Operationalizing Philippine-type Syntax for GRAID System: Clause Structure, Case Marking, and Verb Class in Arta

Kimoto, Yukinori
Nagoya University

This paper concerns the grammatical and annotation notes on the GRAID annotation system, which was extended for the discourse data of Arta, a language spoken in Northern part of Luzon, the Philippines.[1] GRAID (Grammatical Relations and Animacy in Discourse) is an annotation system developed by Geoffrey Haig and Stefan Schnell to explore the relation between argument realization patterns and possible discourse-functional motivations across languages through a qualitative approach. When applying it to Arta and other Philippine languages, it is necessary to calibrate the cross-linguistic concepts by providing some empirical data and evidence. This article first provides evidence for determining the grammatical functions of each argument (S, A, P, and other roles) within each clause. Possible realization patterns of referential expressions are discussed, and some additional information is tagged for capturing these patterns in Arta. The structures of complex sentences, especially those involving a relative clause, are also noted.

Keywords: Arta, GRAID annotation, argument realization patterns, ergativity, gap construction

1. Introduction

This article aims to document the implementation of the Grammatical Relations and Animacy in Discourse annotation system (GRAID) to Arta discourse data. The GRAID glossing conventions are "a system of symbols and conventions for glossing

Kimoto, Yukinori. 2018. "Operationalizing Philippine-type Syntax for GRAID System: Clause Structure, Case Marking, and Verb Class in Arta". *Asian and African Languages and Linguistics* 12. pp.17–35. [Permanent URL: http://hdl.handle.net/10108/91147]

[1] This article is based on a collaborative project on annotating Arta discourse data with the GRAID system. I am grateful to Stefan Schnell, who has been helpful in implementing the system into Arta during and after my stay in University of Melbourne in 2017. I am also thankful to Asako Shiohara, Keita Kurabe, Geoffrey Haig and two anonymous reviewers for their valuable comments and suggestions. Special thanks also go to three Arta speakers: Delia Bueno, Caras Ramos, and Arsenyo Pantalion. Of course any errors that remain are my sole responsibility. I gratefully acknowledge the Japan Society for the Promotion of Science (#16J04401 and #J2801), and Endangered Languages Documentation Programme in SOAS, University of London (#SG0380) for supporting this research financially.

the grammatical relations and overt forms (noun phrases, pronouns etc.) of major clause constituents in texts" developed by Geoffrey Haig and Stefan Schnell (Haig and Schnell 2014).[2] Arta is a Philippine language spoken by the Philippine Negrito people who inhabit Quirino and Aurora provinces in Northeastern Luzon in the Philippines. The language was first reported by Reid (1989) and is being documented by the present author (Kimoto 2014, 2017a, b). This paper begins with an overview of GRAID annotations as applied to Arta, and annotations on clause structure (Section 2), followed by its implementation to nominal structure (Section 3), and sentence (Section 4).

2. Clause structure

2.1. Grammatical relations of core arguments

In GRAID, each annotator is expected to annotate such tags as ⟨:s⟩, ⟨:a⟩, ⟨:p⟩, and ⟨:obl⟩ onto each argument. However, as is widely recognized, there has been a long controversy over the nature of case marking system in Philippine languages, or more broadly, Philippine-type languages, including most Formosan languages in Taiwan and some languages in Sulawesi and Borneo. Although there are admittedly many Philippine-type languages that do not show a clear ergative-absolutive alignment, it is relatively clear that Arta does have the ergative-absolutive alignment in the case marking of full nominals, personal forms, and demonstratives.[3] The following elicited sentences exemplify the typical encoding patterns of single-participant and two-participant events (note that the label **genitive** GEN is used, and has been used in Austronesian linguistics in general, both for possessor role and transitive subject because of their homonymy):

(1) *T<in><um>addyor=di [i babakat=i].*
<PST><INTR>stand=COMP DEF old.woman=SPC
'The old woman stood up.' (elicited)

(2) *B<in>isag=di [ni babakat=i] [i bo:te].*
<PST>break=COMP GEN.DEF old.woman=SPC DEF bottle
'The old woman/women broke the bottle(s).' (elicited)

Regardless of whether the single core argument of the intransitive clause designates an agent or undergoer (cf. split-S system), the nominal is introduced by the determiner *i* in (1), which is identical to the determiner of the undergoer role in the two-participant event in (2), whereas the agent-like argument of (2) receives a distinct marking *ni*.

Note, however, that there is another encoding strategy in which the same prototypical transitive situation is expressed. This kind of construction, which is called "actor-focus

[2] The list of symbols employed in the GRAID convention is given in the appendix of Schnell and Schiborr (2018).
[3] For arguments concerning the ergativity of Philippine languages, see Payne (1982), Shibatani (1988), Mithun (1994), Foley (1998), Liao (2004), and Himmelmann (2005), among others.

construction," has been problematic in Philippine languages in that the actor argument receives the marking identical to the single core argument of the intransitive clause. In Arta, however, this construction is evidently not considered to be a transitive clause on the grounds that the undergoer role is marked by the oblique determiner *ti*. Compare (3) with other examples (4)–(5) shown below, in which the determiner *ti* is used to introduce the temporal phrase 'in March' and the locative phrase 'in the mountains'. Thus, the apparent transitive clause with so-called actor-focus construction is not a true transitive clause in Arta, just like the "conative" construction in English *I kicked at him* where the patient role is encoded by the prepositional phrase.[4]

(3) *Nam-bisag=di [i babakat=i] [ti bo:te].*
PST.INTR-break=COMP DEF old.woman=SPC OBL.DEF bottle
'The old woman/women broke the bottle(s).' (elicited)

