

Jakarta Indonesian first-person singular pronouns

Form, function and variation

Maya Ravindranath Abtahian,¹ Abigail C. Cohn,²

Dwi Noverini Djenar,³ and Rachel C. Vogel²

¹ University of Rochester | ² Cornell University | ³ The University of Sydney

Jakarta Indonesian is a colloquial variety of Indonesian spoken primarily in Indonesia's capital, where it was originally a contact variety between Betawi, the local variety of Malay, and Standard Indonesian. Like other varieties of Indonesian, Jakarta Indonesian is a language with a relatively open system of pronominal reference and multiple forms for self-reference. In this paper we focus on variation in the use of first-person pronouns in Jakarta Indonesian, using two corpora of spoken data collected three decades apart. We employ both quantitative and qualitative methods to examine the form, function and social meaning of 1sg pronouns in Jakarta Indonesian, investigating both inter- and intra-speaker variation over time.

Keywords: Jakarta Indonesian, variation, quantitative methods, qualitative methods, first person pronouns, self-reference

1. Introduction

In Western European languages, pronouns belong to a closed set of semantic and grammatical forms grouped according to person and number and represent paradigmatic oppositions (Frawley, 2003). Following Brown and Gilman's (1960) influential study on the historical shift of address patterns in these languages, numerous studies, including in non-Western European languages, have adopted their analytical insight to analyze the use of second person pronouns in "address" (see Kluge & Moyna, 2019 for a recent collection). The enduring quality of Brown and Gilman's study is evidenced both in studies that adopt their "power" and "solidarity" semantics as their analytical approach and also those that critique the approach (see e.g., Fernández-Mallat, 2020 and Raymond, 2016 for dialects of

Spanish, and Hepburn, Wilkinson, & Shaw, 2012; and Lerner & Kitzinger, 2007, for English).

Compared to addressee-reference, self-reference has received much less attention. Although studies of non-Western European languages that have multiple forms for referring to the self, such as Korean and Japanese (for recent studies, see Christofaki, 2018; Lee, 2018), have helped draw attention to the complex relationship between pronominal and nominal forms and senses of the self that these forms are used to index, self-reference in other languages with similarly rich inventories remains under-explored (although see Khairunnisa, this volume, on Sasak), and few studies have taken a variationist approach (although see Meyerhoff, 2009; Meyerhoff, 2015b; Schnell & Barth, 2018). Indonesian also provides an interesting case because of the added complexity created by language contact.

In Indonesian, language contact has resulted in complexity in the pronominal system in at least two respects. First, although pronouns in Indonesian can be distinguished in terms of number, the loss of number distinction in related languages such as Javanese, Balinese and Sundanese (see Adelaar & Hajek, forthcoming) has influenced pronominal use in Indonesian. For example, the first-person pronoun inclusive *kita* is often used in ways that blur number distinction. Second, Indonesian has been described as a language with an “open” pronominal system in the sense that it allows pronoun borrowing and the use of lexical nouns in self-reference. But pronoun borrowing has also resulted in phonological variation through language contact (e.g., Hokkien *gua* ‘1sg’ has the variant *gue*, and Arabic *ana* ‘1sg’ has the variant *ane*; in both cases the variants emerged via Betawi, the variety of Malay spoken in Jakarta), thereby adding to the pool of forms available to speakers and contributing to inter- and intraspeaker variation.

In this paper we focus on variation in the use of first person pronouns (henceforth, 1sg) in Jakarta Indonesian, an emerging variety of colloquial Indonesian and increasingly the mother tongue of several million Indonesians. In addition to examining variation in self-reference, the study of Jakarta Indonesian also offers an opportunity for examining the interplay between standard language and an urban vernacular. This is an exciting challenge in the enterprise of globalizing sociolinguistics (Meyerhoff & Stanford, 2015, p.3) particularly because this relationship is underexplored in non-WEIRD (Western, Educated, Industrialized, Rich, and Democratic) communities.

This paper presents an analysis of 1sg in Jakarta Indonesian that is preliminary but one that attempts to offer a methodologically broad approach. Our data consist of two corpora collected three decades apart: the data collected by the Jakarta Field Station between 2004–2012 and the Betawi-Jakarta data collected by Stephen Wallace in 1976. We employ both quantitative and qualitative methods to

examine inter- and intra-speaker variation in pronoun use and consider the ways in which pronouns are used in the community to index the self and identities in dynamic ways. We follow Meyerhoff (2015a) in treating identity as a state and a process (cf. Eckert, 2008) and analyzing the relationship between language and identity using both quantitative and qualitative sociolinguistic methods.

We present our analysis in three parts, each focusing on a different level of variation. First, we consider a macro-level analysis of interspeaker variation based on the Jakarta Indonesian corpus, focusing on broad social categories such as language group, age and gender. We then examine 1SG pronoun use at the interactional level, focusing on variation in the speech of a single research assistant in the Jakarta Indonesian corpus as well as that of another research assistant in the older Wallace corpus. These two speakers were both facilitators across multiple conversations that make up the data. This second analysis allows for a narrower focus on intraspeaker variation and provides additional nuance to the initial description of the social factors that constrain the use of 1SG pronouns among Jakarta speakers. Finally, we turn to a qualitative analysis of the specific pronominal forms to show the ways in which our quantitative findings are reflected at the interactional level. Each of these levels of analysis could be developed further in its own right, but as few studies have considered community-wide variation and change alongside pronominal alternation in interaction, we use this an opportunity to contribute to the conversation on the need for this kind of multidimensional analysis.

2. Background

Jakarta Indonesian is a part of the complex, dynamic linguistic landscape of Indonesia, where approximately 700 languages are spoken by 260 million people across an archipelago of over 14,000 islands. During the second half of the 20th century, Standard Indonesian, a variety of Malay, was developed and promoted as the national language for the new nation-state of Indonesia (Sneddon, 2003). Following a series of extraordinarily successful language planning efforts, Indonesian is now the dominant language for a large and growing percentage of the population. As of the 2010 census Indonesian was reported to be the second most widely spoken language at home (Ananta, Arifin, Hasbullah, Handayani, & Pramono, 2015), with 42 million speakers, making it the world's 30th most widely spoken native language worldwide. The use of the label "Indonesian" in the census, however, is somewhat complicated by the fact that the term is used to refer to both the standard variety of the language and colloquial varieties associated with different regions. In recent decades, colloquial varieties of Indonesian have become the native languages of a significant portion of the population. These are some-

times, but not always, distinguished by speakers as varieties distinct from Standard Indonesian (Abtahian, Cohn, White, & Yanti, 2019) and may be more or less mutually intelligible with other colloquial varieties of the language.

One of these colloquial varieties is Jakarta Indonesian (JI), the variety spoken in the nation's densely populated, urban capital of Jakarta, shown in the map in Figure 1. JI is considered a contact variety between Betawi, the local variety of Malay historically spoken in the area, and Standard Indonesian (see Ikranagara, 1975; Sneddon, 2006; Wouk, 1999). It is distinguished from Standard Indonesian by a number of grammatical, lexical and phonological features. It is generally characteristic of contact varieties in that the features that distinguish it from Standard Indonesian are described as showing variable realization for speakers (Cohn & Vogel, 2019; Kurniawan, 2018; Sneddon, 2006). In addition, Jakarta Indonesian has exerted influence on other urban varieties through its use by “students and educated people” (Poedjosoedarmo, 1982, p.142) moving between Jakarta and other cities, and dissemination by various forms of media.



Figure 1. Map of Java (source: CartoGIS Services, ANU College of Asia and the Pacific, The Australian National University)

3. Self-reference in Jakarta Indonesian

In delimiting the variable domain of self-reference in Indonesian, it first bears mentioning that Indonesian's range of forms for referring to speaker and addressee includes lexical nouns (kin terms, name, title and other nouns) as well as pronouns, and unlike in Western European languages generally, pronouns are not always the default forms for indexing speaker- and addressee- roles (Conners, Brugman, & Adams, 2016). The availability of the different forms has posed a challenge for a grammatical description that seeks to determine which items

should be listed in the Indonesian pronominal paradigm. For example, in terms of grammatical category, lexical nouns are clearly not pronouns, yet in terms of grammatical function, they are similar to pronouns. Semantically, nouns such as kin terms denote kin hierarchy and proper names denote a unique identity, whereas pronouns, when used as self- and addressee-reference, denote speaker and addressee roles (see Djenar & Sidnell, forthcoming). But this semantic distinction is complicated by the fact that pronouns used by Indonesian speakers are of different provenance and linked to various social typifications often invoked in interaction (see Djenar, 2015).

