_ds⟩.

(1) Qîzikê gotîye: Perê min pir e.

\[ qizik-ê \quad got-îye: \quad per-ê \quad min \quad pir \quad =e \]
\[ girl-obl \quad say,pst-prez,3sg \quad money-ez \quad 1sg.obl \quad much \quad =cop,3sg \]
\[ # np:h:s_ds \quad ds_v:pred \quad #ds np:s \quad rn_pro.1:poss \quad other:pred =cop \]

‘The girl said, “I have much money.”’

(2) Em ê ji xwe ra xwedî kin.

\[ em \quad =ê \quad ji \quad xwe \quad ra \quad xwedî \quad k-in \]
\[ 1pl \quad =fut \quad for \quad refl \quad pop \quad 0:her \quad raising \quad sbj:do:prs-pl \]
\[ #ds \quad pro.1:a \quad =aux \quad adp \quad refl.1:obl \quad pop_adp \quad 0:pop \quad other:lvc \quad v:pred \]

‘“We will raise (her) for ourselves.”’

1.2 Light verb constructions

Northern Kurdish makes extensive use of light verb constructions: some non-verbal element is combined with one of a small number of so-called light verbs, most commonly ‘do, be, give, strike, bring,’ to create a new predicate. Many of these can conveniently be treated as simple predicates, and in some cases they are orthographically treated as one word. In other cases the borderline to other combinations of generic object plus verb is difficult to draw. For highly frequent combinations, particularly if the non-verbal element does not occur outside of the particular combination, we have glossed the non-verb combination with ⟨:lvc⟩:

(2) Em ê ji xwe ra xwedî kin.

\[ em \quad =ê \quad ji \quad xwe \quad ra \quad xwedî \quad k-in \]
\[ 1pl \quad =fut \quad for \quad refl \quad pop \quad 0:her \quad raising \quad sbj:do:prs-pl \]
\[ #ds \quad pro.1:a \quad =aux \quad adp \quad refl.1:obl \quad pop_adp \quad 0:pop \quad other:lvc \quad v:pred \]

‘“We will raise (her) for ourselves.”’
Light verb constructions raise certain issues with regard to transitivity. As a general rule in Northern Kurdish, lexical verbs are quite strictly sub-categorized for transitivity. Those that belong in the class of transitive verbs will always trigger an ergative subject in past tenses, and this is usually taken as a diagnostic for transitivity (see above). With light verb constructions, the same rule applies: a transitive light verb (e.g. kirin ‘do’, dan ‘give’) triggers an ergative subject in past tenses. However, at the level of syntax many of the light verb constructions created with a transitive light verb do not govern a referential direct object, and in terms of their semantics, they more closely resemble an intransitive predicate. This results in a mismatch of lexical and syntactic transitivity (see Haig 2002 for a discussion). In such cases we have generally followed syntactic transitivity, and have treated the subject as ⟨:s⟩. However, in order to facilitate identification of this class of ⟨:s⟩, we have added a specifier ⟨:s_cp⟩ (‘complex predicate’). This gloss indicates that we are dealing with the subject of a light verb construction, with a transitive lexical verb, but overall syntactic intransitivity; the same convention has also been applied to the MultiCAST Persian corpus (Adibifar 2016, see annotation notes therein). This glossing convention is exemplified in (3).

(3) Bila bibê ewleda me, tu çîma meraq dîkî?

Example (4) illustrates the straightforward case of an intransitive light verb, with the subject treated simply as ⟨:s⟩. (5) shows a transitive light verb in a light verb construction that can govern a direct object; here the subject is glossed as a regular ⟨:a⟩.

(4) Rastê pîrêki tê.

(5) Pîre tîne ji wan ra nan-man hazir dîke.

‘The old lady prepared a meal for them.’
1.3 Adpositions

Northern Kurdish has quite a number of circumpositions consisting of a preposition plus a postpositional clitic particle. The former is glossed ⟨adp⟩, the latter ⟨pop_adp⟩; see (2) above for an example. In some cases, a third person singular pronominal complement of a circumposition is reduced in form and fuses with the preposition. In such examples, we detach the bound pronoun from the preposition, providing it with its own annotation cell. This is illustrated in (6).

