JOURNAL OF MULTILINGUAL AND MULTICULTURAL DEVELOPMENT
2024, VOL. 45, NO. 7, 2503–2521
https://doi.org/10.1080/01434632.2022.2055761

Urbanization, ethnic diversity, and language shift in Indonesia

Thomas B. Pepinsky [a], Maya Ravindranath Abtahian [b] and Abigail C. Cohn[c]

[a]Department of Government, Cornell University, Ithaca, NY, USA; [b]Department of Linguistics, University of Rochester, Rochester, NY, USA; [c]Department of Linguistics, Cornell University, Ithaca, NY, USA

ABSTRACT

Cross-nationally, urbanization is associated with the decline of minority languages and a shift towards national and official languages. But the processes that link urbanization with language shift have not been adequately documented. In this paper we consider the relationship between cities and language shift from a sociolinguistic perspective, focusing our attention on the issue of language use and language shift in Indonesia – a large, ethnically and linguistically diverse, rapidly urbanising country. We use census data to examine how ethnic diversity shapes language shift in the context of urbanicity. We find that in ethnically homogenous regions, urbanicity itself has little relationship with language shift. By contrast, ethnic diversity is consistently associated with a greater probability of speaking Indonesian both among urban and rural Indonesians and in urban and rural areas. These findings contribute to our understanding of language shift and linguistic vitality in diverse, urbanising societies, and highlight the need to distinguish between the process of urbanization and the state of being urban.

ARTICLE HISTORY
Received 23 November 2021
Accepted 15 March 2022

KEYWORDS
Language shift; language maintenance; urbanization; urbanicity; ethnic diversity; Indonesia

Introduction

Many of the seminal synchronic studies of language change focused on cities (Labov [1966] 2006; Sankoff et al. 1976; Trudgill 1974). And yet, the role that the urban environment itself plays in language change is under-documented. What is it about cities, or about people living in cities, that motivates language change? At its surface the answer to this question seems straightforward. Urbanization (i.e. people moving to cities from nonurban areas) leads to linguistic heterogeneity in densely populated areas. Densely populated areas require communication between people who may be speakers of different languages or dialects. These factors – linguistic heterogeneity, densely populated areas, and the necessity of communication – combine to result in language change in the form of dialect levelling, koineization, and/or language shift toward a common dominant language. This progression summarises well-documented cases of the development of *urban vernaculars* (Makoni, Brutt-Griffler, and Mashiri 2007 on Zimbabwe and McLaughlin 2008 on Senegal), *urban contact varieties* (Sneddon 2006 on Jakarta Indonesian), and *multiethnolects* (Clyne 2000; Cheshire et al. 2011; Kircher and Fox 2019) in urban environments around the world. These linguistic outcomes are closely linked to the concept of urbanization and the connections between language and the 'city', and point to cities as loci for the development of new language varieties and, at least in some of the cases described above, corresponding language loss. In Jakarta, for example, the emergence and development of Jakarta Indonesian as a contact variety between

CONTACT Maya Ravindranath Abtahian maya.r.abtahian@rochester.edu Department of Linguistics, University of Rochester, 503 Lattimore Hall, PO Box 270096, Rochester, NY 14627

Betawi Malay and Standard Indonesian is correlated with the endangerment of Betawi (Lauder and Lauder 2017).

At the same time that urbanization generates cross-linguistic interactions that encourage language shift, it is also often the case that – perhaps paradoxically – economic and social segregation within cities produces dense urban areas comprised of multiple non-overlapping social networks. In such contexts, the process of moving to a densely populated urban area may cause people to become more attuned to aspects of their group identity and to maintain linguistic separation as an iconic and indexical marker of ethnic and other aspects of identity. We can see this in the maintenance of language varieties and ethnolects across generations in urban areas. In such cases, we may argue that urbanization highlights the unexpected intersection of some identities.

With respect to language maintenance and shift (LMS) specifically, the study of LMS has from the beginning included an emphasis on creating typologies of LMS by identifying the sociodemographic factors that correlate with either the maintenance of or shift away from minority languages amidst pressure to shift to a dominant language. Conklin and Laurie (1983), and Edwards (1992), present detailed taxonomies of political, social, demographic, cultural, and linguistic factors that are comprehensive and generalisable, and reflect the complexity of language shift scenarios (see also Himmelman 2010). All of these treatments include at least some discussion of the size of the speaker community with respect to the closeness and size of adjacent language communities, and Haarmann specifically lists the 'urban-rural' distinction as a relevant 'ethno-demographic' factor.

Numerous studies since have highlighted a link between urbanization and language shift away from minority languages (for some examples see Lieberson 1970; Cooper and Horvath 1973; Silver 1974; Timm 1980; Tsitsipis and Elmendorf 1983; Nordberg 1994; Bills, Chávez, and Hudson 1995; El Aissati 2001; Söylemez 2004; Makoni, Brutt-Griffler, and Mashiri 2007; Han 2013; Anderbeck 2015). Yet the processes that link urbanization with language shift remain poorly understood. Is language shift a consequence of *urbanization itself*, and the accompanying processes of social change and economic modernisation of everyday life? Or is language shift a consequence of *urbanicity*, that is, particular features of life in urban communities as compared to nonurban ones? In this work we draw on a terminological distinction between *urbanization* and *urbanicity* (Vlahov and Galea 2002), and examine urbanicity and the correlates of ongoing language shift across Indonesia.

We use census data from Indonesia – a large, ethnically and linguistically diverse, rapidly urbanising country – to examine the relationship between urbanicity and language shift in the context of urbanization. Indonesia's unique geographical spread of languages and urban areas, alongside the rapid growth of its national language (Bahasa Indonesia), makes it a particularly useful context for distinguishing the various mechanisms that might explain the relationship between cities and language change. Using a multilevel modelling approach and a 1% sample of Indonesians from the country's 2010 Census (comprising more than 2.1 million out of Indonesia's total population of 230 million people at the time), we show that in ethnically homogenous regions, urbanicity itself has little relationship with language shift. In contrast, ethnic diversity is consistently associated with a greater probability of speaking Indonesian both among urban and rural Indonesians and in both urban and rural areas. We show that the effects of being urban are interactive, strengthening the relationship between diversity and language shift.

Even though our analysis captures a single point in time, documenting the relationship between urbanicity and language shift, our findings provide useful insights into the mechanisms that link urbanization as a sociological phenomenon and language shift, highlighting the complex and conditional effects that urbanization has in a country that is experiencing rapid, large-scale language shift (Errington 1998; Steinhauer 1994; Sneddon 2003; Ananta et al. 2015). Although statistical data from population censuses can never replace the ethnographic insights garnered from careful fieldwork, official census data offer unparalleled insights for assessing the generality of sociological claims about language shift and the local conditions under which it takes place (see also Deumert 2010, 19–20). These findings also provide external validation for recent formal models of language

shift (see e.g. Kandler 2009; Prochazka and Vogl 2017) that hold that local communities interacting regularly are essential for maintaining minority language vitality – our data implies that homogenous urban areas are less likely to promote shift from minority languages to national languages. Finally, our approach suggests new directions for research, both quantitative and qualitative in nature: what is it about cities that heighten the effects of ethnic diversity in urban areas?

The rest of this paper proceeds as follows. We briefly review the classic literature on modernisation and the sociological effects of urbanization on identity, and summarise the evidence from linguistics and adjacent fields that associate urbanization with language shift. We then describe the limits of existing approaches for disentangling urbanization from urbanicity. The subsequent sections introduce first a short summary of the Indonesian linguistic landscape, followed by our data and empirical methods. We then use a graphical approach to characterise the interactive relationships between ethnic diversity, urbanicity, and language shift. The final section concludes with a discussion of the implications of our findings for studies of urbanization, urbanicity, and language shift more broadly.

Urbanization, modernisation, and language shift

The classic literature on modernisation held that urbanization was one of the central processes leading to the emergence of national identities that displaced or even substituted for local or particularistic identities. Nationality, for Deutsch (1953, 101), 'means an alignment of large numbers of individuals from the middle and lower classes linked to regional centres and leading social groups by channels of social communication and economic intercourse'. Although changing linguistic habits were not his main object of inquiry, the link between 'social communication' and the emergence of national identity and solidarity might naturally run through the increased use or development of a common language. Gellner (1983, 41) takes this argument further: in linking industrialisation to the emergence of nationalism, he notes that

> early industrialism means population explosion, rapid urbanization, labour migration, and … that the at least relatively stable and insulated Babel system of traditional agrarian communities, each inward-turned, kept separate by geography sideways … is replaced by a quite new kind of Babel.

That is, he explicitly links urbanization and industrialisation with linguistic heterogeneity. But what does it mean for an individual to be *urban*, and what does this mean for language shift?

