Maya Ravindranath Abtahan*, Abigail C. Cohn and Thomas Pepinsky

Modeling social factors in language shift

DOI 10.1515/ijsl-2016-0036

Abstract: This article introduces a quantitative approach to modeling language shift in communities with millions of speakers. Using Indonesia as a case study, and employing a large body of data from the Indonesian population census, we document how factors such as urbanization, ethnicity, economic development, gender, and religion correlate with the shift from local languages (Javanese, Sundanese, etc.) to the national language, Bahasa Indonesia. Our findings inform ongoing research on the sociological foundations of language shift across both small and large communities. Methodologically, we introduce a statistical approach that borrows from other social sciences, and show how to exploit massive amounts of untapped linguistic, demographic, and sociological data.

Keywords: census data, Indonesia, language shift, endangered languages

1 Introduction

Research on language shift and endangerment worldwide has primarily focused on languages with small speaker populations, partly following Krauss (1992), who defines language obsolescence as an issue affecting non-official languages with speaker populations of under 100,000 (at that time, an estimated 90 percent of the world’s languages). Fishman (1991), by contrast, understands the central problem of language shift to be intergenerational transmission. If intergenerational transmission of a language is not occurring in a particular language community, then that language is potentially endangered, even if that language has millions of speakers. This is the reasoning behind most assessments of language vitality, including Fishman’s (1991) GIDS (Graded Intergenerational Disruption Scale), Lewis and Simons (2010) EGIDS (Expanded Graded Intergenerational Disruption Scale), and UNESCO’s language vitality framework (Moseley 2010) – assessments of language vitality that primarily seek to determine the extent to which children are successfully learning their parents’ languages. It is increasingly understood that in fact a large speaker population does not necessarily

*Corresponding author: Maya Ravindranath Abtahan, University of Rochester, Rochester, NY, USA, E-mail: maya.r.abtahan@rochester.edu

Abigail C. Cohn, Thomas Pepinsky, Cornell University, Ithaca, NY, USA
correlate with successful intergenerational transmission, and indeed, in Indonesia, we find that there is little correlation between a language’s size and its EGIDS vitality score (Ravindranath and Cohn 2014). Moreover we also suspect that this lack of correlation is further masked by researchers’ reluctance to assign a language with a large speaker population a low vitality score.\(^1\) Although the global processes that shape language shift are a subject of ongoing research among linguists (see Grenoble and Whaley 1998; Crystal 2000; Austin and Sallabank 2011; Harbert 2011), they remain poorly understood. Because larger languages are not thought to be critically endangered, they are rarely objects of study. Moreover, most existing work focuses on particular speaker communities, an approach ill-suited to developing comprehensive treatments of the relative prevalence of language shift across communities. Characterizing empirically the nature and scope of language shift in larger speaker communities therefore remains an open field for inquiry.

Accompanying a focus on languages with small speaker populations has been an emphasis on ethnographic and/or small-scale studies at the local speech community level. Volumes of work on language endangerment have largely included case studies and ethnographies of language shift scenarios in local language communities (e. g. in Dorian 1989; Seliger and Vago 1991; Bradley and Bradley 2002; Florey 2010). These provide us with invaluable information on language loss at the individual and community level, and are the most meaningful way of determining anything about individual language attitudes toward language varieties, as well as on the success of language revitalization efforts at the local level. However, once we acknowledge that there is no such thing as “too big to fail” when it comes to language endangerment (Anderbeck 2015; Ravindranath and Cohn 2014), we can complement our understanding of language shift at the local level with analyses of diverse language communities with speaker populations in the millions. Moreover, with the increasing availability of big data sources, we can now use multivariate analyses of social factors to better understand the interplay of a variety of factors in the process of language shift.

\(^1\) An anonymous reviewer for Ravindranath and Cohn (2014) pointed out that despite evidence that larger languages of Indonesia are experiencing loss of speakers, there is a hesitancy to assign EGIDS values less than 6a Vigorous for languages with more than about half a million speakers. Moreover the EGIDS assessments do not include a measure of reliability or an indication of which assessments derive from thorough surveys and which are based on field-worker estimates. It is also likely that the most widely cited current speaker population estimates for local languages of Indonesia (i. e. those in the *Ethnologue 17*) are inflated, since these may reflect cultural rather than linguistic identification.
In this approach we follow Himmelmann (2010), who argues that our understanding of language shift will be best advanced by developing a typology of endangerment scenarios. Based on case studies of eleven Tomini-Tolitoli languages of Sulawesi in Indonesia, he writes that “endangerment is not defined with respect to a list of more or less disparate criteria such as number and age of speakers, homogeneity of the speech community, etc.” (Himmelmann 2010: 45) He points out that no single cause has been shown with certainty to cause language shift, and gives the example of migration. Although migration is often identified as a major factor in language shift (as in his own examples from Sulawesi), Himmelmann notes that there are also numerous examples of cases where immigrant groups have lived in an area for generations with no apparent loss of their heritage language, and gives the example of Javanese and Balinese transmigrants to other regions of Indonesia and Chinese trader families in Sulawesi villages, who have maintained their own languages for multiple generations. Thus, instead of identifying single causes of shift, he proposes that “language endangerment [should be] seen as the possible outcome of an endangerment scenario, i.e. a specific and complex constellation of varied factors, some of which may be conducive to language shift, others to language maintenance” (Himmelmann 2010: 45).

Our current project deepens our understanding of language shift scenarios. In this article, we introduce this approach to the study of large-scale language shift, employing census data available at IPUMS (Integrated Public Use Microdata Series, Minnesota Population Center 2014), a source that to our knowledge has not previously been used to address such questions. The barriers to entry are low, and the methods we use are easy to implement with existing statistical software. Thus our approach invites collaboration between linguists and other social scientists – demographers, sociologists, and political scientists, among others – to interpret language change alongside other, concomitant social processes.

Following work that views language choice as a sociolinguistic variable that is constrained by social factors (e.g. Gal 1978), we treat the use of Indonesian as a primary language as a sociolinguistic variant, and examine the various social factors that constrain this choice. Using Indonesia as a test case, we analyze census data for ten language communities in Indonesia, all with speaker populations of over one million, and investigate a range of social factors (including age,

---

2 Note that an approach such as this necessarily considers language communities (speakers of a language) rather than speech communities (smaller communities in regular contact that share sociolinguistic norms about language use (Labov 1972: 120)). We cannot be sure of the extent to which these overlap in our data.
gender, religion, ethnicity of parents, urbanization, and economic development) and their correlation with shift toward Bahasa Indonesia (Indonesian), Indonesia’s national language. While the effects of the different social factors are largely consistent with existing findings from qualitative studies, we find that their magnitudes are often surprising, and moreover that the effect of urbanization (which can only be examined in large scale studies such as this) amplifies the effect of the other social factors. Indonesia’s sociolinguistic make-up is complex but not exceptional, and similar processes of shift are likely to be taking place in large language communities throughout the world, in any country where one language is accorded official and/or institutional preference over other languages.

