

BIG LANGUAGES AREN'T (NECESSARILY) SAFE: LANGUAGE SHIFT IN THE MAJOR LANGUAGES OF INDONESIA

Abigail C. Cohn, Maya R. Abtahian
Cornell University; University of Rochester
acc4@cornell.edu; maya.r.abtahian@rochester.edu

ABSTRACT

The instatement of Bahasa Indonesia as Indonesia's national language is widely cited as a successful example of language planning; yet it also has implications for endangerment of local languages. With over 700 local languages, Indonesia is one of the most linguistically diverse countries in the world. It is widely acknowledged that “small” languages are at risk of endangerment, but what is the fate of the “big” languages? In a multifaceted project, we address this issue investigating language shift among the “big” languages of Indonesia, those with over 1 million speakers.

INTRODUCTION

The discussion of language endangerment worldwide has focused primarily on languages with very small speaker populations and decreasing numbers of young speakers, partly following Krauss's (1992) prediction that any language with fewer than 100,000 speakers is at risk. This problem is particularly acute in Indonesia, with 706 living languages (accounting for almost 10% of the living languages worldwide) of which close to 90% have fewer than 100,000 speakers (Lewis, Simons and Fennig 2014). The problem may be even greater than we think as rapid changes to intergenerational transmission patterns suggest that even the largest local languages in Indonesia may be at risk. In recent years, linguists working on local languages in Indonesia have observed that the increased use of Indonesian is impacting the domains of use of local languages; however, the degree of impact is not well documented.

In our project, we address this issue by investigating reported language competence and use in the “big” languages of Indonesia, those with a million or more speakers. In this work we seek to find a middle ground between single, global measures of vitality such as the EGIDS score and in-depth ethnolinguistic interviews from different locations across Indonesia. In Section 2, we consider the question of the relationship between speaker population size and risk of endangerment. In Section 3, we turn our attention to the “big” languages of Indonesia. In Section 4, we present our methods and basic results and in Section 5 we return to the relationship between size and risk of endangerment among the largest languages of Indonesia. In Section 6, we conclude and discuss next steps in our project.

METHOD

The relationship between size and risk of endangerment

At the end of the 20th century, attention increasingly turned to the risk of endangerment and loss of minority languages of the world. Krauss (1992), among others, brought awareness to worldwide language endangerment, arguing that any language with fewer than 100,000 speakers should be considered at risk. Based on Krauss's assessment, with only 83 of its 706 living languages estimated to have speaker populations greater than 100,000, 623 or 88.2% of Indonesia's languages are at risk of endangerment. Correspondingly, recent work has attempted to assess the vitality of Indonesian languages (Florey 2010, Anderbeck 2015), largely focusing on languages with small speaker populations. It is now widely agreed that languages with small numbers of speakers are at risk, but does this mean that larger languages are “safe”?

One way this has been evaluated is using the Expanded Graded Intergenerational Disruption Scale (EGIDS) developed by Lewis and Simons (2010), included in *Ethnologue 17th edition*. This 13 step scale going from 0 – International to 10 – Extinct provides an assessment tool. Lewis and Simons (2010) also suggest a coarser categorization of languages as either “safe” or “threatened” where the

authors conclude that languages with a status of 6a – Vigorous or lower are safe and 6b –Threatened and above are at risk.

Assuming the EGIDS scale to be a reliable indicator, we can ask the question to what degree does speaker population correlate with EGIDS status? In Ravindranath and Cohn (2014), we investigated this question for the languages of Indonesia and found that there is not a close correlation. We asked whether there is a significant population difference between those languages that are considered endangered according on this scale and those that are considered safe and found that there is no evidence that the average population size is significantly different for the “more vital” languages in Indonesia than it is for the “less vital” languages. Considering how widely it is assumed that language size and vitality correlate, this is a surprising result. However, there are also issues about the validity of the EGIDS scale that need to be kept in mind.

A small speaker population at some point of course will accompany language death, since the definition of language death includes a dwindling number of speakers. But it is important to distinguish between those factors that result from language death and those that contribute to it. Himmelmann (2010: 46) notes “it is rarely the case that one or two or three causes or factors lead to language endangerment. Instead, language endangerment results from the specific and complex constellation of a variety of such factors...an endangerment scenario.” The fact that language size and vitality do not correlate suggests that a small speaker population is likely a symptom, rather than a cause, of language endangerment. The corollary of this observation is that even the large languages in Indonesia are not necessarily “safe”. As Florey writes, “[r]estricting the definition of ‘endangered language’ to those languages with small speaker populations disguises the extent of the problem” (Florey 2005: 59).