In an attempt to address the problem of categorization, Mahdi (2001) suggests that we treat all forms – pronouns and lexical nouns – as “nominals”. He argues that strictly speaking, pronouns do not constitute a word class in Indonesian and proposes that a distinction be made between what he calls “nonpersonal nominals” and “personal nominals”. Nonpersonal nominals include demonstratives, count and collective nouns, and locatives, whereas personal nominals include proper names, personal pronouns, kin terms and definite articles referring to person and nonhuman animates. With the exception of definite articles (which are only used in third-person reference) personal nominals are used in self-, addressee-, and third-party reference.

Typologically, a language having multiple pronouns for referring to the addressee is not uncommon. However, in terms of 1SG, the range of options in Western European languages is limited compared to East and Southeast Asian languages. English and French, for example, basically have *I* and *je* respectively, though they have inflected forms (*my*, *mine*, *moi*, *mon*, *ma*, *mes*).¹ In contrast, Indonesian speakers can use the standard pronouns, *saya* and *aku*, the standard plural pronoun *kita* ‘1PL’, its variant *kite* ‘1PL’, or pronouns from different vernaculars, such as *gua/gue* (Jakartan Indonesian, borrowed from Hokkien), *beta* (Ambon Malay), and *ane* (from Arabic via Betawi) and *ana* (recent borrowing from Arabic via affiliation with Islam).

Intraspeaker variation, that is, a speaker alternating between different forms for reference to the self and the addressee in speaking to the same or different addressee, has both fascinated and puzzled analysts. Kartomihardjo (1981) argues that, despite the possibility of alternating between different forms, variation is nonetheless largely constrained in any given interaction by participant relationship. He writes that “[t]here is a large variety of TA [Terms of Address] and PR [Pronominals] but there is not a great deal of choice as to which one is used in any given situation. The choice of TA or PR clearly states the speaker’s relation-

1. English *one* and French *on* are used as generic 1SG.

ship vis-à-vis the other participants and thus pre-determines the future course of their dealings” (p. 87).

Later studies show that, while participant relationship is an important consideration, it does not directly correlate with speakers’ choices of form. For example, Djenar (2006) argues that speakers switch between different forms in continual acts of positioning. In a later study, Djenar (2008) examines the shifting pattern in the use of Indonesian first-person pronouns *aku* and *saya* to argue that pronoun alternations can be better understood in terms of flexible self-categorizations than as a contrast between intimacy and formality respectively. The studies by Englebretson (2007) and Manns (2012) show that intraspeaker variation is fundamentally about mobilization of forms in stancetaking (also see Djenar, Ewings, & Manns, 2018). Ewing and Djenar (2019) suggest that variation and sequential placement of pronouns (and lexical nouns) interact in stancetaking work. Other studies describe variation as part of practices of regional differentiation (Djenar, 2015; see also Ewing, 2014) and marking ethnic belonging (Goebel, 2010; for Jakarta, see Grijns, 1980, p. 196).

Previous studies have helped us understand variation by showing how speakers deploy linguistic forms to index interpersonal stances and regional/ethnic orientation. However, because these previous analyses are mainly qualitative, it is difficult to know to what degree social variables such as age, gender, and ethnic background play a role in practices of self-reference involving pronouns. Unlike pronouns, kin terms, which are widely employed in self- and addressee-reference, semantically denote age hierarchy (and gender). Yet, even as the use of kin terms has been noted as common, there has been no study we are aware of that examines their frequency of use compared to pronouns. Similarly, the observation that pronouns are not the default forms for self- and addressee-reference in Indonesian has been based mainly on qualitative analysis. A study that combines quantitative and qualitative methods can inform us not only about how frequently pronouns are used compared to other forms but also the degree of intra- and inter-speaker variation. This study is a contribution toward that endeavor.

We focus on variation in 1sg in this study as a crucial step toward a larger study that considers pronominal and non-pronominal forms in the analysis of self-reference. We focus on Jakarta Indonesian as it is a major variety of the language – major in the sense that it is a prestige urban variety widely employed in mass media and which has had a considerable influence on urban varieties spoken in other parts of Indonesia, including Kalimantan and West Timor (see Errington, 2014). We consider seven variants of the 1sg pronoun: *saya*, *gua*, *gue*, *aku* and its bound variant *-ku*, *kita* and its variant *kite* (see Table 1 below); in different

parts of this analysis we have focused on a subset of these.² Other self-reference forms, such as the 1PL *kami*, which tends to be used when people are speaking in Standard Indonesian, kin terms and names, which form part of a broader set of personal nominals, are excluded from our study. Similarly, pronoun ellipsis is not discussed.

Table 1. Variants of 1sg pronoun under examination, based on Djenar et al. (2018), Englebretson (2007), Ewing (2005), Ewing (2019), Manns (2014), Sneddon (2006)

Pronoun	Previous descriptions
saya	Sanskrit borrowing; public identity; used in interactions between those who are not social intimates
gue	Hokkien borrowing via Betawi; associated with Jakartan youth identity. In Bandung, associated with outspokenness, exaggerated speech, and bravado
gua	Hokkien borrowing, common among Jakarta speakers, less linked with youth identity than <i>gue</i> ; commonly used by ethnic Chinese speakers in other cities
aku	From Malay; used in interactions between social intimates; indexes personal identity, more relaxed and intimate self
-ku	Clitic form of <i>aku</i>
kita	1PL inclusive in Standard Indonesian, inclusive and exclusive in colloquial Indonesian including Jakarta Indonesian; also used to denote first person singular
kite	Betawi Malay variant of <i>kita</i> , mainly used as 1sg

Most of the attention, in our study as well as in previous studies, is on the first four of the 1sg variants. Englebretson (2007) has described *saya* and *aku* in terms of public and non-public contexts, respectively, and the identities linked to these in interaction. *Saya*, a Sanskrit borrowing, tends to be used in public interaction and by speakers who are not social intimates, whereas *aku* (from Malay) indexes a more relaxed and intimate self and tends to be used in interaction between social intimates. Djenar (2008) accounts for intraspeaker variation involving these pronouns in terms of flexible and context-dependent self-categorization, arguing that speakers use *aku* as a strategy for asserting personal identity but not identity as an autonomous self, for when a speaker uses *aku*, it is always relative to some addressee. *Saya*, like *aku*, is a social self-concept, the use of which indexes the self as part of a collective identity.

2. *Gua/gue* and *kita/kite* are sometimes understood to be part of a variable process of a → e raising (see Kurniawan, 2018), although there is evidence that these forms have become lexicalized as well and can take on an independent function.

Gue is strongly identified with a Jakartan youth identity, and outside Jakarta may be perceived to be out of place, as Manns (2014) shows through an analysis of metapragmatic comments by radio broadcasters from Malang, a major city in East Java. Djenar et al. (2018, pp.45–47) and Ewing (2019) mention that in Bandung, a city 150 kilometers from Jakarta, *gue* is stereotypically associated with outspokenness, exaggerated speech and bravado, that is, features that are stereotypically associated with Jakarta. This suggests that, although Jakarta Indonesian has influenced other colloquial varieties, certain elements such as 1SG pronouns remain strongly linked to Jakartan identity. Those who do not identify with Jakarta would be less likely to use it, and if they do, it might be within the context of joking or exaggerating. *Gua*, while also commonly used by Jakarta speakers, is not as strongly linked to youth identity as *gue* is. This pronoun was widely used by ethnic Chinese speakers in major cities such as Surabaya and Malang. *Gua*, like the second person pronoun *lu* ‘2SG’, is a Hokkien borrowing. The variant, *gue*, is borrowed from Hokkien via Betawi.

The last two of the variants, *kita* and *kite*, denote first-person plural (henceforth 1PL) inclusive but are also used as generic ‘we’. *Kite* (also written *kité*) is a variant from Betawi Malay, and in that vernacular, it is mainly used as 1SG reference, while *kita* is a standard Indonesian pronoun. Indonesian has two pronouns denoting 1PL, distinguished in terms ofclusivity: *kami* is inclusive, while *kita* is exclusive. In colloquial Indonesian, *kita* is used for both inclusive 1PL and 1SG (Ewing, 2005), and this is also the case in Jakarta Indonesian (Sneddon, 2006). We include *kita/kite* in our analysis to show that Indonesian pronouns do not always mark number. Speakers use *kita* and *kite* to mark plurality and to refer to the self as a participant in “public” contexts, such as work. These provide little support for Kartomihardjo’s argument that the choice of form is determined and fixed/set by participant relationships.

The body of literature referenced above offers some insight into the different 1SG pronouns, including their use for indexing non-Jakarta vis-a vis Jakarta identity. However, less is known about actual pronoun usage in Jakarta itself. Part of our goal here is to address that gap. We also aim to offer a real-time analysis of pronoun usage in Jakarta Indonesian by drawing on corpora from two different time periods and considering both macro-sociolinguistic and micro-interactional factors. This allows us to examine pronominal variation in Jakarta Indonesian in terms of distribution in the community and its social meanings over time.

