JSEALS Special Publication No. 7

THE MALAY LECTS OF SOUTHERN SUMATRA

*

Jonathan McDowell and Karl Anderbeck

(Edited and Formatted by Paul Sidwell)

UNIVERSITY of HAWAI'I PRESS

OPEN ACCESS – Semiannual with periodic special publications
E-ISSN: 1836-6821
http://hdl.handle.net/10524/52466

Cover photo by Jonathan McDowell and Karl Anderbeck

JSEALS
Journal of the Southeast Asian Linguistics Society

The Journal of the Southeast Asian Linguistics Society publishes articles on a wide range of linguistic topics of the languages and language families of Southeast Asia and surrounding areas. JSEALS has been hosted by the UH Press since the beginning of 2017.

FROM THE JSEALS EDITOR-IN-CHIEF

This volume is the seventh JSEALS Special Publication. The goal of JSEALS Special Publications is to share collections of linguistics articles, such as select papers from conferences or other special research agendas, as well as to offer a way for linguistic researchers in the greater Southeast Asian region to publish monograph-length works.

In this instance, the authors have provided an overview of the tremendous dialectal diversity of Malay spoken throughout the island of Sumatra. It was the result of several years of data gathering in the first decades of the 2000s. The contents include a combination of phonological, lexical, morphological and sociolinguistic data with a synchronic focus but with diachronic points of reference. The work is supportedly generously with maps and comparative tables of linguistic isoglosses and features.

Thus, this is a substantive contribution to Malay linguistics, dialectology, and Southeast Asian linguistics broadly. We are very pleased that JSEALS is able to contribute to the sharing of such valuable linguistic research.

Mark J. Alves

December 1st, 2020

Montgomery College

Rockville, Maryland

Table of Contents

Index of Tables

Index of Figures

Abbreviations

The following table lists the abbreviations used in the report, except for the data site abbreviations which are contained in Table 3.1.

§	section
ISO (639-3)	International Standards Organization 639-3 standard language code set
[xxx]	ISO 639-3 language code
BAN	Basic Austronesian (wordlist) (Blust 1981)
BI	Indonesian (national language; *bahasa Indonesia*)
ILR	Interagency Language Roundtable
id.	*idem* (refers to the last item cited)
IPA	International Phonetic Alphabet
(I)SRT	(Indonesian) Sentence Repetition Test
L2	second language
LWC	language of wider communication
n.a.	not available/applicable
PBh	Pusat Bahasa/Badan Bahasa (Indonesian governmental language department)
p.c.	personal communication
PM	Proto-Malayic (Adelaar 1992)
PMP	Proto-Malayo-Polynesian (Blust 1995)
PP	*Palembang Pasar* ('Marketplace Palembang')
Pal.	Palembang
PSC	percentage of shared cognates
(RA-)RTT	(Rapid Appraisal) Recorded Text Test
SI	Standard Indonesian (see also BI)
SLQ	sociolinguistic questionnaire
SM	Standard Malay
SSML	Southern Sumatran Malay
Sumsel	South Sumatra (Indonesian province, Indonesian abbreviation)
WIST	SIL's West Indonesia Survey Team

Commonly-used Indonesian terms

bahasa	speech
(h)ilir	downstream
(h)ulu	upstream
kabupaten	government regency
kecamatan	government sub-district (sub-regency)
Melayu	Malay
sungai	river
tengah (tg.)	central
bukit	hill, mountain
lama	old, traditional
pasar	marketplace
muara (m.)	river mouth

1 Introduction

History tells us of a powerful Southeast Asian kingdom called Srivijaya. Based in southern Sumatra, this Malay-speaking empire controlled the lucrative maritime trade in the Spice Islands from the seventh to the thirteenth centuries AD and, among other things, spread the Malay language far and wide. Today, Malay boasts millions of native speakers, one of the largest speaker populations of any language in the world. As well as being an official language of Indonesia (*Bahasa Indonesia*), Malaysia, Singapore and Brunei, literally hundreds of non-standard varieties of Malay are spread through Sumatra, Peninsular Malaysia and Thailand, Borneo, and further east in historical trade centers.