To better understand the multiple factors that contribute to a language endangerment scenario, we need a method to model these factors. In Abtahian, Cohn, and Pepinsky (2016) we develop such a method to investigate language shift focusing on the 10 largest non-Malayic languages.

ANALYSIS

The “big” languages of Indonesia

Most accounts consider languages with over a million speakers safe, or at least protected from large-scale language shift. By this measure Javanese, with an estimated 84 million speakers the 10th most widely spoken language in the world, should be safe. Yet as observed by Adelaar (2010, 25) in assessing the languages of Java, “In 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.” Moreover, a number of recent studies demonstrate rapidly shifting language use patterns and lack of intergenerational transmission of Javanese (Smith-Hefner 2009, Setiawan 2012, Kurniasih 2006), including a shift away from use of Krama (Errington 1998, G. Poedjosoedarmo 2006, Setiawan 2012). These observations, mostly based on detailed study of particular communities again call for a method that allows comparison across different languages of Indonesia in order to understand what factors are generalizable.

Based on Lewis et al. (2014), there are 21 languages in Indonesia with over a million speakers (notably all in Western Indonesia, 18 if varieties of Batak are grouped, as we do here). These are listed starting with the largest, Javanese, in Table 1.

Table 1. Spoken languages of Indonesia with over 1 million speakers, based on Lewis et al. 2014 (non-Malayic languages indicated in bold)

Language	Region	Population	EGIDS number
Java	Java and Bali	84,300,000	2 Provincial
Sunda	Java and Bali	34,000,000	5 Developing
Indonesian (Bahasa Indonesia)		22,800,000	1 National
Madura	Java and Bali	6,770,000	5 Developing
Minangkabau	Sumatra	5,530,000	5 Developing
Batak (Toba, Dairi, Simalungun, Mandailing)	Sumatra	5,500,000	5 Developing /6a Vigorous
Betawi	Java and Bali	5,000,000	6b Threatened
Bugis	Sulawesi	5,000,000	3 Wider com.
Malay	Sumatra	4,910,000	6a Vigorous
Aceh	Sumatra	3,500,000	5 Developing
Banjar	Kalimantan	3,500,000	3 Wider com.
Bugis	Sulawesi	3,500,000	6b Threatened
Bali	Java and Bali	3,330,000	5 Developing
Musi	Sumatra	3,105,000	3 Wider com.
Makasar	Sulawesi	2,130,000	6b Threatened
Sasak	Nusa Tenggara	2,100,000	5 Developing
Gorontalo	Sulawesi	1,000,000	6b Threatened
Malay, Jambi	Sumatra	1,000,000	6a Vigorous

The local languages of Indonesia with large speaker populations have EGIDS numbers that range from a 2 (Provincial) to 6b (Threatened), demonstrating both that “big” languages are not necessarily safe (e.g. Gorontalo, Makassar, and Betawi), and that there is not a close correlation between population size and EGIDS number as a measure of vitality.

In the next section, we turn to an analysis of census data to model language shift in ten big languages of Indonesia. In this study we take advantage of the fact that these languages have large and diverse populations, affording us the opportunity to conduct an analysis of language endangerment scenarios that can consider several factors at the same time.

Modeling language shift (Abtahian, Cohn, & Pepinsky 2016)

Our most recent project uses census data to examine language shift scenarios for the 10 largest non-Malayic languages in Indonesia, all with over 1 million speakers. We examined Indonesian use by age as a measure of language shift, using a 1% sample of the 2010 Indonesian census available through the Integrated Public Use Microdata Series. The sub-sample of the census data that we considered consists of the ten areas of Indonesia where a non-Malayic language that has at least 1 million speakers is spoken (those languages in **bold** in Table 1). Our dependent variable is the answer to the question: What language does (RESPONDENT) use daily at home? (“Apakah bahasa sehari-hari yang digunakan (NAMA) di rumah?”). In the pared down census data available through IPUMS the answer is Indonesian or other.

