Multi-CAST

Bora

annotation notes

Frank Seifart

Tai Hong

University of Bamberg

Multi-CAST

*Multilingual Corpus of
Annotated Spoken Texts*

Contents

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 Bora. It corresponds to version 2207 of the annotations, published in July 2022. Unless a more recent version of this document exists, it also applies to any later versions of the annotations.

A comprehensive grammatical description of Bora is Thiesen & Weber (2012). A grammar sketch and a detailed description of the grammar of nominal expressions for the closely related variety Miraña can be found in Seifart (2005). Seifart (2015) discusses argument marking, verb valency, and related issues in Bora.

1.1   Encoding of S and A arguments

Subject core arguments, that is S or A arguments, can be encoded by lexical noun phrases ⟨np⟩ and three types of pronominal elements in Bora. The pronominal elements, which include independent pronouns ⟨pro⟩, pronominal clitics ⟨=pro⟩, and suffixes on verbs ⟨-pro⟩, are not mutually exclusive, and neither do they conflict with overt noun phrases in the same clause. This means cross-references are quite common in this language.

In principle, first and second person subjects are encoded with pronouns that proclitize to the predicate: *o=* 'first person singular' as in (1), *u=* 'second person singular', or *me=* 'first or second person non-singular'. The same behavior is shown by the 'third person same subject' marker *i=*. These pronouns can also occur as free forms. For first or second person non-singular subjects, independent pronouns are used in addition to *me=* to specify person, number, gender, and clusivity, for example *ámupɨ* 'second person dual feminine' or *múha* 'first person plural exclusive', see (2). Third person subjects are in principle expressed by either a free noun phrase, as with *díhjyu* 'your mouth' in (3), or by gender or noun class suffixes on predicates, for example *-:be* 'third person masculine' or *-me* 'third person plural' as in (4), or sometimes by both, as with *mɨamúna* 'people' and *-me* 'third person plural' in (5). In relatively rare cases, S or A arguments are not expressed at all, such as in (6).

(1)   *o péjucóó áádi táhdi núhba éllevu neébe*

	ó=	pee-juco	áádi	táhdií	nuhba	élle-vu	
	1.SG=	go-AFIR	that	grandfather	sun	there-ALL	
## #ds	=pro.1:s	v:pred	other	np.h:other	np:other	np:g	%
	0020				0025		

	nee	-Vbe
	say	-M.SG
	v:pred	-pro.h:s_ds
		0020

'I am going now to see him, the grandfather of the sun, he said.'

[mc_bora_ajyuwa_0031]

(2) *ihjyúcunúhíjcyalle, áyúú llíhíí llihíyómúú, wáháá wahárómúú kiávúamíi ámuha mepéé*

ihjyú-cunu-híjcya	*-lle*		*áyúú*	*llihíyo*	*llihíyo-mu*	*waháro*
cry-MULT.INTR-REP	-F.SG		ah	father	father-PL	mother
% v:pred	-pro.h:s_ds	#ds	other	np.h:voc	np.h:voc	np.h:voc
	0010			0002	0161	0011

waháro-mu	*kiá-vu-ami*	*ámuha*	*me=*	*pee*
mother-PL	where-ALL-irritation	2.PL	1.PL=	go
np.h:voc	other	pro.2:s	=lv_pro_2_s	v:pred
0162		0019	0019	

'That is why she was screaming, Ay father, father, mother, mother, where did you go?'

[mc_bora_ajyuwa_0190]

(3) *áyúú taálléé ɨ̵né imyéjaú díhjyu cuvááhií.*

áyúú	*taalléu*	*ɨ̵ne*	*ímyé-ja-vu*		*di-*	*íhjyu*
ah	grandmother	what	delicious-CLF.cover-ALL		2.SG.POSS-	mouth
##ds	other np.h:voc	other	other		-ln_pro.h:poss	np:s
	0112				0112	0113

cuvaa-hi
smell-PRED
v:pred

'Ah grandmother, your mouth smells delicious.'   [mc_bora_ajyuwa_0253]