We argue for a methodological shift toward examining language shift scenarios more broadly and more quantitatively, for two main reasons: (1) because even a large speaker population does not protect against language shift and endangerment, as we find in this article that all of the language communities that we studied have decreasing numbers of speakers in apparent time; and (2) because large languages, with their diverse speaker populations, offer us an opportunity to quantitatively examine the social factors that contribute to shift and thus contribute to our understanding of the process itself. Our results provide a foundation for modeling the interaction of these factors with the goal of creating more predictive models of language shift.

2 Background

Indonesia (population: 258 million) is the second most multilingual nation in the world, second only to Papua New Guinea. The Ethnologue counts 706 living languages in Indonesia, which accounts for almost 10 percent of the living languages listed worldwide (Lewis et al. 2015). Although the total country population is also large, Indonesia still has a relatively high diversity index (.815), meaning that the likelihood that any two randomly chosen people speak different languages is very high.\(^3\)

\(^3\) That is, there are fewer speakers of every local language we studied in younger age groups than in older ones.


\(^5\) Greenberg’s diversity index is “the probability that any two people of the country selected at random would have different mother tongues. The highest possible value, 1, indicates total diversity (that is, no two people have the same mother tongue) while the lowest possible value, 0, indicates no diversity at all (that is, everyone has the same mother tongue).” (Lewis et al. 2015: 1, “Summary by country”).
Indonesia’s great linguistic diversity is, as with most post-colonial nation-states, a product of its particular and complex socio-history. The language we now know as Indonesian was established as the national language of Indonesia at its founding in 1945. Part of the motivation for this choice came from the fact that a related variety of Malay was at the time widely used as a lingua franca throughout the archipelago. It is currently the sole national and official language of Indonesia. Fishman describes the development and promotion of Indonesian as a “‘miraculous’ process whereby the population was ‘successfully convinced that a particular outside language should become their own integrative, inter-ethnic, unifying tongue’” (1978: 333, as cited by Sneddon 2006). Indeed, the spread of a local, erstwhile minority language as a national language and lingua franca is remarkable, and possibly unparalleled in modern times. Certainly there is no other formerly colonized nation that has promoted a single, non-European language as a national and official language with as much success as Indonesia has, and the results of this policy decision are widely seen to be very positive in terms of unity and nation building (Anwar 1980; Anderson 1991; Dardjowidjojo 1998; Sneddon 2003a; Collins 2004).

Perhaps not surprisingly, the successful spread of Indonesian has led to its increasing use in multiple domains, often at the expense of local languages (Florey 2005, 2010), and it is increasingly apparent that language shift toward Indonesian is affecting all of the local languages of Indonesia, even those with speaker populations in the millions (Anderbeck 2015; Cohn and Ravindranath 2014, and references cited therein). The top three regional/local languages – Javanese, Sundanese, and Madurese – all spoken in Java, account for over half of the population of Indonesia and are of particular interest, given the observation that “[i]n spite of their large speech communities, the Javanese, Sundanese, and Madurese languages are actually endangered in that some of their domains of usage are being taken over by Indonesian, and, to a lesser extent, in that they are not always passed on to the next generation” (Adelaar 2010: 25). To better understand these developments within large language communities that are rarely studied from the perspective of language shift, in this article we focus exclusively on languages with speaker populations of one million or more.

What we refer to as “local” languages are all of the other 700 plus languages of Indonesia that are not Indonesian. Most of these languages (excluding a

---

6 One possible reason for the success of Indonesian is that at the time of its adoption Malay was already an inter-ethnic lingua franca that was not tied to any particular ethnic group. One reader points out that this pre-modern history may even account for today’s urban-rural divide in the use of Indonesian, since as a trade language use of Malay may have taken root in the cities earlier than in the rural areas.
number of small language families, language isolates, and creoles) are members of the Austronesian and Trans-New Guinean language families. All of the Austronesian languages are Malayo-Polynesian; around 30 of these are Malayic languages that are most closely related to (and sometimes mutually intelligible with) Indonesian (Lewis et al. 2015). Moreover, both Errington (1986) and Sneddon (2003b) argue that Indonesian exhibits a clear case of diglossia, with different recognizable varieties of Indonesian (from formal, written Indonesian to colloquial spoken Indonesian) existing along a continuum. Therefore, a complex issue that we do not attempt to resolve here is what exactly is meant by “Indonesian” in speakers’ interpretation of the census question. We do not (because we cannot) distinguish between colloquial spoken varieties of Indonesian and more formal varieties, nor can we distinguish between local varieties of Malay and Malayic languages and Indonesian (but see Adelaar 1992). Thus for methodological and practical purposes we focus in this paper on non-Malay or Malayic local languages in Indonesia, and in contrast treat “Indonesian” as one variety.

3 Data

As a nation, Indonesia includes some of the most densely and least densely populated places in the world. This means that our sample contains respondents living in large urban areas, smaller urban areas, and traditional rural communities. Moreover, Indonesian census data contain rich information on important demographic and socioeconomic characteristics that have been made available to researchers. All of these facts make Indonesian an excellent test case for a multivariate approach to the study of language shift. In this analysis we have used data from the (2010) census only, as expert evaluations of this census consider it to be particularly accurate in terms of coverage (see Hull 2010). See Steinhauer (1994) for analysis of language use in Indonesia using census data from 1971, 1980, and 1990, along with discussion of many of the same social factors we discuss here.

As discussed above, the language communities that we chose for analysis are those languages listed in *Ethnologue 17* with at least a million speakers that are not varieties of Malay or Malayic. By choosing those languages that are distinct structurally and in terms of nomenclature, we avoid the difficult task of interpreting whether speakers in those communities consider local varieties of the national language to be “Indonesian” or some other local variety.