4. Methods

4.1 Corpora

We have taken advantage of two existing corpora collected at two separate time periods. Both corpora consist of conversations between Jakarta residents, and are roughly comparable in terms of topics, although the second corpus is wider-ranging and takes place in a greater variety of settings. The data allow us to describe the use of different 1sg pronoun variants in relatively vernacular speech, while the time depth between the collection of the two corpora allows us to consider change in real time. The first corpus, the Betawi-Jakarta corpus (the BJI corpus), contains audio recordings of conversations collected between 2004–2012 under the auspices of the Max Planck Institute for Evolutionary Anthropology Jakarta field station (Gil & Tadmor, 2014). The recordings were done in informal settings in Jakarta with 143 speakers. They include 28 hours of recorded speech (a total of 75,079 utterances) transcribed by Indonesian linguists using ELAN. Each conversation in the corpus is given a code and segmented into utterances. All utterances are translated, further segmented into morphemes, and glossed by morpheme. Thus, the corpus can be searched by speaker, conversation code, Indonesian word, morpheme, or morpheme gloss.

The second corpus consists of transcribed conversations from recordings collected in the mid-1970's under the direction of Stephen Wallace. The conversations were collected as part of Wallace's 1976 dissertation on phonological variation in JI (which he called Modern Jakarta Malay). A total of 35 hours of recordings involving over 200 adult speakers were collected by fifteen research assistants. These assistants recorded conversations among friends, relatives and neighbors (Wallace was not present for the recordings). The recordings were transcribed in the 1970's after they were collected, and the transcriptions are now publicly available from Cornell University's digital repository through eCommons. The transcripts have been converted to machine-readable text and are searchable by character, word, and string of words (though not all words can always be identified by the search function). For the purposes of this paper, it was necessary to read through the transcripts and identify tokens of 1sg pronouns manually.

4.2 Interspeaker analysis in BJI

We analyzed 1sg pronoun usage in these two corpora in three phases. In the first phase we examined social factors that condition interspeaker variation within the

BJI corpus. For this investigation, we focused on data from a set of 40 speakers³ in the corpus, including 20 who are characterized as Betawi speakers and 20 who are characterized as Jakarta Indonesian speakers. This sample (the same set analyzed by Kurniawan, 2018) was chosen because it is roughly balanced for women and men, with an age range of 17–83 years old at the time of recording and a range of educational backgrounds.

Since we were interested in interspeaker differences and socio-demographic factors for this stage of analysis, we focused on the 1SG variants produced by each of our 40 speakers in the corpus, who were grouped according to their age, gender, and whether they were characterized as a Betawi Malay or a Jakarta Indonesian speaker. To get this data, we extracted all utterances from the corpus with one or more 1SG glossed morpheme and then restricted those utterances to include only those produced by the 40 speakers. This left us with a total of 3795 occurrences of 1SG pronouns. We then determined the total number of 1SG forms that each speaker produced based on a search for 1SG glossed morphemes in the gloss-level data, as well as the number of *gua*, *gue*, *saya*, *aku*, and *kita* based on searches in the morpheme-level data. Although we had initially intended to include clitic *-ku*, we left it out of our analysis as there were few tokens and most of them were possessive. At this stage each variant was a category that included the different pronunciations and transcriptions, that is, the *gua/gue* category included anything transcribed as *gua*, *guá*, *guà*, *gue*, *gué*, and *guè*, and the *kita* category included anything transcribed as *kita* or *kitaq*. The variants and categories are shown in Table 2 below (note that for this initial analysis we considered *gua* and *gue* as a single variant).

Table 2. Variants of 1SG in the BJI corpus

Variant	<i>gua/gue</i>	<i>saya</i>	<i>aku</i>	<i>kita</i>
Transcriptions	<i>gua, guá, guà, gue, gué, guè</i>	<i>saya</i>	<i>aku</i>	<i>kita, kitaq</i>

We then investigated to what extent the 1SG variation observed in our data could be predicted by properties of the speaker such as age, gender, and language variety (Betawi vs. Jakarta Indonesian).

3. This was the set of speakers selected by Kurniawan (2018), who selected a balanced sample of speakers according to social factors who also had long enough recordings and enough tokens for analysis.

4.3 Intraspeaker analysis in both corpora

The second phase of analysis focused on intraspeaker variation exhibited across conversations. In the BJI corpus, we tracked a single male speaker of Jakarta Indonesian, EXPOKK, throughout the 14 conversations he participated in.⁴ This speaker was a research assistant on the project, a conversation participant, and also the person who did the recording. The corpus includes more of his speech than any other single individual. It should be noted that while EXPOKK and many of the other speakers he interacted with were included in the original set of 40 speakers examined in the interspeaker analysis, he also interacted with some additional speakers, so the data examined in the inter- and intraspeaker analyses of the BJI corpus overlap but are not identical. In the Wallace corpus, we also tracked a single male speaker, Edy Setya Budi, a research assistant on Wallace's project and a 25-year old Javanese man who had lived most of his life in Jakarta. Budi recorded 16 conversations with his friends, acquaintances, family members, and neighbors. For fifteen of Budi's conversations, we counted the instances of three 1SG pronouns, *saya* (and *saye*), *gua* (and *gue*) and *aku* for Budi as well as his interlocutors. In this phase of analysis, we were interested in the question of whether any intraspeaker variation is conditioned by properties of individual conversations, either the age and gender of the interlocutors or the variants used by the other interlocutors. We counted the total number of 1SG glossed morphemes and tokens of individual variants produced by each speaker in each conversation separately.

Finally, in the third phase of analysis we focus on specific examples in each of the corpora, using these to illustrate the ways in which speakers use these variants in interaction.

5. Results

5.1 Interspeaker variation in BJI corpus

While null pronouns and pronoun substitutes are possible variants (as described above), in our initial quantitative analysis we considered only four overt full pronoun variants: *saya*, *aku*, *gua/gue*, and *kita*. Table 3 summarizes the representation of each variant in our data, including the number of tokens of each variant and the percentage of total overt 1SG tokens that each variant contributes. As can

4. The BJI corpus also includes child directed speech, but we did not include any of the conversations where our speaker interacted primarily with children.

be seen Table 3, the vast majority of overt 1SG pronouns are *gua/gue* or *saya*, with the remaining two variants (*aku* and *kita*) combined making up less than 3% of the total number of 1SG tokens. For this reason, the rest of our quantitative analysis focuses specifically on *gua/gue* and *saya*.

Table 3. Overt 1SG pronoun usage in the BJI corpus broken down by variant

Variant	Number of tokens	% of total overt 1SG tokens
<i>gua/gue</i>	1525	40.2%
<i>saya</i>	2160	56.9%
<i>aku</i>	23	0.6%
<i>kita</i>	87	2.3%

At this stage of analysis, we explored the role that age, gender, and language variety (Betawi vs. Jakarta Indonesian) play in constraining the variation seen in 1SG pronoun usage. Given the previous characterization of Betawi Malay as highly associated with Jakarta and the most vernacular variety in a continuum from Standard Indonesian to Jakarta Indonesian to Betawi Malay (Ikrangara, 1975; Kurniawan, 2015; Sneddon, 2006), along with previous descriptions of the association of *gua/gue* with Jakartan/Jakartan youth identity (Djenar et al., 2018; Ewing, 2019; Manns, 2014) we expected to find an effect both of language variety and of age. We conducted a mixed-effects logistic regression analysis in R (R Core Team, 2020) using the `glmer()` function in the `lme4` package (Bates, Maechler, Bolker, & Walter, 2015). We tested pronoun variant as the response variable with gender, birth year, and language variety as fixed effects. We also included a random effect of speaker. Since the focus here is on the choice between *saya* and *gua/gue*, we filtered the data to include tokens of just those variants. Birth year was centered and scaled by subtracting the mean and dividing by the standard deviation. The results are summarized in Table 4. The reference levels for gender and language are female and Betawi respectively.

Table 4. Results of mixed-effects logistic regression model. (JI = Jakarta Indonesian; M = male)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.2167	0.6982	-3.175	0.0015
Birth year	1.8864	0.5513	3.422	0.0006
Language = JI	0.9036	0.9680	0.933	0.3506
Gender = M	1.4722	0.8577	1.717	0.0861

Of the three fixed effects, only birth year was found to be significant ($p < 0.001$). (For the other two fixed effects, $p > 0.05$). This result indicates that the only social factor significantly contributing to the variation seen in 1sg usage in this corpus is age. More specifically, *saya* is more likely to be used by older speakers (i.e., those with earlier birth years), whereas *gua/gue* is more likely to be used by younger speakers (i.e., those with later birth years). These patterns can be seen in the histograms in Figures 2 and 3 below, which show the distribution of *saya* tokens and *gua/gue* tokens across birth years respectively.