Although estimates using almost any metric agree that the population of the world is now over 50% urban and growing, there is no standard global metric for what urban means, and every country defines it in different ways – some using simple or relative measures of population size, and others using some combination of size, density, and/or types of infrastructure.[1] As Vlahov and Galea (2002, S4) point out, this lack of uniform definition may hinder investigation of what is unique in urban versus nonurban living but it also highlights the dynamic nature of 'urban' as a construct and the extent to which the process of urbanization in addition to the condition of being urban are important considerations. They make a distinction between *urbanization*, i.e. changes to the size, density and heterogeneity of cities, and *urbanicity,* i.e. the impact of living in urban areas at a given time (Vlahov and Galea 2002). They further sketch out this distinction in terms of the public health outcomes and problems that are associated with urban areas. To give an example from their work: one outcome of *urbanization* may be an influx of impoverished people to a city in search of jobs and services which in turn taxes existing infrastructure and leads to an overtaxed sanitary system that leads to the spread of a disease. In contrast, one health impact of *urbanicity* is pedestrian motor vehicle injuries that result from a relatively high density of motor vehicles. To make the analogy with our work on language shift: an influx of people to a city may result in a sudden change in the racial or ethnic make-up of a city. This may lead to a shift in the primary language of the majority of speakers in a neighbourhood (especially among children in those neighbourhood schools). In this way urbanization changes the linguistic landscape of the city. The complementary

effect of urbanicity is living in a more densely settled locale, with more employment in the formal sector, and in what could be a more ethnically and linguistically diverse neighbourhood.

Urbanization and urbanicity are related but different, and they point to a need to identify both as considerations in the relationship between urban areas and language change. In the same vein, in Lieberson, Dalto, and Johnston's (1975) classic study of the national correlates of linguistic diversity in thirty-five nations, the authors find no relationship between urbanization and linguistic diversity, which they suggest may result from the 'linguistic homogeneity' of particular cities in particular countries; they write, 'the key issue is not urbanization itself, but the conditions under which its consequences will favour or undermine linguistic diversity' (59). Although there are likely to be many sociological processes that produce language shift in the context of cities, comprehensive quantitative evidence on how these processes operate is relatively scarce. Contextual factors related to urbanicity such as residential segregation among migrants (see, e.g. on Deumert 2010 on urban South Africa) may be investigated using a select number of localities, but it is hard to generalise from such results, which also often fail to compare urban to rural localities.

Our specific interest is in the role of ethnic and linguistic diversity in explaining language shift in the context of urbanicity. Urbanization, in principle, can involve a wide range of both social and economic transformations, but it need not feature all such transformations at once. An ethnolinguistically homogenous urban area may find residents engaged in more modern economic tasks, consuming new sources of media in national languages, and participating in new social conventions, but not forced to communicate with speakers of different languages on a daily basis. This possibility suggests that the implications of urbanicity for language shift depend on the precise ethnolinguistic structure of the urban area, and may not be monolithic. Note, also, that diversity may characterise rural areas as well as urban ones. The association of urbanicity with ethnic diversity presumes a distribution of language speakers in a plural society in which rural areas are homogenous ethnolinguistic homelands and urban areas are heterogeneous melting pots. Yet in many places, rural areas are also highly ethnolinguistically diverse (DiCarlo and Good 2020). This raises the possibility that not only do the effects of urbanicity depend on the ethnolinguistic structure, but that the linguistic effects that have been attributed to urbanization or urbanicity may be reducible to diversity alone.

Before proceeding, we pause to note that an influential literature on ethnic politics in urbanising context holds that rather than undermining ethnic attachments, urbanization may actually *catalyze or intensify* ethnic loyalties by making them salient through contact, or more sharply through group-based competition (see Olzak 1983 for an early review). Ethnic solidarity would hence *increase* through urbanization, thus creating incentives for ethnic differentiation rather than national homogenisation (Bates 1983; Eifert, Miguel, and Posner 2010). Some illuminating studies of language revitalisation in urban contexts (e.g. Shulist 2018 on the Brazilian Amazon) document the ways in which 'urban indigeneity' can form the basis for minority language preservation. The preservation of some substratum influence amidst shift to a dominant language (Prince 1988; Sharma 2011) as well as the indexical use of sociolinguistic markers of ethnicity in the form of particular lexical items, phonological features, and prosody (Dubois and Horvath 1998; Newmark, Walker and Stanford 2016; Wolfram and Dannenberg 1999) is arguably also evidence of this. In many cases, however, the explicit and implicit institutional pressure to shift to a dominant language is strong, and although language can be a powerful index of ethnic identify it is also quite possible to index ethnic identity in other ways, or to maintain indexical use of particular linguistic features while also shifting one's primary language of communication. In other words, Javanese migrants to Jakarta may simultaneously become more likely to identify strongly with their Javanese ethnic identity *and* less likely to regularly speak Javanese.

Indonesian context

To tease out the role of diversity in explaining the broader effects of urbanicity, we use the case of Indonesia, a large, multiethnic, linguistically diverse, and rapidly urbanising country. Indonesia is

one of the most multilingual and multiethnic nations in the world, home to 700 of the estimated 7,000 living languages spoken worldwide (Eberhard, Simons, and Fennig 2019) and made up of dozens if not hundreds of ethnic groups, ranging greatly in population size and spread unevenly over the 600+ inhabited islands in the 5,150 kilometer-long Indonesian archipelago islands (Embassy of the Republic of Indonesia). Ananta et al. classify over 1,300 coded categories in the 2010 census into over 600 ethnic (sub-ethnic, sub-sub-ethnic, and sub-sub-sub-ethnic groups, 2015, 41). These groups are defined both etically and emically, and overlap with geographic regions, religion, and language to varying degrees. The country's largest ethnic group, comprising 40% of the population, are Javanese, whose home region spans the provinces of Central and East Java (excluding the island of Madura). Other large ethnic groups include Sundanese (15%), Madurese (3%), Betawi (2.8%), Minangkabau (2.7%), and others.

It is not the case that every ethnic group identified by Ananta et al. speaks a language variety that is mutually unintelligible with others. Nonetheless there is a strong ideological connection between ethnicity and language in Indonesia. In the language communities that we have focused on in our work so far using census data (here, and Abtahian, Cohn, and Pepinsky 2016) this connection is even more pronounced, as we have focused on language varieties that have large speaker populations (of at least one million) and are closely associated with a geographical part of the Indonesian archipelago. Each of the large ethnic groups above (like hundreds of smaller ethnic groups throughout the country) traditionally speaks a language that is mutually unintelligible with Indonesian,[2] and none of these languages has official status in Indonesia.

Indonesia's national language, Bahasa Indonesia (hereafter Indonesian), is a standardised version of a Malay lingua franca originally spoken in parts of Sumatra but which had spread throughout the Malay archipelago prior to and under Dutch, British, and Portuguese colonial rule. Since independence, Indonesian has been taught in schools across this diverse archipelagic state, and nearly all print and broadcast media circulate in Indonesian. A wealth of recent studies has documented the process of language shift in Indonesia (Florey 2005; Goebel 2010; Errington 2014; Abtahian, Cohn, and Pepinsky 2016; Zentz 2017) towards use of the national language Indonesian in most domains.

Auer (2013, 7) points out that 'mobility has become such a central feature of human existence … that any kind of linguistics that is not able to address its effects will be in danger of falling out of step with reality'. Even if this were not the case more globally, no comprehensive linguistic study of an Indonesian community is possible without considering the effects of migration. Internal migration has played an outsized role in Indonesian history and politics, with two types of migration scenarios having particular importance in the Indonesian linguistic landscape: the 'immigration scenario', where a sizeable outside speech community moves to a new community (e.g. due to the *transmigrasi* policies of decades of Indonesian governments that have moved people out of Java and Bali to less populated parts of the archipelago), and the 'emigration scenario', in which some members of the community move out for education or work, in some cases moving back again and bringing the dominant language with them (Ewing 2014).

Indonesia has also seen a rapid rise in urbanization since independence. By 2010, 49.8% of Indonesians lived in urban areas, up from 22.4% as recently as 1980 (Firman, Kombaitan, and Pradono 2007; Firman 2016). Such explosive growth in Indonesia's urban population exceeds other cases of large and rapidly developing economies in the region, such as China, India, the Philippines, and Vietnam (Lewis 2014, 195). Although there is substantial inter-island migration throughout Indonesia, the rate of migration to urban areas far exceeds that of migration to rural areas (or between rural areas) (Muhidin 2014). Although urbanization is most notable in the case of the Jakarta metropolitan area on the island of Java, the growth of urban centres through population increase and rural-to-urban migration has occurred throughout the country. In the years following independence, the establishment of local municipal governments throughout the archipelago led to the rise of small- to medium-sized cities as regional bases for administration and economic activity (Milone 1964). More recently, Fahmi et al. (2014) describe how small cities like Cirebon, in West Java have

seen rapid growth as new decentralisation policies implemented in the early 2000s encourage residents from around the region to settle either in Cirebon's urban core or to cluster right on the outskirts of the city.