Perhaps due to this incredible diversity, large swaths of Malay-speaking country have received little or no linguistic attention. This report concerns varieties of Malay in southern Sumatra, home of ancient Srivijaya (now the city of Palembang).[1] While individual speech varieties (lects)[2] in this area have received scholarly attention, contextualizing these better-known varieties within the Dutch-era lists of ethnic groups, or, harder yet, within a description of the overall language ecology, has been difficult if not impossible. The 14th edition of the Ethnologue (Grimes 2000) contained, for example, a confusing mishmash of language, dialect and even place names in its listing of Sumatran Malay 'languages'. Attempts to prepare materials for fields like education or community development, for example, have been thwarted by lack of clear reference dialects. A dialectology study was needed.

In response to these perceived needs, SIL's West Indonesia Survey Team (WIST) in cooperation with Indonesian government agencies carried out Malay dialectology field research in southern Sumatra from 2003 to 2008. Over the course of the years, a number of recommendations for changes to the language registry were made to the Ethnologue and to the ISO 639-3 code committee (details in §1.5 below). This report details dialect patterns and proposes groupings of Malay lects, presenting the linguistic and sociolinguistic information these conclusions are based on. In the process, changes to the ISO 639-3 language registry are explained and defended.

1.1 Summary of report

This report is organized as follows: Chapter 1 provides some background information on the geography, history and peoples of (southern) Sumatra and particularly its languages. It details previous linguistic research and as well as changes in the Ethnologue's registry (list) of languages and classifications over the past decade. It introduces the clustering system and terminology used in the report, giving a brief preview of our conclusions in the process, and finishes with extended descriptions of the dialect groups of southern Sumatra. Chapter 2, the shortest, discusses the type of research conducted, its rationale, goals, and delimits the area of study. Chapter 3 lists the research sites and outlines the sociolinguistic and linguistic methodology used. Chapter 4 gives the results of the lexical (mostly lexicostatistical) analysis. Chapter 5, the longest of this very long report,[3] details various phonological and morphological innovations that have occurred in southern Sumatran Malay lects since Proto-Malayic. Chapter

---

[1] See §2.3 for the precise borders of this study and the reasoning for excluding areas such as Jambi, West Sumatra, Bangka and Belitung.

[2] In this report we use the neutral term *lect* to denote a given speech variety, agnostic as to where it may fall on the language-dialect continuum. It may be helpful to note that the Malay/Indonesian term *bahasa* (< Sansrit), commonly translated as 'language', is actually better translated as the more noncommittal 'lect', denoting everything from idiolect (*bahasa aku* 'my way of speaking') to a plurality of languages (*bahasa daerah* 'local language(s)') and every communal linguistic aggregation in between. We prefer *lect* over the semantically equivalent but distractingly polysemous term *variety* (Chambers & Trudgill 1998:5) or the somewhat redundant *isolect* (Hudson 1967:12).

[3] The research behind this report is enormous; therefore, the report is enormously long too. It should also be noted that this report was drafted in 2008 and heavily revised in 2014; background research is current to 2014 but no attempt has been made to bring it up to (2020) date.

6 gives our findings on comprehension and attitudes from Recorded Text Testing. Chapter 7 details what things we learned from sociolinguistic questionnaires. Finally, chapter 8 lists our conclusions from the study and lists some areas of possible future research.

1.2 Southern Sumatra – background

The following is a very brief summary of the geography, populations, history and linguistic ecology of Sumatra. Obviously much more could be said; the reader is encouraged to consult a general reference work if more detail in these areas is desired.

1.2.1 Area description

Sumatra is the westernmost major island of Indonesia. See Figure 1.1. With nearly 500,000 square kilometers, it is the sixth largest island in the world, and has a current population of over 50 million. The three southernmost provinces, Lampung, South Sumatra and Bengkulu, make up 31% of its landmass and 31% of its population. See Table 1.1.