(6) Ezê te pê ra bizewicînim.

```
1sg =ê te bi -wê ra bi-zewicîn-im
#ds pro.1:a =lv_aux pro.2:p adp -pro.h:obl pop_adp v:pred

"'I will betroth you with him.'" [mc_nkurd_muserz03_0009]
```

1.4 Verb serialization and the verb ‘bring’

In the dialect of Northern Kurdish recorded for this corpus, we often find tight sequences of motion verb together with another verb, commonly anîn ‘bring, take’. The two verbs generally share a common subject. The problem is to decide whether we are dealing with two distinct predications, or a single item. Related to this is the use of anîn as a kind of aspectual converb, expressing swift and direct action, rather than its lexical meaning of ‘take, bring’. We have been obliged to decide on a case-by-case basis, based on criteria of agreement patterns (sometimes one verb does not carry the expected agreement), and the semantics of the overall expression, that is, to what extent it can be considered a sequence of two distinct propositions or a single event.

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 Northern Kurdish 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 in a relative clause
⟨rel_pro⟩ relative pronoun
⟨pn_np⟩ proper name
⟨ds_v⟩ verb of speaking
⟨intrg_other⟩ interrogative pronoun

Function symbols and specifiers

⟨:lvc⟩ non-verbal complement of a light verb construction
⟨:s_cp⟩ subject of an intransitive complex predicate with transitive light verb
⟨:a_ds⟩ subject of a transitive verb of speech
⟨:s_ds⟩ subject of an intransitive verb of speech

Clause boundary symbols

⟨#rep⟩, ⟨_rep⟩ clause unit is a verbatim repetition of a preceding clause

Subconstituent symbols

⟨_dem⟩ adnominal demonstrative; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_num⟩ adnominal numeral or quantifier; attaches to ⟨ln⟩ and ⟨rn⟩
⟨_vother⟩ various types of non-finite or subordinated verb, most commonly the secondary verb in a serial verb construction; attaches to ⟨lv⟩ and ⟨rv⟩

Other symbols

⟨pop_adp⟩ postpositional particle; co-occurs with a preposition (see (2) above)
⟨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></th>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>first person</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>second person</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>third person</td>
</tr>
<tr>
<td>ADP</td>
<td>ADP</td>
<td>adposition</td>
</tr>
<tr>
<td>COMPL</td>
<td>COMPL</td>
<td>complementizer</td>
</tr>
<tr>
<td>COP</td>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>DEM</td>
<td>DEM</td>
<td>demonstrative</td>
</tr>
<tr>
<td>DRCT</td>
<td>DRCT</td>
<td>directional particle</td>
</tr>
<tr>
<td>EMPH</td>
<td>EMPH</td>
<td>emphasis</td>
</tr>
<tr>
<td>EZ</td>
<td>EZ</td>
<td>ezafe</td>
</tr>
<tr>
<td>EXCL</td>
<td>EXCL</td>
<td>exclamation</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>feminine</td>
</tr>
<tr>
<td>FUT</td>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>IMP</td>
<td>IMP</td>
<td>imperative</td>
</tr>
<tr>
<td>IND</td>
<td>IND</td>
<td>indicative</td>
</tr>
<tr>
<td>INDF</td>
<td>INDF</td>
<td>indefinite</td>
</tr>
<tr>
<td>M</td>
<td>M</td>
<td>masculine</td>
</tr>
<tr>
<td>MOD</td>
<td>MOD</td>
<td>modal</td>
</tr>
<tr>
<td>NEG</td>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>OBL</td>
<td>OBL</td>
<td>oblique case</td>
</tr>
<tr>
<td>PL</td>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POP</td>
<td>POP</td>
<td>postpositional particle</td>
</tr>
<tr>
<td>POSS</td>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>PRF</td>
<td>PRF</td>
<td>perfective</td>
</tr>
<tr>
<td>PROG</td>
<td>PROG</td>
<td>progressive</td>
</tr>
<tr>
<td>PRS</td>
<td>PRS</td>
<td>present</td>
</tr>
<tr>
<td>PST</td>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>PTCP</td>
<td>PTCP</td>
<td>participle</td>
</tr>
<tr>
<td>RED</td>
<td>RED</td>
<td>partial reduplication</td>
</tr>
<tr>
<td>REFL</td>
<td>REFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>SUBJ</td>
<td>SUBJ</td>
<td>subjunctive</td>
</tr>
<tr>
<td>SG</td>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>VOC</td>
<td>VOC</td>
<td>vocative</td>
</tr>
<tr>
<td>NC</td>
<td>NC</td>
<td>not classified</td>
</tr>
</tbody>
</table>
Multilingual Corpus of Annotated Spoken Texts

multicast.aspra.uni-bamberg.de/