

Multi-CAST

Sumbawa *annotation notes*

Asako Shiohara

November 2022
v1.0



ARC CENTRE OF EXCELLENCE FOR
THE DYNAMICS OF LANGUAGE



Australian Government
Australian Research Council



University of Bamberg

DFG

Multi-CAST

*Multilingual Corpus of
Annotated Spoken Texts*

Citation for this document

Shiohara, Asako. 2022. Multi-CAST Sumbawa annotation notes. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/#sumbawa) (date accessed)

Citation for the Multi-CAST collection

Haig, Geoffrey & Schnell, Stefan (eds.). 2015. *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/) (date accessed)

The Multi-CAST collection has been archived at the *University of Bamberg*, Germany, and is freely accessible online at multicast.aspra.uni-bamberg.de/.

The entirety of Multi-CAST, including this document, is published under the *Creative Commons Attribution 4.0 International Licence* (CC BY 4.0), unless noted otherwise. The licence can be reviewed online at creativecommons.org/licenses/by/4.0/.

Multi-CAST Sumbawa annotation notes v1.0 last updated 8 November 2022
This document was typeset by NNS with $\text{X}_{\text{L}}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and the *multicast3* class (v3.2.6).

Contents

1	Notes on the GRAID annotations	1
1.1	Introduction	1
1.2	Referential expressions	1
1.2.1	Pronouns	1
1.2.2	(Forced) zero arguments	1
1.3	Clause structure	2
1.3.1	Intransitive clauses	2
1.3.2	Transitive clauses	3
1.3.3	Auxiliary verbs and serial verbs	4
1.4	Syntactic functions	5
1.4.1	A in <i>ling</i> -constructions: ⟨:a_ling⟩	5
1.4.2	Dislocated topics: ⟨:dt⟩	5
1.4.3	Secondary objects in ditransitive constructions: ⟨:p2⟩	5
1.5	Direct speech	6
2	Notes on the RefIND annotations	6
2.1	Referents in clauses otherwise not considered	6
	References	7
	Appendices	8
A	List of corpus-specific GRAID symbols	8
B	List of abbreviated morphological glosses	9

1 Notes on the GRAID annotations

This document contains notes on the implementation of the GRAID (Haig & Schnell 2014) and RefIND (Schiborr et al. 2018) annotation conventions in the Multi-CAST Sumbawa corpus. It corresponds to version 2211 of the annotations, published in November 2022. Unless a more recent version of this document exists, it also applies to any later versions of the annotations.

1.1 Introduction

Sumbawa (indigenous designation: Samawa) 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 (*Nusa Tenggara Barat*). Sumbawa belongs to the Bali-Sasak-Sumbawa subgroup of the Malayo-Polynesian branch of the Austronesian language family (Adelaar 2005; Mbetse 1990). Among the several dialects of Sumbawa language, the dialect spoken in the texts in this corpus is classified as the Sumbawa Besar dialect, which is distributed in a large area of the western part of the Sumbawa-spoken area (Mahsun 1999).

The transcriptions used for this corpus are largely identical to the IPA, which only the following symbols being different: *e* [ə], *è* [ɛ], *e* [é], *o* [ɔ], *ó* [o], *ny* [ɲ], and *ng* [ŋ]. These follow the orthography and convention of Indonesian, except *è* [ɛ] and *ó* [o], which do not appear in Indonesian. Some lexemes have heavy stress, the presence or absence of which is involved in phonemic distinction. Such stress is indicated by an apostrophe [']; compare *tó* ‘know’ and *tó'* ‘now’.

1.2 Referential expressions

1.2.1 Pronouns

Sumbawa has both free and bound person markers. The third person clitic occurs only with a transitive verb. Table 1 shows the forms of the person markers. The bound clitic *tu=* sometimes refers to general referent.

	<i>free pronoun</i>	<i>clitic pronoun</i>
1SG	<i>aku</i>	<i>ku</i>
1PL.INCL	<i>kita</i>	<i>tu</i>
1PL.EXCL	<i>kami</i>	<i>tu</i>
2SG	<i>kau</i>	<i>mu</i>
2PL	<i>nènè</i>	<i>nènè</i>
3	<i>nya</i>	<i>ya</i>

Table 1 Free and clitic pronouns in Sumbawa. The third-person clitic pronoun *ya* only occurs in transitive clauses.

1.2.2 (Forced) zero arguments

In the relativization of core arguments, which are normally introduced by the relativizer *adè* or its short form *dè*, the constituent that is co-referential with the relativized NP cannot occur within the relative clause, as in (1). These elements are annotated as ⟨f∅⟩ ‘forced zero’, so as to be distinguishable from the regular ⟨∅⟩ zero annotation for a discourse-retrievable entity.