Figure 1.1 Map of Indonesia (CIA Factbook)

Table 1.1 Southern Sumatra basic facts

Province	Population	Area (1000km2)	Capital
Lampung	7.6 million	35	Bandar Lampung
South Sumatra	7.5 million	92	Palembang
Bengkulu	1.8 million	20	Bengkulu City

The *Bukit Barisan* (Row of Mountains) form the western backbone of Sumatra, and descend gradually in the east to often swampy lowlands. Sumatra has been recently subject to heavy deforestation. Typical industries of southern Sumatra include oil palm, rubber and coffee plantations, and natural gas and coal extraction, while typical occupations include subsistence farming, fishing and plantation work. Sumatra is one of the wealthier regions of Indonesia, and education levels are also above average. The preeminent metropolis in the region is the city of

Palembang, bisected by the large Musi River that drains much of South Sumatra Province and flows out to the straits dividing the low lying coastal swampland from the islands of Belitung and Bangka. Palembang, likely the ancient capital of the ancient Srivijaya kingdom, is today the vibrant capital of South Sumatra Province, having much of its income from trade and commerce as well as the energy industry.

1.2.2 History

The first evidence of human habitation in Sumatra dates to perhaps eight thousand years ago. Little is known about these Australomelanesoids, but their Hoabinhian technology is comparable to that found at similar dates on Java (Widianto 2009:34). Austronesian languages (including at least limited migration of related peoples) began to be spoken in Sumatra perhaps as early as 1500 BC (Bellwood 1997:92). It seems likely that some of the Batak and Barrier Island languages date back to the earliest Austronesian settlements (Mahdi 1994:447), while Malay seems a newer arrival (200 BC to 600 AD) which probably leveled previous linguistic diversity in much of Sumatra (Blust 2006:80).

The kingdom of Srivijaya provides us with the first Malay-language inscriptions from the 7th Century AD. This Malay polity (more accurately labeled 'city-state'; Manguin 2009) shifted headquarters over the centuries, from Srivijaya (Palembang) and Jambi in the early centuries, to Malacca and Riau in later centuries, and is probably responsible for much of the spread of Malay to far-off entrepots. Whether Srivijaya can be held responsible for the existence of vernacular Malay(ic) lects in other parts of Sumatra, Peninsular Malaysia and Borneo, relates to a more controversial question that has been explored in detail elsewhere (see, for example, Collins & Sariyan 2006), namely, where is the homeland of Malay? This is discussed briefly in §1.2.3 below.

At least from the time of Srivijaya onward, southern Sumatra has played host to wide-ranging and cosmopolitan influences and been integrated into the world economy (Edwards McKinnon 1985; Munoz 2006). Numerous Buddhist and Hindu remains are scattered throughout the region attesting to its commercial and cultural importance and intimate connections with India. During this period a syllabary called *surat ulu* (upstream/highland writing) was developed (Kozok 2004). Later inscriptions and manuscripts testify to the introduction of Islam (and Jawi script) to Sumatra; today the ethnic groups of the region are nearly 100% Muslim. (Today, both non-Roman writing systems persist only in highly circumscribed domains.)

Besides the sea which links the region to, literally, the world, a crucial geographical feature is the rivers, most notably the Musi, the longest in Sumatra. Southern Sumatra labels itself as The Realm of Nine Rivers (*Batang Hari Sembilan*). As Bronson (1977) postulates, understanding the riverine-dominated nature of social interaction goes a long way towards understanding historical patterns of contact. Later in the report we will see a cluster of phonological innovations which follows the Musi River, for example.

One of the most pervasive outside influences has been from Java, beginning seriously in the 13th century and more-or-less continuing until modern times. Both the Jambi and Palembang royal families are heavily Javanized (Andaya 1993), with Javanese influence seen particularly in the language around the court (Tadmor 2001; Anderbeck 2008); see §1.7.1.

1.2.3 Language family

Malay is a member of the Austronesian language family, in a western Malayo-Polynesian branch. Mid-level classifications are problematic but Malay is clearly linked at a low level with Malay-like languages in Borneo like Kendayan [knx][4] and Iban [iba] in a family labeled

[4] ISO 639-3 3-letter codes will be used in lowercase and enclosed in square brackets at the first introduction of a language within this report.

Malayic, and Malayic and Chamic languages are linked as Malayo-Chamic. This study relies heavily on the reconstruction of Proto-Malayic (PM) (Adelaar 1992 and elsewhere).[5]

The strongest candidate for the homeland of Malayic (and even Malayo-Chamic) is western Borneo (see Collins and Sariyan, eds., id). This is argued largely on the basis of linguistic diversity, arguably greater in Borneo than in Sumatra (Blust 1988). Some argue for a distinction between *Malayic* and *Malay*, where the latter's homeland may be in southern Sumatra (Tadmor 2002). In practice, however, the distinction between Malayic and Malay is difficult to maintain (Adelaar 2008).

