

**A. Kemp Pallesen**Hamilton, New Zealand  
*kemp.pallesen@slingshot.co.nz***Craig Soderberg**Boston, MA, USA  
*craigggg@gmail.com*

Central Sama (ISO code *sml*) is spoken in the Philippines throughout the Sulu Archipelago and the Sibuguey Gulf to the north, and in many scattered communities as far north as Manila Bay. In Sabah, Malaysia, it is spoken primarily in the districts of Semporna and Kunak. For more information regarding the classification of this language within Austronesian, see Gordon (2005).

‘Central’ is not an indigenous label for the language; it simply reflects the central position of the language relative to other members of the Sama-Bajaw subgroup. Most native speakers of the language identify themselves as Sama, and the language they speak as Sinama. They are also known by numerous other names, a fact that has given rise to some confusion both in regard to identity and demographic information. Among these names are: Sama Dilaut (meaning ‘ocean-going Sama’, some semi-nomadic); Sama Pala’u (meaning ‘boat-based Sama’, semi-nomadic); Sama Jengen (meaning ‘house-boat Sama’), found in the Bongao district of the Philippines; Sama Lipid (meaning ‘land-based Sama’), a culturally distinct group of Central Sama speakers found in the Siasi municipality of the Philippines; Sama Pagung and Sama Paosol (meaning ‘floating Sama’ and ‘house-building Sama’, respectively); Badjaw (a sometimes pejorative exonym in the Philippines for migrant Sama Dilaut, but a more positive one for Central Sama speakers in Sabah, Malaysia). In addition, many Central Sama, though few of them own or live on land, choose – like other Sama – to identify themselves by a place name. For example, Sama Siasi, Sama Sitangkai, Sama Kabinga’an, Sama Ubian, and Sama Olutangga.

In Sabah, Malaysia, in addition to the names above, these people are known as Bajau Laut, Sama Laut, Sama Mandelaut, and Bajau Pela’u. English terms used to refer to this people group include both Sea Bajau and Sea Gypsies.<sup>1</sup> In the year 2000, population estimate for the Central Sama in Sabah was 15,000 (Soderberg forthcoming). Gordon (2005) estimates the population of Central Sama in the Philippines to be 90,000.

The ‘North Wind and the Sun’ text was translated and read by Ruth Biral, an unmarried female speaker of Central Sama, 29 years of age at the time of the recording. Ms Biral was born and raised in Siasi, a municipality of Sulu Province in the Philippines. Both her parents are Central Sama. Besides her native Sama, Ms Biral also speaks English, and the Philippine languages Cebuano, Tagalog, Tausug, and Ilonggo.

<sup>1</sup> Some of these designations are from Dunn (1980: 24).

## Consonants

Central Sama has 17 consonant phonemes.

	Bilabial	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive & affricate	p b	t̪	d	dʒ		k g	ʔ
Nasal	m		n		ɲ	ŋ	
Fricative			s				h
Approximant	w				j		
Lateral approximant			l				

PHONEME	PHONETIC FORM	ORTHOGRAPHIC FORM	ENGLISH GLOSS
p	[ <sup>1</sup> paʔok̚]	patuk	‘adze’
b	[ <sup>1</sup> baβäg]	babag	‘cross-piece’
m	[ <sup>1</sup> mayä]	maga	‘morning star’
w	[ <sup>1</sup> waluʔ]	waluʔ	‘eight’
t̪	[ <sup>1</sup> t̪anäm]	tanam	‘comfortable with’
d	[ <sup>1</sup> dagmɪʔ]	dagmit	‘do briefly’
n	[ <sup>1</sup> nadzäʔ]	najaʔ	‘disaster’
s	[ <sup>1</sup> sapid]	sapid	‘braid the hair’
l	[ <sup>1</sup> lapis]	lapis	‘layer’
ɲ	[ <sup>1</sup> naʔäʔ]	nyataʔ	‘see’
dʒ	[ <sup>1</sup> dʒalum]	jalum	‘needle’
j	[ <sup>1</sup> ja]	ya	‘which’
k	[ <sup>1</sup> kap:äl]	kappal	‘ship’
g	[ <sup>1</sup> gaβon]	gabun	‘cloud’
ŋ	[ <sup>1</sup> ŋa:ʔ]	ngäʔ	‘get, take’
ʔ	[ <sup>1</sup> kaʔäm]	kaʔam	‘you plural’
h	[ <sup>1</sup> hapäl]	hapal	‘clear sounding’

Central Sama has three voiced plosives /b d g/, each of which has a fricative allophone [β r ɣ] when it occurs between vowels, either within a word or at word boundaries.