---

7 The numbers in *Ethnologue 17* are based on (2000) census data.
We are left with ten language communities ranging in speaker population from 1 million to 84 million, as summarized in Table 1 (rank refers to rank as defined in Lewis et al. [2013: 142-143]). In general, one language corresponds to a single ethnic group that speaks that language, and there are one or more provinces where that language is predominantly spoken as a majority local language. The exceptions to this pattern are Batak language varieties (Northern Batak, Southern Batak, and Simalungun Batak), which form one broadly defined ethnolinguistic group\(^8\) despite being three non-mutually intelligible languages, and Buginese, which serves as a regional lingua franca in South Sulawesi. Methodologically, the selection of these ten groups was advantageous given restrictions on the microdata that have been made available to researchers: the Indonesian census asks respondents what language they speak at home, but the public-use microdata only include whether the language is Indonesian or a “local language”. With the selection of these ten ethnolinguistic groups we can assume that if a respondent reports speaking a local language, then if he or she is a member of the majority ethnic group living in that ethnic group’s province of origin, that local language is the

Table 1: The ten largest local languages in Indonesia, following *Ethnologue* 17.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Language</th>
<th>Ethnic group</th>
<th># in M</th>
<th>Province</th>
<th>EGIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Javanese</td>
<td>Java</td>
<td>84.3</td>
<td>Central Java, E Java</td>
<td>2 Provincial</td>
</tr>
<tr>
<td>2</td>
<td>Sundanese</td>
<td>Sunda</td>
<td>34</td>
<td>W. Java</td>
<td>5 Developing</td>
</tr>
<tr>
<td>4</td>
<td>Madurese</td>
<td>Madura</td>
<td>6.7</td>
<td>Madura, E Java</td>
<td>5 Developing</td>
</tr>
<tr>
<td>15–18</td>
<td>Batak</td>
<td>Batak</td>
<td>5.5</td>
<td>N. Sumatra</td>
<td>5 Developing/6a Vigorous(^a)</td>
</tr>
<tr>
<td>7</td>
<td>Buginese</td>
<td>Bugis</td>
<td>5</td>
<td>S. Sulawesi</td>
<td>3 Wider com</td>
</tr>
<tr>
<td>9</td>
<td>Acehnese</td>
<td>Aceh</td>
<td>3.5</td>
<td>Aceh</td>
<td>5 Developing</td>
</tr>
<tr>
<td>11</td>
<td>Balinese</td>
<td>Bali/Bali Hindu</td>
<td>3.3</td>
<td>Bali</td>
<td>5 Developing</td>
</tr>
<tr>
<td>13</td>
<td>Makassarese</td>
<td>Makassar</td>
<td>2.1</td>
<td>S. Sulawesi</td>
<td>6b Threatened</td>
</tr>
<tr>
<td>14</td>
<td>Sasak</td>
<td>Sasa</td>
<td>2.1</td>
<td>Lombok WNT</td>
<td>5 Developing</td>
</tr>
<tr>
<td>19</td>
<td>Gorontalo</td>
<td>Gorontalo</td>
<td>1</td>
<td>Gorontalo, Sulawesi</td>
<td>6b Threatened</td>
</tr>
</tbody>
</table>

Note: \(^a\)As discussed in the text, we combine the three major linguistic varieties of Batak and the several ethnic subgroups. All of the linguistic varieties of Batak are listed in the *Ethnologue* as 5 Developing, except for Batak Mandailing that is listed as 6a Vigorous.

\(^8\)These are coded in the Indonesian Census as consisting of the following ethnic subgroups: Angkola, Toba, Mandailing, Tapanuli, Karo, Dairi, Simalungan.
ethnic language. All provinces discussed here also include a major urban area, allowing us to compare urban and rural speakers.

Access to IPUMS for academic researchers opens up the possibility of using census data for the study of language shift in those cases where relevant questions about daily language use are included in the available data. In the case of Indonesia, one or more questions about language use have been included in the last several censuses (although question wordings have not been consistent). In the (2010) census the question asked was “Apakah bahasa sehari-hari yang digunakan (NAMA) di rumah?” [What language does (NAME) use daily at home?] From a linguistic perspective, the oversimplifications are many: this item provides no information about bilingual or multilingual language use, and it cannot classify whether respondents are speaking “real” Indonesian or a local variety of Malay (see work by Errington [2014] on “middle Indonesian”). Yet with the caveats above, for our broad purpose of modeling the predictors of language use, this is a reasonable proxy for whether individuals speak Indonesian or a local language.

4 Methods

From a sociolinguistic point of view, there are three main problems with the use of census data to determine trends in language maintenance and shift. First, when using a single census to study the correlation between age and language use, it is possible that as individuals age their language use patterns may change, their answers to questions about language use may change, and the age-related differences that are apparent in a single survey may be reflective of differences between age groups that do not represent community change over time (see Lieberson 1972, 1980).

The second and third problems arise when using data from multiple censuses as an indicator of real time change. First, it is quite common for census questions to change over time in such a way that respondents’ answers from different censuses cannot be reliably compared. This problem is described by Mesthrie et al. (2009: 381), with the example of significant changes to language-

---

9 For example, we code ethnic Javanese living in the provinces of Central Java, East Java, and Yogyakarta as speaking Javanese if they report speaking a local language. We do not analyze speakers of local languages who live outside of their ethnic group’s province of origin (for example, ethnic Javanese living in Jakarta or outside of Java).
related questions in the Australian census over the course of the twentieth century. This is indeed the case for the Indonesian census data.

Additionally, changes in the socio-historical context in which the censuses are administered may lead to changes in the answers given to similar or even identical questions. This is represented by the well-known situation described by Fishman et al. ([1985], also reported in Mesthrie et al. [2009: 382]), for US Census data collected in 1960 and 1970. Between these two censuses the general population only rose by 13 percent, but the percent of non-English mother tongue claims rose by 70 percent. This is attributed to a change in the socio-historical context of the 1960s, specifically the formation of strong ethnic identities in this time period. All of these problems are addressed by Steinhauser (1994), with analysis and discussion of the (1971), (1980), and (1990) census data. His discussion can offer some assurance that our findings with respect to age reflect change over time, despite the fact that these data are still subject to the second problem, as different questions about language use were asked in each census. In particular, his data clearly show an increase in the number of speakers using Indonesian, but it is more difficult to interpret concomitant patterns of decreasing use of local languages.