(4) *Pam-mula =ami ta pagay [ti Marso].*
PRG.INTR-plant =1PL OBL rice OBL.DEF March
'We will be planting rice in March.' (arta0505)

(5) *Man-di:madima =te: =tid [ti talutalun =i] i be:kut =na.*
INTR-walk =only =3PL OBL.DEF mountain =SPC DEF ghost =3SG.GEN
'Their ghosts will be walking in the mountains.' (arta0111-06)

Pronominal indexes on predicates show the same pattern. The pronominal counterparts of the examples (1–3) are shown in (6–8). Note that prototypical transitive events encoded by actor-voice constructions, as in (8), are not permitted when the arguments are realized pronominally:

(6) *T<in><um>addyor =de: =**tid**.*
<PST><INTR>stand =COMP =3PL
'They stood up.' (elicited)

(7) *B<in>isag =**na** =d =**tid**.*
<PST>break =3SG.GEN =COMP =3PL
'He/she broke them.' (elicited)

(8) *??Nam-bisag =de: =**tid** **did**.*
PST.INTR-break =COMP =3PL 3PL.OBL
'The old woman/women broke the bottle(s).' (elicited)

These data provide a basis for determining S, A, P, and oblique in this language (see Kimoto 2017b for a more extensive discussion). Construction patterns (2, 7) are

[4] This encoding pattern is not preferred in the language in such typical transitive situations such as that of the change-of-state event towards a single undergoer individual, with punctual and completive aspect. This actor-voice construction with a patientive argument appears in such cases as 'to eat food' and 'to hunt for wild pigs', with the object as a mass entity or non-specific/indefinite entity.

regarded as transitive; thus, the actor and undergoer arguments are tagged as A and P, respectively. Construction patterns (1, 6) are regarded as intransitive; thus, the single core arguments are tagged as S. If the clause has two arguments, as in (3), the agentive argument is identified as S, and the undergoer-like argument is identified as oblique. The GRAID annotations are thus:

(9)

	T<in><um>addyor=di	*[i*	*babakat*	*=i].*
	<PST><INTR>stand=COMP	DEF	old.woman	=spc
##	v:pred	ln	np.h:**s**	rn

'The old woman stood up.' (elicited)

(10)

	B<in>isag=di	*[ni*	*babakat*	*=i]*	*[i*	*bo:te].*
	<PST>break=COMP	GEN.DEF	old.woman	=SPC	DEF	bottle
##	v:pred	ln	np.h:**a**	rn	ln	np.h:**p**

'The old woman broke the bottle.' (elicited)

(11)

	?Nam-bisag	*=di*	*[i*	*babakat*	*=i]*	*[ti*	*bo:te].*
	PST.INTR-break	=COMP	DEF	old.woman	=SPC	OBL.DEF	bottle
##	v:pred	other	ln	np.h:**s**	rn	ln	np.h:**obl**

'The old woman broke the bottle.' (elicited)

Many Philippine-type languages, most notably those spoken in the Southern Philippine and in Sulawesi, appear to lack the ergative alignment (see, for example, Shibatani 1988 for Cebuano). In fact, Brickell (2016), on annotating the Tondano discourse in the GRAID system, rejects the ergative analysis, analyzing both actor-voice constructions and undergoer-voice constructions as transitive with A and P arguments within them. To differentiate both constructions, the tag of each argument includes the information of the predicate as in ⟨a_a⟩ (A argument of actor voice), ⟨:a_u⟩ (A argument of undergoer voice), ⟨:p_a⟩ (P argument of actor voice), ⟨p_u⟩ (P argument of undergoer voice). The annotation system of Tondano is incorporated in the present annotation on Arta to enable the comparison of Philippine-type languages that may or may not have clear ergative characteristics. Thus, the modified version of the annotation is shown below:

(12)

	T<in><um>addyor	*=di*	*[i*	*babakat*	*=i].*
	<PST><INTR>stand	=COMP	DEF	old.woman	=spc
##	v:pred	rv	ln	np.h:**s_a**	rn

'The old woman stood up.' (elicited)

(13)

	B<in>isag	*=di*	*[ni*	*babakat*	*=i]*	*[i*	*bo:te].*
	<PST>break	=COMP	GEN.DEF	old.woman	=SPC	DEF	bottle
##	v:pred	rv	ln	np.h:**a_u**	rn	ln	np.h:**p_u**

'The old woman broke the bottle.' (elicited)

(14)

	Nam-bisag	*=di*	*[i*	*babakat*	*=i]*	*[ti*	*bo:te].*
	PST.INTR-break	=COMP	DEF	old.woman	=SPC	OBL.DEF	bottle
##	v:pred	rv	ln	np.h:**s_a**	rn	ln	np.h:**obl_a**

'The old woman broke the bottle.' (elicited)

We have so far used elicited examples for the sake of clarity of presentation. Below are actual discourse data annotated in which S, A, P, and oblique, location,[5] and goal phrases are found:

(15)

	Pang-u:sar-èn	*didi*	*ama*	*=mi*	*=ti*	*ta*	*dutul.*
	PRG-use-TR	PL.GEN.DEF	father	=1PL.GEN	=SPC	OBL	first
##	v:pred	ln	np.h:**a_u**	=pro.1:poss	rn	ln	other

'Our fathers were using them in those days.' (arta0002-43)

(16)

	Saya	*n-inta*	*=ku*	*ti*	*bebbe:*	*=m.*
	DEM.DIST	<PST.TR>-see	=1SG.GEN	PSN	aunt	=2SG.GEN
##	other	v:pred	=pro.1:**a_u**	ln	np.h:**p_u**	=pro.2:poss