Adelaar and Prentice (1996) divide Malay into three sociolinguistic categories:

1. Court Malay – literary language of the Malay courts such as Srivijaya. This is the precursor to the national languages of Indonesia and Malaysia.
2. Pidgin-Malay-derived Malay, such as found in eastern Indonesian entrepots. These lects are often also labeled *trade Malay* or *Eastern Indonesian Malays*.[6]
3. 'Inherited' or regional Malay dialects, what we call here *vernacular Malay*, the subject of this report.

Completing the *Malayic* language family, to this list could be added a fourth (putative) category, Malayic-but-not-Malay (like the Bornean Iban and Kendayan mentioned above, and possibly the Sumatran Duano [dup]).

1.2.4 Dialect diversity

The geographical pattern of vernacular Malayic (categories three and four above) could be generally described as a dialect continuum (Collins 1989) including Sumatra, Peninsular Malaysia and Borneo. In our experience, one finds a dialect pattern strongly suggestive of repeated leveling, with concentric circles largely emanating from the Johor-Riau archipelago. The further one gets from this canonical center, the more divergent the lect (and, in the Ethnologue, more likely to be labeled a distinct language rather than as a dialect of Malay [zlm]). So, Selangor (Peninsular) Malay is much more mainstream than, say, Terengganu Malay further north. Coastal Malay in western Borneo is much more mainstream than the Malayic spoken further inland. In Sumatra, it is not coincidental that the most divergent Malay lects are those spoken furthest west (like Kerinci) from Johor-Riau, or furthest south (HAJI). (The most striking exception to this rule is Duano, extremely divergent yet spoken in the Riau-Johor archipelago.[7]) This hypothesis places southern Sumatran Malay (SSML) in the zone of moderate to heavy variation, a hypothesis we shall revisit in the conclusion.

1.3 Linguistic ecology of southern Sumatra

From a broad linguistic perspective, all the languages of southern Sumatra are Austronesian (with the possible exception of Enggano)[8], and fall into five subgroups: isolates Enggano [eno], Rejang [rej] and Nasal [nsy] in Bengkulu province, various Lampungic lects scattered throughout Lampung and South Sumatra provinces, and Malayic, found in all three provinces. Enggano and Rejang have been described in numerous publications, while Nasal has only recently been documented (Anderbeck & Aprilani 2013). As late as the Ethnologue's 15th edition (Gordon 2005), Lampungic lects were described as nine distinct languages, but field

[5] Notably for this study, the SSML lect Serawai was one of the six main sources for the PM reconstruction.

[6] Within this broad category a spectrum spans from those lects which probably could be best classified as genetically Austronesian, to those which are clearly creoles, such as Sri Lanka Malay [sci].

[7] Anderbeck (2013) argues for a likely non-Malayic substratum in Duano, which would explain the strong linguistic divergence.

[8] Capell (1982:6), as translated by Nothofer (1986:90), states, "Enggano is a remnant of these pre-IN languages, which indeed has IN borrowings, but remains non-Austronesian." Capell's is a minority view but illustrates Enggano's tenuous classification.

research found that Lampungic linguistic diversity was substantially overstated and posited three clusters of mutually comprehensible lects, namely Api [ljp], Nyo [abl] and Komering [kge] (Hanawalt 2007; Anderbeck 2007a; Anderbeck & Hanawalt 2007).

One factor making dialectology problematic is the very high number of named ethnic groups in southern Sumatra, most of which claim their own 'bahasa' or way of speaking. (Actually, it is not uncommon for each village to claim its own 'bahasa'.) The fact that most of the Malayic lects in South Sumatra are named after rivers and other geographic features raises the question of whether these are distinct languages or if they are rather various dialects of Malay which are using place names as a cover term. Upon arriving in the province of South Sumatra, we were frequently told by government officials and local residents that this province possesses the greatest variety of languages (*bahasa*[9]) found in Indonesia. However, the same people reported there was typically mutual comprehension between speakers of these different language communities.