Data and methods

Our data come from a 1% sample of the 2010 Indonesian Census, available from IPUMS-International (Minnesota Population Center 2019). This sample was obtained from the full Indonesian census, which encompasses the full population of persons within Indonesia (both citizens and foreigners). From that full census, IPUMS randomly selected 10% of all individuals using a random sampling technique that ensured equal coverage across Indonesia's provinces. That sample size amounts to 22,928,795 individual records, from which we randomly sampled 10 percent to comprise our analysis sample.

The available census data provide information on individuals' demographic characteristics and household environment, as well as four data points that are of particular use here. First, the data tell us the district (*kabupaten* or *kota*) in which the individual lives; these correspond roughly to cities (*kota*) or counties (*kabupaten*) in a U.S. context. Second, the data tell us the ethnic group with which the individual identifies, which will allow us to construct a measure of district-level ethnic diversity (which we describe in more detail below). Third, the data contain an indicator for whether each individual lives in an urban area or not.[3] Some districts contain only urban residents, others contain only rural residents, and still others lie between these extremes.

Finally, the census data include information on language use. As Ananta et al. (2015, 274–281) explain, the 2010 census tabulated over 1,400 separate languages. These were drawn from question 210 of the Indonesian census, which read 'Apakah Bahasa sehari-hari yang digunakan [Nama] di rumah?' [What is [Name]'s language spoken daily at home?]. Importantly, the census questionnaire allows for multiple answers to this question, and in cases where both Indonesian and another language were supplied, the other language was recorded. As Ananta et al. argue, this means that responses that listed Bahasa Indonesia reflect only those cases in which only Bahasa Indonesia was spoken at home, and therefore comprise good evidence of language shift.

The public use sample, which we employ in our analysis, collapses all of the responses into a binary category: Bahasa Indonesia or some other language. Because we unfortunately lack information on *what* language the individual *does* speak at home in the event that that s/he does not report speaking Indonesian at home, so we are limited in our ability to capture language shift to languages that are not Indonesian, or, conversely, the use or maintenance of a language that is not associated with that individual's ethnic group.[4] Nevertheless these data allow us to probe the individual and contextual factors that correlate with whether an individual reports speaking Indonesian, and the factors that are shaping language shift towards the emerging dominant language. Moreover, because the same methodology was used throughout the census collection process across all of Indonesia, we expect our results to be internally comparable within the country.

One additional concern is that the normative value placed on speaking standard Bahasa Indonesia within the Indonesian political context will create pressure on respondents to say that they speak Indonesian, or on census enumerators to record Indonesian in cases of ambiguity. Although we cannot rule out such concerns empirically, we are confident that this is a minor concern. First, we note that census enumerators had the option to note multiple languages, and listed the respondent's ethnic language in cases where one was Indonesian. Second, the census questionnaire contains a separate question on whether or not the respondent is able to speak Indonesian, which reflects the widespread understanding that Indonesian is not universally spoken and allows respondents the space to report that. And third, our qualitative experience living and working in Indonesia suggests to us that there is little social stigma associated with speaking an ethnic language at home and Bahasa Indonesia in other social settings. The general Indonesian identity does not supplant

ethnic identities, but rather coexists with ethnic identities. One may easily and proudly hold both Javanese and Indonesian identities at the same time.

Because we know the district in which every individual lives, and we know their ethnic group, we can calculate a district-level measure of ethnic diversity using a standard fractionalisation index (Alesina et al. 2003). Given ethnic groups $g \in (1 \ldots G)$, we calculate a district's ethnic diversity as

$$Diversity_d = 1 - \sum_{g=1}^{G} s_{gd}^2 \tag{1}$$

where s_{gd} denotes the population share of ethnic group g in district d. District-level ethnic population shares are calculated directly from the census. Although our measure does not precisely measure ethno*linguistic* diversity (it only does so on the assumption that each ethnic group has its own distinct language, which is only loosely the case), this measure captures the extent of ethnic diversity that should be associated with language shift in the context of urbanicity.

We also use the indicator of whether an individual's residence is urban or not to measure district-level urbanicity by calculating the proportion of individual in each district recorded as urban.

$$Urbanicity_d = \frac{N_{urban,d}}{N_d} \tag{2}$$

Both of these measures – *Diversity* and *Urbanicity* – range from 0 to 1, with larger values corresponding to districts that are more diverse and more urban.

It is important to acknowledge the limitations of an approach that relies on census data rather than close ethnographic insights, or a targeted survey of Indonesians specifically designed to probe respondents about language use. As Abtahian, Cohn, and Yanti (forthcoming) note,

> Censuses are neither objective nor apolitical (Arel 2001), and there are implicit ideologies reflected in the way questions are asked and respondents expected to answer. A census that asks for a single "mother tongue", for example, presumes that all language varieties are named and considered "languages" by speakers and that speakers can name one variety in which they are most proficient, none of which may reflect actual linguistic practices or beliefs in a community.

The Indonesian census, by design, simplifies linguistic reality in order to facilitate classification at scale. In the same way, it also simplifies ethnic reality, encouraging respondents to choose one and only one ethnic identity when many Indonesians are of mixed ancestry. In using a sample of official census data, we are forced to confront the Indonesian state's normative understanding of what identities exist and what constitutes a language. This approach is most suitable for documenting aggregate relationships between urbanicity and the use of Bahasa Indonesia, but other sources of data would be needed to portray the full complexity of language use in contemporary Indonesia (see also Abtahian, Cohn, and Pepinsky 2016).

Diversity and urbanicity in Indonesia

In Figure 1, we present a scatterplot of districts by *Urbanicity* and *Diversity*, which reveals just how diverse Indonesia is across each dimension. We colour districts whose plurality (i.e. most populous) ethnic group is Javanese in blue, and others are in red. Four extreme points are labelled, and are instructive.

Starting at the upper left-hand corner, Kota Blitar (Blitar City) is a highly urban district in East Java that is nearly 100% ethnic Javanese, so it scores very low in terms of ethnic diversity. Moving to the upper right-hand corner, Jayapura is a highly urban district in the eastern province of Papua, a region which is notable for its high levels of ethnic diversity. Looking down the right-hand side we

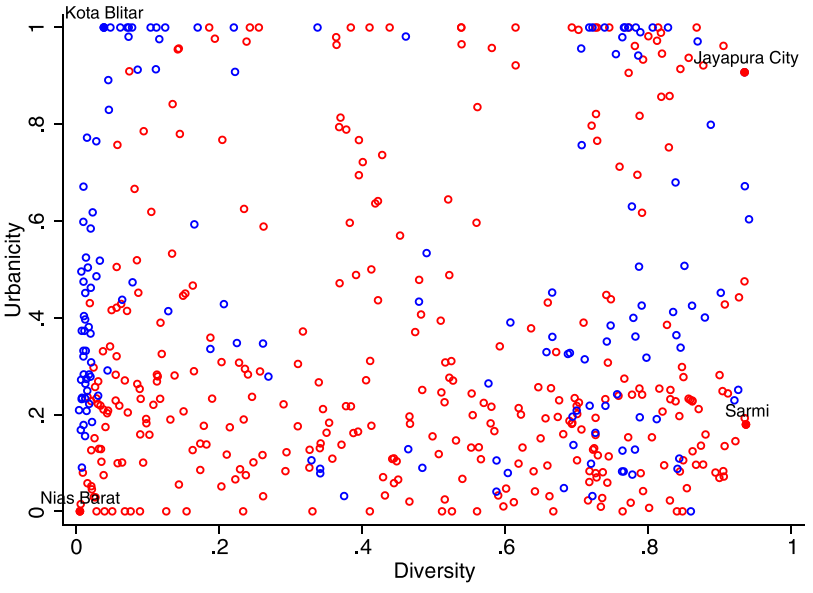

Figure 1. Urbanicity and diversity across Indonesian districts.

Note: This figure compares districts according to their ethnic diversity and level of urbanicity. Districts where the largest ethnic group is Javanese are in blue, others are in red.

discover a range of other districts in eastern Indonesia that are similarly diverse but increasingly more rural; Sarmi, also in Papua province, has relatively low urbanicity but comparable levels of ethnicity diversity to nearby Jayapura. Finally, at the bottom left-hand corner we see the district of Nias Barat, an entirely rural district located on the island of Nias off the southwestern coast of Sumatra, whose population is nearly entirely composed of the Nias ethnic group. Figure 2 shows where these districts are located within Indonesia, and helps to visualise how ethnic diversity and urbanicity vary across the archipelago.

This variation in urbanicity and diversity across Indonesian districts provides us with a unique opportunity to differentiate how each contextual factor shapes language choice. Unlike places where heterogeneous populations are associated only with urban centres, which makes it difficult to disentangle the effects of diversity from those of urbanicity, urbanicity and ethnic diversity vary independently in Indonesia. And indeed, we find wide variation in the use of Indonesian across the four selected districts, as shown in Table 1.