(4) *áánéllihyéváa tétsihvúré ɨ̵hvéjtsoíñúmé dííbyeke*

aa-ne-lliíhye-va-pe	*tee-tsi-vu-re*	*ɨ̵hvejtso-iñu*	*-me*	*di-Vbe-ke*
CON-INAN-BEN-QUOT-REM	3-CLF.place-ALL-REST	leave-DIR.leave	-PL	3-M.SG-ACC
## other	pro:obl	v:pred	-pro.h:a	pro.h:p
			0019	0020

'That is why they left him there.'   [mc_bora_ajyuwa_0024]

(5) *ícyaj neélle múha éhne íílleé, mɨamúnaáubáha muurá illóme*

ícyaj	*nee*	*-lle*		*muha*	*e-ne*	*íílleé*
baj!	say	-F.SG		PN.INTER	that-INAN	here
## other	v:pred	-pro.h:s_ds	#ds	pro.h:s	other	other:pred
		0010				

mɨa-múnaa-uba-haja	*muuré*	*illo*	*-me*
true-people-PROB-curiosity	then	fell	-PL
## np.h:s		other v:pred	-rv_pro_h_s
0065			0065

'Baj, she said, Who is there, suddenly there are people felling [i.e. trees].'

[mc_bora_ajyuwa_0122]

(6)   *teenéváa pácámɨcojááné dille óónováne óónováne.*

	tee-ne-va-pe	*páácamɨ-co-ja-ne*	*di-lle*	*oonóva-ne*
	3-INAN-QUOT-REM	ojé_tree-CLF.pointed-CLF.cover-PL	3-F.SG	paint-INAN
##	other	np:p	pro.h:a	v:pred
		0134	0010	

	oonóva-ne
	paint-INAN
##	0.h:s v:pred
	0010

'She painted that ojé tree, (she) painted.'   [mc_bora_ajyuwa_0290]

When two or more elements encode a core argument, we prioritize whichever is the least bound element and assign it the standard annotation for core argument functions, while the other elements that encode the same argument are annotated as cross-indices of the form ⟨lv/rv_form_animacy/person_function⟩, with all GRAID symbol delimiters replaced with underscores. Examples include ⟨=lv_pro_h_s⟩ in (7), ⟨-rv_pro_h_s⟩ in (8), and similarly ⟨=lv_pro_2_s⟩ and ⟨-rv_pro_2_s⟩ in (9) and ⟨-rv_pro_h_s⟩ in (10).

The hierarchy places lexical noun phrases and independent pronouns at the top, followed by proclitics, and verbal suffixes at the bottom (i.e. ⟨np⟩ and ⟨pro⟩ > ⟨=pro⟩ > ⟨-pro⟩). (8) shows a case when noun phrases and suffixes are simultaneously used for encoding the S argument. Similarly, there are both a NP and a proclitic in (7), a proclitic and a suffix in (10), and all three types of pronominal elements in (9). A zero for an S or A argument is only annotated in the relatively rare case where there is no overt encoding present at all, as in (6) above.

(7)   *teenéváa nújpacyo ditye áraúcúneri dityépɨ ipájtyeki*

	tee-ne-va-pe	*nú-jpacyó*	*di-te*	*ára-úcu-ne-ri*
	3-INAN-QUOT-REM	water-CLF.liquid	3-PL	condense-SG.TR-INAN-INST
##	other	np:p	pro.d:a	v:pred
		0054	0050	

	di-jtépɨ	*i=*	*pajtye-ki*
	3-F.DU	3=	cross-PURP
#ac	pro.h:s	=lv_pro_h_s	v:pred
	0035	0035	

'They lowered the water, so they two could cross at this point.'

[mc_bora_ajyuwa_0095]

(8)   *tsáh ditye íllityétúme*

	tsáha	*di-te*	*íllitye-tu*	*-me*
	no	3-PL	be_afraid-NEG	-PL
##neg	other	pro.h:s	v:pred	-rv_pro_h_s
		0010		0010

'They were not afraid.'   [mc_bora_meenujkatsi_0029]

(9) *ééneúrée ámúhpɨ goocópivu ámuhpɨ*

éée-ne-vu-re	*ámuhpɨ*	*gooco-pi-vu*	*ámuhpɨ*
yes-INAN-ALL-REST	2.F.DU	laugh-in_excess-ALL	2.F.DU
##ds other	ln_pro.2:poss	np:other	pro.2:s
	0035		0035

me=	*úújete*	*-múpɨ*
2.DU=	reach	-F.DU
=lv_pro_2_s	v:pred	-rv_pro_2_s
0035		0035

'That is why, because you laughed, you reach.'   [mc_bora_ajyuwa_0100]

