Lexicon and word formation in Indonesian Bajo

CHANDRA NURAINI

ABSTRACT

This paper[1] deals with the phonology and the lexicology of the Indonesian Bajo language and more specifically with the dialect or variant that can be heard all around the Flores Sea in Kangean, South-East Sulawesi, Sumbawa, and Flores. The phonological survey focuses on vowel lengthening, gemination, pre-nasalized phonemes, and sandhi. The second part of this paper proposes an insight into Bajo lexicology, restricted to nominal and verbal derivation.

KEYWORDS

Sama-Bajau, Bajo, Indonesian Bajo, baun Same, phonology, pre-nasalization, gemination, vowel length, epenthesis, aphaeresis, demarcative glottal stop, sandhi, derivation, causative prefix, locative suffix.

The Sama-Bajau languages are part of the Western Malayo-Polynesian language group, and their speakers may be found in many fishermen villages over an immense area, including the eastern part of Indonesia, the southern Philippines, and North Kalimantan (Sabah). After a short introduction on the Indonesian Bajo in general this paper presents information on aspects of the phonology of their language, especially with regard to the dialect or variant that can be heard around the Flores Sea: Kangean, South-East Sulawesi, Sumbawa, and Flores. In the second part of this paper, I will propose an insight into Bajo lexicology.

In various languages of the Sama-Bajau branch, the lengthening of some vowels and the gemination of certain consonants is perceptible, but their phonological relevance in the variety discussed below is problematic. I will

---

CHANDRA NURAINI, born in 1965, is an associate professor of Indonesian and Civilization of Indonesia at the University of La Rochelle (France). Her research focuses on the language and oral literature of the Bajo people in Indonesia. Chandra Nuraini may be contacted at: cnuraini@univ-lr.fr.

analyse these features and propose a diachronic explanation for the occurrence of the long vowels elsewhere. Furthermore an assessment will be made on the phonological status of consonant sequences beginning with a nasal. Finally, as a transition towards lexicological issues, I shall discuss sandhi phenomena which involve phonemic adjustments between root words and affixes, and which entail aphaeresis and epenthesis.

It would take too much space to examine the Bajo lexicology in its entirety and I will therefore limit myself to nominal and verbal derivation on the basis of root words. In this context causative and locative suffixes will receive special attention.

For more detailed information on the syntax of Indonesian Bajo, with its elaborate voice system, its ergative – absolutive opposition in the pronominal paradigm, and its equal distribution of the syntactic VSO and SVO patterns, I refer to Nuraini 2008.

1 A SHORT PRESENTATION OF INDONESIAN BAJO

The Sama-Bajau languages are spoken by a diaspora of small communities scattered on the shores of the Southern Philippines, Sabah (Malaysia) and a multitude of islands in the eastern part of Indonesia. At one time, the Sama-Bajau constituted one community, but nobody knows from which part of Southeast Asia they originated, and there is no convincing theory about when and why they split up into numerous communities, now separated by several days of navigation.[2]

These people usually call themselves *Sama*, next to the exonym *Bajau* or *Bajo*. In Indonesia, they indeed call themselves *Bajo* when speaking to outsiders. Their language is known as *Bahasa Bajo* in Indonesian or *Baun Same* in their own language. Unsurprisingly, having been separated for centuries, the various Sama-Bajau communities of the different islands have developed their own language varieties. Between the nine Bajo language varieties distinguished by Pallesen (Pallesen 1985: 117) intelligibility is nowadays often difficult or impossible. According to Jun, there are four distinct Sama-Bajau languages in the Sulu archipelago alone (Jun 2005: 379).

In contrast to the situation in the Sulu archipelago, there is only one Indonesian Bajo language. It is spread over a very large area ranging from North-Sulawesi to West Timor, and from Kangean (near Madura) to Maluku and Papua (see Map 1). This maritime region is ten times wider than the Sabah-Sulu Bajau cluster. And more importantly, the Indonesian Bajo language is quite homogeneous: there is mutual intelligibility between Bajo fishermen from around the Flores Sea, and those from Central/North Sulawesi and Maluku. There may be two or three dialects only within Indonesian Bajo, but this issue

---

[2]   A recent report in the Indonesian newspaper *Kompas* (6 April 2010) http://regional. kompas.com/read/2010/04/06/1725400/Kapal.Warga.Bajau.Palau.Segera.Diperbaiki may be exemplary for the pattern of migration of the Bajo: a group of 103 sailors with their families, after having been expelled from Malaysia and the Philippines, ultimately stranded in Indonesia (Berau, Kalimantan Timur).

Map 1. Grey area showing the maximal dispersal of the Bajau diaspora; map by the author.

requires further research. Most of our reference data was obtained from the tiny and over-populated island of Sapeken, Kangean archipelago, Sumenep Regency (eastern tip of Madura), East-Java Province, Indonesia. However, our findings apply to the Bajo language that can be heard all around the Flores Sea: Kangean, South-East Sulawesi, Sumbawa, and Flores.

2 Phonemes

In this section, I will present an inventory of the Bajo phonemes, and discuss the phenomenon of vowel lengthening, and two other problems linked to the description of the consonants.

2.1 Inventory of the vowels

The Bajo phonological system comprises six vowels. The phonologically distinctive Bajo vowels have been identified by Ni Luh et al. (1997: 19-22) by means of this paper. The six vowels are /i/, /e/, /ə/, /a/, /o/ and /u/, as in the following the conventional minimal pairs method, and there is no need to duplicate their work in morphemes:

/i/   *iru* [ʔiru] 'this, that'
/e/   *lé* [le] 'by'
/ə/   *due* [duə] 'two'
/a/   *ai* [ʔaj] 'what'
/o/   *oré* [ʔore] 'over there'
/u/   *uye* [ʔujə] 'song'

Distributional constraints may lead to pronounce /o/ either as [o] or as [ɔ]. As in many other languages, [o] is preferred as the nucleus of an open syllable,

whereas [ɔ] is the preferred realization in closed syllables. The same remark applies to /e/ that will be realized as open [ε] in closed syllables, for instance *sen* [sεn] 'money', as opposed to closed [e] in open syllables.

According to Hinayat (2003: xxv-xxvi), there are in the Bajau language of Sabah (North Kalimantan, Malaysia), "long vowels" for instance in *kook* [kɔ:k] 'head' (Kubang-Semporna dialect), or *nuut* [nu:t] 'come with, follow' (Kota Belud dialect). The phonological relevance of this vowel lengthening seems doubtful, as one cannot find minimal pairs with, for instance, **kok* or **nut*. Hinayat (2003: xxv-xxvi) found only one minimal pair in Kota Belud dialect: *ta'* 'on, at' vs. *taa'* 'be long'. Nevertheless, under the alphabetical entry (Hinayat 2003: 276), one can read that *ta'* is a prefix. The pertinence of a prefix as a term in a minimal pair may be challenged.