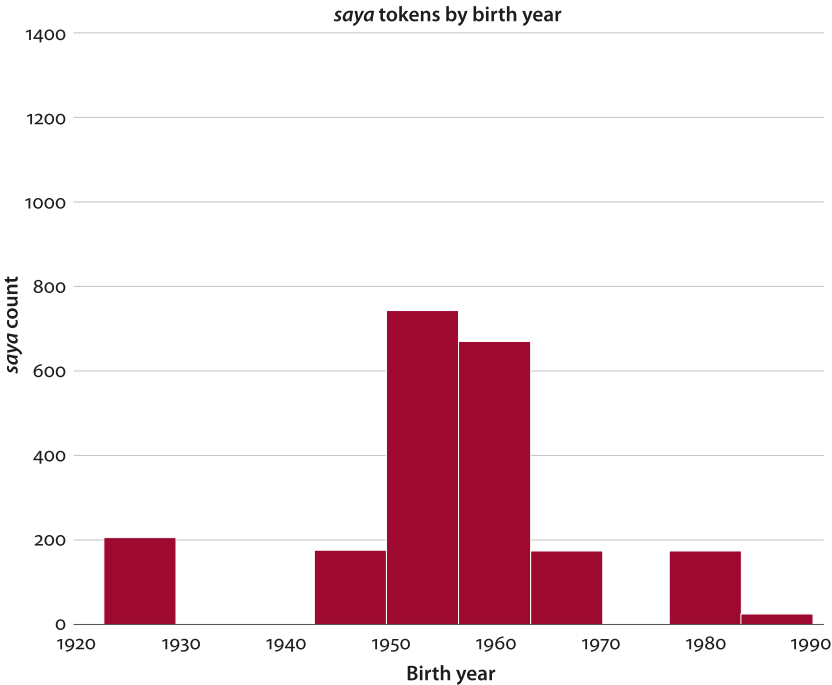


Figure 2. *Saya* usage within BJI corpus by birth year of speaker

It should be noted that when social factors are examined individually, language variety also appears to be important, with Jakarta Indonesian speakers more likely to produce *gua/gue* and Betawi speakers more likely to produce *saya*. That is, out of all overt 1sg pronouns produced by Jakarta Indonesian speakers (1,911 tokens), 69% of them were *gua/gue* and only 28% were *saya*. Out of all overt 1sg pronouns produced by Betawi speakers on the other hand (1,884 tokens), only 11% were *gua/gue* and 86% were *saya*. However, the two sets of speakers exhibit different age distributions. For Betawi speakers, the birth year ranges from 1923–1982 with a mean of 1959, whereas for Jakarta Indonesian

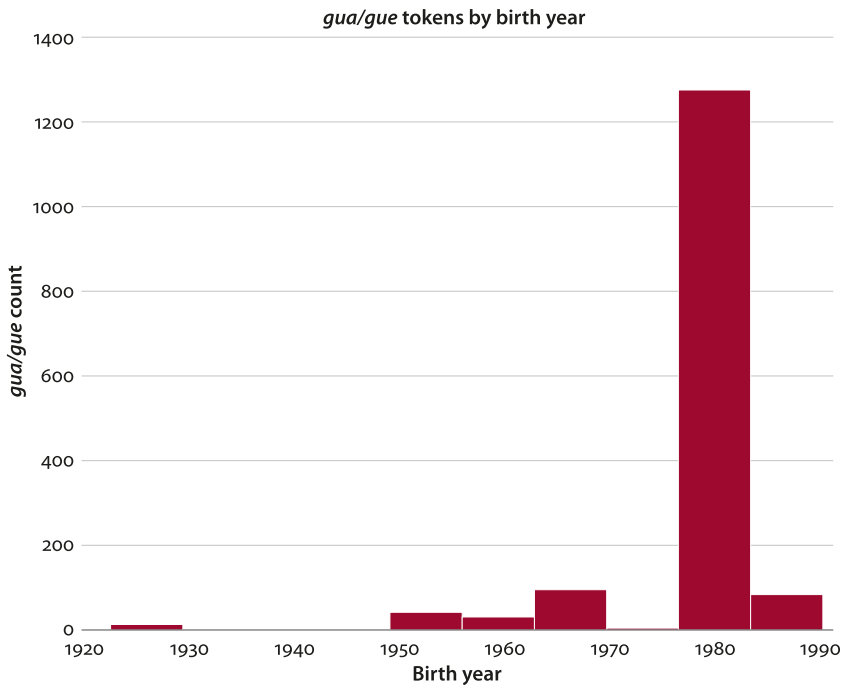


Figure 3. *Gua/gue* usage within BJI corpus by birth year of speaker

speakers, the range is 1959–1989 and the mean is 1975. Not all speakers contributed the same amount of data, but mean birth years weighted by the number of 1SG tokens produced by each speaker reflect the same pattern: 1952.1 for the Betawi speakers and 1974.1 for the Jakarta Indonesian speakers. Thus, the greater rates of *gua/gue* among Jakarta Indonesian speakers can be attributed to the greater proportion of data from younger speakers in that population. Similarly, the greater rates of *saya* among Betawi speakers can be attributed to the greater proportion of that data coming from older speakers.

The scatterplot in Figure 4 presents the rates of *gua/gue* usage for all 40 speakers separately, with % *gua/gue* calculated out of the total number of 1SG tokens produced by each speaker. As can be seen, the rates of *gua/gue* increase as birth year increases (i.e., as the speakers get younger), consistent with the logistic regression findings. It can also be seen that similar distributions hold for speakers of both language varieties, also consistent with the finding that language variety is not a significant factor in the logistic regression analysis.

Figure 4 also illustrates an important point concerning the range in % *gua/gue* found across the 40 speakers more broadly. Not only does the wide range of % *gua/gue* across the set of speakers (spanning from 0 to 100%) reflect substantial



Figure 4. Use of *gua/gue* by speaker within the BJI corpus and year of birth.

interspeaker variation, but the large number of participants with midrange percentages of *gua/gue* also reflects a substantial amount of intraspeaker variation.

That is, while some speakers seem to be nearly categorical users of *gua/gue* (e.g., 80–100% *gua/gue*) or *saya* (e.g., 0–10% *gua/gue*), there are many with percentages between 20 and 80%. This variation suggests that there may be important effects conditioning 1SG pronoun usage beyond demographic factors. In particular, we propose that factors at the conversation level such as the context and interlocutor may play a key role in conditioning the variation for a particular speaker. This would predict further that within any given context, a speaker may exhibit categorical use of one of the pronoun variants even if they exhibit substantial variation (i.e., a midrange percentage) at the macro-level. We test this prediction in the following section, where we turn to a closer analysis of intraspeaker variation within the two corpora.

5.2 Intraspeaker variation in the BJI corpus

The first part of the intraspeaker analysis tracks EXPOKK throughout the 14 adult conversations in the BJI corpus in which he participated. While EXPOKK’s language use is likely influenced by his role as a research assistant in the project, his role in this regard was consistent across conversations. The conversations include anywhere from one to nine other speakers, and we counted the total number of 1sg glossed morphemes produced by each participant in the conversations. Table 5 summarizes these results. Each row corresponds to a different conversation, ordered from highest to lowest number of total 1sg tokens produced by

EXPOKK. Most 1sg tokens in these conversations are either *gua/gue* or *saya*, so we focus on the results for these forms as well as the results for total 1sg pronouns.

Table 5. Rates of 1sg variants in EXPOKK’s conversations, ordered from highest to lowest number of total 1sg tokens produced by EXPOKK. Gu* refers to the total number of *gua+gue*. *This column only includes info for a subset of participants for whom we have birth year.

