

Multi-CAST Uruangnirin

annotation notes

Eline Visser

July 2025 (v1.0)

1 Introduction

This document contains notes on the implementation of the GRAID (Haig and Schnell 2015) and RefIND (Schiborr et al. 2018) annotation conventions in the Multi-CAST Uruangnirin corpus. It corresponds to version 2507 of the annotations, published in July 2025. Unless a more recent version of this document exists, it also applies to any later versions of the annotations. You are welcome to contact the author if you have any specific questions.

2 Typological overview

In this section I will give a brief typological overview of Uruangnirin syntax. I illustrate the analyses with examples from the full Uruangnirin corpus (Visser et al. 2023).

2.1 Word order and verbal clauses

The structure of a basic intransitive, transitive and ditransitive clause in Uruangnirin is as follows. I use the abbreviation A for the subject of a transitive verb, while S is the subject of an intransitive verb. O stands for object, optionally preceded by D for direct.

- intransitive: **S V**
- transitive: **A V O**
- ditransitive: **A V DO other argument**

Uruangnirin also has peripheral arguments, such as locations, comitatives or instrumentals, which are typically but not necessarily introduced by prepositions. Verbs are words that can be indexed with a subject prefix and that typically function as predicates. The subjects of active intransitive and transitive verbs are typically (but not obligatorily, see Section 3.1) indexed on the verb by means of a prefix. Objects are not indexed. In Uruangnirin, the predicate of a clause can also be an item that does not conform to the definition of ‘verb’ just given, for example an adjective, quantifier, pronoun, see Section 2.2 on non-verbal predicates. I refer to these predicates collectively as ‘statives’. In these clauses, there is never any subject indexing.

When arguments are expressed in the clause, subjects precede the verb, while objects follow the verb. Typical intransitive and transitive clauses are given in (1–3).

(1) stative intransitive

<i>[ai</i>	<i>sa</i>	<i>ka]</i> _s	<i>ririn</i>	<i>som</i>
tree	one	dist	tall	already
'That one tree is tall already.'				[s-pohon-kky_175]

(2) active intransitive

<i>[Senja</i>	<i>laman]</i> _s	<i>a-ma</i>	<i>som</i>	<i>to</i>
Senja	father	3-come	already	right
'Senja's father has already come, right?'				[c-tokipala-kb_65]

(3) transitive

<i>[Amin</i>	<i>tatang]</i> _A	<i>a-sei</i>	<i>[sanong]</i> _o
Amin	grandparent	3-sew	already
'Amin's grandmother sewed the thatched roof'			[n-rumah-kky_15–16]

The only syntactically ditransitive verb is *wanggi* 'give'. The theme follows the recipient, and neither are marked with extra morphology or adpositions. An example is given in (4). Other verbs that are trivalent (i.e. requiring three participants, but not all of them expressed as core arguments) are for example *fasa* 'buy', *ftobir* 'send', *(f)penik* 'sell' and *sobang* 'send; order'. They are monotransitive and introduce the oblique participant with a preposition.

(4) ditransitive

<i>[ia]</i> _A	<i>a-wanggi</i>	<i>[sina]</i> _{Recipient}	<i>[wawa</i>	<i>puan</i>	<i>teni]</i> _{Theme}
3sg	3-give	3PL	mango	fruit	tree
'He gave them three mangoes'				[s-pear-kky_67]	

2.2 Non-verbal predicates

Non-verbal clauses are clauses with a predicate that is not a verb. This clause type is common in Uruangnirin, since nouns, demonstratives and quantifiers can act as a predicate. See also Section 3.4.

Example (5) shows two consecutive utterances from the same text. If we take the position in the clause as the parameter for determining which word acts as the predicate, (5a) has a nominal predicate (*ai tultulis* 'tultulis wood'), and (5b) has a demonstrative predicate (*dike* 'PROX').

- (5) a. *ai ke in nggaran ai tultulis*
 wood PROX 3SG.POSS name wood tultulis
 ‘This wood, its name is tultulis wood.’ [n-ujungkampung-in1_28]
- b. *tultulis da dike*
 tultulis REL PROX
 ‘It’s tultulis that’s here.’ [n-ujungkampung-in1_29]

Examples (6) and (7) illustrate clauses with quantifiers as the predicate. In (6), speaker A asks about the number of boats that arrived at an event, using a structure akin to English ‘how many boats were they?’. The answer by speaker B contains two numeral quantifiers as predicates: first *sa* ‘one’ to indicate the number of boats, and then *teni* ‘three’ to indicate the number of people. In (7), both speakers use the non-numeral quantifier *dum* as the predicate of their clause.

