A Progress Report on Sumbawa Annotated-spoken Corpus: Tentative Annotation Notes

Shiohara, Asako
ILCAA, Tokyo University of Foreign Studies

This paper is intended as a progress report on an on-going project of developing an annotated corpus of spoken texts from Sumbawa. It is being developed as part of a larger multi-lingual corpus called Multi-CAST (Multi-lingual Corpus of Annotated Spoken Texts, Schiborr 2016). A common glossing system called GRAID (Grammatical Relations and Animacy in Discourse, Haig and Schnell 2014) is employed in the Multi-CAST corpus, and the main part of this paper is devoted to the notes explaining how the GRAID system is applied to the Sumbawa corpus. As GRAID is a system for glossing major clause constituents in texts with their grammatical relations and overt forms (noun phrases, pronouns etc.), and aims at facilitating cross-linguistic research in corpus-based typology, the present paper also provides a brief sketch of Sumbawa morpho-syntax, based on its spoken corpus.

Keywords: Sumbawa, GRAID annotation, agreement, corpus-based grammar, argument structure

1. Introduction

Sumbawa (indigenous designation: Samawa; ISO-639-3 code: SMW) is a Western Austronesian language spoken in the Western part of Sumbawa Island, Indonesia. Administratively, the area belongs to two districts, namely Sumbawa district (Kabupaten Sumbawa) and West Sumbawa district (Kabupaten Sumbawa Barat), in the province of West Nusa Tenggara[1] (NTB: Nusa Tenggara Barat). Sumbawa belongs to the Bali-Sasak-Sumbawa subgroup of the Malayo-Polynesian branch of the Austronesian language family (Adelaar 2005, Mbete 1990).

Shiohara, Asako. 2018. "A progress report on Sumbawa annotated-spoken corpus: tentative annotation notes". *Asian and African Languages and Linguistics* 12. pp.75–97. [Permanent URL: http://hdl.handle.net/10108/91143]

[1] Nusa Tenggara means 'Southeast' in Malay.

This paper is intended as a progress report on an on-going project of developing an annotated corpus of spoken texts from Sumbawa. It is being developed as part of a larger multi-lingual corpus called Multi-CAST (Multi-lingual Corpus of Annotated Spoken Texts, Schiborr 2016). A common glossing system called GRAID (Grammatical Relations and Animacy in Discourse, Haig and Schnell 2014) is employed in the Multi-CAST corpus, and the main part of this paper is devoted to the notes explaining how the GRAID system is applied to the Sumbawa corpus. For more details of Multi-CAST and GRAID, see Schnell and Schiborr (2018). The Sumbawa corpus include 400 clauses so far, and 1000 clauses are envisaged in the future. These notes may be subject to some revision once more data has been worked on.

The structure of this paper is as follows. In section 2, I will explain the source of the data included. In Section 3–6, a tentative version of annotation notes is given; I survey the clause types and structures in Section 3, in Sections 4 and 5, I examine how referential expressions and the constituents of the predicates can be glossed by GRAID conventions, respectively. In Section 6, I will investigate the types of subordinate clauses and how these clauses and direct speech are glossed by GRAID conventions.

Each example cited from our corpus consists of four lines; the first line shows the Sumbawa sentence[2], the second line shows word-by-word glosses, the third line shows the GRAID annotations, while the fourth line shows the free translations. The list of symbols employed in the GRAID convention is given in the appendix of Schnell and Schiborr (2018).

2. Data included in the corpus

The corpus is a collection of spoken monologues obtained by the present author in 1996 and 1997 from speakers living in Empang, a small town in the eastern part of the Sumbawa district. According to Mahsun (1999), the language spoken in the town is classified as the Sumbawa Besar dialect, which is distributed in a large area of the western part of the Sumbawa-spoken area. At present, the data comprise the five folktales listed below. The forms in parenthesis are used to code the source of each example sentences in sections that follow.

- Tanjung menangis ‘Weeping Cape’ <TM>
- Batu nampar ‘A flat stone’ <FS>
- Tutir Lalu Kurekkure ‘The story of Prince Kurekkure’ <LK>
- Tutir *ba*[3] ‘A story of a flood’ <flood>

[2] The transcription employed here basically follows the orthography of Indonesian, using the following conventions: ng for [ŋ] , ny for [ɲ], c for [tʃ], j for [dʒ], y for [j], and e for [ə]. There are also some additional distinctions in the transcription of some vowels, as in open-mid unrounded front vowel è [ɛ], the close-mid unrounded front vowel é [e], a the open-mid rounded back vowel o [ɔ], and the close-mid rounded back vowel ó [o]. An apostrophe (’) in the word final is used to show stress when it is heavier than usual.

[3] The word *ba* means a flood in Sumbawa.

- Sajara Samawa 'A history of Sumbawa' <history>

The annotation notes given in the following sections are mainly based on the five stories mentioned above. Sentences from another source, however, are cited in Section 5.2, to display a pattern that is not found in the stories above but is often observed in daily conversation.

3. Overview of clause structures

Clauses in Sumbawa may be grouped into two types: verbal clauses (See Section 5.1) and non-verbal clauses (See Section 5.2). A verbal clause has a verbal predicate while a non-verbal clause has a non-verbal predicate as the clause head. In both types, the predicate is the only obligatory constituent. Arguments may not occur when the referent is inferable from the discourse context, as seen in Sentences (3b) and (4c,d) below. Sentences (1) and (2) are examples of non-verbal clauses. Sumbawa does not have a copula; the argument and the predicate are simply juxtaposed in a non-verbal clause. An argument may occur either before or after the predicate. In Sentence (1), the argument, which is glossed by the symbol ⟨`np:s`⟩ (S argument realized as an NP) follows the predicate, which is glossed by the symbol ⟨`np:pred`⟩ (a predicate realized as an NP), while in Sentence (2), the argumen precedes the predicate.

(1)

	yanansi	*penyakit*	*berong*	***singin***.
	that.is	disease	leprosy	name
##	`other`	`np:pred`	`rn`	`np:s`

'That is, the name is leprosy disease.' (TM007)

(2)

	cowèk	*singin*	*talang*	*tau*	*dunung'*.
	cowek	name	dish	people	before
##	`np:s`	`np:pred`	`rn`	`rn`	`rn`

'*Cowek* is a name of dish of people of long time ago.' (BL: 053)

Sentences (3) and (4) are examples of verbal clauses. The single intransitive argument (S) may either follow the predicate as seen in (3f) or precedes the predicate, as seen in (3a). Similarly, the patient (P) may either follow the predicate as seen in (3c) and (3e), or precedes the predicate, as seen in (4b). The core arguments for the transitive agent (A) precedes the predicate, as seen in sentence (4a). The transitive agent may be expressed in the PP with the preposition *ling* 'by', as seen in *ling lala* in sentence (d). A PP with the preposition *ling* 'by' always follows the predicate.