Ni Luh et al. (1997) do not mention a vowel-lengthening feature in their description of the Bajo of East Lombok. I did not hear this feature among the Bajo communities who have settled around the Flores Sea, or in Central Sulawesi. However, vowel lengthening in the Bajau dialect(s) of Sabah is a fact, whose presence can be explained diachronically. Let us compare a few cognates in the two lists below, respectively in Sabah Bajau, quoted from Hinayat (2003), and in Indonesian Bajo (personal field data):

Sabah Bajau		Indonesian Bajo (Kangean)	
tingkoo'	'sit down'	*tingkolo*	'sit down' (root word)
tikook	head'	*tikolo*	'head'
buuh	'hair'	*bulu*	'hair'
boo'	'bamboo'	*bolo*	'bamboo'
daa	'don't do'	*daha*	'don't do'
kee	'hole'	*kéhé*	'hole'

The long vowels in Sabah Bajau correspond to a VCV string where the two vowels are identical in Indonesian Bajo, and the consonant is either /l/ or /h/. Obviously, there is a regular evolutionary pattern, leading to two hypotheses:
1) Indonesian Bajo has developed a systematic epenthesis (intrusion of a consonant) to break apart long vowels, splitting off two identical short vowels: V: > VCV
2) Sabah Bajau has developed a systematic aphaeresis (elision of a consonant) between two identical vowels, merging the two short vowels into a long vowel: VCV > VV > V:

The second hypothesis is obviously the most likely. The consonants /l/ and /h/ are weak (as /s/ in the example below) and easily subject to lenition or elision. Besides, we can see an 'intermediate' state of this evolution in the two examples below, where the lenition of /h/ and /s/ leads to a glottal stop /ʔ/.

Sabah Bajau		Indonesian Bajo (Kangean)	
a'a	'person'	*aha*	'person'
ki'it	'few'	*kisit*	'few'

Even among Sabah Bajau dialects, various states of evolution are manifest, where the /ʔ/ or /h/ is the remnant of a stronger consonant, for instance *sa'ah* and *saa'* 'wrong, error' (Beluran); *toho'* 'dry, ashore' (Kubang-Semporna and other dialects) and *too'* (Kota Belud), see Hinayat (2003). Finally, the aphaeresis in Sabah Bajau is accompanied by a "compensating" consonant epenthesis: /ʔ/ or /h/ as the coda of the last syllable (sometimes written by Hidayat as ' or *k*), see for instance *buuh* and *boo'* in the table above. It could also be interpreted as a metathesis: the weak consonant that separates two identical vowels is moved to the end of the last syllable. However, the aphaeresis of the weak consonants /l/ or /s/ will also lead to the epenthesis of a final /ʔ/ or /h/. In sum, we can reconstruct a probably recent evolution of the Sabah Bajau according to this pattern:

bulu > **buhu* > *buuh* [buːh] 'hair'
bolo > **bo'o* > *boo'* [bɔːʔ] 'bamboo'

Returning to Indonesian Bajo, it is worth noting that for final open syllables, some occurrences of a glottal stop (as onset of the final syllable) are probably weakened remnants of another consonant. For instance, *ba'u* [baʔu] 'new', *ba'i* [baʔi] 'decaying, rotten' and *ma'i* [maʔi] 'come here!' seem to originate from Malay, respectively *baru*, *basi* and *mari*. If this is true, this evolution fits the Sabah Bajau pattern described above. In both cases, one notices the lenition of a medial consonant developing into (/ʔ/ or /h/), which (in Sabah Bajau at least), either fades away or is subject to metathesis with the following vowel becoming the coda of the final syllable.

2.2 INVENTORY OF THE CONSONANTS
Ni Luh et al. (1997) present an inventory of the Bajo consonants based on the minimal pairs method. In this section, I will therefore focus on problems that arise from this inventory: pre-nasalized consonants that could be considered either phoneme sequences or single phonemes; and the phonological relevance of consonant gemination in Bajo.

2.2.1 PRENASALIZED PLOSSIVE OR AFFRICATE CONSONANTS
Like many other Austronesian languages, the Bajo language shows a strong preference for the alternation of consonants and vowels. Actually, borrowed words undergo vowel epenthesis that restores the CVCV alternation, such as in *sepit* < *speedboat*. Glottal stop[3] and approximant ([w] or [j])

---

[3]   The glottal stop /ʔ/ is particularly clear when it occurs between two identical vowels. In this case, there is no syneresis (the two vowels do not merge into a single vowel), for example

epenthesis⁴ are two other means aimed at maintaining a consonant between two vowels. Nevertheless, the preference in Bajo for open CV sequences is not absolute: word-finally syllables may be closed and word-medially there are consonant sequences as in for instance: *cambe* 'tamarind fruit', *gentin* 'tile', *nganjame* 'to work', *kampoh* 'village, hamlet', in which a nasal is followed by a plosive (/p/ /b/ /t/ /d/ /k/ and /g/) or an affricate (/t͡ʃ/ or /d͡ʒ /). All possible combinations are represented in the following list:

m͡p	[tim͡prɔh]	'coconut shell'
m͡b	[ʔum͡bu]	'smoke (n)'
n͡t	[ʔan͡tel:o]	'egg'
n͡d	[man͡di]	'to bath'
ŋ͡k	[taŋ͡kau]	'to steal'
ŋ͡g	[təŋ͡ge]	'high'
ɲ͡tʃ	[bɔɲ͡tʃɔrah]	'ditch, gutter'
ɲ͡dʒ	[ŋaɲ͡dʒama]	'to work'

The consonant sequences in question, the constituents of which are always homorganic, only occur word-medially, with one possible exception, namely the negator *nggai* 'no, not'. However, in less allegro speech this word is often realized bisyllabic as [əŋ͡gai] or [ʔəŋ͡gai].

The question may arise whether we should consider these consonant sequences as "exceptions" to the "general" rule of consonant-vowel (CV) alternation in non-final syllables? Or as single phonemes (pre-nasalized plosives or affricates)? For instance, in *cambe* [t͡ʃam͡bə] 'tamarind fruit', do we hear the random adjacency of the two phonemes /m/ and /b/, or only a single phoneme /m͡b/? Since the audible syllable boundary in such words precedes the consonant sequence rather than separates its constituents, I propose that the Bajo has the following single phonemes: /m͡p/; /m͡b/; /n͡t/; /n͡d/; /ɲ͡tʃ/; /ɲ͡dʒ/; /ŋ͡k/; /ŋ͡g/. These pre-nasalized phonemes are tied to specific distributional

---

to'oh [toʔoh] 'true', *katira'ah* [katiraʔah] 'lazy'. The demarcating role of the glottal stop is also self-evident (and therefore non-phonemic) in lexical derivations, where it marks the border between an affix and the lexical root morpheme, for instance *baséan* [baseʔan] 'covered with water' < *basé* 'wet'; *diintaan* [diʔintaʔan] 'eaten' < *inta* 'to eat'. This demarcating role of the glottal stop between affixes and root morphemes is the counterpart of the sandhi phenomena that we will describe below in section III.

