

Properties of the Active Voice Prefix in Colloquial Jakartan Indonesian: Final Project

Riley Kong

24.909 Field Methods in Linguistics

December 22, 2023

1 Introduction

Colloquial Jakartan Indonesian (CJI) uses an actor voice prefix (AVP) to indicate that a verb is in active voice (AV).

The AVP exhibits a variety of allophones, which behave in complex ways according to the properties of the verb root. In particular, the specific AVP allophone chosen is determined by a combination of the initial segment, transitivity, and presence of affixes on the verb.

Previous work on the AVP within CJI (Sneddon 2006) has focused on describing allophones, but it does not give a full picture of how this allophonic variation arises. In this report, I describe the AVP allophones in full, noting where my account differs from Sneddon 2006.

I give rule- and constraint-based accounts of their derivation, which explain when the AVP applies optionally versus obligatorily. A fuller Optimality Theory treatment of the AVP leads to some theoretical implications for the nature of well-formed constraints.

1.1 Phonological inventory

The consonant and vowel inventories of Standard Indonesian (SI), from Sneddon et al. 2010, are given in Tables 1 and 2, respectively. This inventory seems to be the same as that of CJI: I did not encounter any additional phonemes in my elicitation data.

	Labial	Alveolar	Palatal	Velar	Glottal
Nasal	m	n	ɲ	ŋ	
Stop	p b	t d		k g	(ʔ)
Affricate			tʃ dʒ		
Fricative	(f v)	s (z)	(ʃ)	(x)	h
Approximant	w	l	j		
Trill		r			

Table 1: Consonant inventory of Indonesian

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

Table 2: Vowel inventory of Indonesian

The phonemes /f, v, z, ʃ, x/ are marginal and only used in loanwords (Sneddon et al. 2010). In particular, these phonemes do not appear in any elicited words, so they are not considered in this report.

2 Transitive verbs

With most transitive verbs (but see 6.1), the AVP is optionally applied in active voice.

- (1) a. aku tulis surat ini
I write letter this
- b. aku nulis surat ini
I AV.write letter this
'I write this letter'
- (2) a. aku lagi sapu lantai
I IPFV sweep floor

- b. aku lagi sapu lantai
I IPFV AV.sweep floor
- c. aku lagi ηə-sapu lantai
I IPFV AV-sweep floor
'I'm sweeping the floor'

Examples (1a) and (2a) use the bare form of the verb, while the other examples have the AVP applied.

There exists one special case: AVP + *bəli* 'buy' cannot surface as **ηəbəli* but rather only as *mbəli*. All other verbs elicited beginning with /b/ could take the *ηə-* allomorph, so I believe that this is just a lexical quirk of *bəli*.

2.1 Monosyllabic verbs

Before monosyllabic verb bases, such as *tʃat* 'paint' or *bom* 'bomb', the AVP must take the form *ηə-* and cannot take the NS/NA forms.

- (3) a. aku tʃat rumah-ŋa
I paint house-DET
- b. aku ηə-tʃat rumah-ŋa
I AV-paint house-DET
- c. *aku ŋat rumah-ŋa
I AV.paint house-DET
'I paint the house'
- (4) a. mərəka bom rumah itu
they bomb house that
- b. mərəka ηə-bom rumah itu
they AV-bomb house that
- c. *mərəka mbom rumah itu
they AV.bomb house they
'They bombed that house'

See Section 5 for a full treatment of monosyllabic verbs.

3 Allomorphs of the AVP

The allophones of the AVP are conditioned on the initial segment of the stem. Following Blust 2004, we categorize the allophones of the AVP as follows.

- Before the voiceless obstruents /p, t, k, s, tʃ/, the AVP is expressed as **nasal substitution** (NS), the replacement of the initial consonant with a nasal at the same place of articulation¹.
- Before the voiced obstruents /b, d, g/, the AVP is expressed as **nasal accretion** (NA), the addition of a nasal at the same place of articulation.
- Before the consonants /b, d, g, s, tʃ, h, l, r, w/, the AVP is expressed as a **prefix** $\eta\partial$ -, and before vowel-initial stems, the AVP is a prefix η -. Additionally, before /l/, the AVP can surface as η -, forming the cluster ηl - (see 7.2 for an extended discussion).
- Before the nasals /m, n/, the AVP cannot surface: ie. only *makan*, and not ** $\eta\partial$ makan*, is possible. In my analysis, I choose to analyze this as a zero prefix \emptyset -, which arises because of the impossibility of nasal-nasal clusters (see Section 4).

I was not able to elicit any transitive verbs with a root starting with /dʒ, ŋ, j/.

Note that /b, d, g, s, tʃ/ have two possible allomorphs: $\eta\partial$ - and NA/NS.

This data differs from the CJI data described in Sneddon 2006 in that /s, tʃ/ are not recorded as having two allomorphs, instead only describing the NS allomorph.

¹/s/ is replaced by /p/, which at first glance is in a different place of articulation; but note that the palatal sibilant /ʃ/ doesn't appear in native roots, so /s/ could be filling this gap.