(3) a.

	tapi	*tau*	*ta*	*nongka*	*bernang,*
	but	person	DEM_P	NEG.PST	stop
##neg[4]	`other`	`np:s`	`rn`	`lv`	`v:pred`

[4] A clause boundary is marked by the symbol ⟨#⟩ in the GRAID convention. Main clauses are marked by the symbol

b. *teris* *si,*
go.on DM
0:h:s v:pred rv

c. *ya=surung* *sampan* *ta,*
3=push boat DEM_P
pro.h:a=v:pred np:p rn

d. *ya=turit* *ling* *lala,*
3=follow by princess
0.h:p pro.h:a=v:pred adp np.h:dt_a

e. *dapat* *tenga,*
arrive middle
0.h:a v:pred np:p

f. *balawas* *mo* *tau* *ta.*
recite.a.poem DM person DEM_P
v:pred rv np:s rn

'But the person didn't stop, (he) went on. He pushed the boat. The princess followed. Arriving in the middle (of the sea), the person recited a poem.' (TM062–064)

(4) a. *sópó* *waktu* *ina* *tuja* *padé,*
one time mother pound rice
other other np.h:a v:pred np:p

b. *anak* *ya=satokal* *pang* *bao* *Batu* *Langléló* *ta,*
child 3=sit at above stone Langlelo DEM_P
np.h:p pro.h:a=sit adp np:l rn rn rn

c. *nó.poka* *basebó,*
not.yet have.breakfast
##neg 0.h:s lv v:pred

d. *nó.poka* *mangan,*
not.yet eat
##neg 0.h:s lv v:pred

e. *karing* *beling* *mo* *anak.*
then say DM child
other v:pred rv np:s

'One day, the mother was pounding rice. (She) sat the children on the Langlelo Stone. (They) had not had breakfast, (they) had not had eaten. Then the children said.' (FS006–009)

⟨#⟩, while independent clauses are marked by an additional symbol for distinguishing the types of clause (e.g., #ac for an adverbial clause, and #cc for a complement clause). See the appendix of Schnell and Schiborr (2018) for details.

The relative order of the predicate and the argument is, to a larger extent, determined by the information structure of the clause, the details of which we will not discuss further here. When an A argument and a P argument co-occur before the predicate, the A argument always is rendered by the clitic pronoun, and the P argument precedes it, as seen in (4b).

4. Referential expressions

Glossing of referential expressions is a fundamental part of GRAID annotations. This section, following Haig and Schnell (2014), explores Sumbawa referential expressions.

4.1. Forms of referential expressions

In Sumbawa, as seen in Section 3, core arguments may not occur when the referent is inferable from the discourse context. The symbol ⟨0⟩ is employed to code such core arguments. As seen in 4.1.2 below, A and S function may be coded by a clitic pronoun, which is given the symbol ⟨pro=⟩, as well as free form, which is given the symbol ⟨np⟩ or ⟨pro⟩, depending on whether it is a lexical NP or a pronoun.

4.2. NP

In GRAID annotations, the symbol ⟨np⟩ is given to so-called lexical NPs, or what in literature is labelled 'lexical mention/expression', etc. (Du Bois 1987, Haig and Schnell 2014). In sentence (5), the symbol ⟨np⟩ is given to the NP headed by a common noun, while, in Sentencse (6) and (7), it is given to the NP headed by the proper noun—such as the personal name *Kerekkure* in sentence (6) or the place name *Berangkorong* in sentence (7).

(5)

		menong'	*soara*	*ta*	*ling*	*bidadari*	*pitu*	*ta.*
		listen	voice	DEM_P	by	fairy	seven	DEM_P
##	0.h:a	v:pred	np:p	rn	adp	np.d:dt_a	rn	rn

'The seven fairies heard the voice.' (LK039)

(6)

	ada	*sópó*	*tau,*			*basingin*	*Lalu*	*KerèkKurè.*
	exist	one	person			be.named	prince	Kerekkure
##	v:predex	np:s	rn	##	0.h:s	v:pred	ln	np:other

'There was a man named Lalu Kurekkure.' (LK001)

(7)

	Dadap	*nan*	*bakatokal*	*pang*	*Berangkorong.*
	Dadap	DEM_M	be.located	in	Berangkorong
##	np:s	rn	v:pred	adp	np:l

'Dadap is located in Berangkorong.' (LK003)

In GRAID, glosses are aligned with grammatical words, but they essentially trigger clause level constituent phrases that may in turn have more than one constituent word.

For complex NPs, the form gloss ⟨np⟩ is written underneath of the head noun, and the symbol ⟨ln⟩ (NP-internal subconstituent occurring left of the verb) or ⟨rn⟩ (NP-internal subconstituent occurring right of the verb) is written underneath of other NP-internal sub-constituents—excluding possessors which are specified for their own function by the symbol ⟨poss⟩— depending on their relative position to the head. In Sumbawa, the head noun almost always occurs in the NP initial position followed by adnominal demonstratives, adjectives, and another lexical noun (if any), which are are glossed as ⟨rn⟩. The only exception is the article, which is glossed by the symbol ⟨ln⟩, that indicates the gender or status of a personal name. The default articles are *si*, which is used for a female referent, and *nya*, which is used for a male referent. Other articles include *Lalu* 'prince', as seen in Sentence (6). Numerals may either be the head noun or the post-head noun modifier (⟨rn⟩) according to the information status of the referent of the whole NP; when the referent is newly introduced in the discourse, the numeral occurs as a head noun, as in *sópó tau* 'one person' in Sentence (6), otherwise, the lexical noun occurs as a head noun, as in *bidadari pitu* 'seven fairies' in Sentence (5)[5]. Interrogative pronouns are also glossed with the form gloss ⟨np⟩ when they head NPs, since they are not since these are not 'definite pronoun' in the sense of Lyons (1968: 275ff). Reduplicated forms of interrogatives expressing indefinite referent(s) are glossed with the gloss ⟨np⟩, too. Table 1 shows the list of interrogative pronouns and its reduplicated form.

Table 1 Interrogative pronouns and their reduplicated forms

categories	forms	reduplicated form
person	*sai* 'who'	*sai-sai* 'anyone, whoever'
thing	*apa* 'what'	*apa-apa* 'anything, whatever'
time	*pidan* 'when'	*pidan-pidan* 'anytime, whenever'
amount or number	*pida* 'how many'	*pida-pida* 'any or many number'
manner	*mé* 'how, which'	*mé-mé* 'however, whichever'

Sentence (8) is an example of *sai-sai* 'whoever', and Sentence (9) is an example of *pida-pida* 'many numbers'.