Conversation	EXPOKK total 1sg	EXPOKK % <i>gu</i> *	EXPOKK % <i>saya</i>	Others total 1sg	Others % <i>gu</i> *	Number of males/ females	Others birth year ⁺
BTJ-080807	179	99.4%	0.6%	107	99%	2m	1978–1983
BTW-240707	106	98.1%	1.9%	129	79.1%	2m/2f	1976
BTJ-180807	48	89.6%	10.4%	9	100%	2m	1977–1983
BTJ-220807b	45	91.1%	6.7%	90	22.2%	1m/1f	1976–1977
BTJ-210107	30	96.7%	0%	52	92.3%	2m/1f	1976–1977
BTJ-070109	25	92%	8%	64	23.4%	2m/2f	1976–1980
BTW-270906	11	100%	0%	16	75%	1m/2f	1963–1977
BTJ-220807a	8	100%	0%	1	100%	1m	1977–1978
BTW-191005	8	0%	25%	32	37.5%	2f	1977
BTW-210506	6	0%	100%	15	46.7%	1m/1f	1963–1965
BTW-040606	2	50%	50%	136	8.8%	3m/5f	1955–1977
BTW-051106	2	0%	100%	17	5.8%	3m/3f	1950–1977
BTW-030106A	1	0%	100%	123	2.4%	3f	1923–1977
BTW-030106B	0	n/a	n/a	106	9.4%	3f	1923–1977

Our first observation with respect to the overall use of 1sg pronouns is that unsurprisingly, the number of 1 sg pronouns used by EXPOKK varies substantially across the conversations, ranging from 0–179. Second, in the cases in which EXPOKK uses very few 1sg pronouns, the other participants often use a much higher number of 1sg pronouns than he does (with the one exception of conversation BTJ-220807a). In fact, in three of the seven conversations in which EXPOKK uses fewer than 10 1sg pronouns, the other participants use more than 100 (BTW-040606, BTW-030106A, and BTW-030106B). We interpret these differences as demonstrating the different personae that EXPOKK inhabited in these conversations, sometimes highlighting his role as a research assistant and other times playing down this role and acting more as a participant. That is, in the

conversations in which other participants produced a substantial number of 1SG pronouns but EXPOKK produced very few (or zero), we find that his research assistant role in the conversation is prominent; he spends most of these conversations asking other participants about themselves. In contrast, in the conversations in which EXPOKK also produces many 1SG pronouns, his participant role in the conversation is more prominent, and he spends more time talking about himself.

In terms of his use of specific 1SG variants, EXPOKK's % *gua/gue* also varies substantially, ranging from 0–100. Moreover, almost all of his 1SG pronouns are either *gua/gue* or *saya*, so in conversations in which he uses fewer *gua/gue*, he typically uses more *saya*. Finally, there are not many cases in which he exhibits midrange percentages of *gua/gue*, indicating that typically, he uses almost 100% *gua/gue* or 100% *saya*, with the exception of conversation BTW-040602, in which he produces just two 1SG pronouns, one *gua* and one *saya*. In other words, while EXPOKK exhibits variation in his 1SG pronoun usage broadly, there is relative consistency within a conversation. The most striking pattern that can be seen in EXPOKK's %*gua/gue* results is the relationship between those rates and the rates of his overall 1SG pronoun usage within a conversation. Specifically, in conversations in which he uses a high number of 1SG pronouns, they are always 90–100% *gua/gue*. It is only in the conversations in which he uses fewer than 10 1SG pronouns that we see him use higher rates of *saya*. Following our interpretation of the variation in total number of 1SG pronouns, this suggests that EXPOKK uses mostly *gua/gue* when he is speaking as a main participant in a conversation but uses mostly *saya* when he is acting more as a facilitator or interviewer. The one conversation in which EXPOKK produced fewer than 10 1SG pronouns but 100% of them were *gua/gue* is BTJ-220807a. This is also the one conversation in which EXPOKK produced few 1SG pronouns but his interlocutor produced even fewer.

One question that arose from the interspeaker results was whether a speaker's pronoun choice is conditioned by the preference of their interlocutors within a given conversation. This does not appear to be the case for EXPOKK, since the conversations in which EXPOKK produced more than 90% *gua/gue* included ones in which his interlocutors also produced 90–100% *gua/gue* as well as ones in which they produced fewer than 25% *gua/gue*. A second question is whether his pronoun choice is influenced by social characteristics of his interlocutors, such as age and gender. Unfortunately, it is difficult to evaluate these factors in the conversation-level data. Most conversations include a mix of male and female participants, so variation across conversations cannot necessarily be attributed to gender differences across the conversations. Age presents a somewhat similar challenge, since there is a mix of interlocutor ages within most conversations, even though there is a range in birth years among the participants (1923–1983). We can look at the effect of some older interlocutors, however, since they are not evenly

distributed across conversations. While nearly every conversation includes at least one participant born after 1975, those born before 1960 only participate in the last four conversations in the table and those born before 1950 only participate in two of the 14 conversations. This approach is also difficult to draw conclusions from, however, since the conversations with older participants are a subset of those in which EXPOKK produces fewer than 10 1sg pronouns. Thus, while it may appear that he produces fewer *gua/gue* tokens in conversations that include older people, it is not possible to separate this from the larger pattern in which he uses mainly *saya* in cases where he uses few 1sg pronouns overall. Despite these limitations, this level of analysis offers some possible insights and directions for further study. Moreover, these limitations invite the third stage of analysis presented below as a way to complement what can be understood from this conversation-by-conversation level analysis. In the next section we present the conversation level analysis of the Wallace corpus, before finally turning to an in-depth investigation of conversation-internal patterns.

5.3 Intraspeaker variation in the Wallace corpus

The second part of the intraspeaker analysis tracks a single male Javanese speaker, Edy Setya Budi, in the Wallace Corpus through 15 of the 16 conversations he participated in and recorded. Like EXPOKK in the BJI corpus, Budi was a research assistant who participated in multiple conversations in the corpus. These conversations contain between one and seven participants in addition to Budi. As with the BJI corpus, we counted the total number of 1sg glossed morphemes and total numbers of individual variants produced by each participant in each conversation. Table 6 summarizes the results, with conversations ordered from most 1sg tokens produced by Budi to fewest. As with the results for the BJI corpus, we focus on total numbers of 1sg pronouns and rates of *gua*, *gue* and *saya*.

Overall, the total number of 1sg pronouns is somewhat lower in the Edy Setya Budi conversations than in those with EXPOKK. Nevertheless, there is still a range in the total number of Budi's 1sg pronouns per conversation, with zero being the lowest and 32 the highest. Unlike what we saw in EXPOKK's conversations, there is not a clear relationship between the rates of his 1sg usage and the rates for other participants. Among the conversations in which he produces very few or zero 1sg pronouns, some have high rates of 1sg pronoun usage by other speakers and others have very low rates by other speakers (two or three total 1sg pronouns per conversation). Similarly, among the two conversations in which Budi uses relatively high numbers of 1sg pronouns (13–32 total 1sg pronouns), there is also a large range in how many the other participants use (0–144).

Table 6. Rates of 1sg variants in Edy Setya Budi’s conversations, ordered from highest to lowest number of total 1sg tokens produced by Budi. Gu* refers to the total number of gua+gue. +Dates written with a slash demarcate ranges of birth years given for a single participant.

Conversation	Budi			Others’			Number of males/ females	Others’ birth year ⁺
	total 1sg	Budi % <i>gu</i> *	Budi % <i>saya</i>	total 1sg	Others’ % <i>gu</i> *			
061_Telor	32	0%	100%	144	0%	1m		1915/1920
069_Undangan	22	100%	0%	0	n/a	1m/1f		1957–1975
060_Lawak	17	0%	76.5%	31	3.2%	1m/1f		1950–1952
073_Pak_Marjo	13	0	100%	14	0%	1m		1915/1920
076_Anak_Nakal	3	100%	0%	0	n/a	3m		1948–1951
078_Mbah_Randiman	2	0%	100%	1	0%	1m/2f		1905–1967
024_Anak_Anak_Muda	1	100%	0%	8	100%	5m		1950–1957
094_Kanker	1	100%	0%	7	100%	4m		1950–1960
109_ITA	1	100%	0%	10	100%	2f		1953–1958
070_Die_Die_Juga	0	n/a	n/a	3	0%	1f		1957
068_Ngaku_Bujangan	0	n/a	n/a	19	42.1%	1f		1957
077_Dami	0	n/a	n/a	31	0%	3m/4f		1915–1959
093_Yanto	0	n/a	n/a	2	0%	2m		1910–1930
110_Emi	0	n/a	n/a	2	100%	1f		1951
112_Engkos	0	n/a	n/a	22	100%	2m		1947–1950

Like EXPOKK, Budi uses multiple 1sg pronoun variants, but most of the variation we see is across different conversations. Within a single conversation, he typically uses all *gua/gue* or all or nearly all *saya*. The exception is 060_Lawak, in which Budi uses zero *gua/gue* forms but only 76.5% of the pronouns are *saya* (The rest are *aku*). As in the EXPOKK data, this reflects relative consistency within conversations, suggesting that factors conditioning pronoun usage are set at the conversation level.