- (1)
- nó bau sentèk periuk tódé dè datang ètè ai ta nan*

```

    nó   bau   sentèk   periuk   tódé
    NEG  can   put.on   bucket   child
#neg lv  lv_v  v:pred  np:p   np.h:s
           0079  0077

           dè   datang   ètè   ai   =ta   nan
           REL  come    take  water =DEM_P DEM_M
#rc f0.h:a other lv_v  v:pred np:p =rn_dem % rn_dem %
           0077           0075

```

‘The girl who came to take water couldn’t put the bucket (on her head).’

[mc_sumbawa_kerekkure_0202]

1.3 Clause structure

The predicate is the only obligatory constituent in Sumbawa clauses. Any types of referential expressions may be omitted when the argument is discourse-retrievable, as seen in (2).

- (2)
- tedu pang Dadap bakeban tanam ragan.*

```

    tedu   pang   Dadap
    live  at    Dadap
# 0.h:s v:pred adp pn_np:1 %
    0001           0004

```

```

    bakeban
    farm
# 0.h:s v:pred %
    0001

```

```

    tanam   ragan
    plant  ragam_flower
# 0.h:a v:pred np:p %
    0001           0007

```

‘(He) lived in Dadap. (He) farmed, planting ragam flowers.’

[mc_sumbawa_kerekkure_0011]

1.3.1 Intransitive clauses

In intransitive clauses, the S argument can be encoded by either bare NPs headed by a lexical noun or free pronoun, or by bound pronouns (pronominal clitics). There are two intransitive constructions that takes non-S bare NP; one is a possessive construction with the existential verb *ada* ‘exist’, and the other is a construction with copular verbs such as *dadi* ‘become’ and *basingin* ‘named’. The former is exemplified by (3), in which the possessum is marked as ⟨:s⟩ while the possessor is annotated as a dislocated topic ⟨:dt⟩. The latter is exemplified by (4), in which the domain of the attribute (e.g. the name holder) is marked as ⟨:s⟩ and the constituent representing the attribute (e.g. the name) as ⟨:other⟩. There is no “genuine” copular such as English *be* in Sumbawa. The equational sentence is formed via the juxtaposition of two noun phrases.

- (3)
- dè supu ta enda lamong.*

	<i>dè</i>	<i>supu</i>	<i>=ta</i>	<i>enda</i>	<i>lamong</i>	
	REL	last-born	=DEM_P	NEG.exist	clothes	
#neg	#rc.h:dt	f0.h:s	other	v:pred	% =rn_dem	v:predex np:s %
		0020				0021

‘The youngest one didn’t have clothes.’ [mc_sumbawa_kerekkure_0062]

- (4)
- ada sópó tau basingin Lalu Kerèk-kurè.*

<i>ada</i>	<i>sópó</i>	<i>tau</i>		<i>basingin</i>	<i>Lalu</i>	<i>Kerèk-kurè</i>
exist	one	person		named	prince	Kerek-kure
#	v:predex	ln	np.h:s	#rc	f0.h:s	v:pred ln pn_np:other % %
			0001		0001	

‘There was a person named Lalu Kerèk-kurè.’ [mc_sumbawa_kerekkure_0003]

In anti-passive constructions in which the verb is marked with the nasal prefix, the original P is left implicit, and the original A is encoded as S. (5) exemplifies this.

- (5)
- samong ling ina, na medi anak, ta muntu ku nuja*

<i>samong</i>	<i>ling</i>	<i>ina</i>		<i>na</i>	<i>mudi</i>	<i>anak</i>	<i>ta</i>	<i>muntu</i>
reply	by	mother		ITJ	later	child	DEM_P	time
#	v:pred	adp	np.h:s_ds_ling	#ds	other	other	np.h:voc	dem_pro:s np:pred
			0006				0002	

<i>ku=</i>	<i>nuja</i>	
1SG=	ANTIP.pound	
#ds_rc	=pro.1:s	v:pred % %
	0006	

‘The mother replied “Well, later, kids, now I am ponding rice. (lit. this is the time I pound)”’ [mc_sumbawa_langlelo_0012]

1.3.2 Transitive clauses

The P argument typically occurs in the post-predicate position, encoded by either bare NPs headed by a lexical noun or a free pronoun, or by bound pronouns (pronominal enclitics).