(8)

	isi	*pengumuman*	*nan*	*nè*			*sai-sai*	
	content	announcement	that	ITJ			whoever	
##	np:s	rn	rn	other	#ds_cc:pred		np.h:p	#rc
	adè	*bau*			*saterang*	*anak*	*kaku*	*%*
	REL	can			heal	child	1SG.GEN	
0.h:s	other	v:pred	#cc	0.h:a	v:pred	np.h:p	pro.1:poss	%%

[5] We might be able to see the lexical noun is always the head and the position of the head noun relative to the modifier changes depending on the information structure. I don't take the analysis, however, because the numerals may form an NP by themselves (e.g., *ada dua* [exist two] 'There are two (things, people).'), playing a role as a head noun, and that supports the analysis that the numeral is the head noun in an example such as Sentence (6).

	lamin	*sawai*	%	*ku=ètè*	< ··· >
	if	female		1sg=take	
#ds_ac	0.1:s	other	v:pred	%	pro.1:a=v:pred

'The content of the announcement was "whoever who could heal my child, if she is female, I would take (her) (as my daughter)" ...' (TM011)

(9)

	pida-pida	*mo*	*kena*			*laló*	*ngènèng*	*tulung.*
	many.numbers	DM	distination			go	ask.for	help
##	np:pred	rv	np:s	#rc	0.h:a	v:other	v:pred	np:p

'(She) went to many places asking for help. (*lit.* The destinations to which (she) went asking for help are many.)' (FS071)

In conjunctive coordination of NPs, the form gloss ⟨np⟩ is given to the first co-ordinant, and other co-ordinants to its right are glossed with the form gloss ⟨rn⟩. Coordination within an NP is marked by the coordinator *ké* 'and'.

(10)

	dadi	*pengumman*	*ta*	*ya=menong*		*mo*	*ling*	
	then	announcement	DEM_P	3=hear		DM	by	
##	other	np:p	rn	pro.h:a=v:pred		rv	adp	
sandro-sandro	*pang*	*tana*	*Samawa*	*ké*	*sandro-sandro*	*pang*	*kerajaan*	*lin.*
doctor.RED	in	land	Sumbawa	and	doctor.red	in	kingdom	line
np.h:dt_a	rn	rn	rn	rn	rn	rn	rn	rn

'Then the doctors in Sumbawa and doctors in other kingdoms heard the announcement.' (TM014)

4.2.1. Pronouns

The gloss ⟨pro⟩ is used for personal and demonstrative pronouns. Examples of personal pronouns include:

(11)

	aku	*tedu*	*ké*	*kau.*
	1sg	stay	with	2pl
##	pro.1:s	v:pred	adp	pro.2:obl

"I stay with you.' (LK200)

The distinction between a free pronoun and a clitic pronoun is observable in Sumbawa. Table 2 shows a list of clitic and free pronouns. There is no distinction between singular and plural in the third person. The third person clitic *ya=* occurs only with a transitive verb.

Table 2 Sumbawa free and clitic pronouns

	Free pronoun	Clitic pronoun
1SG	*aku*	*ku=*
1PL.INCL	*kita*	*tu=*
1PL.EXCL	*kami*	*tu=*
2SG	*kau*	*mu=*
2PL	*nènè*	*nènè=*
3	nya	*ya=

*The third person clitic occurs only with a transitive verb.

In addition to the forms shown in Table 1, the humble form *kaji* '1SG.HBL' and the honorific form *sia* '2SG.HON' are used when the addressee is of a higher social status than the speaker, and a special humble form for the first person *kajulin* and a special honorific form for the second person singular *kelépé* and *kelépé-kaji* are used when the addressee is a member of the nobility. Distinction in form between the free pronoun and the clitic pronoun are not exhibited in all these forms.

A free pronoun is used as an argument or the complement of a preposition, while a clitic pronoun is either procliticized to the verb indicating the person of the transitive agent (A) or the single argument of an intransitive predicate (S), or encliticized to the noun indicating the person of the possessor. In Sentences (12)–(14), the clitic pronouns of the first person singular, the first person plural, and the second person singular attached to the verb, respectively.

(12)

	ao'	*anak*	*é*	*ta*	*po*	*muntu*	
	ITJ	child	ITJ	DEM_P	DM	time	
##ds 0.1:s	other	np:voc	other	pro:pred	rn	np:s	#rc:rn

ku=nepé.
1SG=winnow
pro.1:s=v:pred
'Ok, children, just now (finally) I am winnowing rice.' (FS023)

(13)

	ka	*mo*	*suda*	*tu=tuja*	*padé*	*ta*	
	PST	DM	finish	1PL=pound	rice	DEM_P	
##ds 0.h:s	lv	lv	v:pred #cc	pro.1:a=v:pred	np:p	rn	##

ta	*muntu*	*tu=tepé.*
DEM_P	time	1PL=winnow
pro:s	np:pred #rc	0:p pro.1.a=v:pred

'We have finished ponding rice now, now is the time for us to winnow.' (FS024)

(14)

	dadi,	*ao,*	*ba*	*mu=tedu*	*ninta*	*mo*
	then	ITJ	ITJ	2SG=stay	here	DM
##ds_pred	other	other	other	pro.2:s=v:pred	pro:l	other

ling.
words
np:s
'Then (she) said, 'you just stay here. (*lit.* Then (her) words were 'you just stay here.')" (KK: 062)

In Sentence (11) above, the S argument is rendered by a free pronoun. In sentence (15) below, the free form of the first person pronoun occurs as a P argument, while in sentence (16), the free pronoun occurs as an A argument.

(15)

	ta	*luk*	*anak*	*ta*	*é,*		
	DEM_P	way	child	DEM_P	ITJ		
##ds	pro:s	np:pred	np.h:voc	pro.rn	other	##	
ma=mu=bada		*aku*	*tegas*	*ling*	*kau*		*nan.*
let=2SG=tell		1SG	meaning	by	2SG		DEM_M
pro.2:a=v:pred		pro.1:p	np:p2	adp	pro.2:dt_a		rn

'This is the way, boy, please tell me the meaning (*lit.* I hope you would tell me the meaning.)' (flood)

(16)

	kasuda	*nya*	*sedo'*	*lamong,*			*maning.*
	then	3	remove	clothes			take.a.bath
##	other	pro.h:a	v:pred	np:p	##	0.h:s	v:pred

'Then she removed the clothes and took a shower.' (LK030)

A complex distribution is observed as to the form that indicates the possessor within an NP. Either a special possessive form *kaku* '1POSS' or the clitic *ku* is used for the first person singular, while both forms can appear as the possessor constituent in the other categories that exhibit the distinction of the clitic form and free form, that is, the first person plural and second person singular.