'Then I met your aunt.' (arta0601-45)

(17)

	Amma	*atti:*			*um-angay*	*=ti,*
	if	exist			INTR-go	=DEM.PROX.OBL
## #ac	other	other:predex	#rc gap.h:s_a		v:pred	=dem:**g_a**
	d<um>arettyo	*=tid*	*dèn*			
	<INTR>straight	=3PL	1SG.OBL			
%	v:pred	=pro.h:s_a	pro.1:**g_a**			

'If there are people who are coming here, (that is, if) they come straightly to me,' (arta0601-57)

(18)

	Atti:	*=ami*	*ti*	*Aglipay,*	*wa,*
	exist	=1PL	OBL.DEF	Aglipay	PLH
##	other:predex	=pro.1:**s_predex**	ln	np:**l_predex**	other

Disubu.
Disubu
other

'We were in Aglipay, whatchamacallit, Disubu.' (arta0601-06)

The list and exposition of predicate tags are provided in Section 2.4.

Referential expressions may be located in the sentence-initial position, serving as a topic NP. Following the annotation rule, these NPs are glossed as ⟨dt⟩ 'dislocated topic'. The nominals may be co-referential with a clause-internal argument, which is specified, for example, as ⟨dt_s⟩ (dislocated topic corresponding to the S function clause-internally), as shown in:

[5] The location phrase found in (18) constitutes part of an existential construction, whose predicate *atti:* ('exist' ⟨other:predex⟩) is the morphologically irregular verb in that it does not inflect for tense or aspect. The predicate is indexed as ⟨l_**predex**⟩ in this case. See Section 2.3 for further exposition.

(19)

	Asawa	***=ku***	***=ti,***		*awan*	*=kurug*
	spouse	=1SG.GEN	=SPC		NEG	=really
##neg	np.h:**dt_s**	=pro.1:poss	rn	0.h:s_adj	other	other
a	*med-dingatu,*					
LIG	ADJ-tall					
other	adj:pred					

'As for my husband, (he) was not really tall.' (arta0502-02)

(20)

	Tèn,	*awan*	*=muyu*	*=tèn*	*a*	*arig-èn,*
	1SG	NEG	=2PL.GEN	=1SG	LIG	imitate-TR
##neg	pro.1:**dt_u**	other	=pro.2:a_u	=pro.1:p_u	other	v:pred

'(lit.) As for me, you should not imitate me.' (arta0601-74)

2.2. Argument realization patterns

In Arta, an argument is normally realized by either a person index or a full NP; that is, the occurrences of person indexes and full NPs are distributed almost complimentarily. For example, in (21), plural third-person referents are encoded by the person index on the predicate without an independent nominal phrase, and, in (22), by an independent nominal without an index on the predicate:

(21)

	Pab-bi:rè-n	***=di***	*=d*	*tidi*	*a:na:*	*=di,*
	PRG-search-TR	=3PL.GEN	=COMP	PL.DEF	children	=3PL.GEN
##	v:pred	**=pro**.h:a_u	rv	ln	np.h:p_u	=pro.h:poss
	amma	*nap-pati*	*=d*	***=tid.***		
	if	PST.INTR-die	=COMP	=3PL		
#ac	other	v:pred	rv	**=pro**.h:s_a		

'They are looking for their children if they died.' (arta0111-25)

(22)

	Nappati	*=d*	*[tidi*	*ama*	*=na*	*=y*	*aydi:*
	PST.INTR-die	=COMP	PL.DEF	father	=3PL.GEN	=SPC	and
##	v:pred	rv	ln	**np**.h:s_a	=pro.h:poss	rn	rn
ina	*=na].*						
mother	=3SG.GEN						
rn	rn						

'His father and mother died.' (arta0502-22)

These realizations are, according to Haig and Schnell (2014), tagged as ⟨**=pro**⟩ and ⟨**np**⟩ respectively. The equal sign ⟨=⟩ shown in the first case indicates that it is an enclitic attaching to the predicate.

A careful examination reveals that another pattern is occasionally observed; this is the case in which the same role is encoded both by a person index and a full NP. In the example below, the actor role is doubly instantiated by the person index and the independent nominal phrase:

(23)

	Saya	*iggam-an*	***=na***	*a:yi:*	*[ni*	*kanakannak*	*=i].*
	DEM.DIST	hold-TR	=3SG.GEN	DEM.PROX	GEN.DEF	child	=SPC
##	dem:other	v:pred		dem:p_u	ln	bpi_np.h:a_u	rn

'Then the child held this.' (arta0110-114)

This type of argument encoding appears at a relatively low frequency. If this pattern is observed, it is annotated as ⟨bpi⟩ (bound person index) on the gloss of the NP, with the pronominal index left unglossed, as shown in (23).

2.3. Predicate types

This subsection examines the syntactic categories that may function as predicates. First, nouns, adjectives, and verbs can occupy the predicate slot in the language without any formal device such as a copulative formative. In this annotation, these predicates are tagged as ⟨np:pred⟩, ⟨adj:pred⟩, ⟨v:pred⟩ respectively. See the following examples:

(24) ⟨np:pred⟩ (nominal predicates)

a.

	Saya	*ina*	*=de:*	*Mulo,*		*ay*	*ti*
	DEM.DIST	mother	=3PL.GEN	Mulo		PLH	PSN
##	other	np.h:dt_s	other	np.h:poss	0.h:s_np	other	ln

Brida.
Brida
np:pred

'So, as for Mulo's mother, (she is) **Brida**.' (arta0601-41)

b.

	siye:,	*wa*	*=m,*	*kuwarto*	*=m.*
	DEM.PROX	PLH	=2SG.GEN	money	=2SG.GEN
##	dem:s_np	other	other	**np:pred**	=pro.2:poss

'This is your what-cha-ma-call-it (placeholder), **your money**.' (arta0601-94)

(25) ⟨adj:pred⟩ (adjective predicates)

a.