These findings – low levels of Indonesian in homogenous districts, and urbanicity associated with higher levels of Indonesian use in more diverse districts – suggest that both urbanicity and ethnic diversity will shape language shift. Observe, too, that the levels of Indonesian language use in the low diversity districts are very low.

To more formally characterise the relationships among urbanicity, ethnic diversity, and the use of Indonesian, we fit a multilevel logistic regression model (see Gelman and Hill 2007, 301–323) that predicts whether or not an individual speaks Indonesian at home as a function of individual-level demographic characteristics (age and its square, gender, religion, and education), an

Table 1. Indonesian use in four representative districts.

District	Urbanicity	Diversity	% Who Report Speaking Indonesian
Nias Barat	Low	Low	0.9%
Kota Blitar	High	Low	3.3%
Sarmi	Low	High	67.8%
Jayapura	High	High	94.8%

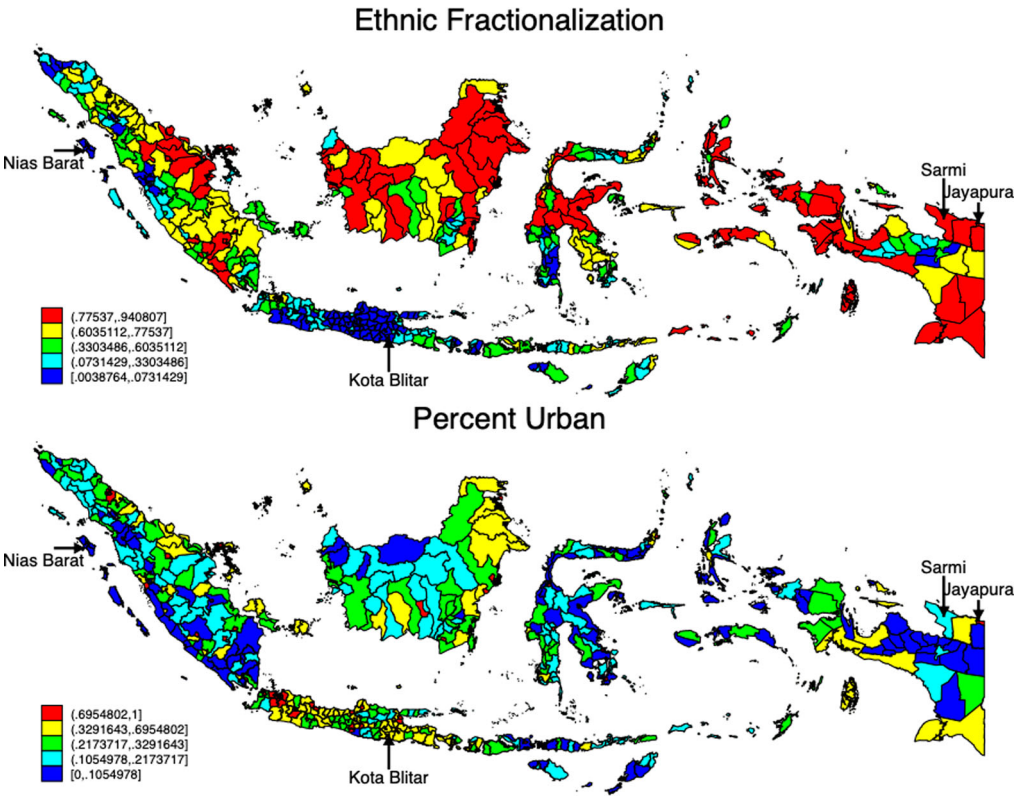

Figure 2. Urbanicity and diversity across Indonesia.

individual-level index of household material development, and district-level urbanicity, ethnic diversity, and their interaction. Specifically, we estimate a model of the following form:

$$\begin{aligned}\text{logit}(\pi_{id}) = {} & \gamma_{00} + \gamma_{01} Urbanicity_d + \gamma_{02} Diversity_d + \gamma_{03} Urbanicity \times Diversity_d + \beta \boldsymbol{X}_{id} \\ & + \delta_{0d} + \varepsilon_{id} \end{aligned} \tag{3}$$

Here, π_{id} is the probability that individual i in district d speaks Indonesian at home, $\boldsymbol{X}_{id}$ is a vector of individual covariates, δ_{0d} captures district-level random effects that are identified by assuming that they are distributed with mean zero and variance to be estimated from the data, and ε_{id} is an error term. The γ terms are the coefficients of interest to explore how urbanicity and diversity shape language choice. Because we do not allow any of the coefficients on the individual-level covariates to vary across districts (an assumption we will relax below), Equation (3) is a straightforward random intercepts model with district-level predictors.

We emphasise that although the statistical associations that we uncover here are consistent with causal arguments about the effects of urbanicity and diversity on language shift, given the absence of a quasi-experimental research design to estimate the causal effects of urbanicity and diversity, we interpret these results as descriptive correlations. Replication materials to recreate these analyses will be made freely available upon publication.

Results

The results for our first regression can be found as Model 1 (first column) in Table A1 in the Appendix. We note that the individual-level results are uniformly consistent with expectations: younger,

urban, more educated, and more socioeconomically privileged respondents are more likely to speak Indonesian at home. So, too, are all religious minorities in Muslim-majority Indonesia with the exception of Hindus, whose population is concentrated on the island of Bali. Although we find that women also more likely to speak Indonesian than men, this effect is substantively very small, a finding whose statistical significance results from the massive amount of data that we employ (for a fuller discussion of these factors, see Abtahian, Cohn, and Pepinsky 2016).

Our main substantive interest, however, is in the district-level coefficients. Because the substantive meaning of interaction terms are difficult to interpret, we adopt a graphical approach to examine on how different combinations of *Urbanicity* and *Diversity* correlate with speaking Indonesian. Using the results of Equation (3), we predict the probability that an individual speaks Indonesian, evaluating our prediction as if their district's ethnic diversity were at the 1st, 5th through 95th, and 99th percentiles of the observed distribution of ethnic diversity, and their district's level of urbanicity were at the 10th, 50th, and 90th percentiles of the observed distribution of district urbanicity.[5] In Figure 3, we plot these predicted probabilities (and their associated 95% confidence intervals).

The results in Figure 3 provide compelling evidence that the effects of urbanicity and ethnic diversity are analytically separable. Among respondents who live in ethnically homogenous districts (those on the left-hand side of the *x*-axis in Figure 3), the probability of speaking in Indonesian at home is very low, no matter how urban that district is. The overlap in confidence intervals indicates that there is no statistically significant difference between urban and rural districts when ethnic diversity is low. But moving along the *x*-axis, as ethnic diversity increases, so does the likelihood of speaking Indonesian – and *especially* so in the most urban districts. Among highly ethnically diverse districts, Indonesians who live in urban districts are statistically significantly more likely to speak Indonesian than are Indonesians who live in rural districts.

These results show very clearly that the relationship between urbanicity and language shift in a diverse developing country like Indonesia depends on whether urbanization comes with increasing ethnic diversity. Put starkly, urbanicity is only correlated with language shift in ethnically diverse

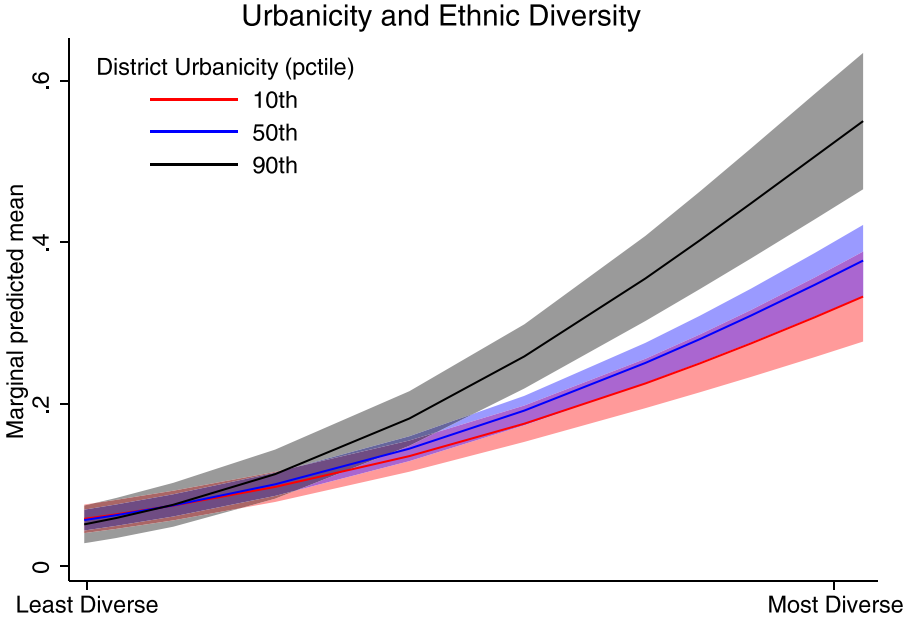

Figure 3. Predicted probabilities, random intercepts model.