We considered a number of independent social factors across the ten language communities. We looked at age, urbanization, development index, education, religion, and gender, and the predictive value of parent’s first language on the first language of their children. 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 particularly in the Indonesian context (Errington 1998, Kurniasih 2006, Smith-Hefner 2009, Florey 2010, Setiawan 2012). Some of these we expected to be generalizable to other language shift scenarios; others we specifically wanted to investigate in Indonesia. There are of course complex relationships between many of these factors, and there are some important factors that this method leaves out, such as attitudes toward the different varieties and language use in different domains. Some of the factors that were necessarily left out of the census work are the focus of our current work using a more detailed sociolinguistic questionnaire (Cohn, Abtahian, Yanti 2017).

In the census project results we find that all of the coefficients are significant at the .001 level;

the magnitude and direction of the effects of different factors can be seen in our visual model (for example in Figure 1 for age, gender, and urbanization). For example, it shows that urban men and women are *far* more likely to speak Indonesian than rural men and women. Age is also an important factor, especially in urban settings, but 10- and 20-year olds still have less than a 40% chance of speaking Indonesian at home. While still statistically significant, we find little effect of gender from this vantage point. Although community level studies often 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 network, access to education, and exposure to Indonesian. We refer the interested reader to Abtahian, Cohn, and Pepinsky (2016) for a fuller discussion of the methods, results, and implications of this study.

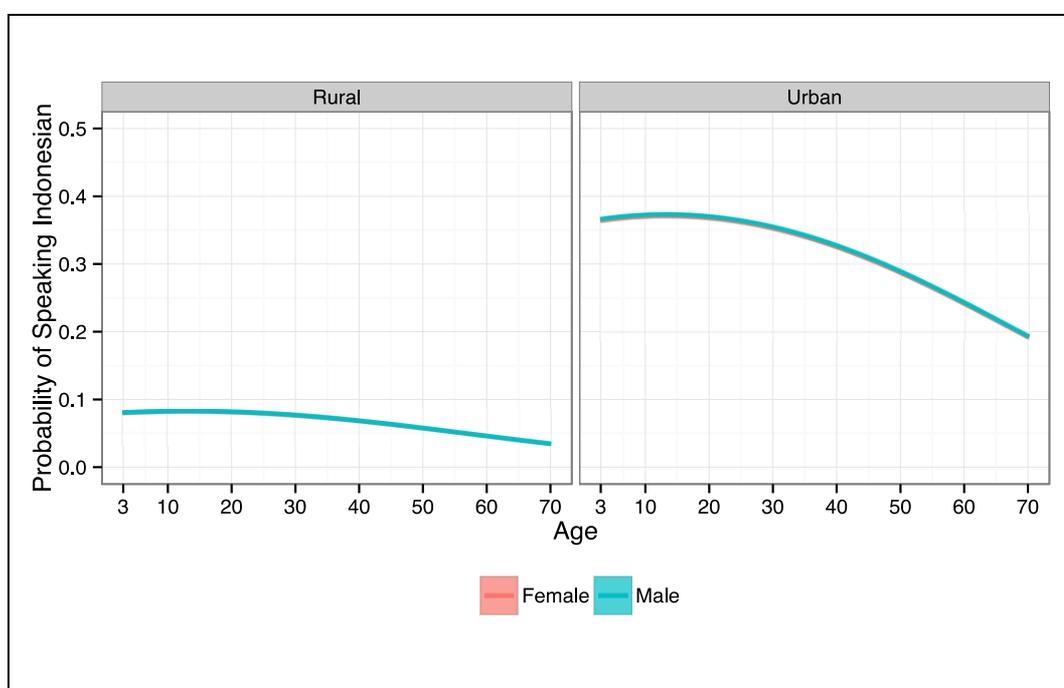


Figure 1: Baseline model for shift

The methodological approach used here involves a tradeoff between the rich qualitative insights that small-scale studies provide, and the panoramic overview of entire linguistic communities that statistical analyses of census data provide. Neither type of analysis can stand alone—in fact, we believe that both are required to understand language shift from the micro- to the macro-level. However, quantitative analyses enable us to ask many different questions using the same dataset. It is not possible for small-scale qualitative studies to compare speakers across multiple demographic categories at once: urban and rural, high and low income, across the age range, for different ethnic groups. In considering multiple factors at once, moreover, we are able to compare the *relative* strength of different factors.

