

# Variation and change in progress

## Evidence from Word-final [-a] and [-e] in Jakarta Indonesian

Ferdinan Okki Kurniawan and

Maya Ravindranath Abtahian

Atma Jaya Catholic University of Indonesia | University of Rochester

This study reports on variation in final [-a] ~ [-e] in Jakarta Indonesian (JI), a new urban variety of Indonesian that emerged from contact between Standard Indonesian (SI) and Betawi, a variety of Malay spoken in the capital of Indonesia. The study incorporates both apparent- and real-time approaches to examining change in this variable, using two large-scale naturalistic speech corpora from three generations of speakers. In the examination of this variable over time, we find that the standard SI form is now predominant in colloquial JI, a change that was led by women and speakers with higher educational backgrounds. The study provides insight into the development of a new language variety within a multilingual environment.

**Keywords:** Betawi, Jakarta Indonesian, corpora, apparent time, real time

### 1. Introduction

This study reports on variation in final [-a] ~ [-e] as in [apa] ~ [ape] ‘what’, [ija] ~ [ije] ‘yes’, and [dia] ~ [die] ‘3sg’ in Jakarta Indonesian (JI), a colloquial variety of Indonesian. The synchronic patterns demonstrated by this variable have roots in the linguistic history of Jakarta, the modern capital of Indonesia, which was established in 1945 on the site of the old city of Batavia,<sup>1</sup> the Dutch settlement at the center of the Dutch East India Company’s trade network from 1619–1945. Final [-e] was the predominant variant used in Betawi, a colonial variety of Malay that emerged in the Dutch settlement, while final [-a] is a Standard Indonesian (SI)

---

1. The Dutch name for Jakarta during the Dutch East Indies colonial time. The name ‘Betawi’ is derived from it.

variant. Variation between the two forms is presumed to have developed with the emergence of JI, a variety which is described to have emerged in the new capital city of Jakarta post-1945 as a result of contact between Betawi (the old town lingua franca) and Standard Indonesian. As is typical of contact varieties, the features of JI that distinguish it from SI are described as showing variable realization for speakers (Cohn & Vogel, 2019; Kurniawan, 2018; Sneddon, 2006), but few studies have examined this variation systematically. The goal of this study is to more closely examine the social factors that predict the use of the different variants of this particular variable, and to examine change in the variable over time. We start with a brief history of the city of Jakarta and some background on the variety of Indonesian spoken there.

## 1.1 A brief history of Jakarta and background on Indonesian

### 1.1.1 *Pre-colonial*

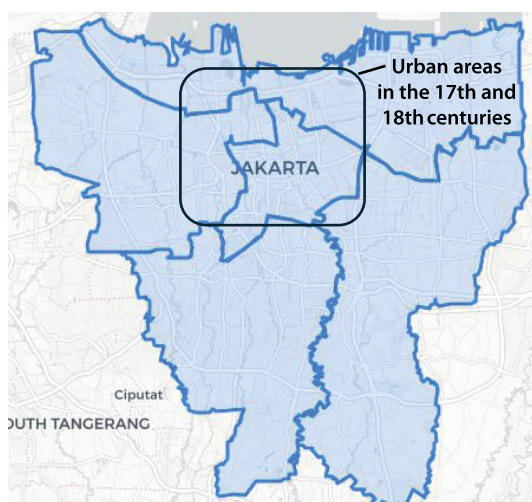
The coastal area and port of Jakarta, situated in northern West Java (see map in Figure 1), have been inhabited since the 4th century BCE. Prior to colonization by the Portuguese and Dutch, several kingdoms held sway over the region. These kingdoms were the Tarumanagara Kingdom, referenced in inscriptions estimated to date from around 358 CE (Iguchi, 2015), the Sunda Kingdom, which ruled between 669 and 1527 CE (Soekmono, 1973), and the Banten Sultanate, which governed from 1527 to 1619 CE. The area was known as Sunda Kalapa (meaning “Coconut Sunda”) during the Sunda Kingdom period. Since the time of the Sunda Kingdom, the port of Sunda Kalapa has been recognized for its strategic importance and thriving trade.



Figure 1. Map of Indonesia and the location of Jakarta (Google, n.d.)

### 1.1.2 *The colonial period and the development of Betawi*

After the Portuguese lost influence in the late 16th century, the Dutch arrived in Sunda Kalapa in the early 17th century. Jakarta's strategic location has always attracted local and foreign traders, leading to a diverse population. Castles (1967) describes Jakarta's diverse population in Dutch colonial times. The population included slaves from various regions in India, Southeast Asia, and Indonesia, with a significant number coming from Bali and South Sulawesi. The city's free inhabitants were primarily Malay, Balinese, Buginese, Ambonese, and South Indian Muslims. These groups, both enslaved and free, resided in the central urban area of Batavia in the 17–18th centuries (Muhadjir, 1981). The approximate location of the city's boundary between urban and rural Batavia is adapted from Muhadjir's map and shown in Figure 2. This diverse group came to recognize itself as a distinct ethnic group; the 1930 census marked the first written documentation of Betawi as such, likely coinciding with the Dutch government's recognition of Betawi as a separate language variety (Castles, 1967, p. 166).



**Figure 2.** The map of modern-day Jakarta (Jakarta Regional Government, 2024) and an approximate location of the urban areas in the 17th–18th centuries based on Muhadjir (1981)

The early population of Batavia spoke a Portuguese-based creole as a *lingua franca* (Castles, 1967). By the beginning of the 19th century, however, the new Malay variety, Betawi, gradually supplanted the use of the creole. Betawi was heavily influenced by Balinese, Javanese, Portuguese, Dutch, Arabic, and Hokkien in its early development (Kahler, 1966). Later, Dutch military expansion into

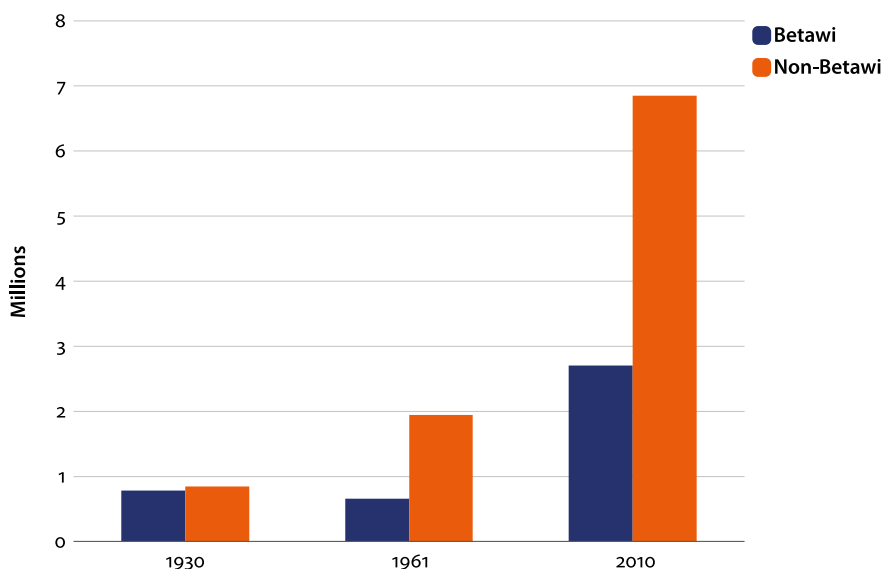
Batavia's rural outskirts bordering Sundanese regions spurred the development of Batavia's initial suburbs in the 19th century (Ikranagara, 1980; Muhadjir, 1981; Wallace, 1976). When the Dutch government expanded city development into rural areas, the influence of Sundanese as a contact language with Betawi grew, particularly in rural communities.

### 1.1.3 *Post-colonial Jakarta and the development of Jakarta Indonesian*

After independence in 1945, Jakarta (renamed during the Japanese occupation) became Indonesia's capital city. The city experienced significant transformation toward modernization during the 1960s under President Sukarno's vision for urban planning (Silver, 2008), and the city's infrastructure development was further fueled by foreign investment under President Suharto's leadership (1967–1998) (Merrillees, 2015). The 21st century has ushered in a period of political stability and economic prosperity, continuously attracting economic migrants from across Indonesia; Jakarta is now one of the most populous urban areas in the world (World Population Review, 2025). By 2010, the number of Jakarta residents had reached 9,988,495 in the urban core area of 664 km<sup>2</sup> and more than 30 million in the greater metropolitan area of 6,392 km<sup>2</sup> (Central Bureau of Statistics, 2025). The population density is 14,469 people per square kilometer in the urban core area, and 4,383 people per square kilometer in the greater metropolitan area. This high population density creates extensive inter-ethnic and linguistic interaction among the inhabitants of this linguistic melting pot.