Note: The solid lines are the predicted probability that an individual speaks Indonesian, for various values of *Urbanicity* and *Diversity*. The shaded areas correspond to 95% confidence intervals.

districts. And although the relationship between ethnic diversity and language shift is *largest* for urban districts, this relationship is substantively quite large in rural districts too. Urbanization strengthens the correlation between diversity and language shift – which is inconsistent with the argument that the effects of urbanization are *reducible* to ethnic diversity – but urbanicity alone has no independent relationship with language shift in homogenous Indonesian districts.

One question that immediately emerges is how to distinguish the effects attributable to urbanicity on those respondents who actually live in urban contexts from the general effects of urbanicity at the district level on all respondents. Individuals in our data may be rural residents who live in districts which are otherwise highly urban; do these relationships still show up among those individuals, and if so, are they comparable to the effects on urban residents in urban districts? To investigate this possibility, we fit a more general version of Equation (3) that allows the effects of urbanicity and diversity at the district level to vary across individuals depending on whether they are themselves living in urban or rural locations within that district. Specifically, we estimate the following model:

$$\begin{aligned}\text{logit}(\pi_{id}) = &\gamma_{00} + \gamma_{01} Urbanicity_d + \gamma_{02} Diversity_d + \gamma_{03} Urbanicity_d \times Diversity_d \\ &+ \gamma_{10} Urban_{id} + \gamma_{11} Urban_{id} \times Urbanicity_d + \gamma_{12} Urban_{id} \times Diversity_d \\ &+ \gamma_{13} Urban_{id} \times Urbanicity_d \times Diversity_d \\ &+ \beta \boldsymbol{X}_{id} + \ \delta_{0d} + \delta_{1d} Urban_{id} + \varepsilon_{id}\end{aligned} \tag{4}$$

Equation (4) simply adds interactions between a binary variable *Urban* (previously captured in $\boldsymbol{X}_{id}$) with *Urbanicity* and *Diversity*, as well as a random slope term δ_{1d} that allows the effects of *Urban* to vary across districts. The addition of the cross-level interactions and the random slopes provide an extremely parsimonious representation of the relationship between urbanicity and diversity across survey respondents.

These results appear as Model 2 (second column) in Table A1 (see appendix). Once again, we follow a graphical approach to interpret these results, this time calculating the predicted probability of speaking Indonesian separately for urban and rural respondents and plotting them next to one another. The top two plots in Figure 4 show the relationship between urbanicity, diversity, and the probability of speaking Indonesian for urban and rural residents separately. The bottom plot shows the estimated difference in the probability of speaking Indonesia for urban versus rural residents.

Together, these results indicate that the correlation between district urbanicity and ethnic heterogeneity applies both to urban and rural Indonesians. The top two plots indicate that it is ethnic diversity that matters the most, both for urban and rural Indonesians. The bottom plot shows that urban residents are always more likely to speak Indonesian than rural residents, and that the size of this difference increases modestly in more diverse districts, but the overlapping confidence intervals across levels of urbanicity reveal that there is no evidence that these differences between urban and rural Indonesians vary according to how urban a district is.

The finding that urbanicity has no relationship with language shift towards Indonesian in ethnically homogenous contexts is striking. One possibility is that these results are driven by the preponderance of ethnic Javanese living in heavily ethnic Javanese districts. Although not a numerical majority, ethnic Javanese are by far the most populous of Indonesia's ethnic groups, and play a particularly prominent role in Indonesian society and politics. Their home regions in Java include some of Indonesia's most densely populated and districts, both urban and rural. Moreover, in these districts, they often form an overwhelming numerical majority, which means that their district's level of ethnic diversity is very low (these are the blue points on the left-hand side of Figure 1). Might it be the case that these results are being driven by a particular tendency of Javanese respondents not to shift to Indonesian?

To test this possibility, we re-analyze the data, distinguishing between Javanese and non-Javanese in two separate ways: by whether or not an individual is Javanese, and whether or not an individual lives in a Javanese-plurality area. This gives us four separate samples (Javanese/non-Javanese

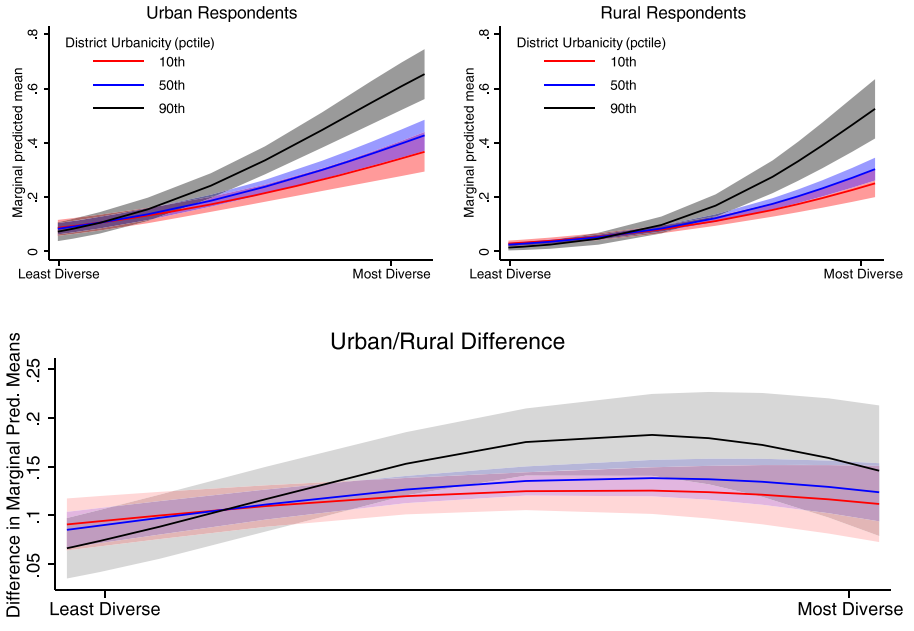

Figure 4. Predicted probabilities, cross-level effects model.

Note: In the top two plots, the solid lines are the predicted probability that an individual speaks Indonesian, for various values of *Urbanicity* and *Diversity*. In the bottom plot, the solid lines represent the difference in predicted probabilities that an individual speaks Indonesian between urban and rural individuals. In all three plots, the shaded areas correspond to 95% confidence intervals.

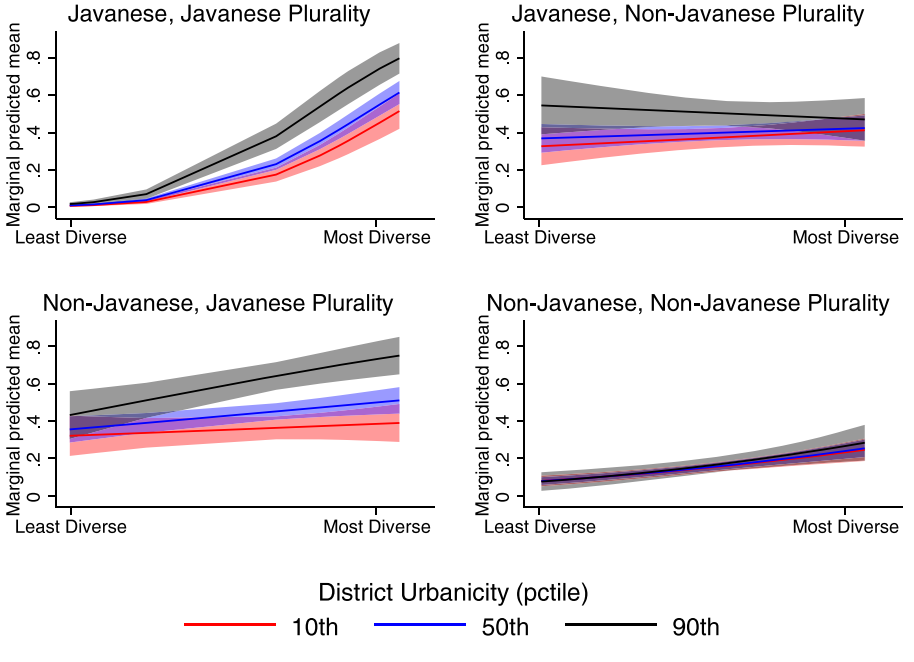

Figure 5. Javanese subgroup analyses.

living in Javanese/non-Javanese plurality areas). We then re-estimate Equation (3) on these four samples and plot the results to see if such analyses uncover similar findings as above. Our graphical results appear in Figure 5 (with numerical results in Table A2 in the Appendix). To avoid

extrapolation, our predictions are evaluated at the 1st, 5th through 95th, and 99th percentiles of the observed distribution of ethnic diversity, and at the 10th, 50th, and 90th percentiles of the observed distribution of district urbanization, for each of the four subsamples.