Ethnolinguistic identity or 'tribe' membership is far from consistent in southern Sumatra. As we queried various people about their ethnicity, a general pattern emerged of highland groups more consistently agreeing with each other on questions of identity than lowland areas, particularly coastal areas. Regarding the former area, Sakai (2006:39), quoting Andaya (1993:17), writes, "Historically, South Sumatran highland societies consisted of groups who defined themselves by descent from common ancestors and by reference to the specific area of land they inhabited". An extended quotation from Lebar (1972:14) describes the latter area well:

> A Pasisir (Malay, "coastal") culture—an amalgam of Malay, Javanese, Macassarese, Arabic, and Indian elements—arose in these trading centers and was subsequently diffused throughout the archipelago; nowadays it is characteristic of most of Indonesia's coastal and urban population centers (H. Geertz 1963). Aside from its strong Islamic content and a generalized orientation to marketing activities, this Pasisir culture type is difficult to characterize in specific terms. Economic pursuits are likely to be highly variable, and social relationships are seldom structured in terms of well-defined categories of kinsmen. This generalized diffuseness pertains also to the physical components of these coastal populations, the result of a long process of "Malayanization" of indigenous peoples—of *masok melayu* ("becoming Malay"), i.e. adopting the Malay language and the Muhammadan religion.

A number of different ethnic lists are extant including one managed by the Indonesian census bureau (Badan Pusat Statistik 2004), but perhaps the most thoroughly vetted and updated, if somewhat clumpish ethnic list, is that maintained by the organization Indonesia Pelangi Nusantara (2010). Table 1.2 provides IPN's list of Malay-speaking ethnic groups in southern Sumatra (including alternate names) placed into clusters of IPN's invention, with population, their primary province and their language/ISO code according to Ethnologue 17 (Lewis, Simons & Fennig 2013). SS = South Sumatra, BK = Bengkulu, LP = Lampung.

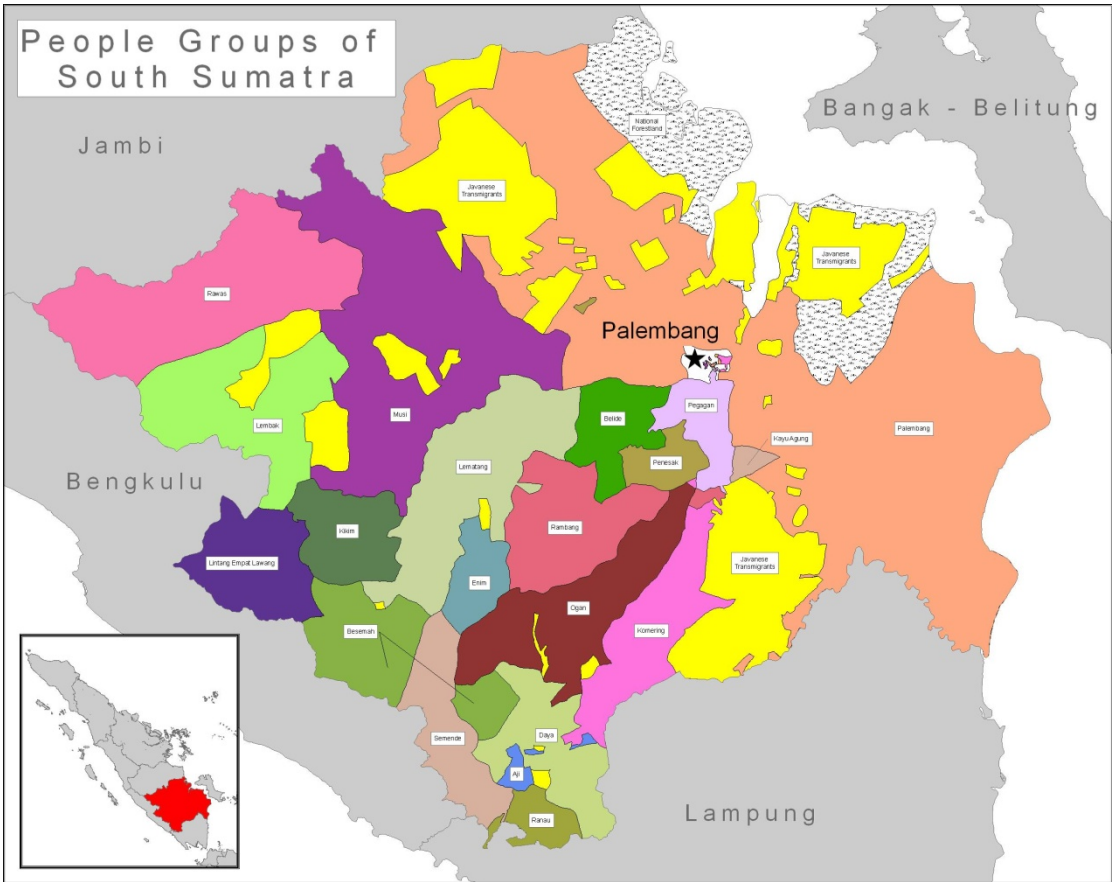
[9] See above for a discussion of the Indonesian term *bahasa*.