PHONETIC FORM	PHONEMIC FORM	ENGLISH GLOSS
[ <sup>1</sup> baβäʔ]	/babaʔ/	‘carry on the back’
[ <sup>1</sup> binäʔyarän]	/binayadan/	‘paid’
[ <sup>1</sup> payäl]	/pagal/	‘fish trap’
[ <sup>1</sup> aku ru]	/aku du/	‘I indeed’

It has three voiceless plosives /p t k/, which are unreleased as [p̚ t̪̚ k̚] in syllable-final position.

PHONETIC FORM	PHONEMIC FORM	ENGLISH GLOSS
[ <sup>1</sup> sakäp̚]	/sakap/	‘prepared’
[ <sup>1</sup> dagmɪʔ̚]	/dagmit/	‘do briefly’
[ <sup>1</sup> paʔok̚]	/patuk/	‘adze’

The lateral approximant /l/ has a retroflexed allophone [ɭ] before a pause or before a consonant other than [l].

PHONETIC FORM	PHONEMIC FORM	ENGLISH GLOSS
[ <sup>1</sup> bulbu <sup>1</sup> ]	/bulbul/	'feather'
[ <sup>1</sup> bulän]	/bulan/	'moon'

The phonemes /j/ and /ɲ/ are uncommon in word-initial position. Out of a corpus of 10,000 words, only five begin with /j/ and six with /ɲ/. The phoneme /j/ occurs frequently word-medially and word-finally.

### Geminate consonants

This section includes notes on the status of the 13 lengthened consonants of Central Sama, in response to suggestions from a colleague that the observed length might be conditioned by a preceding schwa, and therefore not phonemic.

All consonants other than /h ʔ ɲ w/ occur as geminate clusters, word-medially in both stressed and unstressed syllables. They form the coda of the foregoing syllable and the onset of the following one, and are represented orthographically as sequences of two identical consonants.

In the examples below, the schwa is not written orthographically when it occurs word-initially, this being the preference of native speakers. We interpret the cluster [ddʒ] in [<sup>1</sup>buddʒän] 'maiden' as a geminate, with the fricative component lost between the two plosives, perhaps for articulatory reasons. The orthographic representation of the word as *budjang* is the preference of native speakers.

Geminate consonant examples:

PHONEME	PHONETIC FORM	PHONEMIC FORM	ORTHOGRAPHIC FORM	ENGLISH GLOSS
p	[ <sup>1</sup> p:äk]	/əppak/	ppak	'seagull'
	[ <sup>1</sup> kup:ä]	/kuppa/	kuppa	'jump down'
	[ <sup>1</sup> kəp:än]	/kəppang/	k'ppang	'lie flat'
b	[ <sup>1</sup> b:uŋ]	/əbbuŋ/	bbung	'porpoise'
dʒ	[ <sup>1</sup> buddʒän]	/budʒdʒaŋ/	budjang	'maiden'
m	[ <sup>1</sup> kam:uʔ <sup>1</sup> ]	/kammut/	kammut	'gnat'
t	[ <sup>1</sup> t:äp <sup>1</sup> ]	/əttap/	ttap	'open and close lips repeatedly'
d	[ <sup>1</sup> bid:äʔ]	/biddaʔ/	bidda'	'difference'
n	[ <sup>1</sup> n:äʔ]	/ənnaʔ/	nna'	'place inside'
l	[ <sup>1</sup> l:äw]	/əllaw/	llaw	'day, sun'
s	[ <sup>1</sup> s:ä]	/ləssa/	l'ssa	'taste something'
k	[ <sup>1</sup> k:ät <sup>1</sup> ]	/əkkat/	kkat	'to separate'
g	[ <sup>1</sup> g:äk <sup>1</sup> ]	/əggak/	ggak	'belch'
ŋ	[ <sup>1</sup> ŋ:ik <sup>1</sup> ]	/əŋŋik/	ngngik	'to squeak, as a mouse'
j	[ <sup>1</sup> kuj:äʔ]	/kujjaʔ/	kuyya'	'monkey'