⁴   There is a tendency to insert an approximant [w] or [j] between two vowels within a root word. For example, *boé* 'water' is realized [bɔwe] rather than [bɔe], and *tiut* [tijut] 'to blow' (less likely [tiut], but surely not *[tjut]). The bilabial approximant is quite faint in this context, so I propose to transcribe it by the symbol [ʷ], for instance *dambui* [dambuʷi] 'yesterday'. Other examples are *dioh* [dijɔh] 'siren', *due* [duʷ] 'two', *moe* [moʷ] 'to bring', *ayuan* [ajuʷan] 'in-law (father-, sister-, etcetera)'.

constraints: they never occur at the onset of the first syllable or as the coda of the final syllable.[5] If this analysis is correct, the Indonesian Bajo language has 27 consonants, which are all listed in Table 1.

		bilabial	denti-alveolar	alveo-palatal	palatal	velar	glottal
Plosives (occlusives)	unvoiced	p [pore] there	t [taha] long			k [kolɛʔ] to can	ʔ [ʔanaʔ] child
	pre-nasalized unvoiced	m͡p [sam͡paʔ] fence	n͡t [lan͡te] plaited mat			ŋ͡k [taŋ͡kau] to steal	
	voiced	b [bidɔʔ] boat	d [daruwə] identical			g [golə] sugar	
	pre-nasalized voiced	m͡b [m͡boʔ] elder	n͡d [ʔin͡dat] to look at			ŋ͡g [təŋ͡ge] high	
	nasal	m [mələ] to cook	n [nia] there is		ɲ [ɲuloh] green	ŋ [daŋaj] how much	
Fricatives	unvoiced		s [salɔh] tomorrow	t͡ʃ [t͡ʃɔt͡ʃɔtan] snail	ɲ͡t͡ʃ [bɔɲ͡t͡ʃɔrah] steer bar		h [ʔaha] person
	voiced			d͡ʒ [sad͡ʒo] style	ɲ͡d͡ʒ [mɛɲ͡d͡ʒe] where		
	lateral		l [ləl:ə] man, male				
	trill		r [rahat] evil				
	approxi-mant				j [ʔujə] song, sing	w [ʔuwaʔ] father, Mr	

Table 1. The consonants of Indonesian Bajo.

In Table 1 I did not include two phonemes that were borrowed quite recently into Indonesian Bajo: /f/ as in fajar 'dawn' and /d͡z/ as in azan [ʔad͡zan] 'call for prayer'.

2.2.2 CONSONANT GEMINATION IN BAJO

When hearing Indonesian Bajo speakers, one may get the impression that within certain words consonant "lengthening" occurs. Spontaneously, educated Bajo speakers will write these words with a double grapheme, for

---

5   The same remark may apply to Malay-Indonesian, but we admit that this statement does not fit into the traditional description of its phonological system.

example *patappe* [patap:ə] 'to entrust',[6] *abba* [ʔab:a] 'fall down, collapse, lie down', *pammasé* [pam:ase] 'donor', *bette* [bət:ə] 'for long', *matadda* [matad:a] 'pretty', *tasasanne* [tasasan:ə] 'suddenly, without warning', *karesse* [kares:ə] 'sensation', *ballai* [bal:aj] 'rarely', *ellau* [ʔəl:aw] 'day', *lelle* [ləl:ə] 'man, male'. Bajo consonants that may undergo gemination are /p/ /b/ /m/ /t/ /d/ /n/ /s/ and /l/, the last one being the most frequent.[7] Gemination is lexically pegged to a root word, that is to say the geminate consonants are not induced by sandhi.[8] Gemination appears only in medial position, as the onset of a non-initial syllable. More precisely, it affects the second consonant of a CVCV pattern, which we can represent as CVC:V. The words *matadda* 'pretty', *malasso* 'good, well' or *karesse* 'sensation' are in fact frozen prefixed forms with *ma–* (imperfective/stative) or *ka–* (motion/tendency), thus their root word, although never appearing unaffixed, fits into the CVC:V pattern.

Gemination in Bajo cannot always be described as the lengthening of a consonant. Fricative vowels are easy to hold, as long as we have air to flow through the articulatory organs, for instance [s:]. But for a plosive, even if it is physically possible to hold the articulatory organs at the point of articulation, the sound of the occlusion cannot last. The sonograms of one Bajo informant[9] clearly show that for plosives, the gemination is primarily a short pause at the midpoint of an occlusive consonant's realization. The articulatory organs "freeze" at the exact moment of the occlusion, to be released after a short time span. There may be also a little bit more stress on the vowel that precedes the geminated consonant, and a stronger initial glottal stop if applicable, but the most obvious feature is the in-between micro-pause. Compare the sonograms in Figures 1–2.[10]

---

[6]   Of course, this grapheme doubling does not stand for a phoneme doubling: for instance, *patappe* will never be realised *[patapəpə].

[7]   Our data show only two examples of (supposed) gemination of a /r/, which could in fact be of an expressive nature, thus the existence of a lexically geminated /r/ remains doubtful.

[8]   Sandhi phenomena in Bajo are restricted to morpho-phonological adjustments between an affix and a root word (see section III.1 below). Yet, I found an exception to this rule: *masé* "to give" > *pammasé* "giver, donor".

[9]   We thank Mohammad Tamin, a Bajo native speaker, aged 44, born and living in the Kangean archipelago.

[10]   Our sonograms where generated using Amadeus II version 3.8 for Mac OS, from a 44.100 Hz 16 bits sound file. The frequency scale represented in this sonogram is lineary, from 0 to 11.025 Hz.
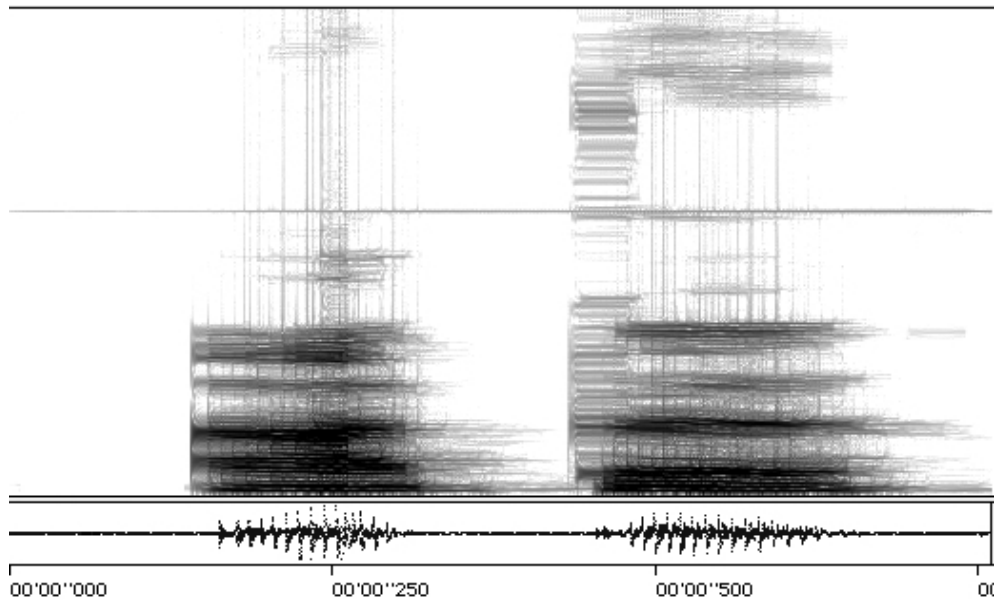
Figure.1. Sonogram of *ate* [ʔatə] 'slave'. Elicited data, the word was recorded in isolation.
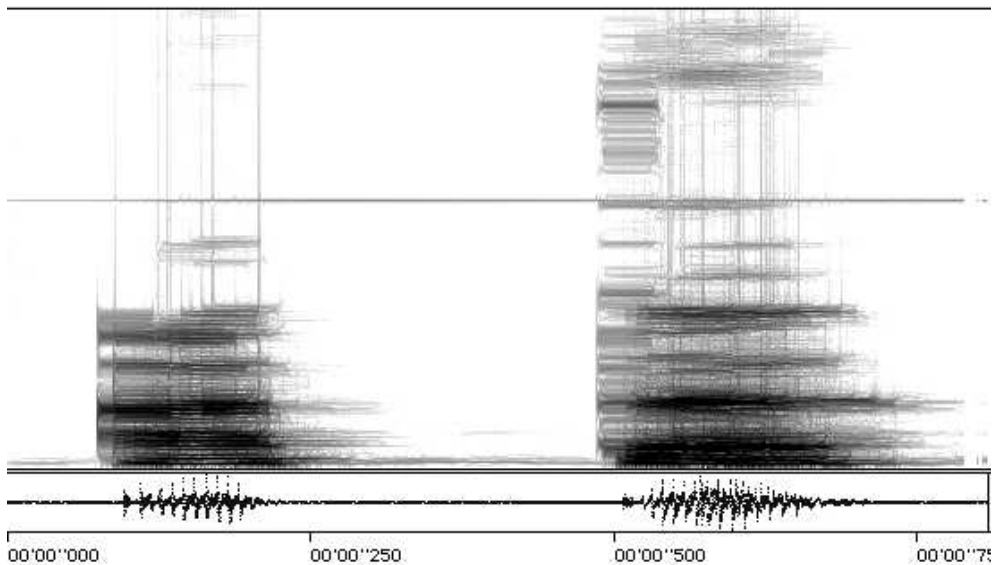