Verb root	Active voice form	AVP allophone
<i>potonj</i> ‘cut’	<i>motorj</i>	NS
<i>tanam</i> ‘plant’	<i>nanam</i>	NS
<i>kasih</i> ‘give’	<i>ηasih</i>	NS
<i>sapu</i> ‘sweep’	<i>ηapu</i> <i>ηəsapu</i>	NS ηə-
<i>tfutfi</i> ‘wash’	<i>ηutfi</i> <i>ηətʃutfi</i>	NS ηə-
<i>banjun</i> ‘build’	<i>mbanjun</i> <i>ηəbanjun</i>	NA ηə-
<i>dənar</i> ‘hear’	<i>ndənar</i> <i>ηədənar</i>	NA ηə-
<i>gigit</i> ‘bite’	<i>ηgigit</i> <i>ηəgigit</i>	NA ηə-
<i>hadijah</i> ‘gift’	<i>ηəhadijah</i>	ηə-
<i>rasah</i> ‘think’	<i>ηərasah</i>	ηə-
<i>wawantfara</i> ‘interview’	<i>ηəwawantfara</i>	ηə-
<i>lukis</i> ‘paint’	<i>ηəlukis</i> <i>ηlukis</i>	ηə- η-
<i>ambil</i> ‘take’	<i>ηambil</i>	η-

Table 3: AVP forms before different verb roots

4 Rule-based analysis

Following a rule-based analysis of the AVP, we posit that the underlying representation of the AVP is / η -/, on which the following rules are applied:

1. $\eta \rightarrow \emptyset / _ [+nas]$ (deletion of η -)
2. $\emptyset \rightarrow \text{ə} / \eta _ [b, d, g, s, tʃ, h, l, r, w]$ (epenthesis of /ə/)
3. $\eta \rightarrow [\alpha place] / _ [-son, \alpha place]$ (assimilation of η - to an obstruent)
4. $[-voi, \alpha place] \rightarrow \emptyset / [+nas, \alpha place] _$ (deletion of unvoiced stop)

Rule 2 is optional. Application will produce the $\eta\text{ə}$ - allomorph, while non-application produces the other forms.

The choice of consonants for Rule 2, /b, d, g, s, tʃ, h, l, r, w/, seems unmotivated. Indeed, we cannot express these consonants completely using traditional distinctive features:

- We cannot separate based on voice, as the set contains both voiced and unvoiced consonants;
- We cannot separate based on place, as every place of articulation is represented in the set;
- We cannot separate based on manner, as excluding voiceless stops like /p/ would necessarily exclude voiced stops like /b/.

However, note that the only consonants for which Rule 2 cannot apply are the unvoiced stops and nasals, which are precisely the consonants that do not involve any airflow through the oral cavity. Hence, while not forming a natural class, they have a phonological basis—the consonants which involve oral airflow. In the constraint-based analysis (see Section 5), $*\text{ə} // \text{CLOSED}$ is an expression of this set of consonants.

To order these rules, we note that:

- (1) bleeds (2). Before nasals, (1) deletes the AVP, so /ə/ will not be inserted.
- (3) feeds (4). Assimilation of the nasal segment to the following consonant produces a context for (4).
- (2) bleeds (3). Application of (2) introduces an intervening segment /ə/ that blocks the application of the assimilation (and deletion) rules.

Hence, (1)-(2)-(3)-(4) form an ordered chain.

We can walk through some verbs to see how the rules generate the correct SR forms.

UR	η-tulis	η-sapu	η-gigit	η-lihat	η-ambil	η-makan
(1)						makan
(2)		ηsapu, ηəsapu	ηgigit, ηəgigit	ηlihat, ηəlihat		
(3)	ntulis	ηsapu, ηəsapu				
(4)	nulis	ηapu, ηəsapu				
SR	nulis	ηapu, ηəsapu	ηgigit, ηəgigit	ηlihat, ηəlihat	ηambil	makan

Table 4: Rules in action

Empty cells represent that the rule is not applied with a specific input form. We see that:

- (1) applies before nasal-initial verbs to delete the *η*- prefix.
- (2) produces the optional *ηə*- allophones for the consonants that allow its application.
- (3) assimilates *η*- before obstruents ...
- and (4) deletes unvoiced obstruents.

One problem with this rule-based analysis is that it cannot adequately explain monosyllabic verbs. Because the initial consonant of *tʃat* is the same as that of verbs like *tʃutʃi* ‘wash’, the rules would predict that both words can undergo nasal substitution, but in reality only *tʃutʃi* can do so. Therefore, we cannot rely on only the immediate context of the AVP to derive the allomorphic variation.

Indeed, this would be difficult to explain from a purely rule-based description. This motivates our analysis of the AVP in a constraint-based approach, where we have access to morphological properties of the verb root.

5 Constraint-based analysis

To model the form of the AVP allomorphs of simple transitive verbs as in 4, I introduce the following constraints, following Pater 2001:

- MINBASE (ie. $\sigma\sigma$): Outputs are minimally disyllabic. Mark one * for a monosyllabic output.