The increase in the migrant population can be observed when we compare the estimates of the Jakarta population in the 1930, 1961, and 2010 censuses as reported in Castles (1967, p. 166) and Ananta, Arifin, Hasbullah, Handayani, and Pramono (2015, p. 106) (Figure 3).

In Figure 3, Non-Betawi includes people who were born in other provinces and moved to Jakarta as well as their descendants who were born and grew up in Jakarta. In the 1930 census, the non-Betawi population in the city (847,175) was roughly similar to the Betawi population (778,953). The censuses in 1961 and 2010 show a sharp increase in the non-Betawi population. Within less than two decades after Indonesia's independence in 1945, more than half of the population of Jakarta was born in provinces outside Jakarta or were children of migrants, with the highest number of migrants coming from other parts of Java (West, Central, and East) and South Sumatra (Castles, 1967). The in-migration of non-Betawi speakers has resulted in a newer lingua franca in Jakarta. Early observers of this new variety labeled it Modern Jakarta Malay (Wallace, 1976) or Colloquial Jakartan Indonesian (Sneddon, 2006). More recently scholars refer to the variety as Jakarta Indonesian (Gil, Tadmor, Bowden, & Taylor, 2015; Hidajat, 2010; Cole et al., 2006; Wouk, 1999).



**Figure 3.** Population estimates in Jakarta’s urban core area from 1930–2010  
(Source: Census data as reported in Castles, 1967, and Ananta et al., 2015)

## 1.2 Standard Indonesian

During this same modern time period, “Indonesian” (*Bahasa Indonesia*) was established as a national and official language of Indonesia, starting with its proclamation as such by the 1928 Youth Council. At the time, Indonesian was no more than an ideal (Alisjahbana, 1971, pp.170–178); the variety of Malay that came to be understood as Standard Indonesian was no one’s native language and in fact was most likely constructed from more than one variety (Ikranagara, 1980). As it was developed as a national language it served as the primary language of education, official institutions, and the media, and existed in a diglossic situation with regional and local languages of Indonesia. Almost a century later it has grown to become the dominant home language for millions of Indonesians (Ananta et al., 2015), replacing even large regional mother tongues such as Javanese, Sundanese, Minangkabau, and Madurese, each spoken by millions. Numerous regional and colloquial varieties of Indonesian have developed in contact with regional languages. As longtime contact varieties with Standard Indonesian spoken in the capital city, both Betawi and Jakarta Indonesian have influenced the forms of standard and colloquial Indonesian that have spread through the archipelago (Poedjosoedarmo, 1982, p. 142).

### 1.3 Word final [-a] and [-e] in Betawi and Jakarta Indonesian

The historical migrations to Batavia and then Jakarta described above ultimately led to three groups with distinguishable linguistic features — urban Betawi, rural Betawi, and non-Betawi. Rural Betawi was in closest contact with Sundanese, from the surrounding Sundanese-speaking areas in West Java, and the effects of this contact distinguish it from the other two varieties. For instance, Muhadjir (1981), Ikranagara (1980), and Wallace (1976) all describe final [-e] as a defining characteristic of urban Betawi, and make a distinction between the final [-e] used by urban Betawi speakers in the 1970s and 80s and the final [-aʔ] and [-a(h)]<sup>2</sup> variants used by speakers of rural Betawi in contact with Sundanese.

Table 1 provides examples of the historical use of the final [-aʔ], [-ah], and [-e] in urban and rural Betawi, based on Wallace (1976). Content and function words behave similarly in urban Betawi, where they appear with final [-e], and in Standard Indonesian/Malay, where they appear with final [-a]. In contrast, rural Betawi makes a distinction between [-aʔ] in content words and [-ah] in function words, with the exception of [ama] and [daripada], two prepositions which are always spoken in sentence-medial position and in which final [-h] is always omitted (see Kurniawan (2023) for a fuller description of final [-h] in JI).

**Table 1.** Final [-a] ~ [-e] in Betawi around the beginning of the nineteenth century

Function words			
Urban	Rural	SI/Malay	Gloss
[ije]	[ijah]	[ija]	yes
[die]	[diah]	[dia]	3sg
[gue]	[guah]	[saja]	1sg
[ape]	[apah]	[apa]	what
[mane]	[manah]	[mana]	where
[ame]	[ama]	[sama]	with
[daripade]	[daripada]	[daripada]	instead of

2. Although at that time final [-e] was also attested in rural Betawi, more recent data from Gil et al. (2015) demonstrate very little use of final [-e] among rural speakers.

Table 1. (continued)

Content words			
Urban	Rural	SI/Malay	Gloss
[mate]	[mataʔ]	[mata]	eye
[(kə)pale]	[(kə)palaʔ]	[kəpala]	head
[taɲe]	[taɲaʔ]	[taɲa]	ask
[puɲe]	[puɲaʔ]	[puɲa]	possess
[mude]	[mudaʔ]	[muda]	young
[gile]	[gilaʔ]	[gila]	crazy

It is worth noting that the adoption of final [-a] would not have been caused by direct lexical borrowings from Sundanese. Although Sundanese and Malay have many cognate forms in content words, the function words in question do not have direct cognates in Sundanese. Rather, it seems that [-aʔ] and [-ah] are modeled on a Sundanese pattern that does not allow open syllables in phrase-final position.

Once *Bahasa Indonesia* was declared and promoted as the national language, the standardized variety of Indonesian (SI) further strengthened by the formation of the National Language Center by the Indonesian government following Independence. Sneddon (2003) reported an increased use of SI during the New Order era under President Soeharto's regime (1966–1998).

In SI, final [-a] is standard, and final [-ʔ] and [-h] sounds in function words like \*[ijaʔ] 'yes' and \*[diah] '3SG' are not permitted.

Table 2. Sources of variation in JI

Rurals Betawi	JI	SI	Gloss
[je] ~ [jaʔ]	[je] ~ [ja]	[ja]	yes
[ije] ~ [ijaʔ]	[ije] ~ [ija]	[ija]	yes
[-ɲe] ~ [-ɲah]	[-ɲe] ~ [-ɲa]	[-ɲa]	determiner
[die] ~ [diah]	[die] ~ [dia]	[dia]	3SG
[ad̪ze] ~ [ad̪zah]	[ad̪ze] ~ [ad̪za]	[sad̪za]	just
[ade] ~ [adaʔ]	[ade] ~ [adaʔ]	[ada]	exist
[ape] ~ [apah]	[ape] ~ [apa]	[apa]	what
[due] ~ [duaʔ]	[due] ~ [duaʔ]	[dua]	two
[gue] ~ [guah]	[gue] ~ [gua]	[saja]; [aku]	1SG

Table 2 provides examples of the variation in function words in the three varieties. As discussed above, final [-ʔ] and [-h] obligatorily occur phrase-finally in Rural Betawi, as in [jaʔ] and [diah], but variably in phrase-medial position with [-e].<sup>3</sup> Final [-ʔ] and [-h] never occur in JI and SI; JI speakers can variably use final [-e] (as in urban and rural Betawi), or final [-a] as in SI.

## 1.4 Objective of the study

The main objective of this study is to describe this variable feature of Jakarta Indonesian quantitatively in both real and apparent time, using two existing corpora of naturalistic conversations in JI. In addition to providing some insight into change over time, the use of these corpora also allow us to consider the extent to which some social factors such as education level and gender may constrain the variation in JI. Our hypothesis is that if the [-a] variant is indeed due to contact with and influence from Standard Indonesian, then speakers with more exposure to SI are likely to produce more final [-a] in function words. As we cannot directly measure speakers' exposure to SI, we use education level as a proxy for exposure (since the medium of education for all Jakartans is SI). We also consider gender as a social factor that we expect may correlate with use of more standard variants, following decades of work in variationist linguistics (see Angle & Hesse-Biber, 1981; Eckert & McConnell-Ginet, 2003; Labov, 2001, 1990; Romaine, 2003; among others).

## 2. Data and methods

The data for this study come from corpora collected at two different time periods almost thirty years apart. The first consists of transcriptions of conversations collected in the mid-1970s by Wallace for a study of phonological variation in JI (Wallace, 1976). The second corpus includes audio recordings of both adult and children speakers in conversational interviews collected between 2004–2012 under the auspices of the Max Planck Institute for Evolutionary Anthropology Jakarta Field Station (Gil et al., 2015) as part of separate but related projects.

