

Multi-CAST

Northern Kurdish annotation notes

Geoffrey Haig

Maria Vollmer

Hanna Thiele

January 2021
v2.1



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

Haig, Geoffrey & Vollmer, Maria & Thiele, Hanna. 2021. Multi-CAST Northern Kurdish annotation notes. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#nkurd) (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The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.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 Northern Kurdish annotation notes v2.1 last updated 27 January 2021
This document was typeset by NNS with $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and the *multicast3* class (v3.2.3).

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.*

<i>qîzik-ê</i>	<i>got-îye:</i>	<i>per-ê</i>	<i>min</i>	<i>pir</i>	<i>=e</i>
girl-OBL	say.PST-PRF.3SG	money-EZ	1SG.OBL	much	=COP.3SG
# np.h:s_ds	ds_v:pred	#ds	np:s	rn_pro.1:poss	other:pred =cop
‘The girl said, “I have much money.”’					[mc_nkurd_muserz01_0115]

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.*

<i>em</i>	<i>=ê</i>	<i>ji</i>	<i>xwe</i>	<i>ra</i>	<i>xwedî</i>	<i>k-in</i>
1PL	=FUT	for	REFL	POP	0_her raising	SBJ.do.PRS-PL
#ds	pro.1:a	=aux	adp	refl.h:obl	pop_adp	0.h:p other:lvc v:pred
‘‘We will raise (her) for ourselves.’’						[mc_nkurd_muserz01_0069]

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 *Multi-CAST Persian* corpus (Adibifar 2016, see annotation notes therein). This glossing convention is exemplified in (3).

(3) *Bila bibe ewleda me, tu çima meraq dikî?*

<i>bila</i>	<i>bi-b-e</i>	<i>ewled-a</i>	<i>me</i>	<i>tu</i>	<i>çima</i>
may	0_she	SBJ-be.PRS-3SG	child-EZ	1PL.OBL	2SG
#ds	other	0.h:s	v:pred	np.h:other	rn_pro.1:poss
				#ds	pro.2:s_cp

<i>meraq</i>	<i>di-k-î</i>
worrying	IND-do.PRS-2SG
np:lvc	v:pred

‘‘May she become our own child — why are you worrying?’’

[mc_nkurd_muserz01_0069]

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 pîrekî tê.*

<i>rast-ê</i>	<i>pîrek-î</i>	<i>tê</i>
0_she	meeting-EZ	old_woman-OBL
##	0.h:s	np:lvc
	np.h:obl	v:pred

‘She met an old lady.’

[mc_nkurd_muserz01_0106]

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

<i>pîre</i>	<i>tîn-e</i>	<i>ji</i>	<i>wan</i>	<i>ra</i>	<i>nan-man</i>	<i>hazir</i>
old_woman	IND.bringPRS-3SG	for	3PL.OBL	POP	food-RED	prepared
##	np.h:a	lv_vother	adp	pro.h:g	pop_adp	np:p
						other:lvc

<i>di-k-e</i>
IND-do.PRS-3SG
v:pred

‘The old lady prepared a meal for them.’

[mc_nkurd_muserz01_0149]

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.*

<i>ex</i>	<i>=ê</i>	<i>te</i>	<i>bi</i>	<i>-wê</i>	<i>ra</i>	<i>bi-zewicîn-im</i>
1SG	=FUT	2SG.OBL	with	-3SG.OBL	with	SBJ-marry.PRS-1SG
##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

- Adibifar, Shirin. 2016. Multi-CAST Persian. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. (<https://multicast.aspra.uni-bamberg.de/#persian>) (Accessed 2019-03-08).
- Haig, Geoffrey. 2002. Noun-plus-verb complex predicates in Kurmanji Kurdish: Argument sharing, argument incorporation, or what? *Sprachtypologie und Universalienforschung (STUF)* 55(1). 15–48.
- Haig, Geoffrey. 2018. Northern Kurdish (Kurmanji). In Haig, Geoffrey & Khan, Geoffrey (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 106–158. Berlin: Mouton de Gruyter.
- Haig, Geoffrey & Öpengin, Ergin. 2018. Kurmanji in Turkey: Structure, varieties, and status. In Bulut, Christiane (ed.), *Linguistic minorities in Turkey and Turkic-speaking minorities of the peripheries*, vol. 111. Wiesbaden: Harrassowitz.
- 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).
- Haig, Geoffrey & Vollmer, Maria C. & Thiele, Hanna. 2019. Multi-CAST Northern Kurdish. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. (<https://multicast.aspra.uni-bamberg.de/#nkurd>) (Accessed 2019-07-05).
- Öpengin, Ergin & Haig, Geoffrey. 2014. Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish Studies* 2(2). 143–176.

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 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_>	<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	NEG	negation
2	second person	OBL	oblique case
3	third person	PL	plural
ADP	adposition	POP	postpositional particle
COMPL	complementizer	POSS	possessive
COP	copula	PRF	perfective
DEM	demonstrative	PROG	progressive
DRCT	directional particle	PRS	present
EMPH	emphasis	PST	past
EZ	ezafe	PTCP	participle
EXCL	exclamation	RED	partial reduplication
F	feminine	REFL	reflexive
FUT	future	SUBJ	subjunctive
IMP	imperative	SG	singular
IND	indicative	VOC	vocative
INDF	indefinite		
M	masculine	NC	not classified
MOD	modal		

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/