We interpret these lengthened consonants as true phonemic geminates for the following reasons:

(i) They occur after each of the six phonemic vowels, not merely after schwa.

Geminate consonants after the six phonemic vowels:

PHONEME	PHONETIC FORM	ORTHOGRAPHIC FORM	ENGLISH GLOSS
ə	[ <sup>1</sup> təp:a]	t <sup>1</sup> ppa	'to pound rice'
e	[ <sup>1</sup> ab:ɔ]	abbu	'pride'
	[ <sup>1</sup> bel:eʔ]	belle'	'sea hawk'
i	[ <sup>1</sup> bid:äʔ]	bidda'	'be different from'
o	[ <sup>1</sup> bok:oʔ]	bokko'	'green turtle'
u	[ <sup>1</sup> bud:i]	buddi	'keep in mind, especially of a debt of gratitude'

(ii) The lengthened consonant is retained even in those Sama-Bajaw languages which have replaced the reconstructed schwa of Proto Sama-Bajaw with a different vowel.

The examples in the following table are phonemic representations. Geminate consonants in Sama-Bajaw languages:

PROTO SAMA-BAJAU	CENTRAL SAMA	YAKAN	ABAKNON SAMA	SOUTHERN SAMA	ENGLISH GLOSS
*kakkal	/kakkal/	/kakkal/			'permanent'
*həlla	/həlla/	/hella/	/alla/	/halla/	'husband'
*təllɯn	/təllon/	/tellen/	/tallun/		'to swallow'
*lappas	/lappas/	/lappas/	/lappas/		'abalone'

(iii) Geminate clusters are consistent with (a) the occurrence of consonant clusters at syllable boundaries, e.g. /dag.mit/ 'do briefly', /bul.bul/ 'feather', /am.bat/ 'do completely', /anak.bi/ 'your (plural) child'; and (b) the occurrence of /ə/ preceding consonant clusters other than the proposed geminates, e.g. /kəmbal/ 'twin', /dənda/ 'woman', /səŋɡgal/ 'wedge'; /tənton/ 'fishing technique' /pənsot/ 'navel'.

(iv) Minimal and near-minimal pairs display a contrast between lengthened consonants and those of normal length, e.g. /agal-agal/ 'seaweed' and /aggal/ 'reluctant'; /bida/ 'to pull' and /biddaʔ/ 'difference'; /kasaʔ/ 'shroud-cloth' and /kassaʔ/ 'glass'; /sabut/ 'understand' and /sabbut/ 'invoke by name'; /tagal/ 'excuse' and /taggal/ 'support'.

(v) There appears to be no set of phonological phenomenon conditioning all the lengthened consonants.

## Vowels

Central Sama has six vowel phonemes, /i e ə a o u/.

	Front	Central	Back
High	i		u
Mid	e	ə	o
Low		a	

PHONEME	PHONETIC FORM	PHONEMIC FORM	ORTHOGRAPHIC FORM	ENGLISH GLOSS
i	[ <sup>1</sup> tɪpɪd]	/tipid/	tipid	'on the edge of'
e	[ <sup>1</sup> tɛpɔ]	/tepo/	tepo	'mat'
ə	[ <sup>1</sup> təp:iʔ]	/təppiʔ/	t <sup>1</sup> ppi'	'approach'
a	[ <sup>1</sup> təpā]	/tapa/	tapa	'to roast'
u	[ <sup>1</sup> tupās]	/tupas/	tupas	'side dish'
o	[ <sup>1</sup> tɔpād]	/topad/	topad	'same age'