---

3. Examples of obligatory [-ʔ] and [-h] in phrase-final position: [ada jaʔ] 'you have it, right?' and [ada diah] 'he's there.' Examples of optional [-ʔ] and [-h] in phrase-medial position: [dua mobil] ~ [duaʔ mobil] 'two cars' and [apa itu] ~ [apah itu] 'what's that.' See further discussion about final [-ʔ] and [-h] in Kurniawan (2023).



## 2.1 The Wallace corpus

The participants in Wallace's recordings were friends, family, relatives, neighbors, and other close acquaintances of Wallace's research assistants, who recorded them talking about everyday topics. Wallace classified these speakers as either Traditional Jakarta Malay (Betawi) or Modern Jakarta Malay (Jakarta Indonesian) speakers (Wallace, 1976, pp.69–70) based on the ethnicity of their parents (Betawi or not). JI (Modern Jakarta Malay) speakers are individuals who were born and raised in Jakarta, but their parents do not have Betawi ancestry. In this analysis, we used Wallace's categorization to include only those he labeled Modern Jakarta Malay. Wallace further classified the speakers in his study into three main socioeconomic groups: low SES, middle SES, and high SES. His low SES speakers were mostly manual laborers, small-scale traders, minor office clerks, and railroad conductors. His middle SES speakers were mostly those who had intermediate positions in government offices or private industries, and his high SES speakers were mostly higher managers and bureaucrats. Wallace did not record the speakers' educational levels. Our assumption is that the speakers who occupied intermediate positions in the civil service or private industries might have had college-level education and the speakers who were non-highly skilled workers most probably attained no more than high school. From this corpus we have included fifteen adult men of lower SES/educational background, nine adult men of higher SES/educational background, nine adult women of lower SES/educational background, and two adult women of higher SES/educational background.<sup>4</sup> Further information about the speakers is in Table A of the Appendix. Although Wallace's analysis was based on the audio recordings of the speakers, those recordings are no longer available, so it was necessary for this study to use the orthographic transcription of this corpus, a phonological transcription based on Indonesian orthography that was completed by research assistants in the 1970s and checked by Wallace, who compared all of the transcriptions with the recordings word-by-word at least twice (S. Wallace, personal communication, August 28, 2013).

## 2.2 The Jakarta field station corpus

For this study we selected the twenty most prolific adult speakers from a larger set of 105 speakers in the Jakarta field station corpus. All of these speakers were born and raised in Jakarta and none of them have Betawi ancestry. Further infor-

---

4. The number of women with higher SES/educational background was limited by the sample – presumably because it was still relatively rare for women to attend college in the 1970s.

mation about the speakers is given in Table B of the Appendix (and for more on these speakers see Kurniawan, 2023). The children were part of Gil et al.'s (2015) child corpus that was collected as part of a longitudinal project to document children's speech over a period of four years. There were a total of ten target children involved in this project whose ages were one to eight when initially recruited. In this study, we included three male and five female pre-adolescent speakers who were mostly siblings or neighbors of the target children and were participants in the conversational interviews. This investigation chose the pre-adolescent speakers rather than the younger children in order to avoid additional complexity with respect to language acquisition in younger children. All of the children in this analysis are elementary school students. In order to categorize their SES/educational background we used their parents' educational background; this is what is listed in the Educational Background of Table 3. Further information on these speakers is included in Table C of the Appendix.

The two corpora together represent three generations of JI speakers (Table 3). The use of the two corpora allows for both a real time and an apparent time investigation of the variable in question. Real time evidence is gathered by comparing the patterns of use of the variants [-a] and [-e] between the 1970s (Wallace, 1976) and the 2000s (Gil et al., 2015) speakers. Additional apparent time evidence is gathered by comparing adult and pre-adolescent speakers in Gil et al.'s 2000s corpus.

**Table 3.** Three generations of speakers

Generation	Gender	Educational background	Number of speakers
The 1970s–adult (born between 1938–1955)	Male	Higher	9
	Male	Lower	15
	Female	Higher	9
	Female	Lower	2
The 2000s–adult (born between 1956–1984)	Male	Higher	5
	Male	Lower	5
	Female	Higher	5
	Female	Lower	5
The 2000s–pre-adolescent (born between 1992–1995)	Male	Higher	1
	Male	Lower	2
	Female	Higher	3
	Female	Lower	2

## 2.3 Data

From these two corpora we extracted all the tokens of the function words presented in Table 4. The decision was made to limit the analysis to function words as both Wallace (1976, pp.69–71) and Grijns (1991, p.203) observe that variation between final [-a] and [-e] occurs only in function words, and this is consistent with our observations of modern spoken Jakarta Indonesian. These nine words are all frequent function words – these are in the top fifty most produced words in the corpus in rank order (Gil et al., 2015) – and each word occurs with both variants in these corpora.

**Table 4.** Function words and their variation

SI written form	Variation		Gloss
<i>nya</i>	[ɲa]	~ [ɲe]	determiner
<i>ya</i>	[ja]	~ [je]	yes
<i>iya</i>	[ija]	~ [ije]	yes
<i>ada</i>	[ada]	~ [ade]	exist
<i>dia</i>	[dija]	~ [dije]	3sg
<i>aja</i>	[ad͡ʒa]	~ [ad͡ʒe]	just
<i>apa</i>	[apa]	~ [ape]	what
<i>gua</i>	[gua]	~ [gue]	1sg

The analysis is further constrained to the patterns of variation in the bare forms, such as in (a), and excludes roots that are combined with suffixes as in (b–d), since the variation in these forms may be morpho-phonologically conditioned, and previous studies have not reached definitive conclusions about their patterns of realization. Thus, the forms of [-a] and [-e] before suffixes are not included.

- a. [apa] ~ [ape] ‘what’
- b. [apa-ɲa] ~ [ape-ɲe] ‘what-determiner’
- c. [ɲ-apa-in] ‘N-what-in’
- d. [apa(?) -an] ‘what-an’<sup>5</sup>

---

5. The examples in (a–d) are taken from Ikranagara (1980); The general pattern in the examples is that variation in the root-final position occurs when it is suffixed with *-nya*, but the variant [-e] does not occur when the root is suffixed with *-an*, or *-in*. It seems that vowel harmony also determines the realization of [-a] in [apa-ɲa] and [-e] in [ape-ɲe]. However, this generalization is called into question by the variation of [ma(ɲ)ka-ɲe] and [ma(ɲ)ke-ɲe] ‘then-nye’ cited by Wallace (1976, p.73). For the suffixes *-an* and *-in*, Wallace agrees with Ikranagara.

After all of the tokens of the nine target function words were extracted from the two spoken corpora, each word was transcribed by two coders. The 1970s corpus search was done manually on the PDF files.

The frequency distribution of the two variants [-a] and [-e] in each social category was calculated and a chi-square test for independence was calculated to determine the correlation between independent sociolinguistic variables (generation, educational background, and gender) and the realization of the two variants [-a] and [-e] in the corpus data. The tests were conducted in R statistical software (R Core Team, 2022).

### 3. Findings and discussion

This section presents both the apparent time findings from the speakers in Gil et al. (2015) as well as a real-time, comparative analysis of the findings from the two datasets. First, however, we discuss the findings from the adult male speakers from the 2000s corpus, as this group demonstrated the highest frequency of use of the variable and offers some insights into the distribution of the variable. We then consider the other speaker groups.

#### 3.1 Adult male speakers in the 2000s corpus: Frequency and linguistic conditioning

The ten adult male speakers in this sample together produced over 4000 tokens of the target function words and across the board the majority of these were produced with word final [-a] (as seen in Table 5). This is consistent with the hypothesis that the Standard Indonesian [-a], as the incoming form, would be prevalent in the modern corpus. Moreover, we find that when we compare the male speakers by educational background, the speakers with lower educational background have, in general, fewer occurrences of the variant with final [-a] than speakers of higher educational background (as shown in Figure 4). The high prevalence of the SI variant [-a] among the speakers from higher educational background provides initial evidence for a correlation between education level and use of the SI variant, and the influence of the standard variety on the emergent variety of Jakarta Indonesian.

---

However, Grijns (1991, pp.203–204) quotes both [diapain] and [diapein] ‘di-what-in’ although the former is of greater frequency than the latter. A reviewer of this article suggests that the difference may be explained by the the different morphosyntactic categories of the affixes: =*nya* is an enclitic, while *-in* and *-an* are suffixes.