Table 3 Forms of pronoun expressing possessors

Categories	Forms and examples
1SG	the clitic *ku* and the possessive form *kaku* (e.g., *anak=ku* 'my child' and *anak kaku* 'my child')
1PL and 2SG	the clitic *tu* or *mu* and the free forms *kita/kami* and *kau* (e.g., *anak=tu* 'our child, *anak kami* 'our child', *anak=mu* 'your (SG) child', *anak kau* (SG) 'your child')
Others	The form *nya* is invariably used. (e.g., *anak nya* 'his/her/their child)

Demonstratives, when they head NPs, are glossed as ⟨dem_pro⟩. They exhibit three-fold distinctions, as shown below, based on the relative distance of a referent from the speaker and the addressee. Each category exhibits a distinction of the basic form and the nominalized form, which are derived with the form *dè*, which is the short

form of the relativiser *adè*. Roughly speaking, the basic form refers to a situation or manner, while the nominalized form refers to entities, that is, things or people.

Table 4 Forms of pronoun expressing possessors

Categories	Forms and examples
ta/*dèta*	nearer the speaker than the addressee
nan/*dènan* or *dèan*	nearer the addressee than the speaker
ana/*dena*	distant from both the speaker and the addressee

4.2.2. Zero

As mentioned in Section 3, the predicate is the only obligatory constituent in a Sumbawa clause. Arguments are often not overt when the referent is inferable from the discourse context. The symbol ⟨0⟩ is used to code unexpressed argument. Since core-arguments, as noted in Section 3, occur either before or after the predicate according to the information status, the symbol ⟨0⟩ are put in the beginning of the clauses as a rule. I will return to this point in 4.4.1, where I will discuss the core argument functions.

4.3. Animacy and person of referential expressions

The symbol ⟨d⟩, which marks anthopomorphized referent in the GRAID convention, is employed for animals when they are personified; it often happens in folktales, as seen in Sentence (17).

(17)

	laló	*gagak*		*ga...*	*ga...*	*%*	*léng*	*gagak*
	go	crow		ga	ga		words	crow
##	v:pred	np.d:s	#ds:pred	other	other	%	np:s	np.d:poss
		keliong	*batu*	*ta*	*ling*	*gagak.*		
		go.around	stone	this	by	crow		
##	0.h:a	v:pred	np:p	rn	adp	np.d:dt_a		

'The crow went. The crow said (*lit.* the crow's words are) 'Ga..., Ga....' The Crow flew around the stone.' (FS064)

4.4. Function of referential expressions

4.4.1. Core argument function

S argument may occur either before or after the predicate. It may be pro-cliticized as seen in sentences (12) and (14) above. A argument may occur only before the predicate. It also may be pro-cliticized as seen in sentence (13) above. As seen in sentences (3d), (5), and (10), a transitive agent may be rendered by the PP with the preposition *ling* 'by', which always occurs after the predicate; the function of the PP is glossed by the symbol ⟨dt_a⟩. It may co-occur with a clitic pronoun, as seen in sentence (3d).)

P argument may occur either before or after the predicate. It always occur in unmarked NPs, as seen in sentences in (4a) and (4b); it never cliticized to the verb.

4.4.2. Non-core argument function

The three symbols, namely ⟨g⟩: goal, ⟨l⟩: locative, and ⟨oblique⟩ are used to encode the non-core argument function.

4.4.2.1. Goal

In Sumbawa, a goal argument of a verb of motion occurs as a complement of the preposition *lakó* 'to' or *kó* 'to'.

(18)

		ya=mólé	*lakó*	*tana*	*Makasar.*
		FUT=return	to	land	Makasar
##	0.h:s	v:pred	adp	np:g	rn

'(He) will return to the land of Makasar.' (TM058)

(19)

	ba	*silamo*	*tu=lalo*	*kó*	*bao*	*bangka*	*kita*
	ITJ	please	1PL=go	to	above	ship	1PL
##	other	other	pro.1:s=v:pred	adp	np:g	rn	pro.1:poss

ta,	*bangka*	*kami*	*ta.*
DEM_P	ship	1PL	DEM_P
rn	np:appos	pro.1:poss	rn

'Let's go onto our ship.' (history069)

A recipient argument of a verb of transfer also may occur as a complement of the preposition lako 'to', and therefore glossed by the symbol ⟨g⟩.

(20)

		lalo	*mo*			*antat*	*lala*	*ta*	*lakó*
		go	DM			bring	princess	DEM_P	to
##	0.h:s	v:pred	other	##	0.h:a	v:pred	np.h:p	rn	adp

kerajaan.
kingdom
np:g

'(He) went and took the princess to the kingdom.' (TM050)

(21)

	mudi	*nó*	*sia=bau'*				*antat*	*mè*	*nan*
	later	NEG	2SG.HON=can				bring	rice	DEM_M
##	other	lv	pro.2:a=v:pred	#cc	0.2:a		v:pred	np:p	rn

kó	*kami*	*ta.*
to	1PL	DEM_P
adp	pro.1:g	rn

'Later you cannot bring the rice to us.' (FS: 022)

As for the verb *bèang* 'give', two patterns are found in our corpus; a recipient argument occurs either in an unmarked NP, as in Sentence (22) or as a complement of the preposition *lakó* 'to', as in Sentence (23). In Sentence (22), the recipient NP is

glossed by the symbol ⟨p⟩ and the theme NP is glossed by the symbol ⟨p2⟩, based on Haig and Schnell (2014: 16).

(22)

	nampo	*ku=bau*			*bèang*	*nènè*	*mè.*
	then	1SG=can			give	2PL	rice
##	other	pro.1:a=v:pred	#cc	0.1:a	v:pred	pro.2:p	np:p2

'Then I can give you(PL.) rice.' (fl: 015)

(23)

	lamin	*kalèpè*	*bèang*	*ijin*	*lakó*	*kajulin*	*lah.*
	if	2SG.HON.(special)	give	permission	to	1SG.HBL	ITJ
#ac	other	pro.2:a	v:pred	np:p	adp	pro.1:g	other

'If you give the permission to me, you know.' (TM027)

An addressee of the speech is expressed by the P argument of the verb *bada* 'tell' in an unmarked NP as seen in Sentence (24).[6] In such an example, the addressee NP is glossed by the symbol ⟨p⟩ and the theme NP is glossed by the symbol ⟨p2⟩, based on Haig and Schnell (2014: 16).