Three of these vowels, /i a u/, exhibit allophonic variation: the tense allophone occurs in stressed syllables, while the lax allophone occurs in unstressed syllables. Vowels displaying allophonic variation:

	STRESSED	ENGLISH GLOSS	UNSTRESSED	ENGLISH GLOSS
/a/	[ <sup>1</sup> t̪ap̪ä]	‘to roast’	[t̪ä'pah̪on]	‘roast it’
/i/	[ <sup>1</sup> sik̪u]	‘elbow’	[sɪ'kuk̪u]	‘my elbow’
/u/	[ <sup>1</sup> bal̪oɬ̪]	‘to hold’	[bä'luɬ̪in]	‘hold it’

All vowels except /ə/ have a lengthened counterpart. We interpret these as sequences of two syllabic vowels because minimal and near-minimal pairs show a semantic contrast between lengthened and unlengthened vowels. Examples:

PHONETIC FORM	PHONEMIC FORM	ORTHOGRAPHIC FORM	ENGLISH GLOSS
[ä'ta:]	/ataa/	atā	‘distant’
[ <sup>1</sup> at̪ä]	/ata/	ata	‘slave’
[sɪ'ɣa:]	/sigaa/	sigā	‘they (two)’
[ <sup>1</sup> siɣä]	/siga/	siga	‘cigarette’
[pinä'te:ʔ]	/pinateeʔ/	pinatē	‘made to count as important’
[pɪ'nateʔ]	/pinateʔ/	pinate	‘broken off’
[pä'ko:]	/pakool/	pakōl	‘cause to agree to’
[ <sup>1</sup> pako]	/pakol/	pakol	‘riding saddle’
[pä'ti:g]	/patiig/	patīg	‘to tilt sideways’
[ <sup>1</sup> patig]	/patig/	patig	‘heavy cloth’
[bɪ'ni:t̪]	/biniit/	binīt	‘lashed in place’
[ <sup>1</sup> piniɬ̪]	/pinit/	pinit	‘house lizard’
[ <sup>1</sup> bu:k̪u]	/buuku/	būku	‘my hair’
[ <sup>1</sup> buku]	/buku/	buku	‘button’

Central Sama phonology demonstrates harmony between front and back vowels in bisyllabic roots. If one vowel is closed, both will be closed, as the examples below demonstrate. Front and back vowel harmony:

CENTRAL SAMA	ENGLISH GLOSS	CENTRAL SAMA	ENGLISH GLOSS
/kuwit/	‘skin’	/kowet/	‘umbilical cord’
/igut/	‘coccyx’	/egot/	‘do relentlessly’
/sipit/	‘clip together’	/sep̪et/	‘spatter’
/butuʔ/	‘eat like an animal’	/botoʔ/	‘penis’

## Stress

Primary stress in Central Sama is not phonemic and falls on the penultimate syllable. When a stem is suffixed, the primary stress shifts so that penultimate stress is maintained, as in the examples below. In words of four syllables or more, secondary stress falls on the syllable two syllables prior to the penultimate. In this section of the paper and in the phonetic transcription which follows, primary stress is marked with [<sup>1</sup>] while secondary stress is marked with [ˌ]. Structural pauses are marked with /.

PHONETIC FORM	ENGLISH GLOSS
[ <sup>h</sup> anäk <sup>ʔ</sup> ]	'child'
[ä <sup>h</sup> nakbɪ]	'your (plural) child'
[kä <sup>h</sup> litän]	'shark'
[ <sub>1</sub> käli <sup>h</sup> tanän]	'shark-infested'
[ <sup>h</sup> baɪɪd]	'matches'
[bä <sup>h</sup> ɪdnä]	'his matches'
[ <sup>h</sup> base <sup>ʔ</sup> ]	'wet'
[mä <sub>g</sub> bäse <sup>ʔ</sup> anän]	'getting wet'