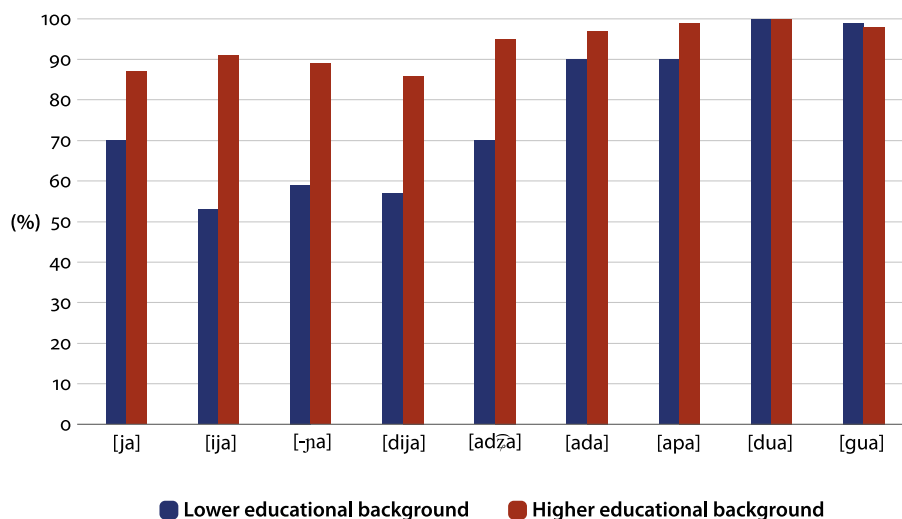


Figure 4. Final [-a] produced by male speakers in the 2000s corpus

Table 5 provides a detailed breakdown of the distribution of the variable in the nine function words for the ten speakers. Again, we see that for most of the lexical items there is significantly more use of [-a] than [-e]. The only context in which [-a] is not significantly more common is in the word [ija] ~ [ije] ‘yes’ amongst the speakers of lower educational background who seem to still maintain more use of the Betawi variant [-e].<sup>6</sup> This table also provides some possible evidence for linguistic conditioning. Preceding bilabial [p], alveolar [d], and high back vowel [u] favour categorically or near categorically [-a] vowel. In contrast, the palatals [j], [ɲ], and [dʒ] exhibit the highest degree of variation. For this reason, in what follows we have further limited our envelope of variation to the variable preceded by palatals, leaving aside those preceded by all other sounds, where [-a] is nearly categorical. A further examination of this linguistic conditioning is left for future investigation, and we now turn to the results from all speakers.

6. The reason for this is still unclear. We speculate that it may be the case that this form has been lexicalized to some degree for some speakers, or that there is a phonological effect and that the preceding vowel [i] favors the Betawi variant [e].

**Table 5.** Final [-a] ~ [-e] produced by male adult speakers in the 2000s corpus

Lower educational background					
	Variation	Final [-a]	Final [-e]	Total	Chi-square test goodness of fit
Palatals	[ja] ~ [je]	219 (70%)	92 (30%)	311	$\chi^2 = 51.86, df = 1, p < .001^{***}$
	[ija] ~ [ije]	91 (53%)	81 (47%)	172	$\chi^2 = 0.58, df = 1, p = 0.44$
	[ɲa] ~ [ɲe]	322 (59%)	224 (41%)	546	$\chi^2 = 17.59, df = 1, p < .001^{***}$
	[dija] ~ [dije]	158 (57%)	117 (43%)	275	$\chi^2 = 6.11, df = 1, p = 0.013^{**}$
	[aḍza] ~ [aḍze]	74 (70%)	32 (30%)	106	$\chi^2 = 17.59, df = 1, p < .001^{***}$
	<b>Total</b>	<b>864 (61%)</b>	<b>546 (39%)</b>	<b>1410</b>	$\chi^2 = 71.719, df = 1, p < .001^{***}$
Bilabial and alveolar	[ada] ~ [ade]	145 (90%)	16 (10%)	161	$\chi^2 = 103.36, df = 1, p < .001^{***}$
	[apa] ~ [ape]	54 (90%)	6 (10%)	60	$\chi^2 = 38.4, df = 1, p < .001^{***}$
	<b>Total</b>	<b>199 (90%)</b>	<b>22 (10%)</b>	<b>221</b>	$\chi^2 = 141.76, df = 1, p < .001^{***}$
High vowels	[dua] ~ [due]	47 (100%)	0 (0%)	47	$\chi^2 = 47, df = 1, p < .001^{***}$
	[gua] ~ [gue]	348 (99%)	2 (1%)	350	$\chi^2 = 342.05, df = 1, p < .001^{***}$
	<b>Total</b>	<b>395 (99%)</b>	<b>2 (1%)</b>	<b>397</b>	$\chi^2 = 389.04, df = 1, p < .001^{***}$
<b>Total (all)</b>					
		<b>1,458 (72%)</b>	<b>570 (28%)</b>	<b>2,028</b>	$\chi^2 = 388.83, df = 1, p < .001^{***}$
Higher educational background					
Palatals	[ja] ~ [je]	465 (87%)	70 (13%)	535	$\chi^2 = 291.64, df = 1, p < .001^{***}$
	[ija] ~ [ije]	216 (91%)	22 (9%)	238	$\chi^2 = 158.13, df = 1, p < .001^{***}$
	[ɲa] ~ [ɲe]	433 (89%)	63 (13%)	496	$\chi^2 = 276.01, df = 1, p < .001^{***}$
	[dija] ~ [dije]	159 (86%)	7 (4%)	166	$\chi^2 = 139.18, df = 1, p < .001^{***}$
	[aḍza] ~ [aḍze]	100 (95%)	5 (5%)	105	$\chi^2 = 85.95, df = 1, p < .001^{***}$
	<b>Total</b>	<b>1373 (89%)</b>	<b>167 (11%)</b>	<b>1540</b>	$\chi^2 = 94.44, df = 1, p < .001^{***}$
Stops	[ada] ~ [ade]	144 (97%)	4 (3%)	148	$\chi^2 = 132.43, df = 1, p < .001^{***}$
	[apa] ~ [ape]	172 (99%)	2 (1%)	174	$\chi^2 = 166.09, df = 1, p < .001^{***}$
	<b>Total</b>	<b>316 (98%)</b>	<b>6 (2%)</b>	<b>322</b>	$\chi^2 = 298.45, df = 1, p < .001^{***}$
High vowels	[dua] ~ [due]	46 (100%)	0 (0%)	46	$\chi^2 = 46, df = 1, p < .001^{***}$
	[gua] ~ [gue]	236 (98%)	4 (2%)	240	$\chi^2 = 224.27, df = 1, p < .001^{***}$
	<b>Total</b>	<b>282 (99%)</b>	<b>4 (1%)</b>	<b>286</b>	$\chi^2 = 270.22, df = 1, p < .001^{***}$
<b>Total (all)</b>					
		<b>1,971 (92%)</b>	<b>177 (8%)</b>	<b>2,148</b>	$\chi^2 = 1498.3, df = 1, p < .001^{***}$

Note on *p*-values: no asterisk shows an insignificant difference, one asterisk represents a significant difference, and two or three asterisks indicate the more significant difference; this note is used for the rest of the article.

### 3.2 Findings from all groups of speakers and social categories

This section presents the total occurrences of final [-a] and final [-e] by speaker, education level and generation, shown in Table 6.