(24)

	ta	*luk*	*anak*	*ta*	*è,*	
	DEM_P	way	child	DEM_P	ITJ	
##ds	dem_pro:s	np:pred	np.h:voc	rn	other	##

ma=mu=bada	*aku*	*tegas*	*ling*	*kau*	*nan.*
let=2SG=tell	1SG	meaning	by	2SG	DEM_M
pro.2:a=v:pred	pro.1:p	np:p2	adp	pro.2:dt_a	rn

'This is the way, boy, please tell me the meaning (*lit.* (I) let you tell me the meaning.)' (flood)

4.4.2.2. Locative

A locative argument typically occurs as a complement of the preposition *pang* 'at,

[6] In other source, such as Jonker (1934), however, examples of the verb beling 'say', in which an addressee constituent occurs as a complement of the preposition *lakó* 'to' or *ké* 'with' are found, as in (i) and (ii), respectively. Such constituents can be considered as goal arguments, which should be glossed by the symbol ⟨g⟩.

(i) a.

	dadi	*beling*	*Salam*	*lakó*	*guru*		*mè*	*ké*	*gula*	*ké*	*nyir*
	then	say	Salam	to	teacher		rice	or	sugar	or	coconut
#	other	v:pred	np.h:s	adp	np.h:g	#ds	np:p	other	rn	rn	rn

tu=bèang.
1PL=give
pro.1:a=v:pred

'Then Salam said to (his) teacher, 'Shall we give rice, sugar or coconut (to the dog)?' (Jonker 1934: 214)

b.

	dadi	*beling*	*nya*	*Salam*	*ké*	*sowai*		*è*	*lalo*	*basió*
	then	say	Mr.	Salam	with	wife		ITJ	go	hide.onself
#	other	v:pred	ln	np.h:s	adp	np.h:g	##ds	other	v:other	v:pred

kau	*lèma.*
2SG	quickly
pro.2:s	other

'Then Salam said to (his) wife, 'Ah, go and hide yourself quickly'.' (Jonker: 222)

in', as seen in sentences (25) and (26) or as an NP attached with nasal prefix 'N-' that is homorganic to the initial sound of the NP, as seen in sentence (27). It also includes a source, which occurs as a complement of the preposition *kaling* or *kalis* 'from', as seen in sentence (28) and (29).

(25)

	ada	*mo*	*sopo*	*kerajaan*	*pang*	*tana*	*Samawa*	*ta.*
	exist	DM	one	kingdom	at	land	Sumbawa	DEM_P
##	v:predex	other	np:s	np:rn	adp	np:l	rn	rn

'There was a kingdom in Sumbawa.' (TM002)

(26)

	anak	*ya=satokal*	*pang*	*bao*	*Batu*	*Langlelo*	*ta.*
	child	3=sit	at	above	stone	Langlelo	DEM_P
##	np.h:p	pro.h:a=v:pred	adp	np:l	rn	rn	rn

'She sat the children onto the Langlelo Stone.' (fl007)

(27)

		óló'	*mo*	*n=dalam*	*cowèk*	*mè*	*ta.*
		put	DM	at=inside	clay.dish	rice	DEM_P
##	0.h:a	v:pred	other	adp=np:l	rn	np:p	rn

'(She) put the rice in the clay dish.' (fl067)

(28)

	tu=ètè	*pènèk*	%	*tu=ètè*	*pènèk*	%	*datang*	*mo*
	1PL=take	short		1PL=take	short	%	come	dm
##ac	pro.1:s=v:pred	rn	%	nc	nc	%	v:pred	other

sandro	*kaling*	*ano-siup*		*adè*	*ya*[7]*=saterang*	*lala*
doctor	from	east		rel	FUT=cure	princess
np.h:s	adp	np:l	#rc	0.h:a other	v:pred	np.h:p

ta.
DEM_P
rn

'To put it briefly, the doctor who are going to cure the princess came from the east.' (TM017)

(29)

	lis	*mo*	*ina*	*nya*	*Lalu*	*KerékKuré*	*ta*	*é*
	go.out	DM	mother	Mr.	Prince	KurekKure	DEM_P	ITJ
##	v:pred	other	np.h:s	ln	ln	np.h:poss	rn	other

kalis	*dalam*	*balé*	*ta.*
from	inside	house	DEM_P
adp	np:l	np:poss	rn

'The mother of Prince KurekKure went out from the house.' (KK: 063)

4.4.2.3. Oblique

In Sumbawa corpus, instrumental and commutative, both of which occur as a complement of the preposition *ké* 'with' are glossed by this symbol.

[7] The clitic *ya=* 'FUT' is homonymous to the third person clitic *ya=*, which was introduced in section 4.2.1.

(30)

	tapi			*lamin*	*salaki*	*%*		*ku=sukat*	
	but			if	male			1SG=marry	
##	other	#ac	0.h:s	other	male	%	0.h:p	pro.1:a=v:pred	
ké	*anak*	*kaku.*							
with	child	1SG.POSS							
adp	np:obl	pro.1:poss							

'But if (he is) male, I will marry (him) with my child.' (TM013)

(31)

		ya=paning	*ké*	*ai'.*
		3=give.shower	with	water
##	0.h:p	pro.h:a=v:pred	adp	np:obl

'(He) gave (her) a shower with water.' (TM042)

(32)

	ai	*ta*	*bacampir*	*ké*	*geti.*
	water	DEM_P	mixed	with	blood
##	np:s	rn	v:pred	adp	np:obl

'The water was mixed with blood.' (TM043)

4.4.3. Dislocated topic

As mentioned in 4.4.1, the PP with the preposition *ling* 'by' indicating an agent is treated as a dislocated topic. (See example (3d).) In addition to that, a dislocated topic may correspond to various relations expressed in the clause that follows. In example (36), the dislocated topic *pantèk* is co-referential to the S argument *nya*.

(33)

	pantèk	*ta,*	*nya*	*singin*	*colo'*	*tau*	*dununng'.*
	flint	this	3	name	match	person	before
##	np:dt_s	rn	dem_pro:s	np:pred	np:poss	np:poss	rn

'Pantek, it is (*lit.* the name of) flint of people long time ago.' (FS032)

The dislocated topic may corresponds to the possessor of the referent of the argument in the clause that follows, as seen in Sentences (34) and (35); the possessor NP of the existential verb *ada* 'exist' almost always occurs in this position.

(34)

	raja	*ta*	*ada*	*sópó*	*anak*	*dadara.*
	king	DEM_P	exist	one	child	girl
##	np.h:dt	rn	v:predex	np.h:s	rn	rn

'The king had a daughter.' (TM004)

(35)

	tapi	*dèan*	*cowèk*	*singin*	*sebab*	*ka=tana'.*
	but	DEM_P	eartheware.dish	name	because	PST=clay
##	other	np:dt	np:pred	np:s	other 0:s	aux=np:pred

'But as for that, "cowek" is the name (*lit.* its name), because it is made of clay.' (FS056)

In Sentence (36), the dislocated topic corresponds to the S argument, *anak nan* 'the child', which is S argument of a complement clause that in turn functions as Argument P of the adverbial clause.