	Asawa	*=ku*	*=ti,*	*(...)*		*apitti*
	spouse	=1SG.GEN	=SPC			short
##	np.h:dt_s	=pro.1:poss	rn		0.h:s_adj	**adj:pred**

=te.
=only
other

'As for my husband, (he) **was short**.' (arta0502-02)

b.

	Med-dès	*i*	*uga:li*	*=ku*	*=y.*
	ADJ-bad	DEF	habit	=1SG.GEN	=SPC
##	**adj:pred**	ln	np:s_adj	=pro.1:poss	rn

'My habit **was bad**.' (arta0601-74)

Existential predicates are one of the items that appear most frequently in the discourse. The positive existential is *atti:* 'exist, there is', and the negative counterpart

is *awan* 'not exist, there is no'. They are tagged as ⟨other:predex⟩ in this annotation system, signaling a non-verbal class predicate functioning as existential (see Haig and Schnell 2014: 21). The negative existential *awan* should be followed by the obliquely-marked subject, annotated as ⟨:ncs_predex⟩:

(26) Positive existential predicate

a.

	Atti	*konta*	*atti:*	*ti*	*ayti*
	exist	but	exist	OBL.DEF	DEM.PROX.OBL
##	**other:predex**	other	**other:predex**	ln	ln

Sinabagan.
Sinabagan
np:l_predex

'They were (here), but they were here in Sinabagan.' (arta0002-32)

b.

	Atti:	*=tep*	*i*	*gilangan*	*=i*	*ta*
	exist	=still	DEF	man	=SPC	OBL
##	**other:predex**	other	ln	np.h:s_predex	rn	ln

ayta	*Danak.*
DEM.DIST.OBL	Danak
ln	np:l_predex

'The man was still there in Danak.' (arta0106-15)

(27) Negative existential predicate

a.

	Awan	*ta*	*Ilongot,*	*awan*	*ta*
	not.exist	OBL	Ilongot	not.exist	OBL
##neg	**other:predex**	ln	np:**ncs**_predex	**other:predex**	ln

agani:,
non-Negrito
np:**ncs**_predex

'There were no Ilongot people (in that place); there were no non-Negrito people (there).' (arta0601-24)

b.

	Ayta	*dutul,*	*awan*	*ta*	*a:nus.*
	OBL	first	not.exist	OBL	kindness
##neg	ln	np:other	**other:predex**	ln	np:**ncs**_predex

'Before, there was no kindness (i.e., no one was kind to others).' (arta0601-53)

Finally, nonfinite verb forms in Arta, defined as non-inflected verb forms for tense or aspect, may appear in some of the subordinate clauses. These are annotated as ⟨vother:pred⟩:

(28) ⟨vother:pred⟩ (nonfinite verb as a predicate)

a.

	Saya	*n-inta*	*=ku*		*a*
	DEM.DIST	<PST.TR>-see	=1SG.GEN		LIG
##	dem:p_u	v:pred	=pro.1:a_u	#rc gap:p_u	other

pang-a:n-èn *=mi.*
NMZ-eat-TR =1PL.GEN
vother:pred =pro.1:a_u
'That is what I saw that we ate,' (arta0002-01)

b.

	Sa:bit-èn	*=di*	*=d,*	#	*ngay*	*ti*
	carry-TR	=3PL.GEN	=COMP	#	go	OBL.DEF
## 0:p_u	v:pred	=pro.h:a_u	other	#	**vother:pred**	ln

bunbun *=mi.*
house =1PL.GEN
np:g_other =pro.1:poss
'They carry it, going (back) to our house.' (arta0002-62, 63)

2.4. Predicate tag on nominals

Table 1 Predicate tags on nominals

tag	predicate class	voice	example	translation
⟨_a⟩	dynamic verb	actor voice	*man-lutu*	'cook'
⟨_u⟩	dynamic verb	undergoer voice	*i-lutu*	'be cooked'
⟨_ap⟩	potentive verb	actor voice	*maka-tim*	'can drink'
⟨_up⟩	potentive verb	undergoer voice	*ma-tim*	'can be drunk'
⟨_stv⟩	stative verb	n.a.	*tit-tuttud*	'be sitting'
⟨_adj⟩	adjective	n.a.	*mep-pullaw*	'be white'
⟨_predex⟩	existential	n.a.	*atti:*	'there is, exist'
⟨_np⟩	nominal predicate	n.a.	*buka:gan*	'be a woman'

Table 2 Dynamic, potentive, stative verbs, and adjectives

	dynamic verb class	potentive verb class	stative verb class	adjective class
progressive	yes	no	no	no
tense (non-past vs. past)	yes	yes	no	no
voice (actor vs. undergoer)	yes	yes	no	no
temporality implication	yes	yes	yes	no
comparative construction	no	no	no	yes
intensifying reduplication 'very X'	no	no	no	yes

In Arta, almost every verb are marked in nature; i.e. morphological roots should take various kinds of verbal (and/or adjectival) affixes to formulate predicates (Table 1). Depending on the possible morphosyntactic behavior, these affixes fall into three verbal classes (dynamic verb, potentive verb, stative verb) and one adjective class. As

shown in Table 2, these predicate classes differ as to (i) whether the class may inflect for progressive, (ii) whether the class has tense distinction between non-past (or present) and past, (iii) whether the class has a productive voice alternation between actor and undergoer voices, (iv) whether the class has an implication of temporality, (v) whether the class can appear in the comparative construction, and (vi) whether the class has the reduplicative morphology signaling such intensification as 'very, too X', all of which are briefly summarized in Table 2 (see Kimoto 2017b for a further examination). Dynamic and potentive verb classes have voice distinction: actor voice and undergoer voice.