Figure 2. Sonogram of *atte* [ʔat:ə] 'to tell lies' (root word). Elicited data, the word was recorded in isolation.[11]

---

[11]   Note also the very sharp beginning of the vowel /a/ which is an acoustic clue to the presence of a glottal stop /ʔ/ as the "default" initial consonant. Its phonological distinctivity is briefly discussed in footnote 3.
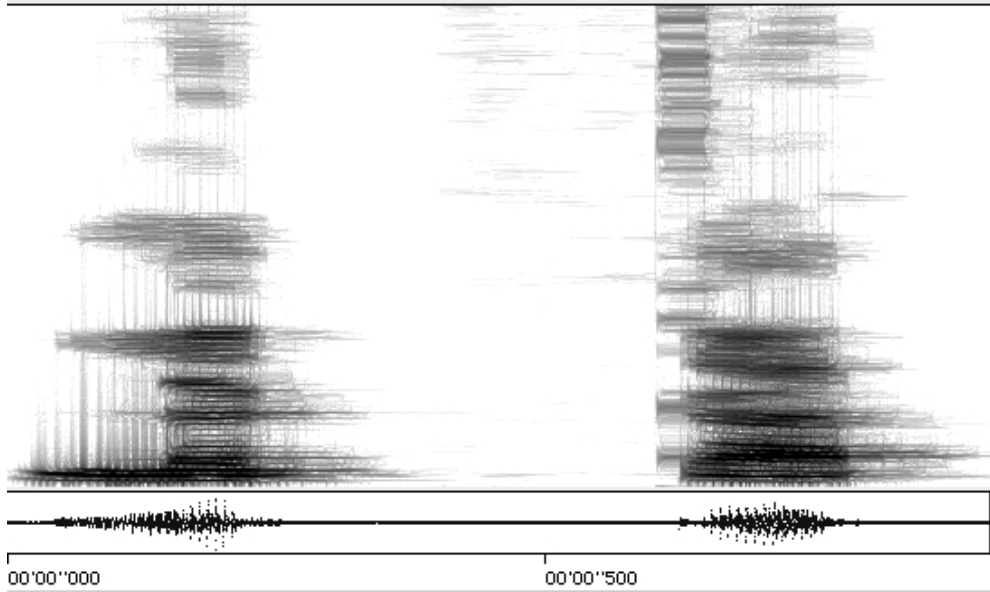
Figure 3. Sonogram of *lekkat* [lǝk:at] 'to separate, to free'. Elicited data. The word was recorded in isolation.
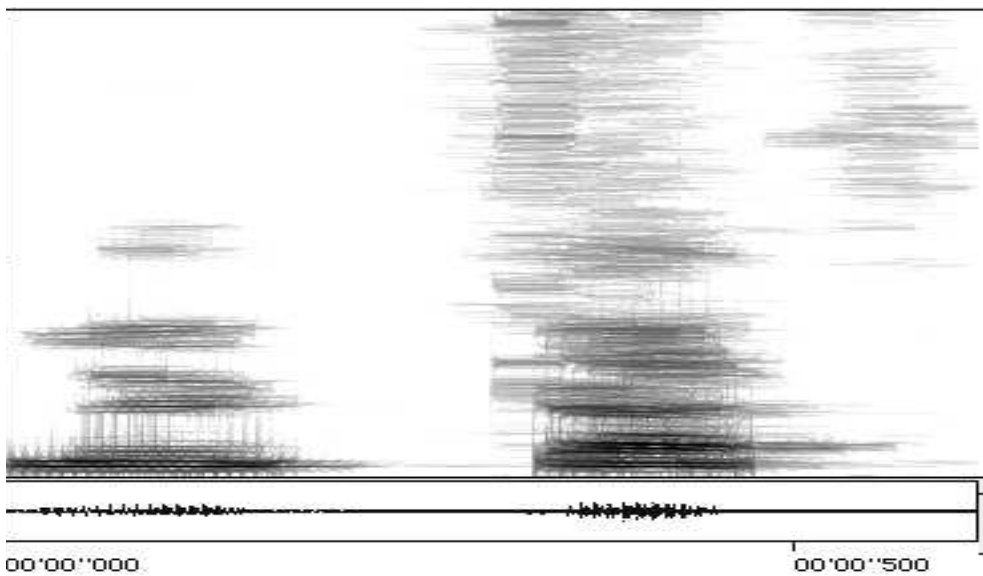

Figure 4. Sonogram of *lekkat*, realized as [lǝkat], 'to separate, to free'. Elicited data. The word was recorded within a sentence, at a normal speech tempo.

The sonogram for *ate* [atə] (see Figures 1–2) shows the average interval of approximately 0,12 seconds between the two vowels. On the other hand, between the two vowels of *atte* [at:ə], we see a micro-pause, lasting roughly 0,25 seconds, that is, twice as long.

However, it seems almost impossible to demonstrate distinctive consonant length in Bajo. In fact, the only pure minimal pair I have attested is the one illustrated in the sonograms above. Consonant gemination in Indonesian Bajo seems to have only marginal or potential phonological relevance. As a matter of fact, I often noticed that in rapid speech (allegro) no gemination can be heard. For instance, in standard conversation, *malasso* [malas:o] 'good, well' will generally be realized [malaso]. If I would ask the same speakers how I should pronounce this word, they will insist that it should be [malas:o]. This can be illustrated by the comparison between two realizations of *lekkat* 'to separate, to free' (see Figures 3–4). The first occurrence was pronounced in isolation, while the second was recorded within a sentence, at a normal speech tempo (andante).