Finally, inferences about language shift over time drawn from comparisons across age groups may be confounded by particularities of census cohorts (Glenn 1977): the probability of speaking Indonesian for a 70-year old in 1960 will differ from the probability of speaking Indonesian for a 70-year old in 2010. This is because a 70-year old in 2010 will have lived through more than six decades of living with Indonesian as a national language (s/he would be have been nine years old at independence), whereas a 70-year old in 1960 will have been 59 years old at independence. It is possible to address age-cohort-time confounding by combining multiple census waves, an exercise that we reserve for future research. In our application, we restrict our conclusions about age effects to capturing synchronic variation across individuals in 2010.

We adopt a flexible empirical strategy to model Indonesian language use. We first define our dependent variable Indonesian as a binary variable that takes the value of 1 when the respondent reports speaking Indonesian at home, and 0 when s/he reports speaking another “local” language at home. This choice excludes speakers of foreign languages such as English, Dutch, or Japanese from our analysis (it is not immediately clear if Chinese languages or Arabic are

---

10 In 1995, for example, respondents were given the choice of “Indonesian”, “Local Language (specify) __ __”, and “Foreign Language”. In 2010, respondents were asked to name the language that they “usually” spoke at home, with no prompt for Indonesian or distinction between local and foreign languages.
excluded under this definition). We emphasize once again that we do not know with certainty which local language the respondent speaks, nor do we have the means to gauge fluency, linguistic competence, or dialectal variation.

We use logistic regression to model the probability that a speaker reports speaking Indonesian as a function of several variables that are also found in the census.\(^\text{11}\) We denote these variables as \(X\), noting the elements of \(X\) will vary across the models that we estimate to reflect different hypotheses that we wish to test. We can express the logistic regression formula as

\[
\Pr(\text{Indonesian} = 1) = \frac{1}{1 + e^{-(a + bX)}}
\]

Here, \(a\) and \(b\) are parameters to be estimated, and the nonlinear functional form constrains the probability that a respondent reports speaking Indonesian to lie between 0 and 1, as is appropriate when modeling binary responses as probabilities. A list of all census items that we employ in our analyses, in the full original Indonesian version that appears in the census instrument and accompanied by English glosses and coding choices, appears in Appendix Table S1.

The results of the logistic regression models are not easily interpretable: whether reported as log-odd or odds ratios, they summarize only the relative probabilities of speaking Indonesian for marginal changes in the predictors. Another problem that we face is the sheer abundance of data. With millions of observations, we have an immense amount of statistical power, and can estimate substantively tiny effects with a high degree of precision. This renders standard tests of statistical significance uninformative, as significance is a function of both sample size and effect size.

We confront both of these issues through a graphical approach that allows us to visualize the complex predicted probabilities that we estimate (see King et al. 2000). We transform the logistic regression results into predicted probabilities for theoretically meaningful combinations of the predictor variables, and then plot them alongside one another. In the Indonesian case, a respondent’s age will always play a central role in the likelihood that s/he speaks Indonesian, so we use this as an illustrative heuristic device through which to explore these interactions. As an illustration of the power of our graphical approach we start with a simple model that includes age, gender, and urban/rural as predictors of

---

\(^{11}\) We choose logistic regressions instead of the commonly used chi-square test to compare the probability that different groups speak Indonesian because chi-square tests are less readily extended to multivariate contexts and analysis of continuous variables than are regression-based tests. See Aldrich and Nelson (1988).
language use. These data come from a 1 percent sample of the (2010) Indonesian census, and include 2,111,288 respondents from across the country.\textsuperscript{12} We display our results in Figure 1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Baseline model results: age, gender, and urban/rural as predictors of language use. The red line (almost completely obscured by the blue line) is our estimate of the probability that a female respondent at any particular age speaks Indonesian at home. The blue line provides the same estimate for male respondents. Shaded areas are confidence intervals.}
\end{figure}

The Y-axis contains the predicted probability that a respondent speaks Indonesian for particular combinations of independent variables. We can learn an immense amount of information from this figure. For example:

1. Urban men and women are \textit{far} more likely to speak Indonesian than rural men and women.
2. Even at the maximum, 10- and 20-year olds still have less than a 40 percent chance of speaking Indonesian at home.
3. The two lines representing predicted probabilities for women and men overlap nearly completely for both urban and rural respondents. This indicates that women and men do not differ much in their probability of speaking Indonesian (although in the logistic regression results these estimates \textit{are} statistically significant).
4. Age effects are larger in urban respondents.

\textsuperscript{12} With such a large sample, all possible combinations of covariate values are represented.
5. 70-year old urban residents are still twice as likely to speak Indonesian as 20-year old rural residents.

With this simple model as a point of departure, we turn to our full results, analyzing a series of sociolinguistic factors that have been shown to correlate with language shift in smaller community studies, treating language choice as a sociolinguistic variable that is constrained by different social factors.

5 Results

The primary social factors that we consider are age, urbanization, socioeconomic development, education, religion, and gender. The choice of these social factors to investigate comes from previous work at the speech community level on the sociolinguistics of language shift and endangerment (e.g. Gal 1978; Dorian 1981), particularly in the Indonesian context (Errington 1998; Kurniasih 2006; Smith-Hefner 2009; Florey 2010; Setiawan 2012), and we discuss our results with respect to those studies and more general findings in the sociolinguistic literature on social factors. We also consider the predictive value of parent’s first language on the first language of their children, as a measure of the success of intergenerational transmission (Fishman 1991).

We begin our discussion with the baseline model, introduced above, that includes age (analyzed both as age and age squared),13 gender, and urban/rural as predictors of Indonesian language use. We include these three predictors in every empirical model that we estimate, so it is useful to begin by analyzing these in more depth. The full logistic regression results for this model appear in Table S2 in the Appendix. Then, in examining the individual social factors more closely we discuss urbanization, development, and education first, as we find that these results are broadly consistent across the ten ethnic groups that we consider. We believe that it may be possible to generalize these findings more widely across different communities outside Indonesia as well. We go on to discuss gender and religion, factors that we argue are not likely to be generalizable across communities in the same way but that nonetheless provide interesting results with respect to the nature of language shift in Indonesia.