The evidence suggests, contrary to expectations, that the effects of urbanicity and diversity on language shift are happening primarily among ethnic Javanese and in Javanese-plurality districts.

Look first to the upper-left hand plot in Figure 5. The least diverse Javanese majority-districts are those in the Javanese heartland; there, use of Indonesian (or language shift away from Javanese) is rare, no matter how urban or rural the setting. But in the most diverse Javanese-plurality districts, use of Indonesian is much more common, and especially in the most urban districts. This pattern is the same as we have uncovered in the main analysis above. Compare this to the upper-right hand plot. The least diverse non-Javanese majority districts are those where non-Javanese ethnic groups predominate, and there, ethnic Javanese are *more* likely to speak Indonesian, especially in the most urban districts, likely because they do not reside in Javanese-dominant subcommunities. We also discover, however, that the independent effect of urbanicity *disappears* in the most diverse non-Javanese plurality districts. In the most diverse districts, ethnic Javanese are quite likely to speak Indonesian, but it does not much matter how urban they are.

Looking to the bottom left-hand plot, we learn that non-Javanese living in Javanese-plurality districts are, like Javanese, more likely to speak Indonesian in more diverse districts. Urbanicity strengthens these correlations. But the bottom right-hand plot confirms that non-Javanese living in non-Javanese-plurality districts, on the whole, are those who are least likely to speak Indonesian. The probability that they speak Indonesian rises modestly as diversity increases, but urbanicity rates do not differentiate among speakers.

Further research is needed to uncover the factors that explain these particular patterns in the data. But the main takeaway point is that the lack of a relationship between urbanicity and language shift in ethnically homogenous districts that we observed in our full sample analysis (see Figure 3) is not driven by Javanese respondents who are not shifting to Indonesian, for we find that this relationship also obtains among non-Javanese in non-Javanese plurality districts. Javanese living in non-Javanese majority districts and ethnically diverse, Javanese-majority urban districts, in fact, are quite likely to shift to Indonesian.

Conclusion

Our analysis of Indonesian census data has provided novel insights into urbanicity and language shift in one of the most linguistically diverse countries in the world. We do not find a consistent relationship between urbanicity and language shift. Instead, we find that ethnic diversity in urban areas is a precondition for urbanicity to correlate with use of Indonesian. In roughly the least diverse tercile of Indonesian districts, there is simply no relationship at all between urbanicity and language shift. Urban respondents are always more likely to speak Indonesian than rural Indonesians, but this difference does not depend on urbanicity alone. It is tempting to summarise these findings as offering evidence against urbanization promoting language shift, but that is not the right inference. Instead, we find that urbanicity matters in more complex ways than is commonly expected, and that ethnic diversity provides a foundation for understanding how. There are a number of ways in which scholars of language shift, urbanization, and Indonesian politics and society can build on these descriptive findings.

First, looking within Indonesia, ethnographic and qualitative evidence – alongside more fine-grained census information than is available here – may be able to shed light on the mechanisms that strengthen the association between diversity and language shift in the most urban parts of Indonesia. It is especially important to consider the social environment of urbanicity, and of the communities in which speakers live and social networks that they participate in. Dense, multiplex social networks, for example, are characteristic of some urban communities, and have been shown to promote language maintenance (Li Wei 1994; Zentella 1997). Sociolinguistic investigations of

urban speech communities in Indonesia would benefit from an approach that goes beyond externally defined models of language choice and examines speakers' use of their multilingual repertoires in practice. Variationist research teaches us that there are no single style speakers (Labov 1972), and that even a single interview (let alone a single census question) is unable to uncover the range of speakers' multilingual/multidialectal competence (Sharma 2011), and their use of their multilingual or multidialectal repertoire in practice.

In order to make any claims about the causal role of urbanization or urbanicity in language change it is crucially important that we also consider the effects of urbanization from the other side – that is, cases of rural multilingualism (see DiCarlo, Good, and Diba 2019 for a discussion of rural multilingualism in Africa). In Indonesia, such research could look selectively at those diverse but largely rural districts to learn the conditions under which language shift is taking place.

Relatedly, future research may be able to show us how various aspects of modernity – from urbanicity to other factors like spatial mobility and information and communication technology – explain the process of language shift. If, for example, mobile communication technology connects people across cities and regions, then this might help to explain why urbanicity itself is unrelated to language shift. It should suggest that there is something about the daily need to communicate with members of other linguistic communities that prompts language shift in the Indonesian context. We see this as a promising direction for future research.

Second, a more disaggregated approach to non-Javanese ethnic groups may be able to examine the mechanisms that allow minority languages to survive. Although many observers of language and development in Indonesia focus on the threats to small languages spoken by small minority communities, we have shown that Javanese in diverse urban contexts are actually quite likely to speak Indonesian; this is also true in homogenous urban areas where they are a minority. Looking more carefully at non-Javanese living in non-Javanese plurality areas – both homogenous and diverse, urban and rural – is essential for further teasing out the mechanisms that explain these findings. Recent work by Kurniasih (2006), Misnadin (2021), Nurani (2015), Travis and Inas (2021), Vanderklok (2019), and Zen and Starr (2021) demonstrate the very local nature of language ecologies in Indonesia and the importance of speech community-level investigations to illuminate the relationships between Indonesian and local languages. In similar work investigating ongoing sound change in Acehnese and Javanese, respectively, for example, Travis and Inas (2021) and Zen and Starr (2021) find that speakers in the two communities (in Banda Aceh and East Java) have very different experiences with the use of Indonesian – most of Travis and Inas's speakers in Banda Aceh reported balanced proficiency in Acehnese and Indonesia, while Zen and Starr's speakers in East Java have largely shifted to Indonesian as a primary and dominant language. Further comparative work is needed to be able to make broader claims about the nature of language shift in Indonesian communities.

Finally, we urge scholars of urbanization, the sociology of language, and language shift to adopt a similar approach in the comparative context. We have identified strong evidence that it is the diversity of urban areas – rather than urbanization itself – that explains the association between urbanicity with language shift. This approach can help to reconcile the inconsistent findings on urbanization from the macro-level analysis of Lieberson, Dalto, and Johnston (1975). But unlike much of the subsequent literature that followed their landmark analysis, which has focused on immigrant inclusion and language shift, our approach encourages us to focus on the processes of internal migration that generate diverse localities that may be either urban or rural. Indonesia is distinctive in that diversity and urbanicity vary independently, but analogous processes of urbanization together with increasing diversity are certainly at play in linguistically diverse, rapidly urbanising countries such as Angola, China, Ethiopia, India, Nigeria, Pakistan, South Africa, Tanzania. Further statistical research looking within such national contexts may help to uncover the specific conditions under which urbanization produces language shift.

Notes

1. United Nations, Department of Economic and Social Affairs, Population Division (2018). World Urbanization Prospects: The 2018 Revision, Online Edition. Available at: https://esa.un.org/undp/wup/
2. The only exceptions are some of the Malayic languages that are natively spoken in the provinces of Riau, Jambi, and elsewhere in the island of Sumatra and its near islands, as well as in parts of Kalimantan (Indonesisan Borneo) as well. The traditional language of the Betawi people is also a hybrid form of Malay spoken around Jakarta and incorporating elements of Hokkien, Portuguese, Javanese, and other elements not found in other Malayic languages (Ikranagara 1980). These languages are mutually intelligible – to greater or lesser degrees – with the standardised Malay form that is Indonesian.
3. The definition of an area as 'urban' in the 2010 Census depends on its population density, the percentage of households in the area engaged in agriculture, and access to 'urban facilities' such as schools, markets, and hospitals. Every lowest-level administrative unit (equivalent to a village or neighbourhood) in Indonesia is given a numerical score based on these factors, which is then used to code that area as urban or rural (see Badan Pusat Statistik 2010 for a description). Individuals are then assigned the status of residing in an urban area or not based on whether their area of residence is urban or rural.
4. The 2010 census itself does record the name of the language that the respondent reports using at home, but those data are not made publicly available for researcher use.
5. These predictions, obtained using the margins command in Stata 17, correspond to 'predictive margins' that average over the observed values of the other covariates in the sample.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Thomas B. Pepinsky http://orcid.org/0000-0002-4000-217X
Maya Ravindranath Abtahian http://orcid.org/0000-0002-3084-0778

References

Abtahian, Maya Ravindranath, Abigail C. Cohn, and Thomas Pepinsky. 2016. Modeling Social Factors in Language Shift. *International Journal of the Sociology of Language* 242: 139–170.

Abtahian, Maya Ravindranath, Abigail C. Cohn, and Yanti. Forthcoming. Language Labeling and Ideology in Indonesia. *International Journal of the Sociology of Language*.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8 (2): 155–194.