Table 1.2 Malay-speaking ethnic groups in southern Sumatra

Cluster	Ethnic group	Pop.	Prov.	Language	ISO
Minangkabau	Muko-muko	60,000	BK	MINANGKABAU	min
Bengkulu	Bengkulu	60,000	BK	BARISAN	pse
	Pekal (Ketahun)	30,000	BK	Pekal	pel
Musi	Musi (Sekayu, Lakitan)	600,000	SS	MUSI	mui
	Palembang (Musi Banyuasin)	3,000,000	SS	MUSI	mui
	Lembak (L. Delapan, Col/Cul, Sindang Kelingi, Saling)	208,000	SS, BK	Col/Lembak	liw
	Rawas (Rupit)	200,000	SS	MUSI	mui
Pasemah	Kaur	40,000	BK	Kaur	vkk
	Kikim	80,000	SS	BARISAN	pse
	Lintang (L. Empat Lawang)	190,000	SS	BARISAN	pse
	Pasemah (Besemah, Kisam)	650,000	SS, BK	BARISAN	pse
	Semendo (Semende)	230,000	SS, LP	BARISAN	pse
	Serawai (Seraway, Selatan)	290,000	BK	BARISAN	pse
Ogan	Aji (Haji)	15,000	SS	HAJI	hji
	Belide	75,000	SS	MUSI	mui
	Enim	120,000	SS	BARISAN	pse
	Lematang (Penukal)	275,000	SS	MUSI	mui
	Ogan	150,000	SS, LP	BARISAN	pse
	Pegagan (Sue Sue language)	225,000	SS	MUSI	mui
	Penesak (Meranjat)	130,000	SS	MUSI	mui
	Rambang	140,000	SS	BARISAN	pse

Modern scholars acknowledge that ethnic identity is typically shifting and multilayered, so we should not take the above as more than a rough guide, especially in the more coastal regions. Additionally, it is an open question how closely (or not) linguistic differences track with ethnic group membership in southern Sumatra, and our sample was not always fine-grained enough to adequately test the correlation. Yet this list can at least serve as a beginning hypothesis about Malay dialects in the region, to be refined later.

Figure 1.2 People groups of South Sumatra (IPN)



Sociolinguistic dynamics of this region are fluid and complex. Alongside indigenous populations, large numbers of *transmigrasi* (government-sponsored internal migrants) from Java, Madura and Bali inhabit the area, as well as less-planned migrants from West Sumatra, North Sumatra and elsewhere. It would be fairly accurate but too simplistic to say that Indonesian is the language of wider communication in this area. To borrow terms from creolistics without implying the presence of creoles, it could be said that Indonesian functions as the acrolect in South Sumatra Province, local languages as basilects, with a range of mesolectal positions which could be broadly labeled *Palembang Pasar* ('Marketplace Palembang'). In general, local language use is vigorous alongside Indonesian and Palembang Pasar, naturally stronger in rural areas than in urban.

More is said about language use, language vitality in the findings section, and about Palembang Pasar in §1.7.1 below.

1.4 Previous research on the Malayic of southern Sumatra

This summary is mostly focused on dialectology and language identification; a comprehensive bibliography of all the ethnographic and linguistic research on this area would probably take several pages.