13 We include both age and age squared in every model because we suspect that there is not a simple linear relationship between respondents’ age and the probability that they speak Indonesian. In results not reported here, we explored cubic and quartic polynomials to allow for even more complex nonlinear effects of age, but they do not appreciably change our substantive conclusions.
5.1 Age

As discussed above, surveying speakers from across the age spectrum can show the relationship between synchronic variation and diachronic change. When only very limited real time data is available to examine diachronic change, sociolinguists rely on what is known as the apparent time construct (Bailey et al. 1991) to surmise the presence of change in progress, taking age as a proxy for the passage of time and concluding that differences across generations of adults from a single speech community mirror actual diachronic developments. As applied to situations of language shift, a decreasing probability of speaking local languages (or increasing probability of speaking Indonesian) across age groups is consistent with the occurrence of language shift. Moreover, measures of language vitality place primary importance on age in that a lack of children speakers is considered the primary predictor of language morbidity (see for example Fishman’s [1991] GIDS and Lewis and Simons [2010] EGIDS). And indeed, the data show a clear correlation between speaker age and use of Indonesian, as seen in Figure 1.

5.2 Urban vs. rural

Although numerous studies of language shift in Indonesia and elsewhere find evidence of shift toward use of the national language or local varieties of Malay even in rural contexts (e.g. Grimes 2010; Himmelmann 2010), we nonetheless expect to find more use of the national language in urban communities than in rural ones, for a variety of reasons not particular to the Indonesian context, including the use of the dominant language as a lingua franca in interethnic communication (as discussed by Batibo [2005] for Africa), the use of the dominant language among more highly educated urban groups, and real or perceived economic advantages to the use of the dominant language (Harbert 2011).

In the modern Indonesian context, Setiawan (2012) examines urban-rural differences in three Javanese communities (city, town and village) and finds that urban children are far more likely than village children to report using Indonesian as their first language. Smith-Hefner (2009) proposes that in fact it is urban educated youth who are actively leading a change toward increased use of the national language, writing that “[i]n Yogyakarta and elsewhere, it is urban, educated youth who have been the most active proponents of the movement towards the greater use of Indonesian” (Smith-Hefner 2009: 58, citing Samuel 2000). Indeed, we see a marked distinction between urban and rural Indonesians in their propensity to use Indonesian in Figure 1.
We expect that these results reflect the effects of: (1) the use of Indonesian as an interethnic lingua franca; (2) differences between urban and rural residents in terms of level of education and access and exposure to mass media, and (3) economic differences between urban and rural residents and differences in the real or perceived economic value of mastery of Indonesian. Historically Indonesian and local varieties of Malay have long been used as interethnic linguae francae. Since there is much greater ethnolinguistic diversity and mixing in urban areas, even in small cities, this is likely to result in more frequent use of Indonesian or a local variety of Malay. Indeed, in the city of Jakarta we find that of 83,764 respondents, only 8.25 percent report speaking a local language, and the remaining 91.75 percent report speaking Indonesian. Although a few studies discuss the use of regional languages in interethnic interactions to indicate familiarity, most earlier generalizations describe Indonesian (and sometimes explicitly not regional languages) as the language of wider communication between different ethnic groups (Goebel [2002] and work cited therein). Finally we expect that it may be the case that in urban settings there is a greater perceived economic and socio-economic value of mastery of the dominant language (see Harbert 2011) and moreover that in urban areas the value of the dominant language in the linguistic marketplace (Bourdieu 1977) is also greater.

5.3 Development index, education, and family

Social class as a factor has been increasingly problematized in sociolinguistic studies (cf. Dodsworth 2009), largely because it relies upon a social construct that may not have universal applicability. One response, often used by sociolinguists working in non-Western communities, is to depend on emic social categories, since it may not be relevant to impose etic class categories on their samples (see Stanford and Preston 2009). Himmelmann (2010: 57), for example, uses “time spent in the garden” as a measure of socioeconomic class in a Sulawesi village, since more well-to-do families may hire labor and therefore spend less time in family gardens. A second response to this problem is to include some type of index of socioeconomic status, like the linguistic market index introduced by Sankoff and Laberge (1978) or the socio-economic index used by Thibault and Vincent (1990) in their Montréal studies.

We follow these and other studies to create an index of socioeconomic development – which we term a “development index” (abbreviated as DEV) – that captures salient features of respondents’ material conditions. The development index is an additive index of eight items included in the census, including whether the respondent: (1) owns their own home; (2) has electricity; (3) has running water; (4) has a sewer; (5) has a flush toilet; (6) has something other than a dirt floor (e. g.
cement, wood, etc); (7) owns a (non-mobile) phone; (8) owns a mobile phone. This index is not a proxy for class, income, or social status, but it will allow us to array respondents along a continuum of material “development” using indicators that are valid in the Indonesian context. The index runs from 0 to 8, with 8 being the highest level of material development. We present the results in Figure 2, plotting only DEV = 1, 3, 5, and 7 to avoid cluttering the figures, again divided by rural/urban and plotted by age. (In Appendix Table S3 we provide full logistic regression results for each individual indicator as well as the summary index.)

![Figure 2: Development and language use.](image)

The red line is our estimate of the probability that a respondent at DEV = 1 speaks Indonesian at home, for each age group. The green line provides the same estimate for respondents at DEV = 3, the blue line for DEV = 5, and the purple line for DEV = 7. Shaded areas are confidence intervals.

The predictive effects of development on Indonesian language use are readily apparent in this figure. The effect across the continuous scale is not linear, with moderate differences between DEV = 1–5 and a noticeably larger effect of DEV = 7. And interestingly, while higher scores on our development index predict higher likelihoods of speaking Indonesian in both urban and rural settings, the substantive effects are much larger in urban areas and the effect of age is also much greater in urban than rural areas. These effects are substantively meaningful: reading the graph in Figure 2 we see, for example, that a twenty-year old urban Indonesian is more than twice as likely to speak Indonesian at DEV = 7 than at DEV = 1 (40 percent versus 17 percent).

We also consider education, with similar results (see Appendix Table S4, Model 1 for regression results). Not surprisingly, especially since formal
education is mandated to be in Indonesian across Indonesia, we find in Figure 3 that there is a clear correlation between level of education and use of Indonesian: across all levels of our education scale, higher levels of education predict a higher probability of speaking Indonesian. That there are effects beyond primary and secondary education is striking. Again, these effects are substantially meaningful: reading the graph in Figure 3 we see, for example, that a thirty-year old urban Indonesian with a post-graduate education has a 65 percent chance of speaking Indonesian, compared to 28 percent for someone with only a junior high school education. Differences at the primary level might simply reflect increased exposure to Indonesian in the classroom. However, this is unlikely to be the explanation for the effect at the highest levels of education. Rather, these effects may instead reflect perceived social, economic and cultural values of using Indonesian that accompany higher education.14

A third factor that has been proposed as an alternative to social class and education in sociolinguistic studies of shift is social network (Aikhenvald 2002; Li 1994; Milroy and Milroy 1978; Zentella 1997). It is clearly not possible in the present study to acquire the type of information about an individual’s social network that these authors use in order to examine the effect of interpersonal relationships on language choice. Nonetheless, it is possible to address questions such as these using census data by determining, for instance, whether an individual lives with a nuclear or extended family and thus has more or fewer regular contacts in different age groups.