The A argument is encoded by either bare NPs headed by a lexical noun or a free pronoun, or by bound pronouns (pronominal proclitics) except in the construction mentioned below. There is generally no cross-reference in a Sumbawa clause, such as the three main clauses in (2) above. One exceptional construction is the ‘*ling*-construction’, in which the A argument occurs in the post-predicate position with the preposition *ling* ‘by’. In this case, a pronominal proclitic on the verb may cross-reference the A argument, as in (6). We have decided to annotate the A argument in this construction as the main A argument with ⟨:a_ling⟩, and to annotate the proclitic with underscore notation, e.g. ⟨=pro_h_a⟩. It is also possible that no cross-reference shows up in a clause with the *ling*-construction. The A argument after the preposition *ling* in this scenario is also annotated as ⟨:a_ling⟩, like in (7), with no additional zero argument annotated ahead of the verb.

- (6)
- sehingga ano yang keempat, ya=èntap ling Lalu Kerèk-kurè ta, sai tau bau ragan ta.*

<i>sehingga</i>	<i>ano</i>	<i>yang</i>	<i>keempat</i>	<i>ya=</i>	<i>èntap</i>	<i>ling</i>	<i>Lalu</i>	<i>Kerèk-kurè</i>
then	day	REL	fourth	3=	peep	by	prince	Kerek-kure
# other	other	other	other	=pro_h_a	v:pred	adp	ln	pn_np.h:a_ling
				0001				0001

<i>=ta</i>	<i>sai</i>	<i>tau</i>		<i>bau</i>	<i>ragan</i>	<i>=ta</i>		
=DEM_P	who	person		get	k.o.flower	=DEM_P		
=rn_dem	#cc:p	np:pred	np.h:s	#rc	f0.h:a	v:pred	np:p	=rn_dem % % %
			0010		0010		0007	

‘Then on the fourth day, Lalu Kerèk-kurè too a peep (to learn) who got the flowers.’

[mc_sumbawa_kerekkure_0020]

- (7)
- karing mentrènè ta, ètè ling nya.*

<i>karing</i>	<i>mentrènè</i>	<i>=ta</i>	<i>ètè</i>	<i>ling</i>	<i>nya</i>
then	termite	=DEM_P	take	by	3
# other	np.n:p	=rn_dem	v:pred	adp	pro.h:a_ling %
	0004				0001

‘So he took the termites.’

[mc_sumbawa_nuntut_0008]

It is safe to say that there is no typical or canonical agreement on Sumbawa verbs. Pronominal clitics, as well as any other person markers, do not occur obligatorily, and the occurrence of above-mentioned cross-reference is pragmatically conditioned.

1.3.3 Auxiliary verbs and serial verbs

In Sumbawa, two verbs can be strung together in a serial verb construction. Some of these can be analyzed as the combination of a modal auxiliary verb and a main verb, such as *bau saterang* ‘can cure’ in the relative clause in (8). Some of them, however, consist of a motion verb and a verb indicating the purpose, such as *lalo antat* ‘go take’ in (9). The first verb in both scenarios is annotated as <1v_v> and the second verb is annotated as the main verb <v:pred>, as the core arguments roles and encoding are determined based on the second verb.

- (8)
- isi pengumuman=nan ne sai-sai adè bau saterang anak kaku*

<i>isi</i>	<i>pengumuman</i>	<i>=nan</i>	<i>ne</i>		<i>sai-sai</i>
content	announcement	=DEM_M	ITJ		whoever
# np:s_ds	rn_np	=rn_dem	other	#ds_cc:pred	np.h:dt_p
	0006				0007

<i>adè</i>	<i>bau</i>	<i>saterang</i>	<i>anak</i>	<i>kaku</i>			
REL	can	cure	child	1SG.POSS			
#ds_rc	other	f0.h:a	v:other	v:pred	np.h:p	rn_pro.1:poss	% % %
		0007		0003	0002		

‘The content of the announcement was that “Whoever has cured my child,...”’

[mc_sumbawa_menangis_0010]

(9) *laló mo antat lala=ta lakó kerajaan*

	<i>laló</i>	<i>mo</i>	<i>antat</i>	<i>lala</i>	<i>=ta</i>	<i>lakó</i>	<i>kerajaan</i>	
	go	DM	take	princess	=DEM_P	to	kingdom	
#	0.h:a	v:other	other	v:pred	np.h:p	=rn_dem	adp	np:g %
	0014			0003			0001	

'The man took her to the kingdom.' [mc_sumbawa_menangis_0047]

1.4 Syntactic functions

1.4.1 *A in ling-constructions*: ⟨:a_ling⟩

Core arguments are normally expressed either by a bare NPs headed by a lexical noun or a free pronoun or a bound pronouns. Only the exception is the A argument in the post-predicate position, which occurs with the preposition *ling*. As mentioned in Section 1.3.2, it is annotated as ⟨:a_ling⟩. It may or may not show cross-reference to the bound pronoun.