Voorhoeve (1955) is a comprehensive summary of pre-World War II research of the languages of Sumatra, while Kennedy's (1962) also comprehensive summary is more ethographically-focused. O.L. Helfrich is the primary linguist whose name can be connected with SSML in that time period; his efforts focused on the highlands lects Besemah and Serawai (labeled 'Midden Maleisch'), and also published 'a few riddles' in Bengkulu Malay. Volume 10/2 of the Holle lists (Stokhof 1987) contains five extended SSML wordlists, from the areas researched by Helfrich.

In search of a broader view, Voorhoeve (p. 19) notes:

> "in 1940 a competition was held in the Palembang residency in order to collect data on the dialects spoken in that area. Twelve useful papers were submitted; if I remember rightly, two of these were in Lampung, one in Kubu, and the remainder in different Malay dialects varying from typical Middle Malay to a Malay which stood curiously close to the language of classical literature. It appeared impossible to draw a sharp dividing line between these two areas. All the papers obtained in this competition were lost, except some Kubu texts…"

Apart from this (abortive) effort, the linguistic landscape of SSML remained foggy.

Since the 1980's the Indonesian Language Center (*Pusat Bahasa*, henceforth PBh) has published over forty monographs on individual SSML lects. A few of these have had a dialectological component, but have suffered either from being quite narrow (Wahab et al. 1990) or methodologically compromised (Kasim et al. 1987) or both.

Collins (1995) is a comprehensive listing of modern research in Sumatran Malay and includes many of the resources mentioned above. One study reviewed by Collins is particularly noteworthy for this report: Mitani's (1980) book section 'Languages of South Sumatra'. Using lexicostatistical analysis of a number of SSML wordlists (which unfortunately were never published), he drew a broad distinction between 'Highland Malay' and 'Lowland Malay'. We will revisit his conclusions later in the monograph. The most comprehensive mapping of SSML lects, however, was William Foley's summary of Sumatran languages (Wurm & Hattori 1983). He mapped a number of Malay dialects for the first time in a remarkably prescient manner. Foley admitted the tentative nature of his information, stating that "[a] full dialect survey of Sumatra Malay has never been attempted." As can be expected, the portrayal included some inaccuracies (as well as omissions), such as mistakenly placing the Lubu group in South Sumatra (actually North Sumatra Province), and misclassifying some Lampungic lects as Malayic (Kayu Agung, Daya and Ranau). Subsequent editions of the Ethnologue for Sumatra relied heavily on Wurm & Hattori including the errors within. However, the Ethnologue compounded the problem by taking nearly all the "Malay dialects" in Wurm & Hattori and christening them distinct languages. (The same thing was done with Lampungic dialects.) The result was that, in the Ethnologue, linguistic diversity for southern Sumatra was overstated by a factor of three times compared to our current understanding.

Most recently (1991-2008), Pusat Bahasa was involved in a massive language identification project called *Penelitian Kekerabatan Dan Pemetaan Bahasa-Bahasa Daerah di Indonesia* ('research on the genetic relationships and mapping of the local languages of Indonesia') using the methodology called dialectometry (Mahsun 2005). Although numerous lengthy SSML wordlists were collected (117 total for South Sumatra province including Bangka and Belitung), no Sumatra-specific monographs have been or likely will be published as they have for some other regions. Sudarmanto (2012) does provide a helpful summary of their conclusions for Sumatra, though, which is discussed in §8.1.1 below, and these conclusions bear at least a passing resemblance to ours here (if you squint). Some of the biggest problems with PBh's approach are lack of attention to historical linguistics/language classification, not seeking sociolinguistic information, and double-counting languages which cross provincial borders.

1.5 Changes to the Ethnologue/ISO 639-3 Language Registry 2001-present

As discussed above, as the result of our background research and fieldwork, the Ethnologue and later ISO 639-3 language registry for SSML lects has undergone significant modifications since the 14th edition (Grimes 2000). The same can be said for the genetic classifications of Malay in general as well as the classification of specific lects. We will discuss the latter first.