A crucial difference between these conversations and those with EXPOKK, however, is that here, the choice of variant seems to be set at the conversation level for all speakers together. That is, in all conversations in which both Budi and the other interlocutors use at least some 1sg pronouns, if Budi uses only *gua/gue*, so do his interlocutors. If he produces no *gua/gue* forms, his interlocutors also produce zero or near zero *gua/gue*. Thus, while the % *gua/gue* rates do vary

for Budi and other participants across conversations, there is a good deal of consistency within each conversation. This raises the question of what the relevant conversation-level factors are that condition pronoun preference for all speakers within a conversation in the Budi data and for just EXPOKK in the BJI corpus.

5.4 Interactional uses of 1SG pronouns in the corpora

Our focus on static socio-demographic factors as well as rates of usage by interlocutors at a conversation-level offer some suggestive differences and beg further analysis. However, they cannot fully account for the widely observed intraspeaker variation seen in the corpus data. We therefore turn in this section to a qualitative investigation in order to achieve a more nuanced understanding of intraspeaker variation. The previous stages of analysis demonstrate that *gua/gue* is the default form at least in the types of situations and registers represented in this corpus, with *saya* serving some particular discourse and socio-indexical functions. We might have expected more use of *aku*, but we find it only rarely in the Wallace corpus and even less frequently in the BJI corpus. Of particular interest in this regard is what conditions the use of the infrequent variants in the corpora. Put another way, what does it mean when JI speakers use anything other than *gua* or *saya*? Given how rare these forms are, comparing overall rates across different conversations does not offer much insight. Rather, it is necessary to examine more closely the contexts within individual conversations in which these rare forms occur. Here we provide some representative examples of use of *saya*, *gua* and *aku*, before turning to the use of *kite*.

The use of *saya* in the corpora broadly accords with previous descriptions of this form (Djenar, 2008; Englebretson, 2007) as a pronoun commonly used by non-intimates. An example of this is shown in (1). EXPOKK is visiting AFRBTJ and is meeting AFTBTJ's wife, TRIBTJ, for the first time. He asks her what her name is, saying he doesn't know her even though he knows her husband.

(1) Symmetrical use of *saya* between non-intimates

(MPI Jakarta Indonesian, BTJ220807b)

EXPOKK: *Ni Mbaqnya namaqnya sapa si?*

'What's your name?'

TRIBTJ: ((laughing))

Heh?

'Huh?'

EXPOKK: *Ngga tauq namaqnya saya.*

'I don't know your name.'

Similarly, in the same conversation, TRIBTJ uses *saya* in talking to EXPOKK, as shown in (2). TRIBTJ is telling EXPOKK that she told her husband, who works as a freelance driver, to charge more if a customer asks him to drive them beyond the city.

- (2) Symmetrical use of *saya* between non-intimates
(MPI Jakarta Indonesian, BTJ220807b)
TRIBTJ: *Iya, (t)erus kan saya bilang, namaqnya luar kotaq.*
'yeah, then I told him, it's out of town.'

By contrast, EXPOKK and ARFBTJ, who already know each other, generally use *gua* when speaking to each other. However, AFRBTJ switches to *saya* when the topic of the talk shifts from one about family members to a work situation, as shown in (3). In the previous talk (not shown) ARFBTJ was talking to EXPOKK about his father-in-law visiting him and his wife in Jakarta before changing the topic to his work situation. He mentions that of late, he hasn't been able to get work easily because there are too many freelance drivers like him in the area so he stops looking for passengers there. He switches to *saya* as he mentions this.

- (3) Switch from *gua* to *saya* (MPI Jakarta Indonesian, BTJ220807b)
AFRBTJ: *Ompréngan uda parah soalnya di sini, jadiq saya uda ngga perna ke sono lagi.*
'There are many more illegal public vehicles now, so I never go there again.'
(referring to his decreasing income as driver of public transport)

The use of *aku* for indexing a personal identity is illustrated in (4). EXPOKK mentions to AFRBTJ and his wife, TRIBTJ, that he is soon to return to the US where he has been studying. TRIBTJ asks if he is taking his wife. EXPOKK responds by saying that he is engaged but not married yet and that, if he were married, he would certainly take his wife along. He switches to *aku* as he refers to himself in this hypothetical situation (recall that he previously used *saya* when telling TRIBTJ that he didn't know her name).

- (4) Use of *aku* in talk about personal situation
(MPI Jakarta Indonesian, BTJ220807b)
EXPOKK: *Kaloq aku dah mérit ma dibawaqa aja.*
'If I were married I would bring her (with me).'

The previous examples show that, while *saya* and *aku* both denote 1sg, their use in interaction broadly accord with previous descriptions (Djenar, 2008;

Englebretson, 2007) that distinguish them in terms of familiarity between the participants and flexible self-categorization.

By contrast, speakers use *kita* and *kite* as 1SG or 1PL reference, lending support to the point made by Adelaar and Hajek (forthcoming) that number distinction is lost in Indonesian. When used as 1SG, the pronouns refer to the speaker involved in a habitual activity or public situations such as work. *Kita* and *kite* in this case are similar to generic 'we'. As 1PL, *kita* and *kite* are used to denote the speaker and addressee (inclusive) or the speaker and another party, excluding the addressee (exclusive). The use of *kite* as 1PL is illustrated in Excerpt (5), from the Wallace corpus. Wina (22 yo) is telling her husband Bambang (25 yo) that Fahri (a married man with three children who lives nearby) was having an affair with Mbak Cut. Wina found out about this from her friends, Trini and Tatik. Her husband Bambang is wondering why he and his wife (*kite* '1PL.INCL') didn't know about the affair (line 1). Wini tells him that she actually knew about it from her friends, Trini and Tatik, referring to herself and her friends with *kite* 1PL.EXCL ('I and my friends but not you'). Wini makes this referential difference by, first, asking Bambang who he was referring to when he uses *kite* (line 2), and secondly, by referring to her husband using the pronoun *kamu* '2sg', excluding him from the exclusive *kite* that is herself and her friends. Wina then points out to her husband that she'd known about the affair even before she met him. Hearing this prompts Bambang to try and establish when it was exactly that his relationship with Wina began (line 3). He uses *kite* once again to refer to himself and Wina (1PL.INCL). Wina follows suit in line 4 in using *kite* inclusively.

(5) *Kite* denoting 1PL inclusive (Wallace corpus 'Ngaku Bujangan')

1 BAMBANG: *Koq kite nggaq taoq, yè?*

'How come we didn't know (about the affair)?'

2 WINA: *Siapa? Kan kite belum kenal sama kamu. Apaqan?Belon kenal deh, taoq yang namanyè Fahri?*

'Who? We didn't know you then. What do you mean? We didn't know each other, do you know Fadilah?'

3 BAMBANG: *Mas Budi pergi kan, ke Ostrali kan kite maèn kesana.*

'Mas Budi was in Australia, remember we visited him there.'

4 WINA: *Iyè, tapi kan, kamu Mei kan? Mei, Juni, Juli, Agustus kurang sebentar Mas Budi maoq pulang. Nah, kitè hubunganyè, nggaq lama Mas Budi pulang kan? Kan?*

'Yeah, but you turned up in May right? May, June, July, August, it wasn't long until Mas Budi went home. So, we got together and not long after that Mas Budi came home right? Right?'

The use of *kita* denoting inclusive ‘we’ is also shown in (7). Mardi (55–60 yo) is talking to Batara (25 yo) about the prospect of Batara and him doing business together, selling eggs. Mardi uses *kita* inclusively to refer to himself and Batara in this imagined future situation.

(6) *Kita* denoting 1PL (Wallace Corpus, ‘Telor’)

MARDI: *Kita langsung, telpun, kirim. Ha, nanti kita jika dapat pesenan langsung. Wa, ini yang jelas itu, ini telur kita molai. Tentang sambilan lainnya yang menguntungkan itu baik jugaq.*

‘We do this quickly, (when someone) rings, (we) send the order.

That’s what we do when we get a direct order. Basically, we’ll start with this egg (business). As for other profitable side businesses, that’s good too.’

BATARA: *Iya.*

‘Yes.’

The use of *kita* denoting 1SG is shown in (7). ARMBTJ runs a small business fixing refrigerators. EXPOKK, who is visiting ARMBTJ, is asking him how he glues the different parts of a fridge he has fixed. ARMBTJ explains that someone else does the gluing, that he only makes the sample and puts his orders in, and the third party does the rest. ARMBTJ refers to the himself as *kita* ‘1PL.EXCL’ and the third party as *dia* ‘3SG’.

(7) *Kita* denoting 1SG (MPI Jakarta Indonesian, BTJ-210107)

EXPOKK: *ni elu ngelémnya diapain, Bot?*

‘how did you glue it, Bot?’

(referring to the rubber webbing)

ARMBTJ: *dia yang ngelém*

‘they glued it.’