**Figure 3**: Education and language use.
The blue line is our estimate of the probability that a respondent with no education speaks Indonesian at home, for each age group. The additional lines reflect increasingly higher levels of education. Shaded areas are confidence intervals.
5.4 Religion

Indonesia is a multi-religious state, and all Indonesians are required by law to be affiliated with one of six religions: Islam, Hinduism, Buddhism, Protestantism, Catholicism, or Confucianism. Islam is the majority religion, with 87 percent of our sample coded as Muslims. A further 10 percent are Christians (we collapse the official distinction between Protestants and Catholics because many respondents are coded simply as “Christians”, which likely refers to Protestants but need not), followed by 1.7 percent Hindus, 0.8 percent Buddhists, and the rest coded as “Other”.

Prior work suggests that religion is an important social factor in language choice in Indonesia, albeit one that is determined by other, local factors. Smith-Hefner (2009), for example, finds that Muslims in Central Java report that marrying another Muslim is more important than marrying another Javanese, so speaking Indonesian offers greater social mobility for educated young Muslim Javanese women. John Bowden compares the role of religion in his own work in the Maluku islands to the role of religion in Florey (2010), suggesting that the role of religion may have more to do with other social forces than religion per se. He gives the example of parts of North Maluku, where riots led to Muslim groups on the west coast fleeing their traditional territory and effectively ending successful transmission of the local language, while the Christian children who remained in the villages continued to use the local language (Bowden, personal communication). Steinhauer (1994: 772) gives the example of Dayaks who shift to the use of Banjarese, the language of Muslim traders and government officials in Kalimantan, as a sign of their total conversion to Islam. We expect that religion plays an important role in language shift, but that the relationship between any particular religion and language choice will not be deterministic.

We find very strong evidence of differences in Indonesian language use by religion, and these are especially pronounced in urban areas as shown in Figure 4 (regression results appear in Appendix Table S4, Model 2).

Christians and especially Buddhists are more likely to speak Indonesian than Muslims. Hindus, on the other hand, are much less likely to speak Indonesian than Muslims. This likely reflects the fact that Hinduism is concentrated in one particular ethnic group, the Balinese. In our data, 94.9 percent of Balinese are Hindu, and 92.6 percent of Hindus in Indonesia are Balinese. Given that the majority of Balinese live on the island of Bali, where they form a large majority, we suspect that low use of Indonesian by Hindus at least partially reflects the particular geographic concentration of Hindus. We will return to this point in greater detail below when we address differences by ethnic group, which we can address systematically given the richness of our data.
Unlike the case of Hinduism in Bali, we cannot attribute the relatively high probability of Christians speaking Indonesian to any particular ethnic group. Rather, this may be the result of a kind of national “identity in formation” among a relatively geographically dispersed religious minority. In the case of Buddhism, ethnic factors might play a role, this time among Chinese Indonesians. Speaking and writing in all Chinese languages was forbidden for decades under the authoritarian New Order regime (1966–1998), and therefore Chinese Indonesians were uniquely forbidden from speaking their own “local” language. And Chinese Indonesians are particularly likely to be Buddhists, representing 79.3 percent of all Indonesian Buddhists.¹⁵ Unlike other ethnic

---

¹⁵ The Chinese factor almost certainly does not explain our results for Christianity, as Chinese form only a small minority – 4.9 percent – of Indonesia’s Christian community as a whole. This does not rule out the possibility that in some parts of Indonesia Chinese Christians influence the relationship between Christianity and speaking Indonesian, since for some parts of Indonesia Chinese are a relatively larger fraction of the Christian population (see Soleiman and Steenbrink 2008 for a discussion). One example is East Java, where Chinese are 21 percent of the Christian population. In provinces such as North Sumatra, Maluku, and East Nusa Tenggara, Chinese comprise less than one percent of the Christian population. On the whole, 84.43 percent of ethnic Chinese Indonesians report speaking Indonesian. This is similar for Christians and Buddhists, and is especially pronounced among those Chinese Indonesians at highest levels of our development index (which is the vast majority of all Chinese).
groups that we will examine separately below, however, Chinese Indonesians do not have a particular “home” region, so we will not be able to examine them any more closely than we have so far.

5.5 Gender

In the sociolinguistic literature numerous studies have shown women’s preference for standard or prestige varieties (Labov 1990). This is often interpreted as women exploiting the social capital of language in social environments where other types of capital are less available to them, and a handful of studies have shown similar patterns in language shift scenarios, with women leading the shift toward the dominant or prestige variety. Gal (1978), for example, shows women leading the shift away from Hungarian in a formerly Hungarian-speaking village in Austria. Most relevant to this study, both Kurniasih (2006) and Smith-Hefner (2009) document the preference of women in Yogyakarta for Indonesian over Javanese (see Cohn and Ravindranath [2014] for discussion). Our expectations were to find the same effect of gender in our data.

At the macro level, we do not find strong evidence of differential use of Indonesian by gender (see Figure 1), and to the extent that we do, we do not find women leading in use of Indonesian, as we might expect. Rather, we find that women are less likely to speak Indonesian than men, though this difference is substantively very small. Our regression results in Table S2 (see Appendix) indicate an odds ratio of .990, and our graphical results above show that probabilities of speaking Indonesian are quite similar. Given that we have two million observations, the statistical significance of our result for gender is relatively modest compared to both age and urban/rural. Indeed, in some of the results in Table S4 (Appendix) that control for additional factors such as development and parent language, the coefficient on gender fails to cross conventional benchmarks of statistical significance. As measured by either substantive or statistical significance, the predictive effects of gender are miniscule.

We find this result to be at the same time surprising and unsurprising. Although community level studies may show an effect of gender, it seems likely that these effects may not be a result of gender per se, but of gender differences in social networks, access to education, and exposure to Indonesian. This follows from previous work on gender and language change, where gender patterns are reflective of other aspects of social life (e.g. in Bakir 1986; Dubois and Horvath 1998; Haeri 1994; Zentella 1997). We find that our results underscore the fact that gender should not be treated as a determiner of language
choice but that gender patterns are more likely to be reflective of gender differences in other aspects of social life (cf. Eckert and McConnell-Ginet 2013).