The sonogram for *lekkat* [lək:at] spoken in isolation (see Figures 3–4) displays a micro-pause between the two vowels of approximately 0,3 seconds, while the one for *lekkat* [ləkat] realized at a normal speech tempo, shows a corresponding span of only 0,1 second. We may notice a slightly stronger stress on the first vowel of *lekkat*, but as this is elicited data, our informant may have hyper-corrected his pronunciation.

In sum, my conclusion is that in Indonesian Bajo, consonant gemination is optionally phonologically distinctive: geminate consonants may be pronounced as single consonants even at a normal speech tempo, but the opposite is not true; phonemically single consonants are not lengthened, not even in slow, careful speech.

3 Word formation in Indonesian Bajo

As a transition between the phonological and the lexical survey of Indonesian Bajo, I will first present a description of Bajo sandhi, a morpho-phonological adjustment between root words and affixes. Secondly, I will describe the main patterns of nominal and verbal word formation.

3.1 Sandhi

Like in Indonesian, sandhi in Bajo entails the insertion and/or substitution of a nasal consonant whose point of articulation approximates the initial phoneme of the root word. Yet, sandhi in Indonesian Bajo is more complex than in Indonesian. I will mainly discuss sandhi which root words undergo when prefixed by the Actor Voice morpheme *N–*. Subsequently, I will briefly describe sandhi between root words and suffixes.

The Actor Voice (av) is indicated by a proclitic morpheme that may have various phonological shapes, always involving nasal consonants and symbolized by the archiphoneme /N/. In his survey of the Bajo dialects

of the lesser Sunda Islands, Verheijen (1986: 10) noticed that "The sounds (phonemes) /h–/, /k–/, /ʔ/ (which is not spelled) are replaced by /ng–/: *kita* > *ngita*, *katonang* > *ngatonang*, *hapus* > *ngapus* [...]". This statement applies to the Indonesian Bajo as a whole: *N–* is realized /ŋ/ before words starting with /ʔ/ (a glottal stop):[12]

Root word		with the Actor Voice proclitic *N–*		
inta	[ʔinta]	*nginta*	[ŋinta]	'to eat'
ala	[ʔala]	*ngala*	[ŋala]	'to take'
upi	[ʔupi]	*ngupi*	[ŋupi]	'to dream'
usek	[ʔusɛʔ]	*ngusek*	[ŋusɛʔ]	'to mix'

Sandhi entails the aphaeresis of the initial glottal stop and likewise of the initial consonants /p/, /b/, /t/, /t͡ʃ/, /s/, /k/, /h/, which are all replaced by the closest nasal consonant: /m/, /n/, /ɲ/ or /ŋ/:

popo	[popo]	*mopoan*	[mopoʔan]	'to wash'
busai	[busaj]	*musai*	[musaj]	'to paddle'
tenun	[tənun]	*nenun*	[nənun]	'to weave'
cabo'	[t͡ʃaboʔ]	*nyabo'*	[ɲaboʔ]	'to push or plunge someone into the water'
soho	[soho]	*nyoho*	[ɲoho]	'to command'
katonan	[katɔnan]	*ngatonan*	[ŋatɔnan]	'to know'
hapus	[hapus]	*ngapus*	[ŋapus]	'to erase'

Root words beginning with /l/, /r/, /d/ and /g/ involve a special sandhi, without aphaeresis, but with the epenthesis of an [a], thus the proclitic *N–* is realized *nga–* [ŋa] or even, for certain speakers, *ngan–* [ŋan] (before /d/) or *ngang–* [ŋaŋ] before /g/), for instance:

lebas	[ləbas]	*ngalebas*	[ŋaləbas]	'to put to the sea to set afloat (esp. from a beach)'
rait	[raʔit]	*ngarait*	[ŋaraʔit]	'to sew'
dapu	[dapu]	*ngadapuan*	[ŋadapuʔan] or	
		ngandapuan	[ŋan͡dapuʔan]	'to own'

---

[12]   I propose that there are no vowel-initial words in Bajo; by default, the initial consonant is /ʔ/, thus the minimal word structure is CV.

gambar   [gambar]   *nganggambar* [ŋaŋ͡gambar]   'to draw'

As expected, the initial nasal consonants /m/, /n /, /ɲ/ or /N/ do not require any sandhi, for instance:

 mesi   [məsi]   *mesi*   [məsi]   'to fish'

Sandhi and affix combinations

Ni Luh et al. (1997 : 29) state that there is a Bajo prefix *paN–* ; their observation may have been inspired, however, by a likeness to Indonesian, which actually owns a nominalization prefix *peN–*. Moreover, their statement seems to accord only with part of the data, as appears from the following example:

 tutut   'to cover, to shut'   *panutut*   'a cover'
 talau   'to fear'   *patalau*   'a fearful person'

In fact, the nominalizing derivation can directly affect a root verb (as in Indonesian), but also a verb in the Actor Voice, with the prefix *N–*. Therefore the Bajo nominalizing morpheme is *pa–*, and the word formation may encompass the prefix *N–*, for instance:

 tutut 'to cover, to shut'   *nutut* 'to cover, to shut' *panutut* 'a cover'

Moreover, there are predictable oppositions between nominalizations with, and without, the Actor Voice prefix *N–*, as illustrated by the opposition between the confixes *pa– –an* and *pa-N– –an*:

 tilau   'to ask'   *patilauan*   'question, request'
 tilau   'to ask'   *panilauan*   'subject of a discussion'

The same pattern appears with the prefix *ma–*, indicating imperfective aspect,[13] that may be combined with the prefix *N–*:

 bunan   'to give'   *munan*   'to provide, give' *mamunan*   'is supplying'
 tuloh   'to help'   *nuloh*   'to help'   *manuloh*   'is helping'

or without the prefix *N–*, directly affixed to the root word:

---

[13]   Holton (2006) analysed the morpheme *ma* in Northeast Halmaheran and some other Eastern Indonesian Austronesian languages. He concludes that "The source of [Northeast Halmaheran] *ma* is likely to be found in [Proto-Malayo-Polynesian] **ma*, originating as a marker of stative verbs". The same remark may well be valid for the Indonesian Bajo proclitic *ma–*.

| *bunteh* | 'to marry' | *mabunteh* | 'is marrying' |
| *tappa* | 'to trust' | *matappa* | 'is believing, is trusting' |

Therefore it would be erroneous to identify a prefix **maN–*. Instead, we see a merging of *ma–* and *N–*, comparable to the combination of *ma–* and *di–* (Undergoer Voice).[14] We summarize the Bajo sandhi system[15] in Table 2.