(referring to the shop where the rubber list is sold)

o, *dia yang ngelém.*

‘oh, they glued it.’

kita bikin

‘we/I make (the sample).’

kita mesen ma dia, kan diukur ama dia, panjang.

‘(then) we/I ordered (it) from them, they measured it, the length.’

This example is similar to (6) in that in both cases, *kita* is used to refer to the self or the self and the addressee within work practices, the difference being in the number of referent(s) denoted.

Kita is also used generically, as in (8). ARMBTJ is telling EXPOKK that he is thinking of renting a BTN house (public housing) because it is cheaper than renting his current house. The generic use of *kita* (written with a glottal stop, *kitaq*, in the example) is ambiguous between 1SG (ARMBTJ referring to himself) and 1PL.EXCL (ARMBTJ referring to himself and his wife).

- (8) Ambiguous reference (MPI Jakarta Indonesian, BTJ220807b)

AFRBTJ: *Daripada ngontrak kéq begini, Ki.*

‘Instead of renting a house like this, Ki.’

EXPOKK: *Iyaq, he-eh.*

‘Yeah, right.’

AFRBTJ: *Mending kitaq ngontrak BTN, iya, kan?*

Better if we rent a BTN, right?

(BTN is a type of public housing)

The foregoing examples illustrate the dynamic use of the pronoun variants and show that they are used as grammatical markers that also do the work of indexing subtle socio-indexical distinctions.

6. Conclusion

In this paper we have investigated 1SG pronoun variation in Jakarta Indonesian in three different stages of analysis. In the first stage we investigated pronominal variation in the Jakarta speech community today, according to broad demographic factors. In the second stage we expanded this analysis to focus on intraspeaker variation for two speakers in the same community at two different time points. Finally we take our investigation of intraspeaker variation and focus more narrowly on the interactional use of the 1SG variants by the same speakers.

In our interspeaker analysis of the BJI corpus we find that the only significant predictor of *gua/gue* is the age of the speaker. *Gua* is more common for younger speakers in the BJI corpus than older, and although an initial apparent time interpretation of the data might suggest change in progress, comparison with the Wallace corpus, where the use of *gua* is also common and frequent among Wallace corpus youth (i.e., those whose birth years are in the same range as the adults in the BJI corpus) leads us to conclude that the age differences represent age-grading in both corpora, suggesting some maintenance of *gua/gue* over time as a marker of Jakarta youth speech. In future work we plan to utilize an interspeaker analysis of the Wallace corpus for a more direct investigation of this preliminary finding.

Our closer examination of intraspeaker variation lends credence to our observations on relative consistency in the pronominal system in Jakarta Indonesian over time, while also providing insights on inter-individual differences in the pragmatic use of the different variants. We gain further insights from examining the use of the pronouns in interaction.

A comprehensive variationist analysis of 1SG pronouns in Indonesian must eventually include linguistic factors that constrain use of the variants, and in our initial analysis we have uncovered some of the factors that will need to be included in further analysis. There are clearly additional linguistic and discourse-level factors to consider, including some that have been commented on in previous work and some that have been illuminated through our examination of these corpora. Englebretson (2007, p.105), for example, makes reference to the “social nature of [the grammatical feature of] Indonesian agent-trigger and patient-trigger clauses,” arguing that “grammatical voice serves as a social resource for stancetaking that speakers can use to position themselves and others in terms of moral agency and responsibility.” Moreover, in our examination of the two corpora we’ve made the as-yet-unquantified observation that *gua* seems to occur more frequently in inverse constructions and subordinate clauses. Agent-trigger vs patient-trigger clauses as well as clausal type are factors we will want to consider more closely in further analysis.

In addition, we aim to expand the quantitative analysis to include features of the interaction, such as the relationship between the participants in the interaction and whether the interaction is dyadic or multi-party. In this we find some limitations in working with corpora in which none of the co-authors of this paper were involved in the collection process – it can be difficult to discern the relationship between the participants in order to code for that relationship. In this our third stage of analysis, with a closer reading of the conversations, has been helpful in providing context and more information about the speakers’ relationships which we can use in further analysis.

In sum, we have provided a preliminary sociolinguistic analysis of the form, function, and interactional uses of 1SG pronouns in Indonesian, drawing on tools from both quantitative and qualitative sociolinguistics. Our close examination of the frequency of the pronominal variants in the corpus, including both inter- and intra- speaker variation, alongside a more nuanced analysis of their use in interaction, provides a model of the use of various sociolinguistic methods for the examination of grammatical features. Finally, this work lays the foundation for looking more fully at second person and third person pronouns in Indonesian, which are equally complex from a grammatical point of view and equally rich in opportunities for sociolinguistic analysis.

Acknowledgements

The authors would like to thank Ferdinan Kurniawan for his assistance with early analysis and with extracting data from the BJI corpus, Edgar Yau for his detailed work coding pronouns from the Wallace corpus, our fellow authors in this volume for helpful feedback and suggestions, and two anonymous referees of this manuscript.

References

- Abtahian, Maya R., Cohn, Abigail C., White, Aaron, & Yanti (2019). *Language ideologies and language shift scenarios in Indonesia*. New Ways of Analyzing Variation 48, Panel on *What's so standard about standards?* University of Oregon, Eugene, October 10–12, 2019.
- Adelaar, Alexander, & Hajek, John (Forthcoming). "Pronouns." In Alexander Adelaar & Antoinette Schapper (Eds.), *The Oxford guide to Malayo-Polynesian languages of South East Asia*. Oxford, UK: Oxford University Press.
- Ananta, Aris, Arifin, Evi Nurvidya, Hasbullah, M. Sairi, Handayani, Nur Budi, & Pramono, Agus (2015). *Demography of Indonesia's ethnicity*. Singapore: ISEAS. <https://doi.org/10.1355/9789814519885>
- Bates, David, Maechler, Martin, Bolker, Ben, & Walker, Steve (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Brown, Roger, & Gilman, Albert (1960). The pronouns of power and solidarity. In Thomas Sebeok (Ed.), *Style in language* (pp. 253–276). Cambridge, MA: MIT Press.
- Christofaki, Rodanthi (2018). Expressing the self in Japanese: Indexical expressions in the service of indexical thought. In Minyao Huang & Kasia M. Jaszczolt (Eds.), *Expressing the self: Cultural diversity and cognitive universals* (pp. 72–87). Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198786658.003.0005>
- Cohn, Abigail C., & Vogel, Rachel (2019). Variation in two patterns of word-initial deletion in Jakarta Indonesian: Insight from naturalistic data. In Sasha Calhoun, Paola Escudero, Marija Tabain, & Paul Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 38–42). Canberra, Australia: Australian Speech Science and Technology Association.
- Connors, Thomas J., Brugman, Claudia M., & Adams, Nikki B. (2016). Reference tracking and non-canonical referring expressions in Indonesian. In Yanti & Timothy McKinnon (Eds.), *Studies in language typology and change, NUSA 60*, (pp. 59–88). Permanent URL: <http://hdl.handle.net/10108/87444>
- Djenar, Dwi Noverini (2006). Patterns and variation of address terms in colloquial Indonesian. *Australian Review of Applied Linguistics*, 29(2), 22.1–22.16. <https://doi.org/10.2104/aral0622>
- Djenar, Dwi Noverini (2008). Which self? Pronominal choice, modernity, and self-categorizations. *International Journal of the Sociology of Language*, 189(1), 31–54. <https://doi.org/10.1515/IJSL.2008.002>
- Djenar, Dwi Noverini (2015). Pronouns and sociospatial ordering in conversation and fiction. In Laure Gardelle & Sandrine Sorlin (Eds.), *The pragmatics of personal pronouns* (pp. 195–213). Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.171.10dj>