- ▶ AGREEPLACE: Mark one * for each pair of consecutive consonants $s_1 - s_2$ where s_1 is [α place] and s_2 is [$-\alpha$ place].
- ▶ *NT: Mark one * for each pair of consecutive consonants $s_1 - s_2$ where s_1 is [+nas] and s_2 is [-voi].
- ▶ *NN: Mark one * for each pair of consecutive consonants $s_1 - s_2$ where s_1 is [+nas] and s_2 is [+nas].
- ▶ *CC: Mark one * for each pair of consecutive consonants $s_1 - s_2$.
- ▶ *ə//CLOSED: Mark one * for each pair of consecutive consonants $\text{ə} - s$ such that s doesn't involve airflow through the oral cavity.
- ▶ DEP ə: Mark one * for each output instance of /ə/ with no input correspondent.
- ▶ MAXROOT: Mark one * for each missing feature of the verb root.

The constraint MINBASE, which marks monosyllabic outputs, has precedent in OT. There is a crosslinguistic tendency for monosyllabic words to be marked in some way (Batais and Wiltshire 2018). In Indonesian, monosyllabic verbs are primarily loanwords, and the vast majority of verbal bases are disyllabic. (In particular, *tʃat* derives from Southern Min *chhat* 'paint', and *bom* from Dutch *bom* 'bomb'.) It would then be logical that monosyllabic verbs would be marked by a constraint. This is similar to the approach taken by Moira 1993 for Cantonese loanwords.

Note that MAXROOT applies to features, not segments. This allows us to eliminate candidates that change only one feature of a consonant. Another equivalent way to express this is by having an IDENTROOT constraint that is ranked lower than MAXROOT; ie. removal of a root segment is worse than removing one feature. In the following analysis, to simply the tableaux, I assume that each segment is composed of exactly two features. (This is not really true, but all that really needs to happen is for segmental loss to be more marked than featural loss.)

Violation of AGREEPLACE is what motivates either the addition of ə (at the cost of DEP ə) or the nasal assimilation of η - (at the cost of MAXROOT), and further deletion of unvoiced stops is driven by *NT, at the cost of Max[-Nasal].

Consider the verb *tulis* ‘write’, where the $\eta\partial$ - allomorph is not licensed:

/ η -tulis/	AGREEPLACE	*NT	*NN	* ∂ //CLOSED	MAXROOT	DEP ∂
a. η tulis	*!	*				
b. ntulis		*!				
c. η nulis			*!		*	
d. η ulis					**!	
e. $\eta\partial$ tulis				*!		*
f. nulis					*	

Table 5: Constraints with *tulis* ‘write’

The candidate ** η ulis* involves removal of both the place and manner features of the root-initial consonant, so it incurs two MAXROOT violations and is filtered out.

Now consider the verb *ba η un* ‘build’, where the $\eta\partial$ - allomorph is licensed:

/ η -ba η un/	AGREEPLACE	*NN	MAXROOT	*CC	DEP ∂
a. η ba η un	*!			*	
b. η ma η un		*!		*	
c. ma η un			*!		
d. $\eta\partial$ ba η un					*
e. mba η un				*	

Table 6: Constraints with *ba η un* ‘build’

*CC and DEP ∂ are variably ranked: *CC » DEP ∂ produces the form *$\eta\partial$ ba η un*, while the opposite ranking produces *mba η un*.

Note that **ma η un* is eliminated in this case, unlike with *nulis*, because there exist candidates which do not modify the base at all and hence incur no MAXROOT violations.

This variable ranking also cleanly explains the special case of *b ∂ li*. This verb imposes the

ranking $\text{Dep } \emptyset \gg *CC$, so $*\eta\text{əb}\eta\text{un}$ violates $\text{Dep } \emptyset$ and falls from consideration.

We unify this analysis with the MINBASE constraint for monosyllabic verbs:

/ η -tʃat/	AGREEPLACE	MINBASE	* \emptyset //CLOSED	MAXROOT	DEP \emptyset
a. η tʃat	*!				
b. η at		*!		*	
η tʃat			*		*

Table 7: Constraints with *tʃat* ‘build’

/ η -bom/	MINBASE	* \emptyset //CLOSED	MAXROOT	*CC	DEP \emptyset
a. mbom	*!			*	
η əbom		*			*

Table 8: Constraints with *bom* ‘bomb’

When MINBASE is introduced, monosyllabic outputs like $*\eta$ at are marked. Therefore the alternative η tʃat is chosen, even though it incurs $*\emptyset$ //CLOSED and DEP \emptyset violations. The power of MINBASE is that it is not active for multisyllabic verb roots, so we can introduce MINBASE without complications to our existing analyses.

In this way, a constraint-based analysis can explain the special case of monosyllables and generalize to all initial segments in a typologically sound way.

We rank these constraints as follows:

- Table 5 implies AGREEPLACE, *NT, *NN, $*\emptyset$ //CLOSED \gg MAXROOT.
- Table 6 implies MAXROOT \gg *CC, DEP \emptyset .
- *CC and Dep \emptyset have a variable ranking, producing two different allomorphs.
- Tables 7, 8 implies MINBASE \gg $*\emptyset$ //CLOSED.