- (6) a. *sina rau fira*
 3PL canoe how_many
 ‘In how many canoes (did they come)?’
- b. *rau sa laf sina teni*
 canoe one only 3PL three
 ‘Just one canoe, they were three [people].’ [c-kampung-sl-lk_12-13]
- (7) a. *puan dum bi tai*
 fruit many or NEG
 ‘Are there many fruits?’
- b. *ino puan dum wana*
 wow fruit many friend
 ‘Wow, there are many fruits, friend.’ [a-panjatpala_130-132]

Finally, (8) has a question word predicate.

- (8) *bunga dike ni penin fira*
 flower PROX POSS price how_much
 ‘He gave them three mangoes’ [c-minang-wn-nk_64]

2.3 Negation

Clauses are negated with clause-final negator *tai*.

- (9) *fisik a-ma tai*
 some 3-come NEG
 ‘Some didn’t come.’ [n-taruhharta-kky_36]

Non-verbal clauses are negated in the same way. Non-existence is expressed with negation of the existential verb *mian* ‘be there; live’. Impossibility is expressed with *eran tai* ‘cannot’. *Eran* cannot be used in positive contexts. (10) illustrates *eran tai* ‘cannot’.

- (10) *kalo timur ke ia a-sa eran_tai*
 if east PROX 3SG 3-ascend cannot
 ‘If [the] east [wind blows] it [the boat] cannot come up.’ [a-masak_167]

2.4 Prepositions

There are three indigenous prepositions, all derived from verbs, and one loan from Papuan Malay. The indigenous prepositions are locative *amina* ‘in; at’ (short *am*), ablative *afara* ‘from’ and comitative/instrumental *nora* ‘with’. The borrowed preposition is *sampe* ‘until’ (*sampai* in Standard Indonesian). Prepositional phrases follow the verb phrase. They are illustrated in (11–14).

- (11) *a-fun kem perperi sa am tawanang ka som*
 3-make tent new one LOC side DIST IAM
 ‘Let’s make a new tent on that side.’ [c-sp-lk-bk_7]

- (12) *adat rwang dike matafasi afar di_luar a-ma*
 custom village PROX women from outside 3-come
 ‘The custom of this village is, [when] a woman from outside comes...’
 [n-ppbaru-kky_3]

- (13) *i-uan seir nora kokok patin*
 1sg-eat fish with chicken egg
 ‘I eat fish with an egg.’ [n-kenari_590]

(14)	<i>ta-desil</i>	sampe	<i>nanam</i>	<i>barsi</i>	
	1PL.EXCL-plane	until	inside	clean	
	‘We plane until the inside is clean.’				[c-minang-wn-nk_99]

3 Annotation notes

3.1 Arguments

Subject arguments can be expressed with a pronoun or noun in the clause and/or with a subject index on the verb. This means there can be:

1. only a subject expression in the clause,
2. a subject referent in the clause and an index on the verb,
3. only an index on the verb or
4. no subject expression in the clause and no index on the verb.

The most common construction is the third. For first and second person, there are subject index variants. Table 1 gives an overview of the subject indexes and pronouns.

	pronoun	subject index(es)
1SG	<i>lau</i>	<i>lai-</i> , <i>li-</i> , <i>i-</i>
2SG	<i>o</i>	<i>mi-</i> , <i>(m-)</i>
3SG	<i>ia</i>	<i>a-</i>
1PL.EX	<i>ami</i>	<i>ama-</i> , <i>am-</i> , <i>ma-</i> , <i>mi-</i> , <i>(m-)</i>
1PL.IN	<i>ita</i>	<i>ta-</i> , <i>(t-)</i>
2PL	<i>imi</i>	<i>im-</i> , <i>mi-</i> , <i>(m-)</i>
3PL	<i>sina</i>	<i>a-</i>

Table 1. Pronouns and subject indexes. Morphophonologically restricted variants are bracketed.

The generalized semantic role for subject and object is indicated at the end of the GRAID annotation following an underscore. Only two possibilities are noted: Agent <_a> or Patient <_p>, following for example Riesberg et al. (2022).¹ I have done this for both subject expressions as argument in the clause, and as prefixes on the verb. Objects are not indexed on the verb, and only appear as arguments in the clause, if at all. In the annotation files, some information about the phonological form of the subject index is prefixed to its annotation, such that for example *a-* and *i-* are marked with <v_>, while *ma-*, *mi-* and *ta-*,

¹ The roles approximately correspond the the macro-roles of Role and Reference Grammar ‘Actor’ and ‘Undergoer’ respectively.

are marked with <cv_>. In the following examples, these latter annotations have been removed to save space.