Sandhi between the prefix *N–* and the root morphemes in Indonesian Bajo			
initial phoneme of the root word	*N–* (Actor Voice prefix)[16]	*pa–* (nominalization prefix) + *N–*	*ma–* (imperfective aspect prefix) + *N–*
consonants /p/ or /b/	*m–* (aphaeresis of /p/ and /b/) *popo > mopoan* 'to wash' *bunan > munan* 'to provide, to give'	*pam–* *popo > pamopo* 'washer' *bunan > pamunan* 'provider'	*mam–* *popo > mamopo* 'is washing' *bunan > mamunan* 'is providing'
consonant /t/	*n–* (aphaeresis of /t/) *tuloh > nuloh* 'to help' *tenun > nenun* 'to weave'	*pan–* *tunggu > panunggu* 'a person who waits'	*man–* *tuloh > manuloh* 'is helping' *tagah > managah* 'is holding'
consonants /t͡ʃ/ or /s/	*ny–* (aphaeresis of /t͡ʃ/ /s/) *cabo' > nyabo'* [ɲaboʔ] 'to push or plunge someone into the water' *soho > nyoho* [ɲoho] 'to command' *salembah > nyalembah* [ɲalɛmbah] 'to arrange'	*pany–* *panyoho* [paɲoho] 'commander'	*many–* *cinggé > manyinggé* 'is bending on his knees' *soho > manyoho* 'is commanding' *sanggut > manyanggut* 'is replying'
consonants /ʔ/, /k/ or /h/	*ng–* (aphaeresis /k/ /h/ /ʔ/) *ala* [ʔala] *> ngala* [ŋala] 'to take' *katonan > ngatonan* [ŋat nan] 'to know' *hapus > ngapus* 'to erase'	*pang–* *inte > panginte* [paŋ intə] 'sight' *katonan > pangatonan* 'knowledge'	*mang–* *ala > mangala* [maŋala] 'is taking' *kusu > mangusu* 'is wiping'

[14]   For instance *aha madikitanu* "the man you are looking at" (*ma-di-kita-nu* : IMPFCT-UV-watch-2SG)

[15]   No sandhi is needed for the flexional proclitic morphemes such as *di–* (UV) and *ta–* (UV, accidental), for instance *popo* "to wash" > *dipopo* "washed". Sandhi is also not required for combinations of prefixes that we will spell out hereafter, *ma– + sa–*, or *pa– + ka–*, for example *sampir > masasampir* "getting closer to each other" or *resah > pakaresahan* "to feel (uneasy)". Obviously, these morphemes preserve the CVCV alternation.

[16]   The information in parentheses throughout this column also holds for the parallel boxes in the other two columns.

Sandhi between the prefix *N–* and the root morphemes in Indonesian Bajo			
initial phoneme of the root word	*N–* (Actor Voice prefix)[16]	*pa–* (nominalization prefix) + *N–*	*ma–* (imperfective aspect prefix) + *N–*
Consonants /l/ or /r/	*nga–* (epenthesis of [a]) *lebas > ngalebas* [ŋaləbas] 'to put to the sea' *rait > ngarait* [ŋaraʔit] 'to sew'	*panga–* *laku > pangalaku* 'candidate for marriage' *rait > pangarait* [paŋaraʔit] 'sewer, taylor'	*manga–* *laris > mangalaris* 'is crossing' *rait > mangarait* [maŋaraʔit] 'is sewing'
consonants /d/ or /g/	*nga–* or *ngan–* or *ngang–* (epenthesis of [a] or [an] or [aŋ]) *dapu > nga(n)dapuan* [ŋadapuʔan] or [ŋandapuʔan] 'to own' *gambar > nganggambar* [ŋaŋgambar] 'to draw'	*panga–* *dapu > panga(n)dapu* 'belongings, property' *gambar > panganggam-bar* [paŋaŋgambar] 'drawer, painter'	*manga–* *manga(n)dapuan* 'is owning' *manganggambar* 'is drawing'

Table 2. The Bajo sandhi system.

Sandhi between root words and suffixes or enclitics

As opposed to Indonesian, in certain contexts Indonesian Bajo shows sandhi between the root word and its suffixes. For instance, with the very frequent suffix or enclitic *–an*,[17] the root-final /n/ must be replaced by /m/:

ansellan	'smooth oil'	*mangansellaman*	'is rubbing smooth oil (on her hair)'
aran	'name'	*diaraman*	'is called'

As for open final syllable roots, a consonant is inserted between the root word and the suffix *–an*. The epenthetic consonants are /h/, the glottal stop /ʔ/ or the approximant /w/.

ala	'to take'	*ngalahan*	'to take'
bele	'to cook'	*belehan*	'cooked food'

Quoted from an *iko-iko* epic song, the following sentence shows an example of complex word formation[18] entailing two distinct sandhi phenomena on the root word *palau* 'call':

---

[17]   There are in fact two homophonous morphemes *–an*: one is a verbal suffix and the other a nominal suffix. We will deal with this in 3.2.

[18]   Gloss: CAUS: causative; BENE: benefactive; FACT: factitive; LOC: locative; AV: actor voice; UV: undergoer voice; IMPFCT: imperfective; EMPH: emphase; PL: plural; SG: singular; ABS: absolutive; ERG: ergative; POL: polite; DET: determiner; PN: proper name; RDP: reduplication.

(1) *Aiko*   *emma'*   *pamalauwante*   *aku?*
     Ai -ko   emma'   *pa- N- palau -an -te*   *aku?*
     Why-2sg   ABS mother   FACT-AV-call-LOC-2sg POL ERG   1sg

     'Why, mother, did you make (someone) calling me?'

The word formation pattern *pa–N–verb–an*, when the verb is *palau* 'to call',
produces *pa–m–alau–wan*. Finally, the use of the glottal stop as the sandhi
consonant is essentially demarcative, for example:

basé	'wet'	[baseʔan]	'covered with water'
inta	'to eat'	[diʔintaʔan]	'eaten (a variety of food)'

The glottal stop may further undergo metathesis, such as in *la'ahan* [laʔahan]
'menstruation' (*laha* 'blood' > **laha'an* > *la'ahan*).

3.2 LEXICAL MORPHOLOGY

The productive affixes in Indonesian Bajo are *pa–, ka–, si–, sa–/da–* and *–an,*
that may be combined with each other in certain contexts. In this section, I
will provide an inventory of lexical affixes, but not of flectional prefixes like
N– (Actor Voice), *ma–* (imperfective aspect), or enclitic personal pronouns
such as *–ku* (1sg).

     In Indonesian Bajo, at least in the Kangean - South-Sulawesi - Lesser Sunda
Islands variety, there are no productive infixes. Lexical affixation (prefixes,
suffixes) represents the main means to form words, beside compounding
and reduplicating root words. A compound or reduplicated root may indeed
undergo further affixation. It is worth noting that the flectional prefix *N–* (Actor
Voice) can be embedded in lexical affixation.

     It would take too long to examine all affixes in detail, thus I summarize
them in Tables 3 and 4, respectively for noun formation and verb formation.
Then I will comment on some affixes that challenge description.