As mentioned in Section 2.1, GRAID annotations on Arta discourses include a cross-reference tag on nominals about the relevant predicate type. Dynamic verb class, a morphologically unmarked category, is tagged simply by ⟨_a⟩ (actor voice (of dynamic verb)) or ⟨_u⟩ (undergoer voice (of dynamic verb)), and potentive verb class, a morphologically marked category, is tagged either by ⟨_ap⟩ (actor voice of potentive verb) or ⟨_up⟩ (undergoer voice of potentive verb). Other predicate categories, which lack voice distinction, are specified simply as abbreviations as shown in Table 1.

(29) ⟨:s_a⟩ (S argument of a dynamic-verb predicate of actor voice)

	Amma	*mam-purab*	*tidi*	*ama*	*=mi*	*=ti,*
	if	INTR-hunt	PL.DEF	father	=1PL.GEN	=SPC
## #ac	other	v:pred	ln	np.h:s_a	=pro.1:poss	rn

'If **our fathers** go hunting' (arta0002-06)

(30) ⟨:a_u⟩/⟨:p_u⟩ (A/P argument of a dynamic-verb predicate of undergoer voice)

	Sa:bit-èn	*=di*	*=d,*	#	*ngay*	*ti*
	carry-TR	=3PL.GEN	=COMP		GO	OBL.DEF
## 0:p_u	v:pred	=pro.h:a_u	rv	#	vother:pred	ln

bunbun	*=mi.*
house	=1PL.GEN
np:g_other	=pro.1:poss

'**They** carry **it**, going (back) to our house.' (arta0002-62, 63)

(31) ⟨0:s_ap⟩ (S argument of a potentive-verb predicate of actor voice)

	Awan	*=tep*	*maka-angay*	*ta*	*ayta*	*lugar.*
	NEG	still	POT-GO	OBL	DEM.DIST.OBL	place
##neg 0:s_ap	other	other	v:pred	ln	ln	np:g_ap

'**They** could not come there to the place yet.' (arta0002-33, 34)

(32) ⟨:s_up⟩ (S argument of a potentive-verb predicate of undergoer voice)

	Saya	*na-pi:piya*	*=d*	*i*	*pamilia*	*=mi.*
	then	PST.POT-good	=COMP	DEF	family	=1PL.GEN
##	other	v:pred	other	ln	np:s_up	=pro.1:poss

'Then **our family** became better.' (arta0601-46)

(33) ⟨:s_**stv**⟩ (S argument of a stative-verb predicate)

	Tit-tèkèk	*=a*	*ta*	*Dios.*
	STV-wish	=2SG	OBL	God
##	v:pred	=pro:s_**stv**	ln	np:obl_stv

'**You** should be praying to God.' (arta0601-89)

(34) ⟨:s_**adj**⟩ (S argument of an adjective predicate)

	Med-dès	*i*	*uga:li*	*=ku*	*=y.*
	ADJ-bad	DEF	habit	=1SG.GEN	=SPC
##	adj:pred	ln	np:s_**adj**	=pro.1:poss	rn

'**My habit** was bad.' (arta0601-74)

(35) ⟨:s_**predex**⟩ (S argument of an existential predicate)

	Atti:	*=tep*	*i*	*gilangan*	*=i*	*ta*	*ayta*
	exist	=still	DEF	man	=SPC	OBL	DEM.DIST.OBL
##	other:predex	other	ln	np.h:s_**predex**	rn	ln	ln

Danak.
Danak
np:l_predex

'**The man** was still there in Danak.' (arta0106-15)

(36) ⟨:s_**np**⟩ (S argument of a nominal predicate)

	siye:,	*wa*	*=m,*	*kuwarto*	*=m.*
	DEM.PROX	PLH	=2SG.GEN	money	=2SG.GEN
##	dem:s_**np**	other	other	np:pred	=pro.2:poss

'**This** is yours, your money.' (arta0601-94)

3. Referential expressions

3.1. Nominal structure

When an independent referential expression is headed by a lexical noun, the noun should be preceded by a determiner that inflects for the number, case, and definiteness. The noun may be followed by a specificity marker, which signals that the referent is a specific object known to the speaker. A determiner is tagged as ⟨ln⟩, and a specificity marker is, when it appears after a noun, tagged as ⟨rn⟩:[6]

[6] In GRAID annotations, ⟨ln⟩ is defined as "NP-internal subconstituent occurring to the left of NP head", and ⟨ln⟩ as "NP-internal subconstituent occurring to the right of NP head" (Haig and Schnell 2014: 9).