We can express this as a Hasse diagram, where \longleftrightarrow represents a variable ranking:

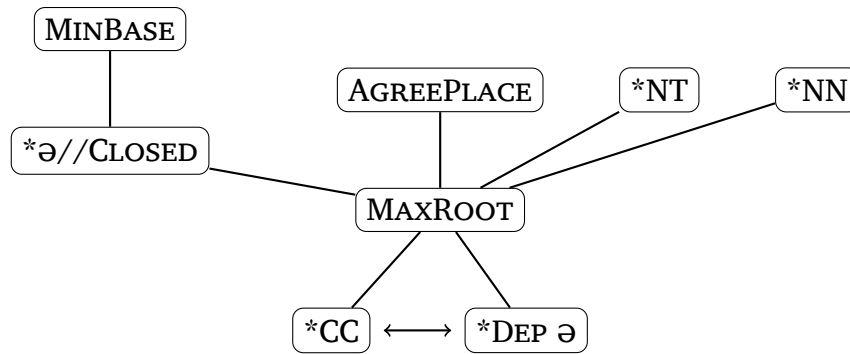


Figure 1: Hasse diagram of constraints (simple transitives)

6 Intransitive verbs

Most intransitive verbs do not accept the AVP, even where the verb is unergative.

- (5) a. aku pərgi ke Indonesia
I go to Indonesia
b. *aku mərgi ke Indonesia
I AV.go to Indonesia
'I'm going to Indonesia'

Some verbs incorporate a nasalized component as in the AVP, like *taŋis* 'cry'.

- (6) dʒaŋan taŋis-in dia
don't cry-BEN her
'Don't cry for her'
(7) a. aku lagi naŋis
I IPFV AV.cry
b. *aku lagi taŋis
I IPFV cry
'I cry'

While the root is *taŋis*, as evidenced by the form *taŋis-in*, we need to add the AVP while in active voice. Other verbs that take an obligatory AVP include *ŋorok* < *orok* 'snore', *ŋərasah* < *rasah* 'feel', and *ŋgonggoy* < *gonggoy* 'bark'. The generalization here is that in intransitive verbs, the AVP is mandatory if it is present.

To shed light on this phenomenon, we turn our attention to other intransitive verbs with an associated prefix. Consider the verb *mə-lətus* 'pop':

- (8) aku lətus-in balon-ŋa
 I pop-CAUS balloon-DET
 ‘I popped the balloon’
- (9) a. * balon-ŋa lətus
 balloon-DET pop
- b. * balon-ŋa ɲə-lətus
 balloon-DET ɲə-pop
- c. balon-ŋa mə-lətus
 balloon-DET mə-pop
 ‘The balloon popped’

The verbal root of this verb is *lətus*, as shown by the fact that *lətus-in* is derived from this verb. However, when used intransitively, the form *mə-lətus* must be used. This parallels the case of *naŋis* ‘cry’, where the prefix is instead *mə-*.

Other verbs with this property include *bər-antəm* ‘fight’ and *kə-luh* ‘be annoyed’. It is the case, in my data as well as in Sneddon et al. 2010, that *bər-* and *kə-* only apply to intransitive verbs.

I posit that the AVP in intransitive verbs is simply one of many prefixes that marks intransitivity. These prefixes are obligatorily applied in active voice.

6.1 Labile verbs

A few transitive verbs behave similarly to intransitive verbs in that they do not display AVP-optionalty.

Some verbs cannot take the AVP, like *bitʃara* ‘speak’:

- (10) a. aku bitʃara bahasa Indonesia
 I speak language Indonesia
- b. * aku mbitʃara bahasa Indonesia
 I AV.speak language Indonesia
- c. * aku ɲə-bitʃara bahasa Indonesia
 I AV-speak language Indonesia
 ‘I speak Indonesian’

While *bitʃara* is transitive because it takes an object, namely *bahasa Indonesia* ‘Indonesian language’, the AVP cannot be applied.

Some verbs like *tari* ‘dance’ instead are obligatorily nasalized:

- (11) a. aku lagi nari waltz
I IPFV AV.dance waltz
b. *aku lagi tari waltz
I IPFV dance waltz
‘I’m dancing a waltz’

We know that *tari* is the verbal root because of forms like *tarian* ‘dance’, where the bare root is nominalized. However, the form that is used in active voice is not *tari* but rather *nari*.

These verbs are termed “pseudo-transitive” in Sneddon et al. 2010. Other examples include *bəladʒar* ‘study’, *main* ‘play’, *mənaŋ* ‘win’, and *tari* ‘dance’.

All of these verbs can also act as intransitive unergatives without an object: they are labile verbs. Indeed, there is something special about the objects of these verbs. They do not pass the objecthood test of passivization, since they cannot be passivized using the passive voice marker *di-*. They can, however, be placed in Object Voice:

- (12) *bahasa Indonesia di-bitʃara aku
language Indonesia PASS.speak I
Intended: ‘Indonesian is spoken by me’
(13) bahasa Indonesia itu bahasa jaŋ aku bitʃara
language Indonesia DEF language REL I speak
‘Indonesian is the language that I speak’
(14) *waltz-ja ditari aku
waltz-DET PASS.dance I
Intended: ‘The waltz was danced by me’

This indicates that these labile verbs act like intransitive verbs with regard to the AVP, and they take an additional argument that is semantically an object but doesn’t behave quite the same syntactically. Sneddon et al. 2010 calls these objects “complements”, and considers them a separate word class.