Ananta, Aris, Evi Nurvidya Arifin, M. Sairi Hasbullah, Nur Budi Handayani, and Wahyu Pramono. 2015. *Demography of Indonesia's Ethnicity*. Singapore: Institute of Southeast Asian Studies.

Anderbeck, Karl. 2015. "Portraits of Language Vitality in the Languages of Indonesia." In *Language Documentation and Cultural Practices in the Austronesian World*, Papers from ICAL 12 (A–PL 019) 4, edited by I. WayanArka, Ni LuhNyoman Seri Malini, and Ida Ayu Made Puspani, 19–47. Canberra: Pacific Linguistics.

Arel, Dominique. 2001. "Language Categories in Censuses: Backward-or Forward-Looking?" In *Census and Identity: The Politics of Race, Ethnicity, and Language in National Censuses*, edited by D. I. Kertzer, and D. Arel, 92–120. Cambridge: Cambridge University Press.

Auer, Peter. 2013. "The Geography of Language: Steps Toward a New Approach." *Freiberger Arbeitspapiere zur Germanistischen Linguistik (FRAGL)* 16: 1–39.

Badan Pusat Statistik. 2010. *Peraturan Kepala Badan Pusat Statistik Nomor 27 Tahun 2010 tentang Klasifikasi Perkotaan dan Perdesaan di Indonesia*. https://jdih.bps.go.id/files/produk_hukum/perka/P08201037.pdf.

Bates, Robert. 1983. "Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa." In *State Versus Ethnic Claims: African Policy Dilemmas*, edited by Donald Rothchild, and Victor A. Olorunsola, 152–171. London: Westview Press.

Bills, Garland D., Eduardo Hernández Chávez, and Alan Hudson. 1995. "The Geography of Language Shift: Distance from the Mexican Border and Spanish Language Claiming in the Southwestern U.S." *International Journal of the Sociology of Language* 114 (1): 9–27.

Cheshire, Jenny, Paul Kerswill, Sue Fox, and Eivind Torgersen. 2011. "Contact, the Feature Pool, and the Speech Community: The Emergence of Multicultural London English." *Journal of Sociolinguistics* 15 (2): 151–196.

Clyne, Michael. 2000. "Lingua Franca and Ethnolects in Europe and Beyond." *Sociolinguistica* 14: 83–89.

Conklin, Nancy R., and Margaret A. Lourie. 1983. *A Host of Tongues: Language Communities in the United States*. New York: Free Press.
Cooper, Robert L., and Ronald J. Horvath. 1973. "Language, Migration, and Urbanization in Ethiopia." *Anthropological Linguistics* 15 (5): 221–243.
Deumert, Ana. 2010. "Tracking the Demographics of (Urban) Language Shift – An Analysis of South African Census Data." *Journal of Multilingual and Multicultural Development* 31 (1): 13–35.
Deutsch, Karl. 1953. *Nationalism and Social Communication: An Enquiry Into the Foundations of Nationality*. Cambridge: The MIT Press.
DiCarlo, Pierpaolo, and Jeff Good, eds. 2020. *African Multilingualisms: Rural Linguistic and Cultural Diversity*. New York: Lexington Books.
DiCarlo, Pierpaolo, Jeff Good, and Rachel Ojong Diba. 2019. "Multilingualism in Rural Africa." *Oxford Research Encyclopedia of Linguistics*. doi:10.1093/acrefore/9780199384655.013.227.
Dubois, Sylvie, and Barbara Horvath. 1998. "Let's Tink about dat: Interdental Fricatives in Cajun English." *Language Variation and Change* 10: 246–261.
Eberhard, David, Gary F. Simons, and Chuck Fennig. 2019. *Ethnologue, Languages of the World*. 22nd ed. SIL International.
Edwards, John. 1992. "Sociopolitical Aspects of Language Maintenance and Loss: A Typology of Minority Language Situations." In *The Maintenance and Loss of Minority Languages*, edited by Willem Fase, Koen Jaspaert, and Sjaak Kroon, 37–54. Amsterdam: Benjamins.
Eifert, Ben, Edward Miguel, and Daniel N. Posner. 2010. "Political Competition and Ethnic Identification in Africa." *American Journal of Political Science* 54 (2): 494–510.
El Aissati, Abderrahman. 2001. "Ethnic Identity, Language Shift, and the Amazigh Voice in Morocco and Algeria." *Race, Gender & Class* 8 (3): 57–69.
Errington, J. Joseph. 1998. *Shifting Languages: Interaction and Identity in Javanese Indonesia*. Cambridge, UK: Cambridge University Press.
Errington, J. Joseph. 2014. "In Search of Middle Indonesian: Linguistic Dynamics in a Provincial Town." In *In Search of Middle Indonesia: Middle Classes in Provincial Towns*, edited by Gerry van Klinken, and Ward Berenschot, 198–219. Boston, MA: Brill.
Ewing, Michael. 2014. "Language Endangerment in Indonesia." *International Journal of Education* 8 (1): 12–22.
Fahmi, Fikri Zul, Delik Hudalah, Paramita Rahayu, and Johan Woltjer. 2014. "Extended Urbanization in Small and Medium-Sized Cities: The Case of Cirebon, Indonesia." *Habitat International* 42: 1–10.
Firman, Tommy. 2016. "Demographic Patterns of Indonesia's Urbanization, 2000–2010: Continuity and Change at the Macro Level." In *Contemporary Demographic Transformations in China, India and Indonesia. Volume 5: Demographic Transformation and Socio-Economic Development*, edited by Christophe Z. Guilmoto, and Gavin W. Jones, 255–269. New York: Springer.
Firman, Tommy, Benedictus Kombaitan, and Pradono Pradono. 2007. "The Dynamics of Indonesia's Urbanisation, 1980–2006." *Urban Policy and Research* 25 (4): 433–454.
Florey, Margaret. 2005. "Language Shift and Endangerment." In *The Austronesian Languages of Asia and Madagascar*, edited by A. Adelaar, and N. Himmelman, 43–64. New York: Routledge.
Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca: Cornell University Press.
Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Modeling*. New York: Cambridge University Press.
Goebel, Zane. 2010. *Language, Migration and Identity: Neighborhood Talk in Indonesia*. Cambridge, UK: Cambridge University Press.
Han, Enze. 2013. "External Cultural Ties and the Politics of Language in China." *Ethnopolitics* 12 (1): 30–49.
Himmelman, Nikolaus P. 2010. "Language Endangerment Scenarios: A Case Study from Northern Central Sulawesi." In *Endangered Languages of Austronesia*, edited by Margaret Florey, 45–72. New York: Oxford University Press.
Ikranagara, Kay. 1980. "Melayu Betawi Grammar." In *Linguistics Studies in Indonesian and Languages of Indonesia*, vol. 9, edited by Amran Halim, Soepomo Poedjosoedarmo, I. Suharno, and John W.M. Verhaar. Jakarta: NUSA.
Kandler, Anne. 2009. "Demography and Language Competition." *Human Biology* 81 (3): 181–210.
Kircher, Ruth, and Sue Fox. 2019. "Multicultural London English and Its Speakers: A Corpus-Informed Discourse Study of Standard Language Ideology and Social Stereotypes." *Journal of Multilingual and Multicultural Development* 40 (10): 847–864.
Kurniasih, Yacinta. 2006. "Gender, Class and Language Preference: A Case Study in Yogyakarta." In *Selected Papers from the 2005 Conference of the Australian Linguistic Society*, edited by Keith Allan, 1–25. Melbourne, Victoria: Australian Linguistic Society..
Labov, William. [1966] 2006. *The Social Stratification of English in New York City*. New York: Cambridge University Press.
Labov, William. 1972. "Field Methods of the Project in Linguistic Change and Variation." In *Language in Use*, edited by John Baugh and Joel Sherzer, 28–53. Prentice-Hall.