In the EAF annotations, we have opted to treat the verbal prefixes as referential items, thus they align with a referent index on the RefIND tier. For ease of processing, these prefixes are therefore glossed in their own annotation cell in the EAF files, but are glossed with a hyphen to indicate their prefixal nature; see e.g. (16) and (17).²

Examples (15)–(18) show the four ways to express subject arguments, while at the same time illustrating object arguments and the extra annotations (see section 3.2 for additional explanations).

(15) type 1: subject argument only

<i>ami</i>	<i>teni</i>	<i>bana</i>	<i>futar</i>	<i>masikit</i>
1PL.EXCL	three	go	turn	mosque
pro.1:s_a	rn	v:pred_u	rv_u	np:other
‘We three went around the mosque.’				[n-iblisemesjid-bk_57]

(16) type 2: subject argument and index

<i>baba</i>	<i>Haji</i>	<i>a-</i>	<i>sobang</i>	<i>ZERO</i>	<i>#</i>
father	Haji	3-	send	ZERO	#
np.h:a_a	rn	-pro_h_a_a	v:pred	0.h:p_p	##
<i>ZERO</i>	<i>a-</i>	<i>ura</i>	<i>lau</i>		
ZERO	3-	call	1sg		
np.h:a_a	-pro_h_a_a	rv	pro.1:p_p		
‘Father Haji sent for me.’				[n-iblisemesjid-bk_2]	

(17) type 3: index only

<i>ZERO</i>	<i>a-</i>	<i>ftonang</i>	<i>wai</i>	<i>am</i>	<i>nan</i>	<i>ke</i>
ZERO	3-	make_sit	thing	LOC	mainland	PROX
0.h:a_a	-pro_h_a_a	v:pred	np:p_p	adp	np:l	rn
‘Put down the stuff at the landslide.’				[n-iblisemesjid-bk_8]		

(18) type 4: nothing

<i>ZERO</i>	<i>por</i>	<i>pangga</i>	<i>tipa</i>	<i>bo</i>
ZERO	PROH	hit	drum	but
0.h:a_a	other	v:pred_u	np:p_p	other
‘Don’t hit the drum, but...’				[n-iblisemesjid-bk_66]

² Note that affixes in GRAID are always marked by *preceding* hyphen, even if they are prefixes; the directionality of affixation can be gleaned off the morphological gloss.

3.2 ‘Unmarked’ verbs

I have used an extra annotation for verbs that are not indexed: <_u> for unmarked. However, when verbs are immediately preceded by a third person pronoun (*ia* or *sina*), it is actually impossible to know whether they are actually unmarked, since the third person index is *a-*, which is the same as the last sound of both third person pronouns. (19) illustrates an ‘unmarked’ verb *ma* ‘come’ where we can actually not know whether or not it is indexed. (20) illustrates an actually ‘unmarked’ verb *pangga* ‘hit’, because there is no pronoun and there is a long pause before the start of this clause.

(19)	<i>jadi</i>	<i>sina</i>	<i>ma</i>	
	so	3PL	come	
	other	pro.h:s_a	v:pred_u	
	‘So they came.’			[n-taruhharta-kky_46]

(20)	<i>ZERO</i>	<i>pangga</i>	<i>kapuni</i>	<i>peta</i>
	ZERO	hit	bamboo	ANA
	0.h:a_a	v:pred_u	np:p_p	rn
	‘Hit that bamboo.’			[n-iblisjesjid-bk_81]

This should be kept in mind during analysis. For example, in a study about differential subject marking (Visser and Faghiri, to appear), we have excluded all clauses with a third person pronoun.

3.3 Inherently marked verbs

A subclass of Uruangnirin verbs has irregular subject indexing. Their subject indexes are glossed as being indivisible from the verb stem, while their GRAID annotation is as for the other verbs (for example <-pro_h_a_a v:pred>, in two cells, cf. above).

3.4 The Uruangnirin predicate

Almost anything can be the predicate in Uruangnirin, so there are examples in the corpus of adjectives, numerals, question words and demonstratives, among others (see Section 2.2). They are all annotated as <other:pred>.

3.5 Adjacent verbs

There are two types of adjacent verb sequences.³ One is treated as monoclausal, while the other is not. When two or more adjacent verbs share the subject argument, they are treated as monoclausal, regardless of whether the verbs carry subject indexes or not. Any verb that is not the first verb in such a construction is annotated as *rv*. When two or more adjacent verbs do not share the same subject argument, they are treated as heading separate clauses. In Uruangnirin, it is common to express actions of movement and caused motion with two adjacent verbs, where the zero-expressed object of the first verb is the subject of the second verb. (21) has three verbs spread over two clauses. The first two verbs are analysed as part of the same clause.