5.6 Parents’ language

Where we do see clear gender effects are when we consider the question of intergenerational transmission, and how the respondent’s primary language is related to the primary language of their parents.

We show in Figure 5 that if both parents speak Indonesian at home, the probability of speaking Indonesian is almost 1, regardless of age or gender (regression results are available in Appendix Table S4, Model 3). If neither parent speaks Indonesian, the chances of speaking Indonesian are very small. The interesting cases are when only one parent speaks Indonesian. Having only a mother who speaks Indonesian raises the probability of speaking Indonesian substantially more than having only a father who speaks Indonesian. This finding is consistent with more general observations of greater influence of mother’s speech on children’s language acquisition. This could be due in part to greater contact between the mother and children in the most cultural settings. But it also points to the profound role mothers play in modeling of linguistic behavior whether for male or female children. Noteworthy in Kurniasih’s (2006) findings is that the mothers’ regularly use local languages with their peers and

![Figure 5: Parents' language and language use.](image-url)
elders, while their linguistic interactions with their children show the shift to Indonesian, which is in turn adopted by the children.

The purple line is our estimate of the probability that a respondent whose mother and father both speak a language other than Indonesian speaks Indonesian at home, for each age group. The red line is the same estimate for respondents whose parents both speak Indonesian at home. The green line captures respondents whose mother speaks a local language but whose father speaks Indonesian, and the blue line captures respondents whose father speaks a local language but whose mother speaks Indonesian. Shaded areas are confidence intervals.

We also find interesting non-linear effects by age. The effect of having one parent who speaks Indonesian is actually higher among the oldest Indonesians than it is for middle-aged Indonesians. This might reflect the residue of the uneven spread of Indonesian as a national language: those families where even just one parent spoke Indonesian four decades ago were probably particularly cosmopolitan or nationally-oriented at a time when this was still relatively rare.

We also can check for differences across genders in the effect of parents’ languages, to test the hypothesis that perhaps the effects of mother’s language are stronger for girls than for boys as suggested by both Kurniasih’s (2006) and Smith-Hefner’s (2009) findings. We do this in Figure 6.

![Figure 6: Parents' language and language use, by gender.](image)

The red line is our estimate of the probability that a female respondent at any particular age speaks Indonesian at home. The blue line provides the same estimate for male respondents. Shaded areas are confidence intervals.
The results provide no evidence that the effects of parental language differ between male children and female children.

5.7 Explaining Indonesian use across ethnic groups

The above results pool all speakers from these ten linguistic groups into a single sample, but we can also explore differences across groups. First, we can ask what explains differential use of Indonesian across ethnic groups? We consider here two explanations. First, although we have argued in this paper that a large speaker population does not protect against language shift, the size of the ethnic group should play a role in how likely its speakers are to be shifting to Indonesian. Second, given the importance of urbanization in predicting shift generally, we consider the level of urbanization across ethnic groups.

To test these claims, we identify members of an ethnic group who live in that ethnic group’s province of origin, which allows us to assume that their own ethnic language is the “local” language that they speak if they do not speak Indonesian. We then calculate the predicted probability that a twenty-year old from each ethnic group speaks Indonesian using our baseline model. This probability ranges from just 0.0026 (95 percent CI = 0.0024, 0.0028) for rural Javanese to 0.815 (95 percent CI = 0.7971, 0.8317) for urban Gorontalo. Finally, we compare that to the total number of respondents in that ethnic group. The results of this appear in Figure 7.

![Figure 7: Ethnic group size and predicted Indonesian language use.](image-url)
We see a strong negative correlation between ethnic group size and Indonesian language use. The correlation between the natural logarithm of ethnic group's size and the probability that a twenty-year old urban resident from that group speaks Indonesian is $-0.53$, with a p-value of 0.11 despite there being only ten observations. For rural Indonesians, the correlation is $-0.53$ with a p-value of 0.12. This finding lends support to the claim that smaller languages are currently at greater risk than larger ones (following e.g. Krauss [1992]).

Another possibility is that different levels of urbanization across ethnic groups explain differences in language use. Figure 8 repeats the above exercise, using the percentage of the ethnic group residing in an urban area in an attempt to explain intergroup variation in the propensity to speak Indonesian. The results do not support an argument that differential patterns of urbanization across ethnic groups explain differences in use of Indonesian. The correlation between urbanization and the probability that a twenty-year old urban resident from that group speaks Indonesian is $-0.33$ (p = 0.36). For rural Indonesians, the correlation is $-0.27$ (p = 0.44).

![Figure 8: Urbanization and predicted Indonesian language use.](image)

Of course, these are only two preliminary explorations for the patterns that we identify, and we emphasize that by looking only at large languages, we will underestimate the effects of language size on language shift (if small languages are most likely to be endangered) or of urbanization on language shift (if small
languages are most likely to be found in rural areas). Nevertheless, these results help to more fully characterize the interaction between individual and group characteristics and language shift in Indonesia.

5.8 Other ethnic differences

Beyond differences in the prevalence of Indonesian across ethnic groups, Indonesia’s ethnic diversity also provides us with an opportunity to examine how the social and demographic factors that we have identified vary across groups. In Figure 9 we examine the urban-rural split by ethnic group, ordered left-to-right by total number of speakers.

![Figure 9: Urbanization by ethnic group (population in parentheses). The red line is our estimate of the probability that a rural respondent at any particular age speaks Indonesian at home. The blue line provides the same estimate for urban respondents. Shaded areas are confidence intervals.](image)

In every case we find that urban Indonesians are more likely to speak Indonesian than rural Indonesians. But more importantly, we find dramatic differences in the effect of urbanization by ethnic group. The probability of speaking Indonesian for Indonesians drawn from the central region of Java and Bali (Javanese, Sundanese, Madurese, and Balinese) is very low even for urban residents, and the same is true on the island of Lombok. Probabilities are
far higher for smaller ethnic groups from the outer islands, and this is especially true of urban residents.

Our analysis also reveals the fragility of Gorontalo as a language: we estimate that there is 75 percent chance that a ten-year old, rural, ethnic Gorontalo respondent living in Gorontalo province speaks Indonesian. This figure is unmatched by the other large ethnic languages we study here. In contrast to Batak, Bugis, and Makassar, which are vibrant in rural areas if rarely spoken at home in urban areas, we reveal that Gorontalo is distinctly threatened among all speakers. Our quantitative approach can be used in this way to create a simple and objective statistic that summarizes the degree to which any language is endangered, a topic which we plan to explore more fully in separate research.