Further evidence comes in the form of labile verb-noun phrases that act as one semantic unit. Consider the phrase *sikat gigi* ‘to brush one’s teeth’:

- (15) a. aku sikat gigi
I brush teeth

- b. *aku *nikat* gigi
 I AV.brush teeth
 ‘I brush my teeth’
- (16) a. aku *sikat* gigi-*na*
 I brush teeth-3.POSS
- b. aku *nikat* gigi-*na*
 I AV.brush teeth-3.POSS
 ‘I brush his teeth’

When used as a normal transitive verb, as in (16), *sikat* can take either the bare form or the AVP form.

However, the phrase *sikat gigi* refers specifically to brushing one’s teeth, and cannot take the AVP. This indicates that *sikat gigi* is being reinterpreted as a labile verb with the incorporation of a verbal complement. In general, labile verbs can be interpreted as forming verbal phrase constituents that then behave similarly to intransitive verbs.

The phenomenon of noun incorporation inducing a transitivity change is well-attested cross-linguistically (see Smirnova and Shustova 2017 for a treatment of English noun incorporation). In this light, labile verbs are an example of this change of valency being explicitly marked as prefixal obligatoriness in active voice.

7 Complex verbs

7.1 The *-in* prefix

The *-in* prefix attaches to verbs and generally has a benefactive or applicative meaning (among other uses). In particular, *-in* often generates a transitive verb, on which the AVP is applied.

- (17) a. *lampu-na mati*
 light-DET off
- b. **lampu-na nə-mati*
 light-DET AV.off
 ‘The lights are off’
- (18) a. aku *mati-in* lampu
 I turn-off-CAUS light

- b. aku $\eta\partial$ -mati-in lampu
 I AV-off-CAUS light
 ‘I turned off the lights’

All of the allomorphs listed in Section 3 are possible with *-in*, with the addition of the $\eta\partial$ - allomorph before all consonants. In other words, rule (1) can now be applied before obstruents. In fact, the $\eta\partial$ - form seems to be the preferred and spontaneously given form in elicitation.

- (19) a. aku tanam buḡa
 I plant flower
 b. aku nanam buḡa
 I AV.plant flower
 c. *aku $\eta\partial$ -tanam buḡa
 I AV.plant flower
 ‘I plant flowers’
 a. aku tanam-in buḡa buat dia
 I plant-BEN flower for him
 b. aku nanam-in buḡa buat dia
 I AV.plant-BEN flower for him
 c. aku $\eta\partial$ -tanam-in buḡa buat dia
 I AV-plant-BEN flower for him
 ‘I plant flowers for him’

Note that vowel-initial verbs cannot take the $\eta\partial$ - allomorph: *adzar-in* ‘teach’ cannot be ** $\eta\partial$ -adzar-in* but rather only *η -adzar-in*.

From a purely phonological perspective, this is surprising, since the initial consonant does not change with the addition of *-in*. We draw on Pater 2001 and claim that the verbal root and *-in* together form a *prosodic word* (Pword), which resists change by external components. Epenthesis of /ə/, as in (1), then becomes possible to ensure that impossible clusters like **/ η t/* are formed.

In a constraint-based treatment, this corresponds to *CRISPEDGE as described in Pater 2001:

- CRISPEDGE: No element within a prosodic word may be linked to a category external to the word.

CRISPEGE then competes with *ə//CLOSED to enforce the *ŋə*- allophone at the cost of introducing a marked /ə/-obstruent sequence.

/ŋ-tanam-in/	AGREEPLACE	*NT	CRISPEGE	*ə//CLOSED	MAXROOT	DEP ə
a. ŋtanamin	*!	*				
b. ntanamin		*!				
☞ c. nanamin			*		*	
☞ d. ŋətanamin				*		*

Table 9: *tanamin* ‘plant for someone’

Note that there is variable ranking of CRISPEGE and *ə//CLOSED. Ranking CRISPEGE » *ə//CLOSED results in the *ŋə*- form, while *ə//CLOSED » CRISPEGE results in the NS/NA form.

7.2 Reduplication

Verbs can be reduplicated to indicate a variety of functions, including habitual or imperfect mood. When reduplicated, the AVP surfaces differently depending on the allomorph. It can surface on both the base and reduplicant, or remain uncopied.

Consider verb stems where the AVP does not surface as *ŋə*- (ie. NS, NA, *ŋ*-):

- (20) a. aku tulis-tulis
I write-write
b. aku nulis-nulis
I AV.write-AV.write
c. *aku nulis-tulis
I AV.write-write
d. *aku tulis-nulis
I write-AV.write

‘I write for a while’

- (21) a. dia ambil-ambil makanan
he take-take food

- b. dia η ambil- η ambil makanan
he AV.take-AV.take food
 - c. * dia η ambil-ambil makanan
he AV.take-take food
 - d. * dia ambil- η ambil makanan
he take-AV.take food
- 'He takes a variety of food (from a buffet)'

(20-21) involve copying of the AVP onto both parts of the verb. If the AVP is present, it must appear on both parts.