(21)	ZERO	<i>i-</i>	<i>tna</i>	<i>i-</i>	<i>ois</i>	<i>ai</i>	#
	ZERO	1-	go	1-	break; cut	wood	#
	0.1:a_a	-pro_1_s_a	v:pred	-pro_1_s_a	rv	np:p_p	##
	ZERO	<i>a-</i>	<i>mno</i>				
	ZERO	3-	come_seawards				
	0:s_p	-pro_s_p	v:pred				
	'I went landwards and cut wood, brought it seawards.'						[n-rumah-kky_20]

In (21), the first two verbs are immediately adjacent and carry the same index (first person), hence the sequence is treated as monoclausal. The third verb is separated from the other two by the object (*ai* 'wood'), and the index switches from first to third person, so it is easy to see this is a separate clause. In (22) it is a bit trickier because there is no overt object and the verbs *fonak* 'put' and *to* 'descend' carry the same index. The preceding clause gives the context to understand the construction. The object of *fonak* is the subject of *to*: *wai noning* 'leaves'.

(22)	<i>ia</i>	<i>sobang</i>	<i>nedin</i>	<i>wai</i>	<i>noning</i>	<i>punit</i>	#
	3sg	send	3.pick	thing	leaf	finished	#
	pro.h:a_a	v:pred_u	rv	np:p_p	rn	other	##
	ZERO	<i>a-</i>	<i>fonak</i>	ZERO	#	ZERO	<i>a-</i>
	ZERO	3-	put	ZERO	#	ZERO	3-
	0.h:a_a	-pro.h_a_a	v:pred	0:p_p	##	0:s_p	-pro_s_p
	'He told [him] to pick some leaves, after that put them down [there].'						[n-iblisemesjid-bk_32]

³Adjacent verbs are verbs, inflected for subject or not, that occur right next to each other on the surface, with no intervening words.

I refrain from referring to any of these constructions as serial verb constructions, partially because it is an ongoing debate how these should be defined, and partly because I haven't explicitly tested which adjacent verb constructions share for example negation and TAM values.

3.6 Non-specific referents

Most texts have one index in the RefIND scheme for a non-specific referent. In the text *urn-n-jualan-an-f-g*, for example, this non-specific referent could be translated as 'people', 'one' or 'we'. Non-specific referents may be used several times throughout the text without referring to the exact same people or things. For example, in (23), the elided subject refers to anyone that is in a similar situation to the speech act participants, namely when they have to talk about things that don't have a name in Uruangnirin. In (24) the subject refers to anyone buying mixed onions and garlic, although it could technically also refer to just the speaker herself. Since the verb does not carry a subject index, it is hard to determine.

(23)	<i>daka</i>	<i>harus</i>	<i>ZERO</i>	<i>a-</i>	<i>paki</i>	<i>bahasa indonesia</i>	<i>pisa to</i>
	DIST	must	ZERO	3-	use	Indonesian	little right
	other	other	0.h:a_a	-pro.h:a_a	v:pred	np:p_p rn	other other
	'There one has to use a little Indonesian, right.'						[n-jualan-an_27]

(24)	ZERO	fasa	bawang	campur	ripuca	
	ZERO	buy	onion	mix	10 000	
	0.h:a_a	v:pred_u	np:p_p	rn	other	
	'You buy mixed onions/garlic, 10.000 [rupiah].'					[n-jualan-an_59]

Abbreviations

Following the Leipzig Glossing Rules, non-standard abbreviations only given here.

ANA	anaphoric demonstrative-like word
IAM	imitative

References

- Haig, Geoffrey, and Stefan Schnell. 2015. *Annotations using GRAID: Grammatical Relations and Animacy in Discourse*. Manual version 7.0.
- Riesberg, Sonja, Maria Bardají i Farré, Kurt Malcher, and Nikolaus P. Himmelmann. 2022. Predicting voice choice in symmetrical voice languages: All the things that do not work in Totoli. *Studies in Language* 46:453–516.
- Schiborr, N., Stefan Schnell, and Hanna Thiele. 2018. *RefIND—Referent Indexing in Natural-language Discourse*. Annotation guidelines v1.1.
- Visser, Eline, and Pegah Faghiri. Subject realization in Uruangnirin. To appear in *STUF*.
- Visser, Eline, Kamal Kanabara, Rustam Kanabara, Salis Maswatu Kanabara, and Safrudin Tianotak. 2023. *Uruangnirin language documentation*. Berlin: Endangered Languages Archive. (<http://hdl.handle.net/2196/dd459a7c-2493-4d26-9801-cc071a72bd34>)