Semantic features	affixes	examples of noun formation				
human	*pa–*	*ringit*	'be angry'	>	*paringit*	'an angry person'
		kikire	'be meticulous'	>	*pakikire*	'a greedy person'
		talau	'be afraid'	>	*patalau*	'a fearful person'
human or instrument	*pa-N–w*	*talau*	'be afraid'	>	*pangitalau*	'a frightening creature, a ghost'[19]
		indat	'to see'	>	*pangindat*	'an onlooker' or 'a fortune-teller'
		bunan	'to provide, to give'	>	*pamunan*	'a provider'
		tutut	'to close, to shut'	>	*panutut*	'a cover'
result of (or affected by) a process	*–an*	*inta*	'to eat'	>	*intahan*	'food'
		baun	'to speak, language'	>	*baunan*	'statement, utterance'
		upi	'to dream'	>	*upian*	'dream, ambition'
location	*pa– –an*	*dayah*	'to eat'	>	*padayahan*	'fish auction platform'
		bulut	'to speak, language'	>	*pabulutan*	'chain of mountain'
		atai	'heart'	>	*pa'ataian*	'emotion, feeling'
nominalization of a process	*pa– –an*	*tilau*	'be good'	>	*patilauan*	'question, request'
		baun	'to speak, language'	>	*pabaunan*	'discussion'
nominalization of an actualized process	*pa– N– –an*	*tilau*	'to ask'	>	*panilauan*	'subject of a discussion'
		baun	'to speak, language'	>	*pamaunan*	'reported speech'
		bunan	'to provide, to give'	>	*pamunanan*	'donation'
		boe	'to brimg'	>	*pamoehan*	'transport, carriage'
abstraction of a (mostly human) property	*ka– –an*	*malasso*	'be good'	>	*kamalassoan*	'kindness, goodness'
		gage	'be strong'	>	*kagagean*	'strongness'

Table 3.

semantic features	affixes	Examples of verb formation				
locative	*–an*	*laris*	'to cross, to move along'	>	*ngalarisan*	'to cross (a place)'
		ala	'to take'	>	*ngalahan*	'to shift, to move (something for/ to somebody)'
		ala	'to take'	>	*alahan(-nu)*	'(you), take (something)'
		popo	'to wash'	>	*mapopoan*	'bring (something) to have it washed (by somebody)'
causative/ factitive	*pa–* *pa–N–*	*bagal*	'be big'	>	*dipabagal*	'is made bigger'
		pedi	'to feel pain, hurt'	>	*papedi*	'to cause pain, to hurt'
		katonan	'to know'	>	*pangatonan(- nu)*	'to inform (you)'
causative + motion	*pa–ka–* *pa–ki*	*dia*	'down'	>	*pakadia*	'to descend, to bring down'
		(ma)laso	'good, well'	>	*pakilaso (-nu)*	'(you must) ameliorate!'
detrimental	*ka– –an*	*uran*	'rain'	>	*kauranan*	'be sprayed by the rain'
		pedi	'pain, hurt'	>	*kapedian*	'feel pain, be hurt'
		talau	'fear'	>	*katalauan*	'be frightened'
excessive intensity	*ka– –an*	*bagal*	'be tall, big'	>	*kabagalan*	'be too tall, too big'
iterative aspect	RDP *–an*	*apo*	'to cut'	>	*ma'apo-apoan*	'cutting repeatedly, to chop'
		palu	'to hit'	>	*mapalu-paluan*	'hitting repeatedly'
reciprocal	*si–*	*bakus*	'to kiss'	>	*sibakus*	'to kiss each other'
		palu	'to hit'	>	*sipalu*	'to hit each other'
	*si-*RDP	*tunduk*	'to incline the head'	>	*situnduk- tunduk*	'to incline the head repeadly to each other'

Table 4.

The prefix pa–

There are two homonymous morphemes *pa–*. The nominalization *pa–* prefixed to stative verbs or dynamic verbs used statively (without intervening *N–*) conveys the meaning of 'someone who is', 'someone who has the property of', 'someone who usually performs an action, has an occupation', for example *talau* 'be afraid' > *patalau* 'a fearful person'. When referring to a dynamic process, the prefix *N–* (Actor Voice) is used, resulting in a compound prefix

pa–N– conveying the generally meaning 'a person who is [verb]ing' or 'an instrument for [verb]ing', for instance *tutut* 'to close, to shut' > *panutut* 'a cover'. Readers who are familiar with Indonesian will recognize that the use of the Malay-Indonesian prefix *peN–* encompasses the two Bajo nominalizing prefixes *pa–* and *pa–N–*.

Like in Tagalog and other Austronesian languages, the verbal prefix *pa–* has a causative meaning. In Indonesian Bajo, *pa–* is a frequently used prefix whose meaning fits into the broad definition of causativity (the subject makes that someone or something acquires a new property), including factitivity (the subject makes that someone performs the action of the base [verb]).

(2) *Lamun na   patedeku   soppét itu   ka kite…*
     lamun na   pa– tede –ku soppét itu   ka kite
     if   will   CAUS- pass-1SG boat   DET   to 2SG POL
     'If I hand over this boat to you …'

(3) *emma é...   pasinggete   mbo' Panai.*
     emma é...   pa– singge –te mbo' Panai
     mother emph   FACT- visit   -2SG POL ERG elder PN
     'Mother, oh … Let grand-father Panai come in.'

The verbal prefix *pa–* is often lexicalized; i.e. some root words never appear unaffixed, for instance *pabolo* 'to wake up', *paléa* 'to lay down, to rest', *palau* 'to call, to cry out'. The prefix *pa–* also appears in a few verbs of perception, such as *pakalé* 'to listen', from *kalé* 'to hear'.

It may be worth noting that for stative root verbs, causative *pa–* will occur only in the undergoer voice (prefix *di–*). For instance, from the stative verb *bagal* 'be big', one can form *mabagal* 'enlarging', but not **mapabagal* '*making enlarged'. For the undergoer voice, however, the form *dipabagal* 'enlarged' is correct, but not **dibagal*.

(4) *Foto   iru na   dipabagalne.*
     Foto   iru na   di– pa– bagal –ne.
     photo   DET will   UV- CAUS- big   -3SG ERG
     'This photography will be enlarged by him.'

The causative prefix *pa–* may appear jointly with the suffix *–an*, especially if this suffix carries a benefactive meaning.

inun   'to drink'   *dipanginuman* 'is provided (with …) to drink'
aran   'name'   *dipangaraman* 'is named, is called'

| *buntéh* | 'to marry' | *dipabuntéhan* | 'has been married ("united in holy matrimony")' |

Causative *pa–* is also compatible with the reciprocal prefix *si–*, and the prefix *ka–* or *ki–* (motion, tendency):

temu	'to meet'	*mapasitemu*	'making (people) meet'
temu	'to meet'	*dipasitemuan*	'be met with, introduced to each other'
dilaut	'at the sea'	*dipakadilaut*	'made afloat, launched to the sea'
talau	'be afraid'	*pakitalauan*	'make (people) afraid'

It is doubtful that the Bajo prefix *pa–ka–* originates directly from the Proto-Austronesian causative prefix **paka–*. Rather, I believe that it embeds the preposition *ka* 'to, towards'. Moreover, I did not find any satisfactory explanation regarding *pa–ki–*, which is a rare compound prefix but for the fact that it usually occurs within an irrealis predicate (that is, in imperative and interrogative sentences).

(5) *Sesemé Same   pakiemmanu nyo'   ana'kode Boloh.*
     sesemé Same   pa– ki– emma –nu nyo'   ana'kode Boloh
     torn cloth   pn caus- to- mother -2sg   say captain   pn
     'You, Sama in torn clothes, go to your mother!' said captain Boloh.'