We can look closer at reduplication of verb bases starting with /l/:

- (22) a. aku η lukis-lukis lukisan
I AV-paint-paint painting
 - b. * aku η lukis- η lukis lukisan
I AV-paint-AV-paint painting
 - c. aku η lukis- η lukis lukisan
I AV.paint-AV.paint painting
- 'I've been painting a lot of paintings'
- (23) a. aku η lukis-lukis-in lukisan
I AV-paint-paint-BEN painting
 - b. * aku η lukis- η lukis-in lukisan
I AV-paint-AV-paint-BEN painting
 - aku η lukis-lukis-in lukisan
I AV.paint-paint-BEN painting
 - aku η lukis- η lukis-in lukisan
I AV.paint-AV.paint-BEN painting
- 'I painted a lot of paintings for (someone)'

When the η - allomorph is chosen, it must appear on both parts, as in *η ambil- η ambil* in (21). However, with the $\eta\partial$ - allomorph, the $\eta\partial$ - allomorph appears only initially and does not copy to both bases.

This phenomenon parallels the reduplication of intransitive verbal bases with a prefix, like *mə-lədak* 'explode':

- (24) a. *bapak bom jaŋ mə-lədak-lədak*
many bomb COMP mə-explode-explode

- b. *banak bom jaŋ mlədak-mlədak*
many bomb COMP *mə.explode-mə.explode*
 - c. * *banak bom jaŋ lədak-mə-lədak*
many bomb COMP *explode-mə-explode*
 - d. * *banak bom jaŋ mlədak-lədak*
many bomb COMP *mə.explode-explode*
 - e. * *banak bom jaŋ mə-lədak-mə-lədak*
many bomb COMP *explode-mə-explode*
- ‘A lot of bombs exploded (one after another)’

The *mə-* prefix usually surfaces as *mə-*, but it can also be “blended” into the verb base to produce *m-*.²

This pattern holds true for most verbal bases, with one caveat. When monosyllabic verbs are reduplicated, the *ŋə-* allomorph can be copied. The form without copying is not possible here:

- (25) a. *Budi ŋə-las-ŋə-las pipa-ŋa*
Budi AV-weld-AV-weld pipe-DET
- b. * *Budi ŋə-las-las pipa-ŋa*
Budi AV-weld-weld pipe-DET
- c. * *Budi ŋlas-ŋlas pipa-ŋa*
Budi AV.weld-AV.weld pipe-DET

‘Budi keeps on welding pipes’

This stands in opposition to the multisyllabic forms described above.

7.3 Constraint-based analysis

Because the entirety of the root is copied, we arbitrarily declare the reduplicant to be a suffix of the base.

MINBASE deserve further attention. In verbs without reduplication, *textscMinBase* prohibits monosyllabic outputs. With reduplication, we need to ensure that the entirety of the root is copied to the reduplicant, and also that a constraint can force the incorporation of *ŋə-* to the reduplicant in monosyllabic bases.

²Note that *mə-* is also the Standard Indonesian equivalent of the AVP, and in a few verbs it is carried into CJI as-is (Sneddon 2006).

To accommodate this, we can leverage the notion of Pwords. We assume that the base and reduplicant form separate Pwords. Then, we refine the definition of MAXROOT to be applied on any segment corresponding to the verb root:

- MINBASE: Mark one * for each Pword that contains exactly one syllable.

When reduplication is not applied, we recover our original constraint: in monosyllabic bases, because the output of a phonological process itself is a Pword, MINBASE will apply if $\eta\partial$ - is not recruited to form a disyllabic output.

We introduce a base-reduplicant Faithfulness constraint to enforce similarity in reduplication, following McCarthy and Prince 1995. This ensures that the reduplicant takes the same form as the base, even under phonological changes that might induce dissimilarity. Which constraint to adopt is a tricky question. We cannot simply adopt a hypothetical IDENT BR which preserves just segments between the base and reduplicant, because then forms like ** $\eta\text{om}\eta\text{-om}\eta$* or ** $mb\grave{a}li\text{-}b\grave{a}li$* would be preferred to minimize violations of DEP.

Crucially, notice that $\eta\partial$ - is the only allomorph that doesn't get copied at the expense of DEP, and this allomorph introduces another syllable to the verb. Therefore, I propose a constraint that preserves entire syllables between B-R.

- IDENT SYLLABLE BR (IDSYL BR): Mark one * for each corresponding base-reduplicant pair of syllables $s_1 - s_2$ such that $s_1 \neq s_2$.

We can now consider reduplicated forms:

/ η -tulis-RED/	IDSYL BR	AGREEPLACE	* ∂ //CLOSED	MAXROOT
a. η tulis-tulis	*!	*		
b. nulis-lis				4*!
c. $\eta\grave{a}$ tulis-tulis			*!	
d. $\eta\grave{a}$ tulis- $\eta\grave{a}$ tulis			*!*	
e. nulis-tulis	*!			*
f. nulis-nulis				**

Table 10: *tulis-tulis* ‘write for a while’

MAXROOT enforces similarity to the verb root. With *nulis-nulis*, IDSYL BR and * ∂ //CLOSED

leave no choice but to incur violations of MAXROOT, in this case with one segment on each base. Hence *nulis-nulis* is the preferred candidate.