(37)

	Atti:	*=tep*	*[i*	*gilangan*	*=i]*	*ta*	*ayta*
	exist	=still	DEF	man	=SPC	OBL	DEM.DIST.OBL
##	other:predex	other	**ln**	np.h:s_predex	**rn**	ln	ln

Danak.
Danak
np:l_predex
'The man was still there in Danak.' (arta0106-15)

When a specificity marker appears within a noun phrase, it should occupy the slot immediately after the first lexical element. For example, when a modifier appears before a head noun, the specificity marker should no longer follow the noun but should be encliticized to the modifier:

(38)

	Tidi	*tallip*	***=i***	*a*	*buka:gan,*	*awan*	*=tid*
	PL.DEF	two	=SPC	LIG	woman	NEG	=3PL
##	ln	ln	**ln**	ln	np.h:dt_s	other	=pro.h:s_ap

naka-panga:dal
PST.POT-learn
v:pred
'As for the two women, they were not able to go to school.' (arta0110-046)

Some optional elements may modify a head noun with the intervening connective *a* (LIGature), as shown by the numeral quantifier in (38). Such elements within the nominal are also annotated as ⟨ln⟩ or ⟨rn⟩ based on the relative position to the head noun. The examples below illustrate the cases in which the adjectives, quantifiers, and/or demonstratives modify the head nouns.[7]

(39) (adjective+noun)

	Basta	*in-an'anu:s-an*	*=mi*	*=tèddi*	*ay*	*[ka:man*	*=i*
	just	PST-tolerate-TR	=1PL.GEN	=only		big	=SPC
##	other	v:pred	=pro.1:a_u	other	other	**ln**	**ln**

a *to:luda].*
LIG tent
ln np:p_u
'We just tolerated a big tent (instead of their own houses).' (arta0007-08)

(40) (quantifier+adjective+noun)

	Um-angay	*=de:*	*=tèn*	*=ti,*		*man*
	INTR-go	COMP	=1SG	DEM.PROX.OBL		as.if
##	v:pred	other	=pro.1:s_a	=dem:g_a	#ac 0.1:a_up	other

[7] It is sometimes difficult to determine which element is the head of the nominal among several words because alternative ordering of elements is possible in Arta, and, in fact, in Philippine languages in general.

na	*ne:but*	*[attanan*	*a*	*med-dès*	*a*	*uga:li].*
GEN	PST.POT.lose	all	LIG	ADJ-bad	LIG	custom
other	v:pred	**ln**	**ln**	**ln**	**ln**	np:p_up

'After I came here, it seems that (I) have lost every bad custom.' (arta0601-55)

(41) (quantifier+noun), (demonstrative+adjective+noun)

èns i:na	*di:san*	*=na*	*[i*	*gissa*	*=y*	*a*	*lingo]*	*[aynina*	*a*
so.that	reach	=3SG.GEN	DEF	one	=SPC	LIG	week	DEM.MED	LIG
## other	v:pred		**ln**	**ln**	ln	**ln**	np:p_u	**ln**	**ln**

me"a:du	*a*	*baggat].*
ADJ-plenty	LIG	rice
ln	**ln**	dpi_np:a_u

'so that this plenty of rice will last for one week.' (arta0515-107)

Note that the same constructional template [modifier *a* head] or [head *a* modifier] is employed both for adjective modifications and relative clauses; i.e., both of them could be described as instances of the single constructional template "adnominal modification." For the purpose of cross-linguistic comparisons with non-Philippine-type languages, adnominal modifications exclusively by means of adjectives, quantifiers, and demostratives are treated as ⟨ln⟩ or ⟨rn⟩, while adnominal modifications by means of verbs (or more precisely verb-headed clauses) are treated as relative clauses and annotated separately as ⟨#rc⟩, which will be noted in Section 4.

3.2. Possessive construction

Possessive constructions, composed of the possessed (or possessum) and the possessor, are structurally parallel to argument realizations composed of a transitive verb and an A argument in that both the possessor and the A argument receive a genitive marking. The possessor may be encoded by (i) a bound person form on the possessed noun, (ii) a (bound) demonstrative form, (iii) a full NP, (iv) a demonstrative form with a bound person index on the possessed noun, or (v) a full NP with a bound person index on the possessed noun. In this annotation, the possessive constructions are tagged as ⟨:poss⟩, and those different formal strategies are glossed as follows:

Table 3 Annotations of possessive forms

tag	description
⟨=pro:poss⟩	bound person form
⟨(=)dem:poss⟩	(bound) demonstrative form
⟨np:poss⟩	full NP
⟨bpi_dem:poss⟩	demonstrative form with bound person index
⟨bpi_np:poss⟩	full NP with bound person index

Following the glossing rules of argument realization patterns, if the possessor role is doubly marked by a bound person index as well as a demonstrative or a full NP, the bound person index is glossed as ⟨**bpi**_dem/np:poss⟩. Some of the actual annotations of possessive constructions are shown below.

(42) ⟨=pro:poss⟩ (the person index only)

	Amma	*mam-purab*	*tidi*	*ama*	***=mi***	*=ti,*
	if	INTR-hunt	PL.DEF	father	=1PL.GEN	=SPC
## #ac	other	v:pred	ln	np.h:s_a	**=pro**.1:poss	rn

'If **our** fathers go hunting' (arta0002-06)

(43) ⟨np:poss⟩ (the genitive NP only)

	Punan	*=na*	*ayni*	*babakat*	*=i,*	*a*
	say	=3SG.GEN	GEN.DEF	old.woman	=SPC	LIG
## 0:p_u	v:pred		ln	=pro_np.h:a_u	rn	rn

ina	***ni***	***buka:gan***	***=i.***
mother	GEN.DEF	woman	=SPC
np.h:appos	ln	**np**.h:poss	rn

'this old woman, (who is) the mother **of the woman**, said it.' (arta0106-25)

(44) ⟨=dem:poss⟩ (bound demonstrative form) and ⟨dem:poss⟩ (demonstrative NP)

	Satidi:na	*ne:nan*	*=mi*	*=ta*	*a*	*pare:ho*
	DEM.MED	PST-GO-TR	=1PL.GEN	=DEM.DIST.OBL	LIG	same
##	other	v:pred	=pro.1:a_u	=dem:g_u	other	other