Next we turn to religion. Recall that in Figure 4, we provided evidence that Christians and Buddhists are more likely to speak Indonesian, and Hindus were less likely to speak Indonesian, relative to Indonesian Muslims. In Figure 10 we repeated this analysis by ethnic groups. One important difference, however, is that we cannot estimate effects by religion for every ethnic group due to the complete non-overlap of religion and ethnicity in certain cases (there are no Christian Acehnese, Buddhist Madurese, and so forth). We therefore restrict the analysis to religions in which there are at least fifty respondents in the sample for each particular ethnic group. It is also important to highlight that although the vast majority of Indonesians are Muslims, many ethnic groups are predominantly non-Muslim, a point to which we return below. Among our ten ethnic groups, Islam is the majority religion for eight: Javanese, Sundanese, Madurese, Sasak, Acehnese, Bugis, Makassar, and Gorontalo. Batak are majority Christian, and Balinese are majority Hindu.

We see here a very interesting pattern that the predictive effect of religion depends on whether the religion is that ethnic group’s majority religion. Among Muslim-majority ethnic groups, non-Muslims are more likely to speak Indonesian. Among Balinese (majority Hindu), non-Hindus are more likely to speak Indonesian. Among Bataks, where Christians form a small majority, Muslims are more likely to speak Indonesian. The only exception is Bugis.

16 We leave it for further work to discuss in more depth why Gorontalo is more progressed in language shift than the other languages (see also Anderbeck 2015). One obvious fact is that Gorontalo has a smaller speaker population than the other languages, but this is not the only difference. In contrast to the other languages, Gorontalo is notable for not having a written literature. Other sociolinguistic factors are also probably at play, as Anderbeck (2015: 20) notes that “[i]ronically, since Gorontalo split from North Sulawesi province in 2000, many Gorontalos no longer feel the need to assert their distinct identity by using their language.”
among whom Hindus are less likely to speak Indonesian than the majority Muslim community—but this may reflect the fact that the category “Hindu” on Sulawesi captures indigenous religious practices whose origins predate the arrival of both Islam and Christianity. These findings are consistent with our broader argument that religious minorities are more likely to speak Indonesian as a form of minority identity in formation; the analysis from these subgroups suggests that this process also operates at the province level as well.

6 Discussion and conclusions

We have used Indonesia as a test case for a quantitative approach to modeling language shift based on census data. Indonesia is a particularly good test case

Figure 10: Religion by ethnic group.
The green line is our estimate of the probability that a Muslim respondent speaks Indonesian at home, for each age group. The blue line provides the same estimate for Buddhists, the red line for Christians, and the purple line for Hindus. Shaded areas are confidence intervals.
for the use of these methods, given its large population, its linguistic diversity, and the range of variation it exhibits in terms of types of language communities. Moreover, our findings with respect to age, and more importantly, the effect of parent’s language on children’s language (two indicators of the success of intergenerational transmission) suggest that language shift away from the large local languages of Indonesia is indeed underway. Similar processes of shift are likely to be taking place in other large language communities throughout the world, in any country where one language is accorded official and/or institutional preference over other languages. As the methods we discuss here are easily replicable for any country whose census data is available in IPUMS, we encourage others to exploit these opportunities for analysis of language shift scenarios in other places.

Although we find that a large population does not protect against language shift, by comparing different ethnolinguistic groups within Indonesia we do find that size matters: in Indonesia at least, ethnolinguistic groups with smaller populations are more likely to use the dominant national language, and in those groups the effect of having one parent who speaks the dominant language is stronger than in larger groups. This finding alone has implications for the ways that we think about language revitalization and maintenance programs.

Grenoble and Whaley write that “[t]here is a commonality to the general circumstances that bring about language endangerment, yet regionally specific, or even community specific, factors dictate the ultimate effect of these circumstances” (1996: 211). We do not take issue with this basic point about the local nature of language shift (and in fact we find some support for it). Nevertheless, we argue for a renewed focus on the commonality of the general circumstances that bring about language endangerment worldwide, even among large languages. A shift toward examining language shift scenarios more broadly and more quantitatively helps to distinguish between the symptoms and the causes of language shift, where factors such as the number of children speakers and the number and quality of domains in which a language is used should be seen as symptoms of language shift that are caused by other factors (Himmelmann 2010). Moreover, large languages with their diverse speaker populations offer us an opportunity to quantitatively examine the social factors that contribute to shift and thus contribute to our understanding of the process itself.

Our findings with respect to both gender and religion reflect the complementarity of the examination of language shift at both national and local levels. Our results with respect to these social factors underscore the local nature of some aspects of language shift. At this macro level, we notably find that there is no effect of gender on the probability that one speaks Indonesian, despite previous reports on gender differences in use of Indonesian in particular
communities. Similarly, we find that the effect of religion is not the same across groups, but rather depends on whether the religion is that ethnic group’s majority religion. Namely, the minority religious group is more likely to speak the national language than the majority group. At the local level we hypothesize that social factors such as gender and religion are significant only in that they are indicative of other social processes. With an analysis like this we can only guess at the identity functions of language use, although in forthcoming work we also show that it is possible to examine these factors in more detail than we have done here.

Our results provide a foundation for modeling the interaction of the different social factors with the goal of creating more predictive models of language shift. We, along with many others, see a need for language revitalization efforts that address the primary causes of shift, beyond a focus on language-related activities such as documentation and education (see Henderson et al. 2014). We see our work as contributing to such activities in that it offers a more expansive description of those primary causes. Moreover, we feel that work at this level of abstraction from local communities may actually help to “deepen our understanding of the economic factors and marginalization of indigenous peoples that are intrinsic to language shift situations” (Florey 2010: 6).

Acknowledgements: The authors would like to thank J. Joseph Errington, Wayne Harbert, Daniel Kaufman, Naomi Nagy, John Wolff, three anonymous reviewers, and the audience at NWAV 44 in Toronto for their helpful comments on previous versions of this article. We would also like to thank the Fulbright Foundation for funding Cohn’s research in Indonesia, Atma Jaya Catholic University for hosting Cohn during her recent sabbatical, and both the Mario Einaudi Center for International Studies at Cornell University and the Center for the Humanities at the University of New Hampshire for funding portions of this research.

References


Supplemental Material: The online version of this article (DOI: 10.1515/ijsl-2016-0036) offers supplementary material, available to authorized users.