(10) *aane diille íeeválleke neélle*

aa-ne		*di-lle*	*i=*	*eeva*	*-lle*	*-ke*
CON-INAN		3-F.SG	3=	be_pregnant	-F.SG	-ACC
## other	#rc.h:p	other	=pro.h:s	v:pred	-rv_pro_h_s	-other %
	0010		0010		0010	

nee	*-lle*
say	-F.SG
v:pred	-pro.h:a
	0060

'And she told to the pregnant one'   [mc_bora_ajyuwa_0109]

1.2   Encoding of P arguments

Animate P arguments in Bora are usually encoded by lexical noun phrases or pronouns with an accusative marker *-ke,* as in (11). In subordinate clauses, verbal suffixes may also encode P arguments, as is the case in (12) and (13). Unlike S and A arguments, cross-references for P arguments by means of person suffixes on verbs are only possible in subordinate clauses, where person suffixes function like relative pronouns.

(11) *muhdúami ɨdsɨhɨɨváneri llihíyó meke nujpáñutso*

muhdu-ami	*llihíyo*	*me-ke*	*nú-jpacyó-nu-tso*
how-irritation	father	1.PL-ACC	water-CLF.liquid-VBZ.do-CAUS
##ds other	np.h:a	pro.1:p	v:pred
	0002	0035	

'How can our father send us to get water with this thing?'   [mc_bora_ajyuwa_0066]

(12) *aallévaa ílluréjuco píívyeténé íñe úwáhllojɨ menééilléréjuco*

aa	*-lle*	*-va-pe*	*íllu-re-juco*	*píívyete-ne*
CON	-F.SG	-QUOT-REM	SO-REST-AFIR	turn_into-INAN
## other	-pro.h:s	-other	other	v:pred
	0010			

í-ne	*úwáhllojɨ*	*me=*	*nee-i*	*-lle*	*-re-juco*
this-INAN	tuayo_bird	1.PL=	say-FUT	-F.SG	-REST-AFIR
#cc:other other	np.n:p2	=pro.1:a	v:pred	-pro.h:p	-rv
	0029			0010	

'And she turned into what we will call a tuayo.'   [mc_bora_ajyuwa_0045]

(13) *diibyéjuco o ámejcánúíyoobe, ehdu tsúúca méhdityu pééneé*

di-Vbe-juco	*ó=*	*ámejcánu-i-ró*	*-Vbe*	*e-du*
3-M.SG-AFIR	1.SG=	to_appoint-FUT-FRUS	-M.SG	that-CMPV
##ds pro.h:s	#rc =pro.1:a	v:pred	-pro.h:p other	%
0020	0002		0020	

tsúúca	*me-di-tu*	*pee-ne*
already	1.PL-ANIM-ABL	go-INAN
other	pro.1:obl	v:pred
	0033	

'For him I was going to throw a feast, and now he is leaving us.'

[mc_bora_ajyuwa_0055]

1.3   Verbal suffixes

As discussed in Section 1.1, verbal number-gender/class suffixes in Bora may encode S and A argu-ments, as well as P arguments in subordinate clauses (see Section 1.2). They are always attached directly to the verb stem. A very similar set of number-gender/class suffixes is also used on the conjunction *aa*, forming an anaphoric pronoun (see Seifart 2010 for discussion). In this case, the suffix is usually annotated as a dislocated topic ⟨:dt⟩ as seen in (14).

(14) *aamévaa tsúúca teene walléémú íjtyamu ujcújucóóhii*

aa	*-me*	*-va-pe*	*tsúúca*	*tee-ne*	*wallee-mu*	*íjtya-mu*	*ujcu-juco-hi*
CON	-CLF.PL	-QUOT-REM	already	3-INAN	woman-PL	starch-PL	get-AFIR-PRED
## other	-pro.h:dt_a	-other	other	ln	np.h:a	np:p	v:pred
	0007				0007	0008	

'And they, the women got the starch.'   [mc_bora_ajyuwa_0007]

1.4   Possessive markers

Possessive relations are usually encoded by the possessive prefixes on nouns, which are *ta-* for the first person singular, *di-* for the second person singular, *i-* for third person singular and plural, and *me-* for first person plural. Two of the four prefixes, *i-* and *me-*, share the same form as the proclitic indices for subjects. Examples of their usage are given below in (15–17).