- Djenar, Dwi Noverini, Ewing, Michael C., & Manns, Howard (2018). *Style and intersubjectivity in youth interaction*. Berlin: De Gruyter.
<https://doi.org/10.1515/9781614516439>
- Djenar, Dwi Noverini, & Sidnell, Jack (Forthcoming). Introduction. In Dwi Noverini Djenar & Jack Sidnell (Eds.), *Signs of deference, signs of demeanour: Interlocutor reference and self-other relations across Southeast Asian speech communities*. Singapore: National University of Singapore (NUS) Press.
- Eckert, Penelope (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4): 453–476. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Englebretson, Robert (2007). Grammatical resources for social purposes: Some aspects of stancetaking in colloquial Indonesian conversation. In Robert Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction*. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.164.05eng>
- Errington, Joseph J. (2014). In search of middle Indonesian: Linguistic dynamics in a provincial town. In Gerry van Klinken & Ward Berenschot (Eds.), *In search of middle Indonesia: Middle classes in provincial towns* (pp. 199–220). Leiden: Brill.
- Ewing, Michael C. (2005). Colloquial Indonesian. In Alexander Adelaar & Nikolaus P. Himmelman (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 227–254). London: Routledge.
- Ewing, Michael C. (2014). Motivations for first and second person subject expression and ellipsis in Javanese conversation. *Journal of Pragmatics*, 63, 48–62.
<https://doi.org/10.1016/j.pragma.2013.09.023>
- Ewing, Michael C. (2019). Localising person reference among Indonesian youth. In Zane Goebel, Deborah Cole, & Howard Manns (Eds.), *Contact talk: The discursive organization of contact and boundaries* (pp. 140–159). London: Routledge.
<https://doi.org/10.4324/9780429427848-9>
- Ewing, Michael C., & Djenar, Dwi Noverini (2019). Address, reference and sequentiality in Indonesian conversation. In Paul Bouissac (Ed.), *The social dynamics of pronominal systems* (pp. 253–287). Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.304.11ewi>
- Fernández-Mallat, Víctor (2020). Forms of address interaction: Evidence from Chilean Spanish. *Journal of Pragmatics*, 161, 95–106. <https://doi.org/10.1016/j.pragma.2020.03.006>
- Frawley, William J. (2003). *International encyclopedia of linguistics (2nd ed.)*. Accessed on August 8, 2020, from <https://www.oxfordreference.com/view/10.1093/acref/9780195139778.001.0001/acref-9780195139778>
- Gil, David, & Tadmor, Uri Tadmor (2014). The MPI-EVA Betawi-Jakarta database. *A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and the Center for Language and Culture Studies*, Atma Jaya Catholic University. Available at <https://archive.mpi.nl/tla/islandora/search/Jakarta%20Indonesian?type=dismax>
- Goebel, Zane (2010). *Language, migration and identity: Neighbourhood talk in Indonesia*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511778247>
- Grijns, Cornells Dirk (1980). Some notes on Jakarta Malay kinship terms: The predictability of complexity. *Archipel*, 20, 187–212. <https://doi.org/10.3406/arch.1980.1599>
- Hepburn, Alexa, Wilkinson, Sue, & Shaw, Rebecca (2012). Repairing self- and recipient reference. *Research on Language and Social Interaction*, 45(2), 175–190.
<https://doi.org/10.1080/08351813.2012.673914>

- Ikranagara, Kay (1975). *Melayu Betawi grammar*. Doctoral dissertation, University of Hawaii, Manoa.
- Kartomiharjo, Suseno (1981). *Ethnography of communicative codes in East Java*. Doctoral dissertation, Research School of Pacific and Asian Studies, Australian National University, Canberra.
- Kluge, Bettina, & Moyna, María Irene (2019). *It's not all about 'you': New perspectives on address research*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tar.1>
- Kurniawan, Ferdinan (2015). Nasal assimilation in Jakarta Indonesian. In Amber Camp, Yuko Otsuka, Claire Stabile, & Nozomi Tanaka (Eds.), *AFLA 21: The Proceedings of the 21st Meeting of the Australian Formal Linguistics Association* (pp. 149–166). Canberra: Asia-Pacific Linguistics.
- Kurniawan, Ferdinan (2018). *Phonological variation in Jakarta Indonesian: An emerging variety of Indonesian*. Doctoral dissertation, Cornell University.
- Lee, Hye-Kyung (2018). Self-referring in Korean, with reference to Korean first person markers. In Minyao Huang & Kasia M. Jaszczolt (Eds.), *Expressing the self: Cultural diversity and cognitive universals* (pp. 58–71). Oxford: Oxford University Press.
- Lerner, Gene H., & Kitzinger, Celia (2007). Extraction and aggregation in the repair of individual and collective self-reference. *Discourse Studies*, 9(4), 526–557. <https://doi.org/10.1177/1461445607079165>
- Mahdi, Waruno (2001). Personal nominal words in Indonesian: An anomaly in morphological classification. In Joel Bradshaw & Kenneth L. Rehg (Eds.), *Issues in Austronesian morphology: A focusschrift for Byron Bender* (pp. 166–192). Pacific linguistics 519. Canberra: Pacific Linguistics.
- Manns, Howard (2012). First-person pronominal variation, stance and identity in Indonesia. *Australian Journal of Linguistics*, 32(4), 435–456. <https://doi.org/10.1080/07268602.2012.744265>
- Manns, Howard (2014). Youth radio and colloquial Indonesian in Urban Java. *Indonesia and the Malay World*, 42(122), 43–61. <https://doi.org/10.1080/13639811.2014.876156>
- Meyerhoff, Miriam (2009). Animacy in Bislama? Using quantitative methods to evaluate transfer of a substrate feature. In James N. Stanford & Dennis R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 369–396). Amsterdam: Benjamins. <https://doi.org/10.1075/impact.25.19mey>
- Meyerhoff, Miriam (2015a). All these years and still counting: Why quantitative methods still appeal. In Dwi Noverini Djenar, Ahmar Mahboob, & Ken Cruickshank (Eds.), *Language and identity across modes of communication* (pp. 61–82). Berlin: De Gruyter. <https://doi.org/10.1515/9781614513599.61>
- Meyerhoff, Miriam (2015b). Turning variation on its head: Analysing subject prefixes in Nkep (Vanuatu) for language documentation. *Asia-Pacific Language Variation*, 1(1), 78–108. <https://doi.org/10.1075/aplv.1.1.04mey>
- Meyerhoff, Miriam, & Stanford, James N. (2015). “Tings change, all tings change”: The changing face of sociolinguistics with a global perspective. In Dick Smakman & Patrick Heinrich (Eds.), *Globalising sociolinguistics: Challenging and expanding theory* (pp. 1–15). London: Routledge.
- Poedjosoedarmo, Soepomo (1982). *Javanese influence on Indonesian*. Materials in languages of Indonesia, No. 7. Series D. Canberra: Pacific Linguistics.

- Raymond, Chase Wesley (2016). Linguistic reference in the negotiation of identity and action: Revisiting the T/V distinction. *Language*, 92(3), 636–670.
<https://doi.org/10.1353/lan.2016.0053>
- Schnell, Stefan, & Barth, Danielle (2018). Discourse motivations for pronominal and zero objects across registers in Vera'a. *Language Variation and Change*, 30 (1), 51–81.
<https://doi.org/10.1017/S0954394518000054>
- Sneddon, James N. (2003). *The Indonesian language: Its history and role in modern society*. Sydney: UNSW Press.
- Sneddon, James N. (2006). *Colloquial Jakartan Indonesian*. Canberra, Australia: Pacific Linguistics.
- Wallace, Steven (1976). *Linguistic and social dimensions of phonological variation in Jakarta Malay*. Unpublished doctoral dissertation, Cornell University. Available at <https://ecommons.cornell.edu/handle/1813/45568>
- Wouk, Fay (1999). Dialect contact and koineization in Jakarta, Indonesia. *Language Sciences*, 21, 61–86. [https://doi.org/10.1016/S0388-0001\(98\)00013-8](https://doi.org/10.1016/S0388-0001(98)00013-8)

Abstract (Indonesian)

Bahasa Indonesia yang digunakan di Jakarta, yang sering disebut ‘bahasa Indonesia dialek Jakarta’, merupakan variasi bahasa Indonesia yang muncul dari persentuhan antara bahasa Betawi, dialek Melayu setempat, dan bahasa Indonesia baku. Seperti variasi bahasa Indonesia lainnya, variasi Jakarta memiliki sistem pronomina yang relatif terbuka dan memiliki beberapa bentuk pronomina orang pertama. Makalah ini berfokus pada variasi pemakaian pronomina orang pertama, dengan data dari dua korpora bahasa lisan yang dikumpulkan dalam jarak waktu tiga dekade. Dengan menggunakan metode kuantitatif maupun kualitatif, bentuk, fungsi dan makna sosial pronomina orang pertama dikaji, dan variasi antar-penutur serta intra-penutur selama jarak waktu tersebut dianalisa.

Kata kunci: Bahasa Indonesia dialek Jakarta, variasi bahasa, metode kuantitatif, metode kualitatif, pronomina, acuan orang pertama

Address for correspondence

Maya Ravindranath Abtahian
 Department of Linguistics
 University of Rochester
 PO Box 270096
 Rochester, NY 14627
 USA
maya.r.abtahian@rochester.edu

Co-author information

Abigail C. Cohn
Cornell University
Department of Linguistics
acc4@cornell.edu

Rachel C. Vogel
Cornell University
Department of Linguistics
rcv4@cornell.edu

Dwi Noverini Djenar
The University of Sydney
Indonesian Studies Department
novi.djenar@sydney.edu.au