Risk of endangerment and the effect of ethnic group

Ethnicity and correspondingly ethnolinguistic group is an important part of group identity in Indonesia, as it is in many multi-ethnic and multi-linguistic societies. One of the interesting aspects of using Indonesia as a case study is that, with so many groups varying in size and many ethnolinguistic groups with populations in the millions, we can examine ethnic group as a social factor independent of our other social factors. In the census study cited above, each one of our ten language community groups corresponds with one major province that has one major urban area, and each one of these is also associated with one major ethnolinguistic group.

Figure 2 displays our baseline analysis, divided by ethnic group, looking at just urbanization and age. The groups are laid out by size of population, left to right and top to bottom. In every case we find that urban residents are more likely to speak Indonesian than rural residents. But more importantly, we find dramatic differences in the overall probability of speaking Indonesian by ethnic group, and we hypothesize that this is influenced by a few factors. First we see a general trend by size,

where the languages with larger speaker populations are more likely to be maintained. However that alone does not account for the ethnic group differences we see. Javanese, Sundanese, Madurese, Balinese and Sasak all have both more maintenance of the local language and less difference between urban and rural speakers than Batak, Bugis, Aceh, Makassar, and Gorontalo which are uniformly further progressed in shift away from local languages. We hypothesize that this is due to an orientation of “inner island” (Java, Bali, and interestingly Lombok) vs. “outer island” (Sumatra and Sulawesi). Four out of five of the “outer island” languages still show a marked distinction between urban and rural speakers as well. The one exception to this is Gorontalo, which is clearly in the most vulnerable situation. We suspect that the vulnerability of Gorontalo may be related to factors beyond size or inner vs outer island. Anderbeck (2015) makes the point that since Gorontalo split from North Sulawesi province in 2000, many Gorontalo no longer feel the need to assert their distinct identity by using their language.

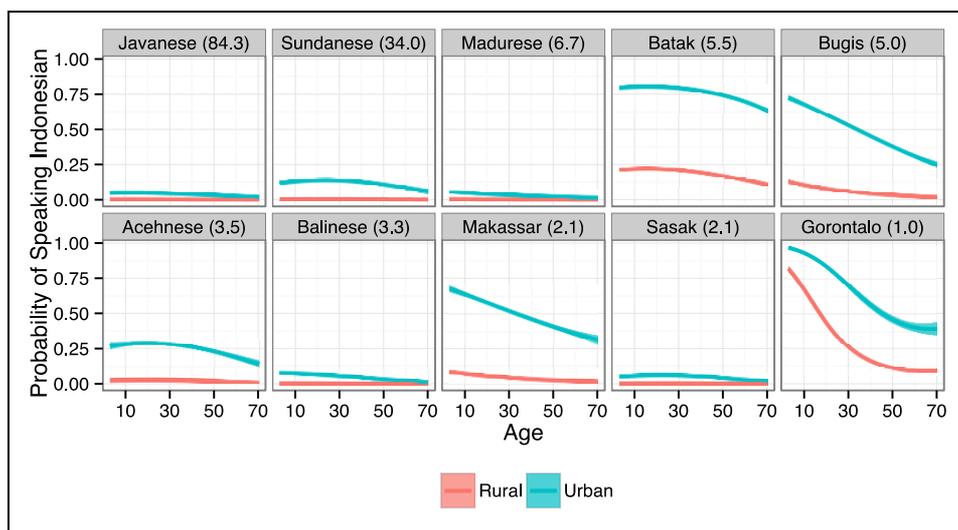


Figure 2: Ethnic group analysis

CONCLUSIONS AND NEXT STEPS

As it becomes increasingly clear that a small population is a symptom and not a cause of language shift, and that even large languages experience language shift, multivariate quantitative studies of big languages undergoing shift allow us to create a better picture of language shift scenarios. Quantitative studies such as ours, in turn, provide a baseline for community level studies of shift, where local social factors, domains of use, and language attitudes may be more closely examined and triangulated by observations of speakers’ actual language use.