**Table 6.** Final [-a] and [-e] by gender and educational background

Corpora	Categories	Final [-a]	Final [-e]	Total	Chi-square test goodness of fit
The 1970s*	Male-adult-lower education	40 (5%)	720 (95%)	760	$\chi^2 = 608.42$ , $df=1$ , $p < .001^{***}$
	Male-adult-higher education	10 (5%)	179 (95%)	189	$\chi^2 = 151.12$ , $df=1$ , $p < .001^{***}$
	Female-adult-lower education	6 (5%)	127 (95%)	133	$\chi^2 = 110.08$ , $df=1$ , $p < .001^{***}$
	Female-adult-higher education	33 (26%)	94 (74%)	127	$\chi^2 = 29.29$ , $df=1$ , $p < .001^{***}$
The 2000s	Male-adult-lower education	1,458 (72%)	570 (28%)	2,028	$\chi^2 = 388.83$ , $df=1$ , $p < .001^{***}$
	Male-adult-higher education	1,971 (92%)	177 (8%)	2,148	$\chi^2 = 1498.3$ , $df=1$ , $p < .001^{***}$
	Female-adult-lower education	882 (95%)	47 (5%)	929	$\chi^2 = 750.51$ , $df=1$ , $p < .001^{***}$
	Female-adult-higher education	1,167 (98%)	27 (2%)	1,194	$\chi^2 = 1088.4$ , $df=1$ , $p < .001^{***}$
	Male-preadolescent-lower education	26 (100%)	0 (0%)	26	$\chi^2 = 26$ , $df=1$ , $p < .001^{***}$
	Male-preadolescent-higher education	37 (100%)	0 (0%)	37	$\chi^2 = 37$ , $df=1$ , $p < .001^{***}$
	Female-preadolescent-lower education	670 (100%)	0 (0%)	670	$\chi^2 = 670$ , $df=1$ , $p < .001^{***}$
	Female-preadolescent-higher education	666 (100%)	0 (0%)	666	$\chi^2 = 666$ , $df=1$ , $p < .001^{***}$

\* The tokens produced by the female speakers are fewer than those produced by the male speakers since the male speakers mostly dominated the conversations. In addition, the occurrences produced by the nine female speakers of lower educational backgrounds are comparable with those produced by the two female speakers of higher educational backgrounds since the speakers of higher educational backgrounds participated in most recording sessions.

The results of the chi-square test in Table 6 show highly significant differences across the board in the occurrence of final [-a] and [-e] across speakers and social categories. Most notably, we see a major shift in real time in the patterns of use between the 1970s corpus and the 2000s corpus, indicated in Table 6 with bold lines. In the 1970s corpus the occurrence of final [-e] is much more common than that of final [-a], while final [-a] is the predominant variant in the 2000s corpus. Figure 5 shows the percentage of final [-a] and [-e] variants produced by JI speakers from the two corpora, classified by generation, gender, and education level.

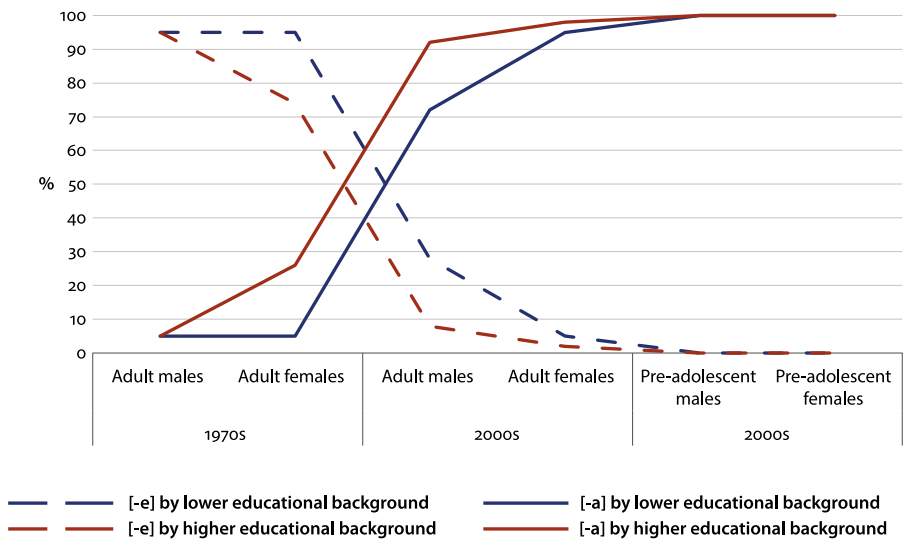


Figure 5. Final [-a] and [-e] across genders, educational backgrounds, and generations

The figure highlights a demonstrable generational shift between the 1970s and 2000s cohorts, but relative stability across age groups in the 2000s. The results of the chi-square test for independence are presented in Table 7.

Table 7. Differences between generations, education, and genders

Differences between generations			Chi-square test for independence
Male-adult-higher education-70s	Male-adult-higher education-2000s	Male-preadolescent-higher education-2000s	$\chi^2 = 1025.6, df = 2, p < .001^{***}$
Male-adult-lower education-70s	Male-adult-lower education-2000s	Male-preadolescent-lower education-2000s	$\chi^2 = 1010.8, df = 2, p < .001^{***}$



Table 7. (continued)

Differences between generations			Chi-square test for independence
Female-adult-higher education-70s	Female-adult-higher education-2000s	Female-preadolescent-higher education-2000s	$\chi^2 = 1098.4$ , $df = 2$ , $p < .001^{***}$
Female-adult-lower education-70s	Female-adult-lower education-2000s	Female-preadolescent-lower education-2000s	$\chi^2 = 1174.8$ , $df = 2$ , $p < .001^{***}$
Differences between educational backgrounds			
Male-adult-higher education-70s	Male-adult-lower education-70s		$\chi^2 = 2.6149\text{e-}28$ , $df = 1$ , $p = 1$
Male-adult-higher education-2000s	Male-adult-lower education-2000s		$\chi^2 = 278.94$ , $df = 1$ , $p < .001^{***}$
Female-adult-higher education-70s	Female-adult-lower education-70s		$\chi^2 = 21.84$ , $df = 1$ , $p < .001^{***}$
Female-adult-higher education-2000s	Female-adult-lower education-2000s		$\chi^2 = 11.34$ , $df = 1$ , $p < .075$
Male-preadolescent-higher education-2000s	Male-preadolescent-lower education-2000s		n/a <sup>†</sup>
Female-preadolescent-higher education-2000s	Female-preadolescent-lower education-2000s		n/a
Differences between genders			
Male-adult-higher education-1970s	Female-adult-higher education-1970s		$\chi^2 = 25.936$ , $df = 1$ , $p < .001^{***}$
Male-adult-higher education-2000s	Female-adult-higher education-2000s		$\chi^2 = 46.826$ , $df = 1$ , $p < .001^{***}$
Male-adult-lower education-1970s	Female-adult-lower education-1970s		$\chi^2 = 0.022285$ , $df = 1$ , $p = .881$
Male-adult-lower education-2000s	Female-adult-lower education-2000s		$\chi^2 = 203.57$ , $df = 1$ , $p < .001^{***}$
Male-preadolescent-lower education-2000s	Female-preadolescent-lower education-2000s		n/a
Male-preadolescent-higher education-2000s	Female-preadolescent-higher education-2000s		n/a

† No  $p$ -value is generated from the test as one of the cell values (the occurrence of final [e]) equals to zero.

### 3.3 The role of gender and education

If we assume, as we have above, that final [-a] is a feature adopted from Standard Indonesian, then the limited occurrences of the variant with final [-a] in the 1970s results across speakers suggests that the influence of SI on JI was more limited at that time. This follows from the fact that during the 1970s SI was still in the early stages of its spread and influence through formal education (Steinhauer, 1994).

Correspondingly, the robust occurrences of the final [-a] variant in the 2000s results, which appears across educational background, gender, and age groups (adult and pre-adolescent), aligns with the hypothesis that in the intervening decades SI exercised a broad and pervasive influence on JI. The abrupt increase in the final [-a] variant in only one generation of speakers is entirely in line with the success story of SI cultivation, and follows from Sneddon's (2003, p.7) observation that the most successful period in the growth of SI occurred under President Soeharto's leadership (1966–1998), when the government viewed the standardization and modernization of Indonesian as essential to support the stability of the nation. Accelerating the spread and influence of SI through formal education was seen as fundamental to the nation's political stability, and an absolute necessity to support the government's economic development program. Also of note is that one of the major programs of Indonesian standardization was spelling reform. The standardized spelling system (*Ejaan Yang Disempurnakan*) was introduced in 1972 and is obligatorily used in schools. It is also used widely in other official matters, printed and electronic media, and more recently in internet media. We hypothesize that standardized spelling with final [-a] has promoted the use of final [-a] even in spoken language, especially among more educated (and thus highly literate) JI speakers.

The patterns that we see among women in the two corpora lend support to these hypotheses. The higher frequency of final [-a] among women with more education in the 1970s suggests that this group may have started to lead the change in progress. If we interpret the incoming use of final [-a] as a change from above due to contact with SI, these findings are in line with previous literature that demonstrates women leading in both the maintenance of prestige forms and the adoption of changes from above (Gordon, 1997; Holmes, 2013; Labov, 1990, 2001; Meyerhoff, 2011; Trudgill, 1972, 1986). This pattern of women, particularly highly educated women, adopting use of SI and SI forms has also been demonstrated more recently in Indonesia (Smith-Hefner, 2009; Travis & Ghina, 2021; Zen & Starr, 2021).