This constraint ranking satisfactorily handles the other allophones of of the AVP. Consider η -, NA, and $\eta\partial$ -:


/ η -ambil-RED/	IDSYL BR	DEP ∂
a. η ambil-ambil	*!	
b. $\eta\partial$ -ambil-ambil		*!
 c. η ambil- η ambil		

Table 11: *ambil-ambil* ‘write for a while’



/ η -gigit-RED/	IDSYL BR	DEP ∂	*CC
a. η gigit-gigit	*!		*
b. $\eta\partial$ gigit- $\eta\partial$ gigit		**!	
 c. η gigit- η gigit			**
 d. $\eta\partial$ gigit-gigit		*	

Table 12: *gigit-gigit* ‘bite in succession’

IDSYL BR removes all candidates that do not share identical syllables. Then, the two allophonic choices are chosen, as before, by a variable ranking of DEP ∂ and *CC.

In monosyllabic bases, violations of MINBASE drive the copying of $\eta\partial$ -:


/ η -bom-RED/	IDSYL BR	MINBASE	DEP ∂	*CC
a. mbom-bom	*!	**		*
b. mbom-mbom		**!		**
c. $\eta\partial$ bom-bom		*	*!	
 d. $\eta\partial$ bom- $\eta\partial$ bom			**	

Table 13: *bom-bom* ‘bomb in succession’

Regardless of the ranking of DEP \emptyset and *CC, the correct candidate will be chosen.

Our new rankings are CRISPEDGE \longleftrightarrow * \emptyset //CLOSED, as in 7.1, and IDSYL BR \gg MAXROOT.

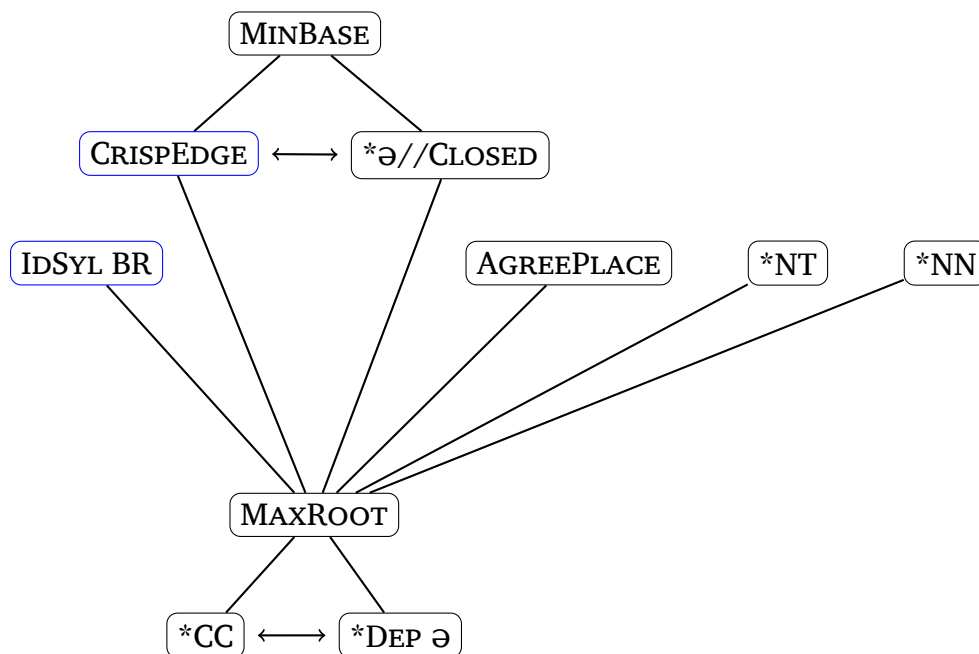


Figure 2: Hasse diagram of constraints (new constraints in blue)

7.4 Implications

The OT approach taken here to resolve reduplication uses the approach to reduplicative identity described in McCarthy and Prince 1995, assuming that a B-R Faithfulness constraint, considered jointly with other constraints, drives overapplication of phonological change.

This is not without controversy. Inkelas and Zoll 2005 describe an alternative approach, Morphological Doubling Theory (MDS), which claims that phonological alternations arise from truncation of one of a pair of identical underlying copies. These copies undergo phonological changes independently, as opposed to interacting with each other as in McCarthy and Prince 1995.