## Recorded passage

### Phonetic transcription

na də<sup>h</sup>kaju<sup>ʔ</sup> <sup>h</sup>l:äw mä<sub>g</sub>käl<sup>h</sup>u<sup>ʔ</sup>kas:ä si hä<sup>h</sup>baɪät<sup>ʔ</sup> <sup>h</sup>makä si <sup>h</sup>l:äw / ja pä<sub>g</sub>käl<sup>h</sup>u<sup>ʔ</sup>kä<sup>h</sup>s:ahän <sup>h</sup>siyə:m baɲ sai <sup>h</sup>kono<sup>ʔ</sup> ä<sup>h</sup>kosog min <sup>h</sup>siyə:m kä<sup>h</sup>r<sup>h</sup>u<sup>ʔ</sup>waɲän / pä<sup>h</sup>sal<sup>h</sup>tä<sup>ʔ</sup> ä<sup>h</sup>nijä<sup>ʔ</sup> də<sup>h</sup>kaju<sup>ʔ</sup> <sup>h</sup>bijä<sup>h</sup>hero mä<sup>h</sup>r<sup>h</sup>na<sup>ʔ</sup>än <sup>h</sup>aŋkän <sup>h</sup>siyə:m mä<sub>g</sub>sul<sup>h</sup>u<sup>ʔ</sup> baɲ sai <sup>h</sup>kono<sup>ʔ</sup> mä<sup>h</sup>käpäh<sup>h</sup>rusän <sup>h</sup>səm:ek <sup>h</sup>bijä<sup>h</sup>hero <sup>h</sup>r<sup>h</sup>na:n / ijä ja ä<sup>h</sup>ŋän<sup>h</sup>da<sup>ʔ</sup>u<sup>ʔ</sup> / män<sup>h</sup>jar<sup>h</sup>i pä<sup>h</sup>rä<sup>h</sup>hu: na si hä<sup>h</sup>baɪät<sup>ʔ</sup> <sup>h</sup>ämä<sup>h</sup>luwäs ko<sup>h</sup>sognä bo<sup>ʔ</sup> ijä ä<sup>h</sup>nij<sup>h</sup>u<sup>ʔ</sup>p pä<sup>h</sup>kosog / <sup>h</sup>pin<sup>h</sup>us<sup>h</sup>pusan <sup>h</sup>to<sup>ʔ</sup>o<sup>h</sup>ŋän <sup>h</sup>he<sup>ʔ</sup>nä ko<sup>h</sup>sognä <sup>h</sup>sayo<sup>ʔ</sup> <sup>h</sup>jampä na ijä ä<sup>h</sup>nij<sup>h</sup>u<sup>ʔ</sup>p pä<sup>h</sup>kosog gom <sup>h</sup>pa<sup>ʔ</sup>in na <sup>h</sup>sa:b ä<sup>h</sup>ŋəm:os <sup>h</sup>dinä <sup>h</sup>bijä<sup>h</sup>hero <sup>h</sup>ina:n / na <sup>h</sup>pagkä <sup>h</sup>haläm kä<sup>h</sup>h<sup>h</sup>rusän e<sup>ʔ</sup> si hä<sup>h</sup>baɪät<sup>ʔ</sup> <sup>h</sup>səm:ek <sup>h</sup>bijä<sup>h</sup>hero <sup>h</sup>r<sup>h</sup>na:n / <sup>h</sup>pinä<sup>h</sup>səl:e<sup>ʔ</sup> e<sup>ʔ</sup>nä si <sup>h</sup>l:äw <sup>h</sup>ämä<sup>h</sup>luwäs ko<sup>h</sup>sognä / na pä<sup>h</sup>siläk mä<sub>g</sub>t<sup>h</sup>u<sup>ʔ</sup> si <sup>h</sup>l:äw <sup>h</sup>ä<sup>h</sup>mus<sup>h</sup>pusän pä<sup>h</sup>su<sup>ʔ</sup>nä / <sup>h</sup>landu<sup>ʔ</sup> <sup>h</sup>to<sup>ʔ</sup>o<sup>h</sup>ŋän ä<sup>h</sup>pasu<sup>ʔ</sup> <sup>h</sup>sampäj <sup>h</sup>mbal <sup>h</sup>tä<sup>h</sup>sandäl e<sup>ʔ</sup> <sup>h</sup>bijä<sup>h</sup>hero <sup>h</sup>r<sup>h</sup>na:n <sup>h</sup>aŋkän ijä ä<sup>h</sup>ŋ<sup>h</sup>rusän <sup>h</sup>səm:eknä <sup>h</sup>in<sup>h</sup>u<sup>ʔ</sup> <sup>h</sup>in<sup>h</sup>u<sup>ʔ</sup> / mä kä <sup>h</sup>tä<sup>h</sup>pu<sup>h</sup>sannä <sup>h</sup>pagkä mbäl na <sup>h</sup>käsän<sup>h</sup>dalän <sup>h</sup>pasu<sup>ʔ</sup> <sup>h</sup>r<sup>h</sup>na:n / ä<sup>h</sup>nant<sup>h</sup>ä<sup>h</sup> <sup>h</sup>bija<sup>h</sup>hero <sup>h</sup>it<sup>h</sup>u: bo<sup>ʔ</sup> pä<sup>h</sup>kuppä ni <sup>h</sup>tigbäw ja baj mä<sup>h</sup>r<sup>h</sup>na<sup>ʔ</sup>än ä<sup>h</sup>sa:l ma lä<sup>h</sup>bajän / hä<sup>h</sup>tinä ä<sup>h</sup>ha:p <sup>h</sup>layi<sup>ʔ</sup> <sup>h</sup>kitä ä<sup>h</sup>hä<sup>h</sup>nun<sup>h</sup>u<sup>ʔ</sup> min ä<sup>h</sup>moyos