In the 2000's corpus, we observe a prevalence (95%) of the final [-a] even among female speakers with lower educational attainment. However, this may also follow from our previous hypothesis that SI and SI forms are adopted as

changes from above, in one of at least two possible ways. One possibility is that although this group has less access to education, there is still an incentive to adopt SI and SI forms as a type of symbolic capital (Eckert, 1989), and that for women in the 2000s corpus, even those with less education, this variant was already known and available for this use. A second possibility is that as the change has spread through the speech community, both gender and education differences have been blurred. Similar patterns have been reported in many communities elsewhere (see Satyanath, 2021), where social differences are neutralized as a change reaches completion.

#### 4. Conclusion

The synchronic patterns that we find in Jakarta Indonesian have roots in the linguistic history of Indonesia's capital and its complex language ecology. Table 8 summarizes this chronological history with respect to the particular variable in question.

**Table 8.** A summary of chronological developments

Late 18th–early 19th centuries	Late 19th century	End of WW II / Indonesian independence (1945)–present time	Present time
<ul style="list-style-type: none"> <li>– Portuguese based creole was replaced by Betawi Malay</li> <li>– The emergence of Urban Betawi: [-e]</li> </ul>	<ul style="list-style-type: none"> <li>– Expansion of Batavia</li> <li>– The emergence of Rural Betawi in contact with Sundanese: [-aʔ]</li> </ul>	<ul style="list-style-type: none"> <li>– The development of SI and orthographic standardization: [-a]</li> <li>– The emergence of JI (an admixture of Betawi and SI): The increased use of [-a] in JI (influence from SI)</li> </ul>	<ul style="list-style-type: none"> <li>– Widespread adoption of SI</li> <li>– Prevalence of SI [-a] in colloquial JI</li> </ul>

In this study we focused on the emergence of the final [-a] in Jakarta Indonesian. We demonstrated that this variant is rapidly replacing final [-e], a variant which had been primarily associated with urban Betawi speakers, and which was still produced in the 1970's by non-Betawi Jakarta Indonesian speakers (Wallace, 1976) due to the influence of Betawi on the emerging variety Jakarta Indonesian. We find that in the 2000s both adult and pre-adolescent speakers have shown a sharp decline in use of the final [-e] variant in function words in favor of final [-a], and we hypothesize that this is due to contact with Standard Indonesian. The fact




that this change seems to be led by more highly educated speakers lends support for this hypothesis, as these are the speakers most likely to be regular users of SI.

This study serves as an example of the insights to be gained from the use of corpora of naturalistic speech, particularly in complex language contact scenarios such as that of Jakarta; there are few studies that have been able to take advantage of corpora such as these (but see Satyanath, 2018, 2024; Suokhrie, 2016; Imchen, 2017; and studies cited in Satyanath, 2025), and even fewer that have had multiple corpora of colloquial speech from different time periods, allowing for real-time analyses of changes in progress. These two corpora in particular can give us insight into how a complex admixture of varieties such as Jakarta Indonesian may emerge. In this case our findings provide evidence for the increasing influence of SI on JI, but previous studies have not demonstrated the same: Abtahian et al. (2021), for example, in a study of first-person pronouns in Jakarta Indonesian, find a great deal of variation in the pronominal system, but no evidence of change away from Betawi forms or toward SI forms. Further studies can examine more closely the trajectory of other linguistic forms in Jakarta Indonesian and determine whether there is indeed an ongoing shift from a variety that is closer to Betawi to a variety that is more influenced by Standard Indonesian.

## Acknowledgments

The first author would like to thank Abby Cohn, John Wolff, and Draga Zec for providing insightful guidance and encouragement throughout the duration of this study, as well as two anonymous reviewers and the editor for valuable comments and suggestions on earlier versions of this manuscript.

## References

-  Abtahian, Maya R., Cohn, Abigail C., Djenar, Dwi Noverini, & Vogel, Rachel C. (2021). Jakarta Indonesian first-person singular pronouns: Form, function, and variation. *Asia-Pacific Language Variation*, 7(2), 186–215.
- Alisjahbana, Sutan Takdir (1971). Some planning processes in the development of the Indonesian-Malay language. In Joan Rubin & Bjorn Jernudd (Eds.), *Can language be planned?* (pp. 170–178). Honolulu: University Press of Hawai'i. <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/b048eb9c-d2e8-49aa-84ad-d41b27c65299/content>
-  Ananta, Aris, Arifin, Evi Nurvidya, Hasbullah, M. Sairi, Handayani, Nur Budi, & Pramono, Agus (2015). *Demography of Indonesia's ethnicity*. Singapore: Institute of Southeast Asian Studies.
-  Angle, John, & Hesse-Biber, Sharlene N. (1981). Gender and prestige preference in language. *Sex Roles*, 7, 449–461.

- doi Castles, Lance (1967). The ethnic profile of Djakarta. *Indonesia*, 3, 153–204.
- Central Bureau of Statistics (2025, February 12). Census Population 2010. <http://sp2010.bps.go.id/index.php/site/index>
- Cohn, Abigail C., & Vogel, Rachel C. (2019). *Variation in two patterns of word-initial deletion in Jakarta Indonesian: Insight from naturalistic data*. Proceedings of the 19th International Congress of Phonetic Sciences. The Australasian Speech Science and Technology Association (ASSTA). [https://assta.org/proceedings/ICPhS2019/papers/ICPhS\\_87.pdf](https://assta.org/proceedings/ICPhS2019/papers/ICPhS_87.pdf)
- doi Cole, Peter, Hermon, Gabriella, & Tjung, Yassir N. (2006). Is there *pasif semu* in Jakarta Indonesian? *Oceanic Linguistics*, 45, 65–90.
- doi Eckert, Penelope (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, 1(3), 245–267.
- doi Eckert, Penelope, & McConnell-Ginet, Sally (2003). *Language and Gender*. Cambridge: Cambridge University Press.
- doi Grijns, Cornelis D. (1991). *Jakarta Malay I: A multidimensional approach to spatial variation*. Leiden: KITLV Press.
- Gil, David, Tadmor, Uri, Bowden, John, & Taylor, Bradley (1999–2015). Item “Jakarta Field Station” in collection “MPI EVA corpora”. *The Language Archive*. <https://hdl.handle.net/1839/3a60a94c-0b69-4513-bf28-de0dbea23471>
- Google. (n.d.) [Google Maps Indonesia]. Retrieved September 27, 2024, from <https://maps.app.goo.gl/HAKPH3kfKbf8o6j77>
- doi Gordon, Elizabeth (1997). Sex, speech, and stereotypes: Why women use prestige speech forms more than men. *Language in Society*, 26(1), 47–63.
- Hidajat, Lanny (2010). *The acquisition of verb argument structure in basilectal Jakarta Indonesian* [Unpublished doctoral dissertation]. University of Delaware.
- doi Holmes, Janet (2013). *An Introduction to Sociolinguistics*. London: Pearson.
- Iguchi, Masatoshi (2015). *Java Essay: The history and culture of a southern country*. Market Harborough: Troubador Publishing Ltd.
- Ikranagara, Kay (1980). Melayu Betawi Grammar. *NUSA: Linguistics Studies in Indonesian and Languages in Indonesia*, 9, 0–150.
- Imchen, Yutensingla (2017). *Bidialectalism in Mokokchung Town* [Unpublished M.Phil. dissertation]. University of Delhi.
- doi Suokhrie, Kelhouvinuo (2016). Clans and clanlectal contact: Variation and change in Angami. *Asia-Pacific Language Variation*, 2(2), 188–214.
- Jakarta Regional Government (2024, November 1). *Jakarta Area*. Official Portal of DKI Jakarta Province. <https://www.jakarta.go.id/>
- Kähler, Hans (1966). Worterverzeichnis Des Omong Djakarta [Vocabulary of Jakarta talk], *Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg* [Publications of the Seminar for Indonesian and South Sea Languages of the University of Hamburg], 5.
- Kurniawan, Ferdinan (2018). *Phonological variation in Jakarta Indonesian: An emerging variety of Indonesian* [Unpublished doctoral dissertation]. Cornell University.
- doi Kurniawan, Ferdinan (2023). Variation and change in Jakarta Indonesian: Evidence from final glottals. *Oceanic Linguistics*, 62(2), 267–288.