=mi	*tidi:na*	*arta*	*aydi*	*ina*	*=ni,*	*ama*
=1PL.GEN	DEM.MED	person	and	mother	=DEM.PROX.GEN	father
=pro.1:poss	ln	np.h:p_u	rn	rn	**=dem.h:poss**	rn

ni	*ayni.*
GEN.DEF	DEM.PROX.GEN
ln	**dem.h:poss**

'We went to those people and to the mother **of this (guy)** and the father **of this (guy).**' (arta0114-052)

(45) ⟨bpi_np:poss⟩ (a demonstrative form with a bound person index)

	Ti	*Lenin*	*ama*	***=na***	***ni***	***ayni.***
	PSN	(personal.name)	father	=3SG.GEN	GEN.DEF	DEM.PROX.GEN
##	ln	np:pred	np.h:s_np		ln	**bpi_np**.h:poss

'The father **of this (guy)** is Lenin.' (arta0114-053)

(46) ⟨bpi_np:poss⟩ (full NP with a bound person index)

	konta	*ay*	*ngadin*	***=na***	*=te*	***[ni***	***wanga:r***
	but		name	=3SG.GEN	=only	GEN.DEF	stream
##	other	other	np:s_np		rn	ln	**=bpi_np:poss**

=i], *Dikerawyan.*
=SPC (place.name)
rn np:pred
'But as for the name of the stream, it is Dikerawyan.' (arta0114-049)

4. Complex sentence: the treatment of gap constructions

The structure of complex sentences is relatively simple, so it is unproblematic to apply the GRAID annotation rules to our data. However, the treatment of gap constructions employed for relative clauses is worth noting. Consider the following excerpt from a discourse, in which the complex nominal phrase is headed by the head noun *ka:huy* 'sweet potato' and immediately followed by a relative clause *a nimulamula=mi* 'that we planted':

(47)	*(...)*	*i*	*ka:huy*	*#*	*a*	*n-i-mulamula*	*=mi*
		DEF	sweet.potato	#	LIG	PST-TR-plant	=1PL.GEN
		ln	np:p_up	#rc gap:p_u	other	v:pred	=pro.1:a_u

'(and) the sweet potatoes we planted' (arta0007-21)

Although the transitive verb *nimulamula* 'planted' within the relative clause creates the expectation for two arguments to occur, the undergoer argument cannot appear within the embedded clause. This is not because the argument is pragmatically inferable but because the construction does not allow overt appearance of the argument. In GRAID annotations, the zero realization caused by grammatical suppressions should be specified differently from the one caused by pragmatic conditions. It is thus specified as ⟨gap:p_u⟩, implying a **gap** argument with **no pronominal index** within a relative clause, in which the gap functions as the **P** argument of the **undergoer voice** of a dynamic verb.

One of the prominent features of Philippine languages is that a relative clause itself may function as an argument of another clause (i.e. the headless relative clause). As illustrated in the examples below, the relative clauses serve as the S argument (48), the A argument (49), the P argument (50), the oblique argument (51), and the predicate (52) of the higher-order clauses respectively:

(48)		*Maski*	*adin*	[**S ARGUMENT**:	*i*	*e:n-an*	*=mu*
		even	where		DEF	GO-TR	=2SG.GEN
	## #ac	other	np:pred	#**rc:s**_np gap:p_u	ln	v:pred	=pro.2:a_u
	=y],						
	=SPC						
	rn						

'Wherever you go' (lit. 'Even where the place you go to is') (arta0601-90)

(49) *Saya* *i* *n-i-bud*
DEM.DIST DEF PST-TR-say
dem:s_np #rc:pred gap:p_u ln v:pred
[**A ARGUMENT** *na* *na-dupu:* *=ya* *(...)*]
3SG.GEN PST.POT-old.man =DEM.DIST
#**rc:a**_u gap_dem.h:s_up ln v:pred =rn
'That is what the one who got old told (to me).' (arta0565-18)

(50) *Awan* *=mi* *ta:tataw* [**P ARGUMENT** *i* *e:n-an*
NEG =1PL.GEN know DEF go-TR
##neg other =pro.1:a_u v:pred #**rc:p**_u ln v:pred
=mi]
=1PL.GEN
=pro.1:a_u
'We did not know where to go.' (arta0007-04)

(51) *Awan* *=am* *pe:-barka:da* [**OBLIQUE** *didi*
NEG =2PL INTR-company PL.OBL.DEF
##neg other =pro.2:s_a v:pred #**rc:obl** gap.h:s_a ln
mantatim *=i*]
INTR-drink =SPC
v:pred rn
'You should not be in company with those who drink habitually.' (arta0601-78)

(52) *Saya* [**PREDICATE** *i* *an-èn* *=mi.*]
DEM.DIST DEF eat-TR =1PL.GEN
dem:s_other #**rc:pred** gap:p_u ln v:pred =pro.1:a_u
'That is what we used to eat.' (arta0601-18)

In this GRAID annotation, this type of headless relative clause in Arta is differentiated by specifying the grammatical relation with the predicate type after ⟨#rc⟩, as in ⟨#rc:a_u⟩ (the relative clause functions as the A argument of the undergoer voice of a dynamic verb). This annotation is exactly the same as that of other simpler referential expressions, which makes it possible to compare the distributions of virtually all kinds of referential strategies in the language.

5. Concluding remarks

In this short paper, the annotation notes were provided for implementing the general GRAID glossing system (Haig and Schnell 2014) for the particular purpose of annotating Arta discourse. First, some exposition was made concerning the case marking system and argument structure, which have been problematic in Philippine languages and necessary for identifying S, A, and P roles in this glossing system. Different realization patterns of referential expressions in this language were also

documented, and a particular way to annotate them was proposed. A predicate tag on nominals is added to this annotation for comparing it with other Philippine-type languages with particular reference to the glossing convention in Tondano. Some exposition was also given of complex sentences, especially of gap constructions, which is crucial for understanding relative clauses and complex nominal structure.