1.4.2 *Dislocated topics*: ⟨:dt⟩

Any type of bare NP that does not have an argument relation in the clause is annotated as a dislocated topic ⟨:dt⟩. This NP may precede the predicate, as in (10), or follow it, as in (11).

(10) *pantèk ta nya singén colo' tau dunóng'*

	<i>pantèk</i>	<i>=ta</i>	<i>nya</i>	<i>singén</i>	<i>colo'</i>	<i>tau</i>	<i>dunóng'</i>	
	flint	=DEM_P	3	name	lighter	person	before	
#	np:dt	=rn_dem	pro:s	np:pred	rn_np:poss	rn_np.h:poss	rn	%
	0011		0011			0012		

“Pantek” is the name for a lighter of people from before.’ [mc_sumbawa_langlelo_0032]

(11) *ta=nya ku=mongka' aku.*

	<i>ta</i>	<i>nya</i>	<i>ku=</i>	<i>mongka'</i>	<i>aku</i>	
	DEM_P	3	1SG=	ANTIP.cook	1SG	
#ds	other	other	=pro_1_s	v:pred	pro.1:dt	%
			0006		0006	

“Now I am cooking.” [mc_sumbawa_langlelo_0030]

1.4.3 *Secondary objects in ditransitive constructions*: ⟨:p2⟩

In clauses of some “three-place” verbs, such as ‘tell’ or ‘tie’, the recipient NP or instrument NP may behave in the same way as the patient NP in the typical transitive construction, that is, as a bare NP occurring in the post-predicate position. These NPs are annotated as ⟨:p2⟩. (12) is an example of a recipient, (13) of an instrumental.

(12) *ma=ku=bau' bèang' mè anak ku ling*

	<i>ma=ku=bau'</i>	<i>ku=</i>	<i>bau'</i>	<i>bèang'</i>	<i>mè</i>	<i>anak</i>	<i>ku</i>	
	let	1SG=	can	give	rice	child	1SG	
#ds_ac	lv	=pro.1:a	lv	v:pred	np:p	np.h:p2	rn_pro.1:poss	% %
		0006			0007	0002	0006	

ling
words
np:s %

‘(She) said, “So that I can give rice to my children.”’

[mc_sumbawa_langlelo_0063]

(13) *ka=ya=antar benang pisak nè.*

	<i>ka=</i>	<i>ya=</i>	<i>antar</i>	<i>benang</i>	<i>pisak nè</i>	
	PST=	3=	tie	string	black ITJ	
#	0.h:a	0:p	=lv	=pro.h:a	v:pred	np:p2 rn other %
	0020	0065	0020		0066	

‘(She) tied (the ring) with piece of black string.’

[mc_sumbawa_kerekkure_0182]

1.5 Direct speech

Direct speech is often followed by the noun *ling* referring to the speaker’s ‘words’ or ‘what is told’. In this case, the whole sentence is taken as an equational sentence in which the direct speech is understood as a complement clause serving as the predicate ⟨#ds_cc:pred⟩ of the framing clause, of which the noun *ling* is the S (i.e. ⟨np:s⟩). See (14) for an example.

(14) *na sia tomas-tomas ina, ling, ta yakubalawas, ling.*

	<i>na</i>	<i>sia</i>	<i>tomas-tomas</i>	<i>ina</i>	<i>ling</i>
	PROH	2SG.HON	noisy	mother	words
#	#ds_cc.neg:pred	lv	pro.2:s	v:pred	np.h:voc % np:s %
			0002		0002

	<i>ta</i>	<i>ya=</i>	<i>ku=</i>	<i>balawas</i>	<i>ling</i>
	DEM_P	FUT=	1SG=	recite.a.poem	words
#	#ds_cc:pred	other	=lv	=pro.1:s	v:pred % np:s %
			0001		

‘He said (lit. (his) words were), “Don’t be noisy”. He said (lit. (his) words were), “Now I am going to recite a poem”’

[mc_sumbawa_kerekkure_0065]

2 Notes on the RefIND annotations

2.1 Referents in clauses otherwise not considered

Segments that have not been annotated for whatever reason, be that because they are incomplete or not syntactically well-formed, or because they are taken out of the normal flow of narration (e.g. because they address the listener, directly reply to the interviewer’s questions, or are not produced by the primary speaker), are marked as ⟨#nc⟩ ‘not considered’, and all of the elements they contain are glossed ⟨nc⟩.