(15)   *tábállííjyúhdené o úújeíñúnáa*

	ta-	*bállií-jyu-hde-ne*	*ó=*	*uuje-iñu-náa*
	1.SG.POSS-	ash-CLF.powder-possibility-REC	1.SG=	go-DIR.leave-SIM
##ds	-ln_pro.1:poss	np:p	=pro.1:a	v:pred
	0065	0085	0065	

‘I forgot my ashes.’   [mc_bora_ajyuwa_0164]

(16)   *díuubálléré o lléébohíjcyalle*

	di-	*uubálle-re*	*ó=*	*lleebo-híjcya-lle*
	2.SG.POSS-	story-REST	1.SG=	listen-REP-F.SG
##ds	-ln_pro.2:poss	np:p	=pro.1:a	v:pred
	0010		0100	

‘I only heard your story.’   [mc_bora_ajyuwa_0202]

(17)   *teene íeevácoréjuco ditye cóeváné pátsoodónoho.*

	tee-ne	*i-*	*éévaco-re-juco*	*di-te*	*coéva-ne*
	3-INAN	3-	pregnancy-REST-AFIR	3-PL	be_left-INAN
##	other	-ln_pro.h:poss	np:p	pro.d:a	v:pred
		0010	0156	0147	

pa-tsoodónoho
all-CLF.big_person
other

‘They left only her pregnancy.’   [mc_bora_ajyuwa_0353]

1.5   Headless relative clauses functioning as complements

It is common in Bora for a subordinate verb that includes a number-gender/class suffix to be placed at the beginning of a complex sentence and function as a subject complement clause for the following verbs, which then lack any overt subject indexing. It is often not easy to translate such structures into English, and they can easily be mistaken for instances of zero subjects, as shown in the examples below.

(18)   *dsɨɨneebe péérónáa dííbyé úmɨwari diitye wáárɨmu, kiávú wáárɨmu*

	dsɨɨne	*-Vbe*	*pee-ró-náa*	*di-Vbe*	*húmɨwa-ri*	*di-te*
	run	-M.SG	go-FRUS-SIM	3-M.SG	front-LOC	3-PL
##	#rc.h:s	v:pred	-pro.h:s %	v:pred	#ac ln_pro.h:poss	np:pred	other
	0065	0065	0065			0065

wáárɨa-mu *kiá-vu* *wáárɨa-mu*
ant-PL   where-ALL ant-PL
np.n:s   other   np.n:other
                     0084

‘He ran but could not proceed because in front of him was a slab of ants. [lit. ‘The one who was running could not proceed...’]’   [mc_bora_ajyuwa_0163]

(19)   *íjyu nééllere téhulle ɨɨtécunú.*

	í-ju	*nee*	*-lle*	*-re*
	this-CLF.path	say	-F.SG	-REST
##	#rc.h:s other	v:pred	-pro.h:s_ds	-rv %
	0124 0132		0124	

tee-hulle	*ɨɨte-cunu*
3-CLF.further_place	look-MULT.INTR
other	v:pred

'This way, she said and looked in that direction. [lit. 'The one who said, This way! looked in that direction'.]'   [mc_bora_ajyuwa_0272]

1.6   Determiner-like uses of third person demonstratives and other third person pronouns

Third person pronouns can be referential, but can also be used as determiners when they immediately precede corresponding noun phrases, in which case they are not referential. See, for example, *diille* in (20) and *teeja* in (21).