Abbreviations

1	first person	NMZ	nominalizer
1+2	first-second person	O	transitive object
2	second person	OBL	oblique case
3	third person	PL	plural
A	transitive subject	PLH	placeholder (filler-like element)
ABS	absolutive	POT	potentive verb
ADJ	adjective	PRG	progressive aspect
COMP	posterior phase 'already'	PROX	proximal (demonstrative)
DEF	definite	PSN	personal nominal
DIST	distal (demonstrative)	PST	past tense
GEN	genitive case	S	intransitive subject
INTR	intransitive verb	SG	singular
LIG	ligature (linker)	SPC	nominal specificity
MED	medial (demonstrative)	STV	stative verb
NEG	negation	TR	transitive verb

Appendix: Person forms, determiners, and demonstratives

The following tables show the paradigm of person forms (including enclitic forms and independent pronouns), determiners, and demonstratives (including enclitic and independent forms). Note that in the grammatical glosses in the second line of each example, "absolutive" and "singular" are omitted for the sake of simplicity.

Table A Person forms

PERSON	TOPICAL	ABSOLUTIVE	GENITIVE	OBLIQUE
1SG	*tèn*	*=tèn*	*=ku*	*dèn*
1PL	*tami*	*=ami*	*=mi*	*dami*
2SG	*taw*	*=a, =taw*	*=mu*	*daw*
2PL	*tam*	*=am*	*=muyu*	*dam*
1+2SG	*tita*	*=ita*	*=ta*	*dita*
1+2PL	*titam*	*=itam*	*=tam*	*ditam*
3SG	*siya*	*Ø*	*=na*	*dya*
3PL	*tidi*	*=tid*	*=di*	*did*

Note: "1+2SG/PL" is the person category "we" which includes "both speaker and hearer (and other(s) in plural)." It differs from "1PL" which includes "speaker and other(s), not hearer," or "2PL" which includes "hearer and others (not speaker)."

Table B Determiners

			ABSOLUTIVE	GENITIVE	OBLIQUE
INDEFINITE			*Ø*	*na*	*ta*
DEFINITE	SINGULAR	(COMMON)	*i*	*ni*	*ti*
		(PERSONAL)	*ti*	*ni*	*ni*
	PLURAL		*tidi*	*didi*	*didi*

Table C Demonstratives

		TOP	ABS	GEN/ERG	OBL
PROXIMAL	SG	*si:yèy*	*a:yi:*	*ni/na a:yi:/ayni*	*ti/ta a:yi:*
			=i	*=ni*	*=ti*
	PL	*satidi:*	*(ay)tidi a:yi:*	*(ay)didi a:yi:*	*(ay)didi a:yi:*
MEDIAL	SG	*sayna*	*a:yina,*	*ni/na ayna*	*ti/ta ayna*
			=ina	*=nina*	*=tina*
	PL	*satidi:na*	*(ay)tidi:na*	*(ay)didi:na*	*(ay)didi:na*
DISTAL	SG	*saya*	*a:ya:,*	*ni/na a:ya:*	*ti/ta a:ya:*
			=ya:		*=ta*
	PL	*satiddya:*	*(ay)tiddya*	*(ay)didi a:ya:*	*(ay)didi*
					a:ya:

References

Brickell, Timothy C. 2016. Multi-CAST Tondano annotation notes. In Geoffrey Haig and Stefan Schnell (eds.) *Multi-CAST (Multilingual Corpus of Annotated Spoken Texts)*. (accessed 2017/11/19)

Foley, William A. 1998. Symmetrical voice systems and precategoriality in Philippine languages. Paper presented at the workshop on voice and grammatical functions.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In *The Austronesian languages of Asia and Madagascar*. Abingdon: Routledge. pp.110–181.

Kimoto, Yukinori. 2014. A preliminary report on the grammar of Arta. *Proceedings of the Thirty-Eighth Annual Meeting of The Kansai Linguistic Society (KLS34)*, pp.85–96.

———. 2017a. Vowel length, mora, and diachrony: The case for Arta, a Philippine Negrito language. In Hsiu-chuan Liao (ed.) *Issues in Austronesian historical linguistics. JSEALS 10.3, Special Publication No. 1*. Honolulu: University of Hawai'i Press. pp.1–22

———. 2017b. A grammar of Arta: A Philippine Negrito language. Ph.D dissertation, Kyoto University.

Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine languages. Ph.D. dissertation, University of Hawai'i.

Mithun, Marianne. 1994. The implications of ergativity for a Philippine voice system. In Barbara A. Fox and Paul J. Hopper (eds.) *Voice: Form and function*. Amsterdam: John Benjamins. pp.247–277.

Payne, Thomas E. 1982. Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6(1). pp.75–106.

Reid, Lawrence A. 1989. Arta, Another Philippine Negrito language. *Oceanic Linguistics* 28(1). pp.47–74.

Rubino, Carl Ralph Galvez. 1997. A reference grammar of Ilocano. Ph.D. dissertation, University of California, Santa Barbara.

Schnell, Stefan and Nils N. Schiborr. 2018. "Corpus-based typological research in discourse and grammar GRAID and Multi-CAST". *Asian and African Languages and Linguistics* 12. pp.1–16.

Shibatani, Masayoshi. 1988. Voice in Philippine languages. In Shibatani, Masayoshi (ed.) *Passive and Voice*. Amsterdam: John Benjamins. pp.85–142.