However, these segments may still contain identifiable discourse references, which are presumably registered by the listener even in cases where the clause in question is abandoned part-way through. So as to preserve the genuine sequence of references in the annotations, mentions in <#nc> segments are indexed with RefIND, even though they do not receive meaningful GRAID annotations. This is true of all Multi-CAST corpora with RefIND.

For the Tabasaran corpus, however, we have attempted to go one step further by adding form and person/animacy glosses back onto those <nc> elements that have referent indices. The glosses are added as specifiers to the righthand side of the <nc> symbol, yielding, for instance, <nc_np> or <nc_pro.h>. Grammatical functions are not glossed.

(15) *pina kerè apa-apa ling sia papin nè ning*

	<i>pina</i>	<i>kerè</i>	<i>apa-apa</i>	<i>ling</i>	<i>sia</i>		<i>papin</i>		<i>nè</i>	<i>ning</i>
	make	cloth	whatever	by	2SG.HON		grand.father		ITJ	words
#nc	nc	nc	nc	nc	nc		nc_np.h		nc	nc %
							0016			

‘Please forgive us! (lit. (his) words were “you, grandfather, can make whichever cloth.”)’
[mc_sumbawa_flood_0036]

This approach makes it clear that while some information can be gleaned from these elements, one should not rely on being able to retrieve full information from the rest of the <#nc> segment. For most types of analysis, the <nc_> glosses should not be conflated with related GRAID symbols, given their uncertain status.

References

- Adelaar, Alexander. 2005. Malayo-Sumbawan. *Oceanic Linguistics* 44(2). 357–388.
- 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).
- Mahsun. 1999. *Variasi dialektal Bahasa Sumbawa: Kajian dialektologi Diakronis*. Lombok: University of Mataram Unpublished manuscript.
- Mbete, Aron Meko. 1990. *Rekonstruksi protobahasa Bali-Sasak-Sumbawa*. Jakarta: University of Indonesia Ph.D. dissertation.
- 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).
- Shiohara, Asako. 2022. Multi-CAST Sumbawa. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. (<https://multicast.aspra.uni-bamberg.de/#sumbawa>) (Accessed 2022-11-07).

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 Sumbawa 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
⟨dem_pro⟩	demonstrative pronoun (see morphological gloss to distinguish proximal, medial, distal demonstratives)
⟨pn_np⟩	proper name

Person/animacy symbols and specifiers

⟨.n⟩	non-human animate
------	-------------------

Function symbols and specifiers

⟨:a_ling⟩	A argument occurring with the preposition <i>ling</i>
⟨_ds⟩	<i>specifier</i> : subject of a verb of speech; attaches to ⟨:s⟩ and ⟨:a⟩

Subconstituent symbols

⟨_adj⟩	attributive adjective; attaches to ⟨1n⟩ and ⟨rn⟩
⟨_dem⟩	demonstrative determiner; attaches to ⟨1n⟩ and ⟨rn⟩
⟨_aux⟩	auxiliary; attaches to ⟨1v⟩ and ⟨rv⟩

Other symbols

⟨_1⟩	<i>on cross-indices</i> : first person
⟨_2⟩	<i>on cross-indices</i> : second person
⟨_h⟩	<i>on cross-indices</i> : human
⟨_d⟩	<i>on cross-indices</i> : anthropomorphized entity
⟨_n⟩	<i>on cross-indices</i> : non-human animate
⟨_s⟩	<i>on cross-indices</i> : subject of an intransitive clause
⟨_a⟩	<i>on cross-indices</i> : subject of a transitive clause
⟨nc_⟩	<i>specifier</i> : marks form glosses with RefIND indices in segments otherwise not considered (i.e. those marked ⟨#nc⟩)

B List of abbreviated morphological glosses

1	first person	M	medial
2	second person	NEG	negative
3	third person	P	proximal
ANTIP	anti-passive	PL	plural
ART	article	POSS	possessive
CAUS	causative	PROG	progressive
COMP	complementizer	PROH	prohibitive
D	distal	PST	past
DEM	demonstrative	PTC	particle
DES	desiderative	Q	question marker
DM	discourse marker	REDPL	reduplicant
EXCL	exclusive	REFL	reflexive
FUT	future	REL	relativizer
HBL	humble	SG	singular
HON	honorific		
ITJ	interjection	NC	not classifiable

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts



multicast.aspra.uni-bamberg.de/