(36)

	dadi	*lala*	*ta*			*ling*	*to'*	*ling*
	then	princess	DEM_P			because	know	by
##	other	dt	rn	#ac	0.h:a	other	v:pred	adp
raja		*ta*		*anak*	*nan*	*sakit,*		
king		DEM_P		child	DEM_M	sick		
np.h:dt_a		rn	#cc	np.h:s	rn	v:pred	##	0.h:a
ya=utus			*mo*	*pengawa*			*ya=umumkan*	*lakó*
3=order			dm	vassal			FUT=announce	to
pro.h:a=v:pred			other	np.h:p	#cc	0.h:s	v:pred	adp
rakyat.								
people								
np.h:g								

'Then the princess, as the king knew that the daughter was ill . . .' (TM009)

5. Predicates

The predicate constituents vary depending on whether it is a verbal predicate or a non-verbal predicate. In Sumbawa, the noun and verb as word category are distinguished by whether they can co-occur with a clitic pronoun. For example, the word *gera* 'beautiful' may be attached to the clitic pronoun, as in the sentence *mu=gera* [2SG=beautiful] 'you are beautiful', while the word guru 'teacher' may not be attached to the clitic pronoun, thus, the sentence *mu=guru* [2SG=teacher] '(intended meaning) You are a teacher.' is not accepted by the speakers. On the basis of this distinction, the word *gera* is classified into the verb, while the word *gera* is classified into the noun.

5.1. Verbal predicates

Verbal predicates have a lexical verb as the predicate head. (37) shows the structure of the predicate; only the verb as predicate head is an obligatory element in the predicate. Elements in parenthesis are optional constituents in a clause.

(37)

constituents:	(negator, TAM marker)	(pronominal clitic (A/S)=)	verb
symbols:	⟨lv⟩	⟨pro:a/s=⟩	⟨v:pred⟩

The negators are glossed by the symbol, ⟨lv⟩, i.e., 'subconstituent of the verb complex occurring to the left of the verbal head' (Haig and Schnell 2014: 9).

(38)

	dadi	*peno*	*mo*	*sandro*		*adè*		*datang,*
	then	many	DM	doctor		REL		come
##	other	v:pred	rv	np:s	#rc	other	0:s	v:pred

	serèa	*nó*	*bau*			*saterang*	*lala*	*ta.*
	all	NEG	can			cure	princess	DEM_P
##	np:s	lv	v:pred	#cc	0.h:a	v:pred	np:p	rn

'Then many doctors came, but all (of them) couldn't cure the princess.' (TM019)

The pronominal clitics for S and A are glossed by the symbols, ⟨pro:s=⟩ or ⟨pro:a=⟩ according to their function, as seen in 4.2. In addition to the constituents mentioned above, a discourse particle (DM) may occur after the first constituent of the predicate. They are glossed either by the symbol ⟨lv⟩ before the verb, as in Sentence (13) above, or by the symbol ⟨rv⟩ after the verb, as seen in sentence (38) above.

5.2. Non-verbal predicates

Non-verbal predicates may include an NP or PP as head, and co-occur with an NP argument in the clause, which is glossed by the symbol ⟨s⟩. Non-verbal predicates normally consist of only the predicate head, but they may include the negator *siong*, which is marked by the symbol ⟨lv⟩. Sentences (39) is an examples of the negator *siong*.

(39)

		tau	*Indonesia*	*si*			*tapi*	*siong*
		person	Indonesia	DM			but	NEG
##ds	0.h:s	np:pred	np:rn	rn	##ds.neg	0.h:s	other	ln

tau	*Samawa.*
person	Sumbawa
np:pred	np:rn

'(They) are Indonesians, but not Sumbawans.' (conversation)[8]

6. Complex sentences and direct speech

6.1. Relative clauses

When one of the core arguments are relativised, relative clauses are introduced by a relativiser *adè*, as seen in Sentence (40) and (41).

(40)

	dadi	*peno*	*mo*	*sandro*			*adè*	*datang,*
	then	many	DM	doctor			REL	come
##	other	v:pred	rv	np:s	#rc	0:s	v:pred	

	serèa	*nó*	*bau*			*saterang*	*lala*	*ta.*
	all	NEG	can			cure	princess	DEM_P
##	np:s	lv	v:pred	#cc	0.h:a	v:pred	np:p	rn

'Then many doctors came, but all (of them) couldn't cure the princess.' (TM019)

[8] This is an example cited from the video in which conversation is recorded, which is mentioned in section 2.

(41)

	dadi	*pengawal*	*adè*	*baèng*	*ka=gita'*	*karing*	*sisi*
	then	guard	REL	PFT	PST=see	from	side
##	other	np:s #rc	other	lv	v:pred	adp	np:l

tampar	%	*kamelas,*
seashore		be.surprised
np:poss	%	v:pred

		ling	*gita'*		*tau*	*ta*	*dadi*	*tau*
		because	see		person	DEM_P	become	person
#ac	0.h:a	other	v:pred	#cc	np:s	rn	v:pred	np.h:other

gera	*kapasir.*
beautiful	clean
rn	rn

'Then the guard who had seen from the seashore was surprised, as he saw the person become a handsome and clean guy.' (TM071)

When the head nouns are nouns such as *pang* 'place', *ling* 'words, what is said', *seda* 'voice', *muntu* 'time', *luk* 'manner, way', a relative clause directly modifies the head noun without a relativiser.

(42)

	mé	*pang'*		*tedu*	*kau.*
	which	place		stay	2SG
##	np:pred	np:s #rc	0:l	v:pred	pro.2:s

'Where do you live? (*lit.* Which is the place you live?)' (TM030)

(43)

	Batu	*nampar*	*Batu=Langléló*	*ta*	*pang'*
	stone	flat	stone=Langlelo	DEM_P	place
##	np:s	rn	np:appos	rn	np:appos #rc_rn

ka=sia=satokal	*kami*	%	*ta,*
PST=2SG=make.sit	1PL.EXCL		DEM_P
pro.2:a=v:pred	pro.1:p	%	rn

pang'		*kami*	*bakedèk*	%	*ta,*	*narang*	*lè*
place		1PL.EXCL	play		DEM_P	the.more	long
np:appos	#rc_rn	pro.1:s	v:pred	%	rn	other	other

narang	*tingi.*
the.more	high
other	v:pred

'The flat stone, the Langlelo stone, the place onto which you made us sit, the place on which we are playing is getting higher.' (fl: 017–019)

(44)

	nan	*nya*	*seda*			*beling.*
	DEM_M	3	sound			say
##	dem_pro:s	pro:appos	np:pred	##rc	0.h:s	v:pred

'That was the voice (he) said.' (FS68)

(45)

		ao'	*anak*	*é*	*ta*	*po*	*muntu*	
		ITJ	child	ITJ	DEM_P	DM	time	
##	0.1:s	other	np:voc	other	dem_pro:pred	rn	np:s	#rc:rn

ku=nepé.
1SG=winnow
pro.1:s=v:pred
'Ok, children, just now (finally) I am winnowing rice.' (FS023)

The relativiser *adè* may introduce a so-called headless relative clause (e.g., *adè datang nan* [REL come this] 'the one coming', which is glossed by the symbol ⟨#rc⟩ and is assigned the function, as in ⟨#rc:s⟩, in example (46), following Haig and Schnell (2014: 24).