- Labov, William (2001). *Principles of Linguistic Change: Social Factors*, 2. Malden: Wiley-Blackwell.
-  Labov, William (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2), 205–254.
- Merrillees, Scott (2015). *Jakarta: Portraits of a Capital 1950–1980*. Sheffield: Equinox Publishing.
- Meyerhoff, Miriam (2011). *Introducing sociolinguistics*. New York: Routledge.
- Muhadjir (1981). Morphology of Jakarta dialect, affixation and reduplication. *NUSA: Linguistics Studies in Indonesian and Languages in Indonesia*, 11, 0–117.
- Poedjosoedarmo, Soepomo (1982). Javanese influence on Indonesian. *Pacific Linguistics series D* (38), 148–154. Canberra: Australian National University.
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.Rproject.org/>
-  Romaine, Suzanne (2003). Variation in Language and Gender. In Janet Holmes & Miriam Meyerhoff (Eds.), *The Handbook of Language and Gender* (pp. 98–118). Hoboken: Wiley-Blackwell.
- Satyanath, Shobha (2018). Kohima: Language variation and change in a small but diverse city in India. In Dick Smakman & Patrick Heinrich (Eds.), *Urban Sociolinguistics: The city as a linguistic process and experience* (pp. 95–112). Milton Park: Routledge.
-  Satyanath, Shobha (2021). Genealogies of Sociolinguistics in India. *Journal of Sociolinguistics*, 25(5), 762–784.
-  Satyanath, Shobha (2024). Contact, diffusion and divergence: Classifiers in Assamese and its two contact varieties – Divergence in classifier systems. *Journal of Language Contact*, 16(1), 104–139.
-  Satyanath, Shobha (2025). Sociolinguistics of multicultural societies: Implications for data and methodology. In Christopher Cieri, Lauren Hall-Lew, Katie Drager, & Malcah Yaeger-Dror (Eds.), *Dimensions of linguistic variation* (pp. 70–93). Oxford University Press.
- Silver, Christopher (2008). *Planning the megacity: Jakarta in the twentieth century – Planning, history and environment series*. Milton Park: Routledge.
-  Smith-Hefner, Nancy (2009). Language shift, gender, and ideologies of modernity in Central Java, Indonesia. *Journal of Linguistic Anthropology*, 19(1), 57–77.
- Sneddon, James (2003). *The Indonesian language: Its history and role in modern society*. Sydney: UNSW Press.
- Sneddon, James (2006). *Colloquial Jakartan Indonesian*. Canberra, ACT: Pacific Linguistics.
- Soekmono, R. (1973). *Pengantar Sejarah Kebudayaan Indonesia [Introduction to History and Culture of Indonesia]* (2nd ed.). Yogyakarta: Penerbit Kanisius.
-  Steinhauer, Hein (1994). The Indonesian language situation and linguistics: Prospects and possibilities. *Bijdragen tot de taal-, land-en volkenkunde*, 150(4), 755–784.
-  Travis, Catherine E., & Ghina, Inas (2021). Gender, mobility and contact: Stability and change in an Acehnese dialect. *Asia-Pacific Language Variation*, 7(2), 142–167.
-  Trudgill, Peter (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, 1(2), 179–195.
-  Trudgill, Peter (1986). Social identity and linguistic sex differentiation. In Harold B. Allen & Michael D. Linn (Eds.), *Dialect and language variation* (pp. 395–402). Amsterdam: Elsevier.

- Wallace, Stephen (1976). *Linguistic and social dimension of phonological variation in Jakarta Malay* [Unpublished doctoral dissertation]. Cornell University.
- World Population Review (2025, January 13). *Largest Cities by Population 2024*. <https://worldpopulationreview.com/cities>
- Wouk, Fay (1999). Dialect contact and koineization in Jakarta, Indonesia. *Language Sciences*, 21, 61–86.
- Zen, Evynurul Laily, & Starr, Rebecca Lurie (2021). Variation and contact-induced change in Javanese phonology among multilingual children in Indonesia. *Asia-Pacific Language Variation*, 7(2), 95–19.

## Appendix

**Table A.** JI adult speakers and their backgrounds (the 1970s corpus)

Speakers	SES <sup>^</sup>	Gender	Mother's ethnicity	Father's ethnicity	Age	Word count
FAL-S6 <sup>*</sup>	Low	Female	Javanese	Javanese	21	62
FAL-S7	Low	Female	Javanese	Sundanese	35	120
FAL-S8	Low	Female	Javanese	Javanese	37	243
FAL-S9	Low	Female	Batak	Batak	20	92
FAL-S10	Low	Female	Sundanese	Sundanese	38	129
FAL-S11	Low	Female	Chinese Jakarta	Javanese	23	330
FAL-S12	Low	Female	Javanese	Javanese	28	378
FAL-S13	Low	Female	Javanese	Javanese	31	411
FAL-S14	Low	Female	Javanese	Sundanese	32	312
FAH-S6	Mid/High	Female	Javanese	Javanese	27	215
FAH-S7	Mid/High	Female	Javanese	Sundanese	30	313
MAL-S6	Low	Male	Javanese	Javanese	21	93
MAL-S7	Low	Male	Sundanese	Sundanese	24	222
MAL-S8	Low	Male	Sundanese	Sundanese	32	34
MAL-S9	Low	Male	Chinese Jakarta	Chinese Jakarta	38	57
MAL-S10	Low	Male	Javanese	Sundanese	34	356
MAL-S11	Low	Male	Sundanese	Javanese	21	214
MAL-S12	Low	Male	Ambonese	Javanese	37	143
MAL-S13	Low	Male	Javanese	Javanese	33	111
MAL-S14	Low	Male	Sundanese	Sundanese	25	98
MAL-S15	Low	Male	Sundanese	Sundanese	33	67
MAL-S16	Low	Male	Javanese	Javanese	21	90
MAL-S17	Low	Male	Javanese	Javanese	42	187
MAL-S18	Low	Male	Flores	Javanese	38	164
MAL-S19	Low	Male	Javanese	Javanese	34	133
MAL-S20	Low	Male	Sundanese	Javanese	34	123
MAH-6	Mid/High	Male	Javanese	Javanese	29	11

Table A. (continued)

Speakers	SES^	Gender	Mother's ethnicity	Father's ethnicity	Age	Word count
MAH-7	Mid/High	Male	Chinese	Javanese	22	123
MAH-8	Mid/High	Male	Minang	Minang	23	344
MAH-9	Mid/High	Male	Batak	Batak	33	54
MAH-10	Mid/High	Male	Javanese	Javanese	22	61
MAH-11	Mid/High	Male	Ambonese	Ambonese	38	345
MAH-12	Mid/High	Male	Sundanese	Javanese	21	24
MAH-13	Mid/High	Male	Javanese	Javanese	36	92
MAH-14	Mid/High	Male	Javanese	Javanese	38	104

\* M: male; F: female; A: adult; H: higher education; L: lower education; S1: speaker one, S2: speaker two, and so on.

^ SES: socio-economic status.

Table B. JI adult speakers and their backgrounds (the 2000s corpus)

Speakers	Educational background	Gender	Mother's ethnicity	Father's ethnicity	Age	Word count
FAL-S1	Secondary	Female	Sundanese	Javanese	24	1,300
FAL-S2	Secondary	Female	Javanese	Javanese	29	1,940
FAL-S3	Secondary	Female	Javanese	Chinese Jakarta	47	2,969
FAL-S4	Secondary	Female	Javanese	Javanese	26	1,907
FAL-S5	Secondary	Female	Javanese	Javanese	21	2,134
FAH-S1	Tertiary	Female	Javanese	Flores	25	675
FAH-S2	Tertiary	Female	Javanese	Javanese	26	1,624
FAH-S3	Tertiary	Female	Chinese Javanese	Chinese Javanese	23	2,004
FAH-S4	Tertiary	Female	Chinese Javanese	Chinese Javanese	22	1,271
FAH-S5	Tertiary	Female	Chinese Javanese	Chinese Javanese	34	4,474
MAL-S1	Secondary	Male	Chinese Jakarta	Chinese Manado	34	4,540
MAL-S2	Secondary	Male	Javanese	Eastern Indonesia	49	5,442
MAL-S3	Secondary	Male	Sundanese; Dutch	Ambonese	29	759
MAL-S4	Secondary	Male	Javanese	Javanese	29	3,119
MAL-S5	Secondary	Male	Javanese	Javanese	26	1,557
MAH-S1	Tertiary	Male	Sundanese	Sundanese	27	4,158
MAH-S2	Tertiary	Male	Chinese Jakarta	Chinese Javanese	35	1,154