Inkelas and Zoll 2005 use data from Javanese, which features similar AVP phenomena as CJI. In this data, AVP + *tulis-tulis* ‘write aimlessly’ must be nasalized on both copies, while AVP + *baliq-baliq* ‘turn around’ may surface as *mbaliq-baliq* or *mbaliq-mbaliq*:

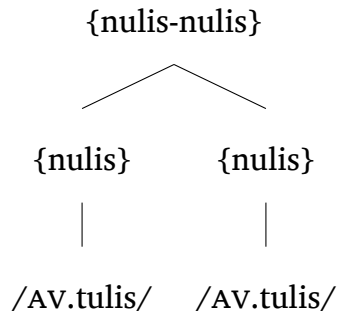


Figure 3: Analysis of *nulis-nulis* in Javanese, from Inkelas and Zoll 2005

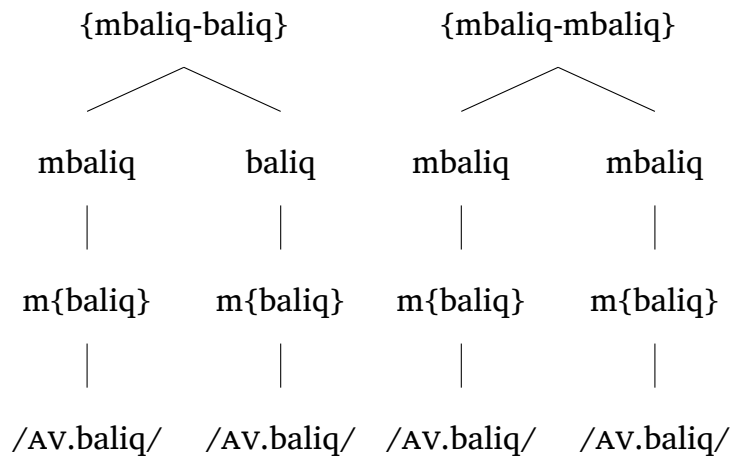


Figure 4: Analysis of *mbaliq-baliq* ~ *mbaliq-mbaliq* in Javanese, from Inkelas and Zoll 2005

According to this analysis, the AVP may be truncated if it does not retain segmental integrity (ie. is not incorporated into the verbal root). Figure 3 involves preservation of the AVP segment on both copies, while 4 optionally truncates the second AVP. It is unclear to me how this analysis eliminates forms like **baṅun-ṅabaṅun*, where the first but not the second copy is truncated.

However, CJI differs from this data in that forms like **mbaṅun-baṅun* aren't possible. The account cannot be as simple as that in Javanese. In particular, this analysis fails to explain why forms like **mbaṅun-baṅun* or **ṅabaṅun-ṅabaṅun* cannot be produced, because while the AVP retains segmental integrity in both cases, the AVP either must be or cannot be truncated.

In our approach, this is handled by the IDS_{YL} BR constraint, which will mark **mbaṅun-baṅun*, and extra violations of DEP \emptyset eliminate **ṅabaṅun-ṅabaṅun* (cf. Table 12). In MDS, reduplicated copies undergo phonological changes independently, which is problematic

because it cannot invoke an output-wide DEP \emptyset that will mark two violations for each /ə/ in the AVP.

8 Conclusion and further directions

In this report, we explored the allophonic variation of the AVP with regards to the properties of the verb root. With a proper selection of constraints, we were able to explain phenomena that would be difficult to explain using rule-based approaches.

The constraint system in this report involves some unmotivated constraints. Specifically, *ə//CLOSED is *ad-hoc* and doesn't apply to any conditions outside of the AVP, given that sequences of /ə/ followed by a “closed” consonant abound in CJI (eg. in roots such as *latus* ‘break’, *mənaŋ* ‘win’). However, I was not able to figure out a way to restrict the epenthesis of /ə/ without invoking a markedness constraint on /ə/-closed sequences. A future analysis could invoke more sophisticated OT theories like Harmonic Grammar (cf. the treatment of Indonesian loanwords in Batais and Wiltshire 2018) to provide a more convincing set of constraints.

A Data

A full list of verbs collected, as well as relevant judgements, can be viewed at https://docs.google.com/spreadsheets/d/1nUc2IE4PDMouq0Hfhz9PpkSk_LsMBHkoCC1CQB_NsBI/edit?usp=sharing.

References

- Batais, Saleh and Caroline Wiltshire (2018). “Indonesian borrowing as evidence for Harmonic Grammar”. In: *Journal of Linguistics* 54.2, pp. 231–262. DOI: 10.1017/S0022226717000317.
- Blust, Robert (2004). “Austronesian Nasal Substitution: A Survey”. In: *Oceanic Linguistics* 43.1, pp. 73–148. DOI: 10.1353/ol.2004.0004.
- Inkelas, Sharon and Cheryl Zoll (2005). *Reduplication: doubling in morphology*. Vol. 106. Cambridge Studies in Linguistics. Cambridge University Press.
- McCarthy, John and Alan Prince (Jan. 1995). “Faithfulness and Reduplicative Identity”. In: *Papers in Optimality Theory* 18.
- Moira, Yip. (1993). “Cantonese loanword phonology and Optimality Theory”. In: *Journal of East Asian Linguistics* 2.3, pp. 261–291.
- Pater, Joe (2001). “Austronesian Nasal Substitution Revisited”. In: *Segmental Phonology in Optimality Theory: Constraints and Representations*, p. 160.
- Smirnova, Elizaveta and Svetlana Shustova (2017). “Noun-incorporation in English as a valency-changing device”. In: *Lingua* 194, pp. 15–25. ISSN: 0024-3841. DOI: 10.1016/j.lingua.2017.05.005.
- Sneddon, James (2006). *Colloquial Jakartan Indonesian*. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Sneddon, James et al. (2010). *Indonesian: A Comprehensive Grammar*. Routledge.