(46)

		adè	*terahir*	*maning'*	%	*ta*	*dèan,*	
		REL	finaly	take.shower		DEM_P	DEM_M	
##	#rc:s	other	other	v:pred	%	rn	dem_pro:s	
			dè	*paling*	*balong.*			
#rc:appos			REL	most	good			
#rc:appos	0.h:s	other	other	v:pred				

'That, the one who was most beautiful was the one who took the shower at last.' (LK032)

6.2. Complement clauses

In examples (47) and (48), the complement clause occurs after the main verb, such as *gita*' 'see' or *to*' 'know' without any complementiser. In the examples given below, the complement clauses correspond to Argument P semantically and are tentatively glossed by the symbol p (thus ⟨#cc:p⟩), though we do not have enough examples to observe their syntactic function. (See Haig and Schnell (2014: 47) for a discussion on the grammatical function of complement clauses.)

(47)

		ina'	*ina'*	*ina'*	*né*	*sia=gita'*				*mo*
		mother	mother	mother	ITJ	2SG=see				DM
##	np.h:voc	np.h:voc	np.h:voc	other	pro.2:a=v:pred					rv
		batu	*ta*	*narang*	*lè'*	*narang*	*tingi*	< ··· >		
		stone	DEM_P	the.more	long	the.more	high			
#cc:p	np:p	rn	other	other	other	v:pred				

'Mother, mother, mother, please see that the stone is growing up ...' (FS049)

(48)

		dadi	*lala*	*ta*			*ling*	*to'*	*ling*	
		then	princess	DEM_P			because	know	by	
##	other	dt	rn	#ac	0.h:a		other	v:pred	adp	
raja		*ta*		*anak*	*nan*	*sakit,*				
king		DEM_P		child	DEM_M	sick				
np.h:dt_a	rn	#cc:p	np.h:s	rn	v:pred	##	0.h:a			

ya=utus	*mo*	*pengawa*			*ya=umumkan*	*lakó*
3=order	dm	vassal			FUT=announce	to
pro.h:a=v:pred	other	np.h:p	#cc	0.h:s	v:pred	adp

rakyat.
people
np.h:g

'Then the princess, as the king knew that the daughter was ill ...' (TM009)

6.3. Adverbial clauses

Adverbial clauses are glossed with the symbol ⟨#ac⟩. In our corpus, the clause that is introduced by a conjunction that indicates its semantic relation to the adjacent main clause is glossed as an adverbial clause; adverbial clauses cannot be distinguished from the main clause by their morpho-syntactic features.

Sentences (49) and (50) are examples of adverbial clauses indicating a time relation between the two events expressed.

(49)

	ina'	*ina'*		*ènèng*	*mè*	*gama*
	mother	mother		ask.for	rice	please
#	np.h:voc	np.h:voc	0.1:a	0.h:a=v:pred	np:p	other

ina'	*muntu*	*nó.poka'*	*tingi*	*batu=Langléló'.*
mother	when	not.yet	high	stone=Langlelo
np.h:voc	#ac	other	lv	v:pred	np:s

'Mother, mother, give (us) rice, please, when the stone has not got high, yet.' (FS026)

(50)

	tapi		*lamin*	*salaki*	*%*		*ku=sukat*	*ké*
	but		if	male			1SG=marry	with
#	other	#ac	0.h:s	other	male	%	0.h:p	pro.1:a=v:pred	adp

anak	*kaku.*
child	1SG.POSS
np:obl	pro.1:poss

'But if (he is) male, I will marry (him) with my child.' (TM013)

Sentence (51) is an example of an adverbial clause of reason, while Sentence (52) is an example of an adverbial clause of condition.

(51)

	apa	*lagi*	*datu*	*ta*	*makin*	*ka=susa*		*ling*
	what	more	king	DEM_P	the.more	PST=suffer		because
#	other	other	np.h:s	rn	other	v:pred	#ac	other

lala	*ta*	*baè*	*anak*	*soai*	*datu*	*ta.*
princess	DEM_P	only	child	female	king	DEM_P
np:s	rn	other	np.h:pred	rn	np.h:poss	rn

'Needless to say, the king's agony increased, because the princess was the king's only daughter.' (TM022)

(52)

	lamén	*ka*	*mo*	*tingi*	*batu=Langléló'*	*%*	*mudi*	*nó*	
	if	PST	DM	high	stone=Langléló'		later	NEG	
##	#ac	other	lv	lv	v:pred	ln=np:s	%	other	lv

sia=bau'			*antat*	*mè*	*nan*	*kó'*	*kami*
2SG.HON=can			bring	rice	DEM_M	to	2PL
pro.2:s=v:pred	#cc	0.2:a	v:pred	np:p	rn	adp	pro.1

ta.
DEM_M
rn

'If the Langlelo stone got high, you cannot bring the rice to us later.' (FS021–022)

6.4. Direct speech

Direct speeches are glossed by the symbol ⟨#ds⟩. Typically, a direct speech clause follows the verb *beling* 'say', as in Sentence (53)

(53)

	karing	*beling*	*mo*	*anak*			*ina*	*ina*
	then	say	DM	child			mother	mother
#	other	v:pred	other	np:s	#ds	0.h:a	np.h:voc	np.h:voc

ènèng	*mè*	*gama*	*ina.*
ask.for	rice	please	mother
v:pred	np:p	other	np:voc

'Then the child said, 'Mother, Mother, I beg rice, please, Mother.'

A direct speech clause may occur as a predicate of non-verbal clause, in which the NP headed by the noun *ling* 'words, what is said' is S argument.