(20)   *diille wahráúba mééwadívú úújetéllé.*

	di-lle	*wára-vu-ba*	*mewa-di-vu*	*úújete*	*-lle*
	3-F.SG	polish-ALL-CLF.object	wife-ANIM-ALL	reach	-F.SG
##	ln	ln_np	np.h:g	v:pred	-pro.h:s
			0114		0010

'She reached the polishing woman.'   [mc_bora_ajyuwa_0256]

(21)   *pácallúrɨɨjɨ teeja ihjya pácallúrɨɨjɨ ímamúíhcyu.*

	pa-cállúrɨ́jɨɨ	*tee-ja*	*i-*	*ja*
	complete-bend	3-CLF.cover	3-	house
##	v:pred	ln	-ln_pro.h:poss	np:s
			0118	0120

pa-cállúrɨ́jɨɨ	*i-*	*mamu-ihcyu*
complete-bend	3-	tripod-CLF.frame
## v:pred	-ln_pro.h:poss	np:s
	0118	0119

'Bent was her house, bent was her tripod.'   [mc_bora_ajyuwa_0263]

References

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).

Seifart, Frank. 2005. *The structure and use of shaped-based noun classes in Miraña (North West Amazon)*. Nijmegen: Radboud University Nijmegen Ph.D. dissertation. (`http://hdl.handle.net/11858/00-001M-0000-0013-1E5B-0`).

Seifart, Frank. 2010. The Bora connector pronoun and tail-head linkage: A study in language-specific grammaticalization. *Linguistics* 48(4). 893–918. (`https://doi.org/10.1515/ling.2010.028`).

Seifart, Frank. 2015. Valency classes in Bora. In Malchukov, Andrej & Comrie, Bernard (eds.), *Valency classes in the world's languages, volume 2*: *Case studies from Austronesia, the Pacific, the Americas, and theoretical outlook*, 1481–1514. Berlin: Mouton de Gruyter.

Seifart, Frank & Hong, Tai. 2022. Multi-CAST Bora. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST*: *Multilingual Corpus of Annotated Spoken Texts*. (`https://multicast.aspra.uni-bamberg.de/#bora`) (Accessed 2022-07-09).

Thiesen, Wesley & Weber, David. 2012. *A grammar of Bora with special attention to tone*. Dallas, TX: SIL International.

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 Bora 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

Person/animacy symbols and specifiers

⟨.n⟩   non-human animate

Function symbols and specifiers

⟨_ds⟩   *specifier:* subject of a verb of speech; attaches to ⟨:s⟩, ⟨:a⟩, and ⟨:ncs⟩

Other symbols

⟨_1⟩   *on cross-indices:* first person
⟨_2⟩   *on cross-indices:* second person
⟨_h⟩   *on cross-indices:* human
⟨_d⟩   *on cross-indices:* anthropomorphized entity
⟨_n⟩   *on cross-indices:* non-human animate

⟨_s⟩   *on cross-indices:* subject of an intransitive clause
⟨_a⟩   *on cross-indices:* subject of a transitive clause
⟨_p⟩   *on cross-indices:* direct object

⟨nc_⟩   *specifier:* marks form glosses with RefIND indices in segments otherwise
                        not considered (e.g. those marked with ⟨#nc⟩)

B   List of abbreviated morphological glosses

1	first person	M	masculine
2	second person	MULT	multiple action
3	third person	NEG	negation
ABL	ablative case	NMZ	nominalizer
ACC	accusative case	PERF	perfective
AFIR	affirmative	PERT	pertaining to
ALL	allative case	PL	plural
ANIM	animate	PN	third person pronoun
AUG	augmentative	POSS	possessive
BEN	benefactive case	PRED	predicate
CAUS	causative	PROB	probability
CLF	classifier	PROX	proximal
CMPV	comparative	PURP	purposive
CON	sentence connector	Q	interrogative particle
COND	conditional	QUOT	quotative
COP	copula	REC	recent past
DIM	diminutive	RECP	reciprocral
DIR	directional	REFL	reflexive
DU	dual	REM	remote past
EXCL	exclusive	REP	repetitive
F	feminine	REST	restrictive
FRUS	frustrative	SG	singular
FUT	future	SIM	similar to
HAB	habitual	SOC	sociative case
IFR	inferential evidentiality	STA	stative
IMP	imperative	TR	transitive
INAN	inanimate	VBZ	verbalizer
INST	instrumental	VOC	vocative
INTER	interrogative		
INTR	intransitive	NC	not classified
LOC	locative case		

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts

multicast.aspra.uni-bamberg.de/