Table B. (continued)

Speakers	Educational background	Gender	Mother's ethnicity	Father's ethnicity	Age	Word count
MAH-S3	Tertiary	Male	Chinese Jakarta	Chinese Manado	27	4,076
MAH-S4	Tertiary	Male	Javanese	Javanese	27	1,740
MAH-S5	Tertiary	Male	Javanese	Javanese	34	1,732

Table C. JI pre-adolescent speakers and their backgrounds (the 2000s corpus)

Speakers	Parents' educational background	Gender	Mother's ethnicity	Father's ethnicity	Age	Word count
FPL-S1	Secondary	Female	Javanese	Javanese	10–13	6,500
FPL-S2	Secondary	Female	Javanese	Sundanese	10	397
FPL-S3	Secondary	Female	Javanese	Javanese	10–11	570
FPH-S1	Tertiary	Female	Chinese Jakarta	Chinese Jakarta	10–11	6,482
FPH-S2	Tertiary	Female	Chinese Jakarta	Chinese Jakarta	10	1,895
MPL-S1	Secondary	Male	Javanese	Javanese	11	62
MPL-S2	Secondary	Male	Javanese	Javanese	13	223
MPH-S1	Tertiary	Male	Javanese	Javanese	10	613

\*P: pre-adolescent.

Table D. Final [-a] ~ [-e] produced by female adult speakers in the 2000s corpus

Lower educational background				
Word form	Final [-a]	Final [-e]	Total	Chi-square test goodness of fit
<i>ya</i>	228 (96%)	9 (4%)	237	$\chi^2 = 202.37, df = 1, p < .001^{***}$
<i>iya</i>	138 (88%)	19 (12%)	157	$\chi^2 = 90.19, df = 1, p < .001^{***}$
<i>nya</i>	377 (7%)	12 (3%)	389	$\chi^2 = 342.48, df = 1, p < .001^{***}$
<i>dia</i>	72 (94%)	5 (6%)	77	$\chi^2 = 58.29, df = 1, p < .001^{***}$
<i>aja</i>	67 (97%)	2 (3%)	69	$\chi^2 = 61.23, df = 1, p < .001^{***}$
<b>Total</b>	<b>882 (95%)</b>	<b>47 (5%)</b>	<b>929</b>	$\chi^2 = 750.51, df = 1, p < .001^{***}$
Higher educational background				
<i>ya</i>	343 (97%)	12 (3%)	355	$\chi^2 = 308.62, df = 1, p < .001^{***}$
<i>iya</i>	146 (97%)	5 (3%)	151	$\chi^2 = 131.66, df = 1, p < .001^{***}$
<i>nya</i>	499 (99%)	6 (1%)	505	$\chi^2 = 481.29, df = 1, p < .001^{***}$
<i>dia</i>	121 (98%)	3 (2%)	124	$\chi^2 = 112.29, df = 1, p < .001^{***}$
<i>aja</i>	58 (98%)	1 (2%)	59	$\chi^2 = 55.06, df = 1, p < .001^{***}$
<b>Total</b>	<b>1,167 (98%)</b>	<b>27 (2%)</b>	<b>1,194</b>	$\chi^2 = 1088.4, df = 1, p < .001^{***}$

Table E. Final [-a] ~ [-e] produced by male adult speakers in the 1970s corpus

Lower educational background				
Word form	Final [-a]	Final [-e]	Total	Chi-square test goodness of fit
<i>ya</i>	16 (10%)	151 (90%)	167	$\chi^2 = 109.13, df = 1, p < .001^{***}$
<i>iya</i>	5 (3%)	182 (97%)	187	$\chi^2 = 167.53, df = 1, p < .001^{***}$
<i>nya</i>	18 (7%)	225 (93%)	243	$\chi^2 = 176.33, df = 1, p < .001^{***}$
<i>dia</i>	1 (1%)	99 (99%)	100	$\chi^2 = 96.04, df = 1, p < .001^{***}$
<i>aja</i>	0 (0%)	63 (100%)	63	$\chi^2 = 63, df = 1, p < .001^{***}$
Total	40 (5%)	720 (95%)	760	$\chi^2 = 608.42, df = 1, p < .001^{***}$
Higher educational background				
<i>ya</i>	5 (21%)	19 (79%)	24	$\chi^2 = 8.16, df = 1, p < 0.01^{**}$
<i>iya</i>	0 (0%)	12 (100%)	12	$\chi^2 = 12, df = 1, p < .001^{***}$
<i>nya</i>	4 (4%)	96 (96%)	100	$\chi^2 = 84.64, df = 1, p < .001^{***}$
<i>dia</i>	1 (3%)	33 (97%)	34	$\chi^2 = 30.11, df = 1, p < .001^{***}$
<i>aja</i>	0 (0%)	19 (100%)	19	$\chi^2 = 19, df = 1, p < .001^{***}$
Total	10 (5%)	179 (95%)	189	$\chi^2 = 151.12, df = 1, p < .001^{***}$

Table F. Final [-a] ~ [-e] produced by female adult speakers in the 1970s corpus


Lower educational background				
Word form	[-a]	[-e]	Total	Chi-square test goodness of fit
<i>ya</i>	3 (9%)	30 (91%)	33	$\chi^2 = 22.09, df = 1, p < .001^{***}$
<i>iya</i>	1 (7%)	14 (93%)	15	$\chi^2 = 11.26, df = 1, p < .001^{***}$
<i>nya</i>	2 (4%)	55 (96%)	57	$\chi^2 = 49.28, df = 1, p < .001^{***}$
<i>dia</i>	0 (0%)	16 (100%)	16	$\chi^2 = 16, df = 1, p < .001^{***}$
<i>aja</i>	0 (0%)	12 (100%)	12	$\chi^2 = 12, df = 1, p < .001^{***}$
Total	6 (5%)	127 (95%)	133	$\chi^2 = 110.08, df = 1, p < .001^{***}$
Higher educational background				
<i>ya</i>	11 (31%)	25 (69%)	36	$\chi^2 = 5.44, df = 1, p = .019^{**}$
<i>iya</i>	6 (32%)	13 (68%)	19	$\chi^2 = 2.57, df = 1, p = .10$
<i>nya</i>	9 (24%)	29 (76%)	38	$\chi^2 = 10.52, df = 1, p < 0.01^{**}$
<i>dia</i>	4 (20%)	16 (80%)	20	$\chi^2 = 7.2, df = 1, p = .007^{**}$
<i>aja</i>	3 (21%)	11 (79%)	14	$\chi^2 = 4.57, df = 1, p = .032^{**}$
Total	33 (26%)	94 (74%)	127	$\chi^2 = 29.29, df = 1, p < .001^{***}$

## Abstract (Indonesian)

Penelitian ini melaporkan variasi bunyi akhir [-a] ~ [-e] dalam Jakarta Indonesian (JI), sebuah ragam urban baru bahasa Indonesia yang muncul dari percampuran antara bahasa Indonesia dan Betawi dan dituturkan di ibu kota Indonesia. Investigasi ini menggabungkan pendekatan apparent-time (waktu tampak) dan real-time (waktu nyata) untuk mengkaji perubahan variabel dari waktu ke waktu dengan menggunakan dua korpora ujaran naturalistik berskala besar dari tiga generasi penutur. Ketika mengamati perubahan variabel dari waktu ke waktu, kami menemukan adanya kecenderungan penggunaan bentuk standar yang lebih sering ditemukan di antara penutur perempuan dan penutur dengan latar belakang pendidikan yang lebih tinggi. Kajian ini memberi sumbangsih pengetahuan tentang bagaimana sebuah ragam bahasa baru dapat muncul secara bertahap dalam lingkungan multibahasa.

**Kata kunci:** Betawi, Jakarta Indonesian, korpora, waktu tampak, waktu nyata

## Address for correspondence

Ferdinan Okki Kurniawan  
Atma Jaya Catholic University of Indonesia  
Jl. Jendral Sudirman No 51  
Jakarta, 12930  
Indonesia  
ferdinanokki@gmail.com  
ferdinan.kurniawan@atmajaya.ac.id  
 <https://orcid.org/0000-0003-3591-2766>

## Co-author information

Maya Ravindranath Abtahian  
Department of Linguistics  
University of Rochester  
maya.r.abtahian@rochester.edu

## Publication history

Date received: 15 May 2023  
Date accepted: 15 April 2024  
Published online: 25 March 2025