(54)

		kajulin=tedu	*pang*	*sisi*	*olat*	*anosiup*	*ana*	
		1SG.HON=stay	at	side	mountain	east	DEM_D	
##	#ds_pred	pro.1:s=v:pred	adp	np:l	np:poss	rn	rn	

<ling	*sa>*[9]	*%*	*ling*			*samong*	*ling*	*tau*
<words	?>		words			answer	by	person
nc	nc	%	np:s	#rc	0.p	0.h:a	v:pred	adp	np.h:dt_a

loka	*ta.*
old	DEM_P
rn	rn

'I stay at the mountain side of the east', the old man answered. (*lit.* The words the old man answered was 'I stay at the mountain side of the east.' (TM030–032)

(55)

	tapi			*lamin*	*salaki*[10]	*%*	*ku=sukat*	*ké*
	but			if	male		1SG=marry	with
#	other	#ac	0.h:s	other	v:pred	%	pro.1:a=v:pred	adp

[9] It seems that this is a false start and we do not consider this part. This type of constituent is glossed by the symbol ⟨nc⟩ 'not considered' in the GRAID convention.

anak *kaku.*
child 1sg.poss
np:obl pro.1:poss
'But if (he is) male, I will marry (him) with my child.' (TM013)

(56)

		dadi,	*ao,*	*ba*	*mu=tedu*	*ninta*
		then	itj	itj	2sg=stay	here
##	#ds_pred	other	other	other	pro.2:s=v:pred	dem_pro:l

mo, *ling.*
dm words
other np:s
'Then (she) said, 'you just stay here. (*lit.* Then (her) words were 'you just stay here.')" (KK: 062)

In some clauses, a direct speech clause occurs without a quotative frame. Sentence (57)(=Sentence (12) given above) is a mother's reply to what her son said in Sentence (52) given above. Here, two direct speeches are simply juxtaposed like lines in a drama.

(57)

	ao'	*anak*	*é*	*ta*	*po*	*muntu*	
	itj	child	itj	dem_p	dm	time	
## 0.1:s	other	np:voc	other	dem_pro:pred	rn	np:s	#rc:rn

ku=nepé.
1sg=winnow
pro.1:s=v:pred
'Ok, children, just now (finally) I am winnowing rice.' (FS023)

6.5. Coordination

Each coordinated clause is treated as an independent clause, and therefore glossed by the symbol ⟨##⟩ in our corpus. It includes clauses either simply juxtaposed without any conjunctions, which often indicate a series of events in a narrative, as seen in Sentence (3) given above, or clauses co-ordinated by what Haig and Schnell (2014: 24) call 'neutral' coordinators, such as *karing* 'and then', *dadi* 'and then', and *tapi* 'but'. These conjunctions indicate a semantic relation not only to a specific adjacent clause, but also to the situation expressed by a series of clauses or a situation not explicitly mentioned. We do not observe any formal dependence between the two clauses linked by such coordinators, which Haig and Schnell (2014) mention as a possibility. Sentence (58) is an example of *tapi* 'but'. Sentence (4e) is an example of *karing* 'and'.

(58) a.

	cowèk	*singin*	*talang*	*tau*	*dunung'.*
	cowek	name	dish	people	before
##	np:s	np:pred	rn	rn	rn

[10] The word *salaki* belongs to the word class of verb in Sumbawa.

b. *cowèk* *nan* *ka=tana',* *ka=tana'.*
cowek DEM_M PST=clay PST=clay
np:s rn np:pred np:appos

c. *jadi* *samacam* *talang,* *nan.si* *samacam* *piring.*
then a.kind.of dish that.is a.kind.of dish
0:s other other np:pred other other np:appos

d. ***tapi*** *dèan* *cowèk* *singin* *sebab*
but DEM_P eartheware.dish name because
other np:dt_poss np:pred np:s #ac other 0:s
ka=tana'.
PST=clay
np:pred

'*Cowek* is a name of dish of people of long time ago. *Cowek* is made of clay. So, (it is) a kind of dish (*talang*), that is, a kind of dish (*piringè*). But as for that, "cowek" is the name (*lit.* its name), because it is made of clay.' (FS053–056)

7. Final remarks

In this paper, I have given a progress report on an on-going project of developing a Sumbawa annotated-spoken corpus. After a brief summary of the language, the project, and the nature of the data included in the corpus in Sections 1 and 2, an annotation note is provided that explains how the GRAID annotation system is applied to the Sumbawa corpus. First, we survey the clause types and structures in Section 3. Sumbawa clauses may be grouped into two categories: the verbal clause and the non-verbal clause. The predicate is the only obligatory constituent in each clause, and the arguments, if any, may follow the predicate, or only one argument may precede the predicate. In Sections 4 and 5, we examine how referential expressions and the constituents of the predicates can be glossed by GRAID conventions, respectively. In Section 6, we study the types of subordinate clauses and how these clauses and direct speech are glossed by GRAID conventions.

Abbreviations

1, 2, 3	the 1st, 2nd, 3rd person	ITJ	interjection
DEM_D	distal demonstrative	ITR	interrogative
DEM_M	medial demonstrative	NEG	negation
DEM_P	proximal demonstrative	PL	plural
EXCL	exclusive	POSS	possessive
FUT	future	PST	past
HBL	the humble form (of the 1st person pronoun)	DM	discourse marker
		RED	reduplication
HON	honorific	SG	singular

References

Adelaar, Alexander. 2005. "Malayo-Sumbawan". *Oceanic Linguistics* 44(2). pp.357–388.

Du Bois, John W. 1987. "The discourse basis of ergativity". *Language* 63(4). pp.805–855.

Haig, Geoffrey and Stefan Schnell. 2014. *Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators (version 7.0)*. (https://lac.uni-koeln.de/en/multicast/) (accessed 2018-01-14)

Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

Mahsun. 1999. *Variasi Dialektal Bahasa Sumbawa-Kajian Dialektologi Diakronis [Dialect Variation in the Sumbawa Language - A Study of Diachronic Dialectology]*. Unpublished manuscript, University of Mataram, Lombok.

Mbete, Aaron Neko. 1990. *Rekonstruksi protobahasa Bali-Sasak-Sumbawa [Reconstruction of the proto-language Bali-Sasak-Sumbawa]*. PhD thesis. Universitas Indonesia, Jakarta.

Schiborr, Nils N. 2016. "Multi-CAST corpus overview and description". In Geoffrey Haig and Stefan Schnell (eds.) *Multi-CAST (Multilingual Corpus of Annotated Spoken Texts)*.
https://www.uni-bamberg.de/fileadmin/aspra/Multi-CAST_corpus-overview.pdf (Accessed 2018-01-14)

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.

Jonker, J. C. G. 1934. "Soembawareesche Teksten met Vertaling [Sumbawa text with translation]". *Bijdragen tot de Taal-, Lan- en Volkenkunde* 92(1). pp.211–335.