The current orthography is a work in progress, with the full impact of Tagalog (effectively the national language) and growing literacy yet to be felt. We acknowledge that the use of the symbol [ʔ] for both glottal stop and schwa is not an elegant solution to the problem of representing two non-standard phonemes without resorting to the use of diacritics. In its defence as an interim orthography, however, we note that native speakers of Central Sama who have been previously taught to read Tagalog and English, read material in this orthography with no further instruction.

### Orthographic transcription

Na, dakayu' llaw, magkalukassa si Habagat maka si Llaw. Ya pagkalukassahan sigām bang sai kono' akosog min sigām karuwangan. Pasalta' aniya' dakayu' biyahero maina'an angkan sigām magsulut bang sai kono' makapahurusan s'mmek biyahero inān, iya ya anganda'ug. Manjari parahū na si Habagat amaluwas kosogna bo' iya aniyup pakosog. Pinuspusan to'ongan he'na kosogna sago' yampa na iya aniyup pakosog, gom pa'in na isab ang'mmos di-na biyahero inān. Na, pagka halam kahurusan e' si Habagat s'mmek biyahero inān, pinas'le' e'na si Llaw amaluwas kosogna. Na, pasilak magtūy si Llaw, amuspusan pasu'na. Landu' to'ongan apasu' sampay mbal tasandal e' biyahero inān.

Angkan iya angurusan s'mmekna inut-inut. Ma katapusanna, pagka mbal na kasandalan pasu' inān, anatang biyahero itū bo' pakuppa ni tigbaw ya bay maina'an asal ma labayan. Hatina, ahāp lagi' kita ahanunut min amogos.

## Acknowledgements

We wish to thank David Mead, Steve Parker, Pete Unseth, Scott Youngman and two anonymous reviewers for helpful comments and suggestions. All errors are our responsibility.

## References

- Dunn, Phyllis A. 1980. Sabah peoples' who's who: A glossary of the terms used for the people and languages of Sabah, East Malaysia. *Sabah Museum Annals* 2, 19–61.
- Gordon, Raymond G., Jr. (ed.). 2005. *Ethnologue: Languages of the world*, 15th edn. Dallas, TX: SIL International. [Online version: <http://www.ethnologue.com/>]
- Soderberg, Craig. Forthcoming. Intelligibility prediction models applied to the coastal languages of Sabah, Malaysia. Ms., Boston, MA.