Lauder, Multimia RMT, and Allan F. Lauder. 2017. "Language Change and Endangerment in West Java: Recent Dialectology Research." Plenary talk at the *International Seminar on Sociolinguistics and Dialectology*. https://linguistik.fib.ui.ac.id/wp-content/uploads/sites/46/2018/01/8-18-Multamia-RMT-Lauder-dan-Allan-F.-Lauder.pdf Last consulted 10/9/2021.
Lewis, Blane D. 2014. "Urbanization and Economic Growth in Indonesia: Good News, Bad News and (Possible) Local Government Mitigation." *Regional Studies* 48 (1): 192–207.
Lieberson, Stanley. 1970. *Language and Ethnic Relations in Canada*. New York: Wiley.
Lieberson, Stanley, Guy Dalto, and Mary Ellen Johnston. 1975. "The Course of Mother-Tongue Diversity in Nations." *American Journal of Sociology* 81 (1): 34–61.
Makoni, Sinfree, Janina Brutt-Griffler, and Pedzisai Mashiri. 2007. "The Use of 'Indigenous' and Urban Vernaculars in Zimbabwe." *Language in Society* 36 (1): 25–49.
McLaughlin, Fiona. 2008. "The Ascent of Wolof as an Urban Vernacular and National Lingua Franca in Senegal." In *Globalization and Language Vitality*, edited by Cécile Vigouroux, and Salikoko Mufwene, 142–170. London: Continuum.
Milone, Pauline D. 1964. "Contemporary Urbanization in Indonesia." *Asian Survey* 4 (8): 1000–1012.
Minnesota Population Center. 2019. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS. doi:10.18128/D020.V7.2.
Misnadin. 2021. "Voice Onset Time and Vowel Quality in Madurese: The Role of Gender and Dialect." *Asia Pacific Language Variation* 7 (2): 168–184.
Muhidin, Salut. 2014. "Migration Patterns: People on the Move." In *Regional Dynamics in a Decentralized Indonesia*, edited by Hal Hill, 317–341. Singapore: ISEAS–Yusof Ishak Institute.
Newmark, Kalina, Nacole Walker, and James Stanford. 2016. "The rez accent knows no borders': Native American ethnic identity expressed through English prosody." *Language in Society* 45 (5): 633–664.
Nordberg, Bengt, ed. 1994. *The Sociolinguistics of Urbanization: The Case of the Nordic Countries*. New York: Walter De Gruyter.
Nurani, Lusia Marliana. 2015. "Changing Language Loyalty and Identity: An Ethnographic Inquiry into the Societal Transformation of the Javanese People." PhD Dissertation; Arizona State University, Tempe, Arizona, USA.
Olzak, Susan. 1983. "Contemporary Ethnic Mobilization." *Annual Review of Sociology* 9: 355–374.
Prince, Ellen. 1988. "On Pragmatic Change: The Borrowing of Discourse Functions." *Journal of Pragmatics* 12: 505–518.
Prochazka, Katharina, and Gero Vogl. 2017. "Quantifying What Drives Language Shift." *Proceedings of the National Academy of Sciences* 114 (17): 4365–4369.
Sankoff, David, Gillian Sankoff, Suzanne Laberge, and Marjorie Topham. 1976. "Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale." *Cahier de linguistique de l'Université du Québec* 6: 85–125.
Sharma, Devyani. 2011. "Style Repertoire and Social Change in British Asian English." *Journal of Sociolinguistics* 15 (4): 464–492.
Shulist, Sarah. 2018. *Transforming Indigeneity: Urbanization and Language Revitalization in the Brazilian Amazon*. Toronto: University of Toronto Press.
Silver, Brian. 1974. "Social Mobilization and the Russification of Soviet Nationalities." *American Political Science Review* 68 (1): 45–66.
Sneddon, James. 2003. *The Indonesian Language; Its History and Role in Modern Society*. Sydney: University of New South Wales Press.
Sneddon, James N. 2006. *Colloquial Jakartan Indonesian*. Canberra, Australia: Pacific Linguistics.
Söylemez, Ümit. 2004. "Urbanization and Language Shift in Turkey: The Change Processes at Work in the Transition from Rural to Urban Settings." *International Journal of the Sociology of Language* 2004 (165): 93–119.
Steinhauer, Hein. 1994. "The Indonesian Language Situation and Linguistics; Prospects and Possibilities." *Bijdragen tot de Taal-, Land- en Volkenkunde* 150: 755–784.
Timm, Lenora A. 1980. "Bilingualism, Diglossia and Language Shift in Brittany." *International Journal of the Sociology of Language* 1980 (25): 29–42.
Travis, Catherine, and Ghina Inas. 2021. "Gender, Mobility and Contact: Stability and Change in an Acehnese Dialect." *Asia Pacific Language Variation* 7 (2): 142–167.
Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
Tsitsipis, Lukas D., and William W. Elmendorf. 1983. "Language Shift among the Albanian Speakers of Greece." *Anthropological Linguistics* 25 (3): 288–308.
Vander Klok, Jozina. 2019. "The Javanese Language at Risk? Perspectives from an East Java Village." *Language Documentation and Conservation* 13: 300–345.
Vlahov, David, and Sandro Galea. 2002. "Urbanization, Urbanicity, and Health." *Journal of Urban Health* 79 (4): S1–S12.
Wei, Li. 1994. *Three generations, two languages, one family*. Clevedon: Multilingual Matters.

Wolfram, Walt, and Clare Dannenberg. 1999. "Dialect Identity in a Tri-Ethnic Context: The Case of Lumbee American Indian English." *English World-Wide* 20 (2): 179–216.
Zen, Evynurul, and Rebecca Starr. 2021. "Variation and Contact-Induced Change in Javanese Phonology among Multilingual Children in Indonesia." *Asia Pacific Language Variation* 7 (2): 95–119.
Zentella, Ana. 1997. *Growing up bilingual*. Oxford: Blackwell.
Zentz, Lauren. 2017. *Statehood, Scale, and Hierarchy: History, Language, and Identity in Indonesia*. Bristol, UK: Multilingual Matters.

Appendix

Table A1. Regression results for Models 1 and 2.

Variable	Model 1	Model 2
Age	−0.0591	−0.0595
	(−92.14)	(−92.31)
Age × *Age*	0.000542	0.000547
	(63.03)	(63.26)
Female	0.0461	0.0466
	(8.84)	(8.90)
Urban	1.396	1.797
	(170.65)	(10.69)
Level of Development	0.247	0.243
	(129.42)	(125.34)
Level of Education	0.268	0.269
	(151.71)	(151.03)
Religion: Other	0.789	0.779
	(14.00)	(13.76)
Religion: Christian	0.830	0.855
	(83.53)	(84.82)
Religion: Buddhist	2.058	2.068
	(57.04)	(56.89)
Religion: Hindu	−1.992	−1.980
	(−57.94)	(−57.62)
District Urbanization	−0.405	−0.662
	(−0.80)	(−1.20)
Ethnic Fractionalization	2.902	3.320
	(6.73)	(7.65)
District Urbanization × *Ethnic Fractionalization*	2.322	2.789
	(2.62)	(2.95)
Urban × *District Urbanization*		−0.0152
		(−04)
Urban × *Ethnic Fractionalization*		−1.193
		(−4.02)
Urban × *District Urbanization* × *Ethnic Fractionalization*		0.476
		(0.71)
Constant	−5.826	−6.003
	(−24.35)	(−24.71)
Var(δ_{0d})	3.452	3.337
	(14.85)	(14.26)
Var(δ_{1d})		1.112
		(12.43)
Observations	2111288	2111288
Log-likelihood	−650180.2	−648074.8
Chi-square statistic	315465.2	321016.7

Cells report logistic regression coefficients with *t* statistics in parentheses.

Table A2. Regression results, Javanese subgroups.

Variable	Model 1	Model 2	Model 3	Model 2
Age	−0.0738	−0.0518	−0.0487	−0.0668
	(−25.30)	(−11.31)	(−17.21)	(−34.83)
Age × *Age*	0.000687	0.000440	0.000464	0.000611
	(17.82)	(7.06)	(12.21)	(23.62)
Female	0.106	0.0657	0.0839	0.0643
	(4.46)	(1.82)	(3.62)	(4.16)
Urban	0.195	0.247	0.204	0.277
	(21.10)	(16.69)	(23.78)	(50.81)
Level of Development	0.284	0.200	0.248	0.289
	(35.87)	(15.96)	(29.48)	(55.01)
Level of Education	0.856	−0.805	0.408	0.861
	(2.16)	(−1.68)	(1.60)	(5.89)
Religion: Other	1.412	0.807	0.383	1.147
	(26.34)	(7.80)	(11.12)	(36.17)
Religion: Christian	2.253	0.830	1.455	2.259
	(8.22)	(1.22)	(11.98)	(23.45)
Religion: Buddhist	1.440	−0.267	−0.365	−2.058
	(4.98)	(−0.80)	(−2.55)	(−21.42)
Religion: Hindu	1.301	0.977	1.196	1.564
	(33.53)	(15.34)	(33.36)	(64.71)
District Urbanization	0.895	0.732	0.556	−0.0575
	(1.64)	(0.91)	(0.86)	(−0.07)
Ethnic Fractionalization	6.450	−0.642	−0.140	2.321
	(11.55)	(−1.06)	(−0.21)	(4.10)
District Urbanization × *Ethnic Fractionalization*	1.229	0.0540	2.553	0.461
	(1.28)	(0.04)	(2.30)	(0.34)
Constant	−7.235	−2.146	−2.913	−5.911
	(−22.33)	(−5.76)	(−7.54)	(−18.67)
Var(δ_{0d})	1.454	2.662	1.826	4.054
	(8.10)	(8.73)	(7.87)	(11.65)
Observations	193355	22313	63127	248868
Log-likelihood	−31085.9	−12471.3	−29182.1	−75885.7
Chi-square statistic	10514.9	4808.5	10258.1	38345.9

Cells report logistic regression coefficients with *t* statistics in parentheses.