Building on these findings we are currently analyzing results from a questionnaire which was developed to collect detailed data from subjects from different language backgrounds, allowing cross-group comparison of the interaction of a complex set of variables (Cohn et al. 2013). The questionnaire provides an intermediary level of inquiry, allowing collection of more data in a shorter period of time from more locations than would be possible with detailed interviews in individual communities, but also offering far more detail and insight into individual language choices than can be gained from census data. It includes questions about respondent’s background (geographic, ethnic, religious, educational and linguistic); language proficiency; language use in 34 different domains; and language background, proficiency, and use of their parents, grandparents, spouse and children (if relevant). It also includes several questions about language attitudes and use of technology. To date it has been conducted with 548 participants in 11 locales. Through the analysis of batches of questionnaires from different locations we can consider both intra-location and cross-location comparisons, furthering our goal of building more predictive models of the interaction of language background, language mastery, use, and attitude with potential language shift. In this way we contribute to the larger goal of a better understanding of language endangerment scenarios.

REFERENCES

- Abtahian, M., A. Cohn, and T. Pepinsky (2016). Modeling social factors in language shift. *International Journal of the Sociology of Language*, Volume 2016, Issue 242: 139–170.
- Adelaar, K. A. (2010). Language Documentation in the West Austronesian World and Vanuatu: An Overview. In *Endangered Languages of Austronesia*, ed. by Margaret J. Florey, 12-41. Oxford: Oxford University Press.
- Anderbeck, K. (2015). Portraits of language vitality in the languages of Indonesia. In I Wayan Arka, Ni Luh Nyoman Seri Malini, and Ida Ayu Made Puspani (eds.) *Pacific Linguistics 019: Language documentation and cultural practices in the Austronesian world: Papers from 12-ICAL, Volume 4*, pp. 19-47.
- Cohn, A., J. Bowden, T. McKinnon, M. Ravindranath, R. Simanjuntak, B. Taylor, Yanti (2013). Kuesioner Penggunaan Bahasa Sehari-hari. <http://lingweb.eva.mpg.de/jakarta/kuesioner.php>
- Cohn, A., M. Abtahian, and Yanti (2017). Inner islands and language shift in Indonesia. ISLOJ6, Semarang, Indonesia.
- Errington, J. (1998). *Shifting Languages: Interaction and Identity in Javanese Indonesia*. Studies in the social and cultural foundations of language. Cambridge: Cambridge University Press.
- Florey, M. J. (2005). Language shift and endangerment. In K. Alexander Adelaar and Nikolaus P. Himmelmann (eds.) *The Austronesian languages of Asia and Madagascar*. Routledge Language Family Series. London: Routledge, pp. 43-64.
- Florey, M. J. (ed.) (2010). *Endangered Languages of Austronesia*. New York: Oxford University Press.
- Himmelmann, N. P. (2010). Language endangerment scenarios: A Case Study from Northern Central Sulawesi. In *Endangered Languages of Austronesia*, ed. by Margaret Florey, 45-72. New York: Oxford University Press.
- Krauss, M. (1992). The world’s languages in crisis. *Language* 68. 4-10.
- Kurniasih, Y. (2006). Gender, Class and Language Preference: A case study in Yogyakarta. In Keith Allan (ed), *Selected papers from the 2005 Conference of the Australian Linguistic Society*. <http://www.als.asn.au>.
- Lewis, M. P., and G. F. Simons (2010). Assessing endangerment: expanding Fishman’s GIDS. *Revue Roumaine de Linguistique* 55. 103-120.
- Lewis, M. P., G. F. Simons, and C. D. Fennig (eds.) (2014). *Ethnologue: Languages of the World, Seventeenth edition*. Dallas, Texas: SIL International. Online: <http://www.ethnologue.com>.
- Poejosoedarmo, G. (2006). The Effect of Bahasa Indonesia as a Lingua Franca on the Javanese Speech Levels and Their Functions. *International Journal of the Sociology of Language* 177(1):111–121.
- Ravindranath, M. and A. Cohn (2014). Can a language with millions of speakers be endangered? *Journal of the Southeast Asian Linguistics Society (JEALS)* 7: 64-75.
- Setiawan, S. (2012). *Children’s language in a bilingual community in East Java*. PhD dissertation, The University of Western Australia, Perth Australia.

Smith-Hefner, N. (2009). Language shift, gender, and ideologies of modernity in Central Java, Indonesia. *Journal of Linguistic Anthropology* 19.1: 57-77.