The suffix –an

There are two homophonous morphemes *–an* : a nominal suffix and a verbal one. The first one nominalizes the root word and expresses the result of a process or an entity that is affected by the process. It also signals that the noun refers to a collection of entities, reference to which generally remains unexpressed.

bele	'to cook'	*belehan*	'cooked food'
baun	'to speak; language'	*baunan*	'statement, utterance'
upi	'to dream'	*upian*	'dream, ambition'

The suffix *–an* is frequently combined with the compound prefix *pa-N–*:

bunan	'to give'	*pamunanan*	'donation, present'
(ng)anjame	'work'	*pangajamean*	'profession, job'
loros	'to harvest'	*pangalorosan*	'crop, harvest'

It also appears in the confix *pa– –an*, without intervening *N–*, to form nouns that refer to places, such as *pinang* 'betel nut' > *papinangan* 'betel tray'. In Indonesian Bajo (but apparently not in other Sama-Bajau languages), there are *ka– –an* confixes, probably borrowed from Malay-Indonesian *ke– –an*, and similarly indicating excessiveness of a property, detrimental event or expressing an abstract notion, for instance:

pare	'a lot of'	*kaparehan*	'too many'
lumpan	'a speck of dust'	*kalumpanan*	'to have a speck of dust in the eye'
nia	'there is'	*kaniahan*	'execution, achievement'

The verbal suffix *–an* also conveys a locative meaning; it is however a very broadly defined locative. For any transitive verb of which the object can be seen as a location or destination of the action, the verbal root will be affixed with *–an*, see example (1). Therefore, it is highly frequent,[19] encompassing other semantic values including benefactive. However, in Kangean Bajo, the strictly locative use of *–an* (when it indicates a point in space) tends to fade because it is often reinforced or replaced by the preposition *ma* "at, in, into". We leave the challenge of explaining this optional presence of the locative preposition *ma* within an object phrase to syntacticians. The suffix *–an* cannot be employed with intransitive verbs, as in the second sentence of example (8).

(6) *Iye ngalarisan (ma) romah Kabaena.*
     3SG AV-cross-LOC (in) forest   NP
     'He crossed the forest of Kabaéna.'

(7) *Taguanku   gole   (ma) gelas.*
     pour-LOC-1SG ERG   sugar   (in) glass
     'I pour sugar into the glass.'

(8) *Iye nganjame ma Jakarta.*   **Iye nganjamean (ma) Jakarta.*
     3SG AV-work in Jakarta.   3SG AV-work-LOC (in) Jakarta
     'He works in Jakarta.'

The locative meaning of *–an* often applies to animates, including humans, as in examples (1) and (9), or to inalienable parts of the body, as in example (10).

---

[19]   Donohue (1996: 788) comments on the locative meaning of *–an* in Tukang Besi Bajo.

(9)   *Iye nekahan mbo'ne.*   *Mbo'ne ditekahanne.*

   Iye   n– (t)eka–an   mbo'–ne.   *Mbo'–ne   di– teka –an -ne.*

   3SG AV-visit -LOC   elder-3SG   elder-3SG UV-visit-LOC-3SG

   'He visits his grand-father.'   'His grand-father has his visit.'

(10)   *Lolohne magagalunsuran tanganne ma cocorah.*

   Loloh –ne   ma–   ga– galunsur–an   tangan–ne ma cocorah.

   Follow-3SG ERG IMPFCT-RDP-descend-LOC   hand -3SG on ladder frame

   'He follows the ladder frame, by lowering step-by-step his hand.'

Finally, the following excerpt from an *iko-iko* epic song displays several examples of the affixes that we have presented.

(11)   *Mbo' Panai, pakadianu lélépenu beke busaiannuku ka darat.*

   Mbo' Panai, pa– ka– dia –nu   lélépe –nu

   elder PN   CAUS- to- below-2SG ERG proa -2SG

   beke   busai –an   –nu   –ku   ka darat.

   with Ø- paddle -BENE -2SG ERG -1SG ABS to ground

   'Grand-father Panai, put your proa to the sea then paddle (bring) me to the shore.'

4 CONCLUSION

The Sama-Bajau languages are spoken by a diaspora of small communities scattered on some shores of the Southern Philippines, Sabah (Malaysia) and many islands in the eastern part of Indonesia. The data presented in this paper were collected in Indonesia, from several Bajo communities around the Flores Sea (Kangean, Lesser Sunda Islands, South-East Sulawesi).

   As opposed to various languages of the Sama-Bajau branch, no vowel lengthening is perceptible in Bajo. The gemination of certain consonants is audible, but it is only marginally and potentially phonemic. Through sonogram interpretation, the acoustic features of this gemination have been described. Pre-nasalized consonants can be analysed as single phonemes. Sandhi proves to be more complex in Bajo than in Indonesian, because it involves phonetic adjustments between root words and suffixes as well as prefixes, and entails some aphaeresis and epenthesis.

   At first glance, Bajo affixes seem to have cognate counterparts in Indonesian. However, this impression does not hold when we confront the data. For instance, the compound prefix *pa–N–* should not be confused with the Malay-Indonesian *peN–*. The verbal prefix *pa–N–* is in effect a combination of the causative prefix *pa–* and the actor voice prefix *N–*. Besides the verbal *pa–N–* there is also a compound nominalizing prefix *pa–N–* in Bajo. The same

kind of homophony is observed with regard to the suffix *–an*, which may concern noun formation and verb formation.

References

Donohue, Mark. 1996. "Bajau: a symmetrical Austronesian language", *Language - Journal of the Linguistic Society of America* 72(4): 782-793.

Himmelmann, Nikolaus P. 2002. "Voice in western Austronesian: an update", in: F. Wouk and M. Ross (eds), *The history and typology of Western Austronesian voice system*, pp. 7-16. Canberra: Pacific Linguistics.

Himmelmann, Nikolaus P. 2005. "The Austronesian languages of Asia and Madagascar: typological characteristics", in: A. Adelaar and Nikolaos P. Himmelmann (eds), *The Austronesian languages of Asia and Madagascar*, pp. 110-181. New York: Routledge.

Hinayat, Mohamad Said. 2003. *Glosari Bahasa Bajau/Sama - Bahasa Melayu.* Kinabalu: Persatuan Seni Budaya Bajau, Sabah.

Holton, Gary. 2006. "The relational noun marker in Tobelo (Northeast Halmaheran)". Paper, The Tenth International Conference on Austronesian Linguistics (10-ICAL) Puerto Princesa City, Palawan, the Philippines.

Jun, Akamine. 2005. "Sama (Bajau)", in: A. Adelaar and Nikolaos P. Himmelmann (eds), *The Austronesian languages of Asia and Madagascar*, pp. 377-396. New York: Routledge.

Ni Luh, Komang Candrawati et al. 1997. *Struktur Bahasa Bajo,* Jakarta: Pusat Bahasa.

Nuraini, Chandra. 2008. *Langue et production de récits d'une communauté Bajo des îles Kangean. Thèse de doctorat en ethno-linguistique, soutenue le 14 juin 2008*. PhD thesis, Université de La Rochelle.

Pallesen, Kemp. 1985. *Culture contact and language convergence.* Manila: Linguistic Society of the Philippines.

Verheijen, Jilis A.J. 1986. *The Sama/Bajau language in the lesser Sunda islands.* Canberra: Pacific Linguistics.