1.5.1 Classification changes

It would be safe to say that the Ethnologue classifications of Malayic in recent history have been rather turbulent. In Ethnologue 14, most SSML lects were classified like this:

Austronesian, Malayo-Polynesian, Western Malayo-Polynesian, Sundic, Malayic, Malayan, Local Malay

(The final node in the case of Muko-Muko and Pekal was *Para-Malay* instead of *Local Malay*.) In Ethnologue 17 (Lewis, Simons & Fennig 2013), they are classified thusly:

Austronesian, Malayo-Polynesian, Malayo-Chamic, Malayic, Malay

How did they get from there to here? For the 15th edition (Gordon 2005), the outdated *Sundic* node from Dyen (1965) finally died a happy death, as did the controversial *Western Malayo-Polynesian*. In the 16th edition (Lewis 2009), based on Adelaar (2005), a *Malayo-Sumbawan* node was added between *Malayo-Polynesian* and *Malayic*, encompassing Malayic, Chamic, Balinese, Sasak, Sumbawan, Sundanese and Madurese. This change proved to be premature, however, as scholarly support for Malayo-Sumbawan never solidified. Therefore, in Ethnologue 17, this was replaced with the much more conservative and accepted *Malayo-Chamic* node.

The membership of Malayic has also narrowed since the term was originally proposed by Dyen (id.); see Blust (1988). In Ethnologue, Rejang was belatedly evicted from the Malayic group as of the 15th edition, as were the Lampungic lects Ranau and Kayu Agung [previously rae and vky; now merged with other Lampungic languages]. By the 16th edition it was recognized that Daya was not a Malay dialect but was instead Lampungic.

The distinction between a *Malayic* node and a *Malayan* node originates with Nothofer (1988), but Adelaar (1993), specifically addressing Nothofer's classification, argued against it. The *Malayan* term has since dropped off the radar. The *Para-Malay* category had long been done away with in the wider linguistic world. The term was introduced by Leach (1950) but it never had any linguistic currency, and Collins (1985b:478) discusses why this term should not be used. Both *Malayan* and *Para-Malay* were therefore removed from Ethnologue classifications in the 15th edition.

Finally, the retention of the *Malay* node under *Malayic* (an abbreviated continuation of the *Local Malay* node) acknowledges the sociolinguistic reality that some Malayic lects have a historical and/or social connection with *alam Melayu* ('the Malay world') and some do not. Whether this node actually belongs within a genetic classification is an unresolved question.

1.5.2 SSML Language registry changes

The Ethnologue applies the following (sometimes conflicting) criteria in defining a language (Gordon 2005:8):

- Two related varieties are normally considered a variety of the same language if the speakers of one variety have inherent understanding of the other variety at a functional level (that is, can understand based on knowledge of their own variety without needing to learn the other variety).
- Where spoken intelligibility between varieties is marginal, the existence of common literature or of a common ethnolinguistic identity with a central variety that both understand can be a strong indicator that they should nevertheless be considered varieties of the same language.
- Where there is enough intelligibility between varieties to establish communication, the existence of well-established distinct ethnolinguistic identities can be a strong indicator that they should nevertheless be considered to be different languages.

Table 1.3 summarizes the changes in the Ethnologue/ISO 639-3 language registry from the 14th edition to the present, using those criteria. As can be seen, seventeen listed languages were merged into five.

Table 1.3 Language registry changes 2000-present

Old entry	ISO	New Entry	ISO	Action
Pasemah	pse	S. BARISAN MALAY	pse	rename S. BARISAN MALAY
Serawai	srj			merge with S. BARISAN MALAY
Bengkulu	bke			merge with S. BARISAN MALAY
Semendo	sdd			merge with S. BARISAN MALAY
Lintang	lnt			merge with S. BARISAN MALAY
Ogan	ogn			merge with S. BARISAN MALAY
Enim	eni			merge with S. BARISAN MALAY
Musi	mui	MUSI	mui	
Sekayu	syu			merge with MUSI
Rawas	rws			merge with MUSI
Palembang	plm			merge with MUSI
Lematang	lmt			merge with MUSI
Penesak	pen			merge with MUSI
Lembak	liw	Col	liw	name change
Sindang Kelingi	sdi			merge with Col
MINANGKABAU	min	MINANGKABAU	min	
Muko-Muko	vmo			merge with MINANGKABAU
(new)		HAJI	hji	new language
Malay	mly	Malay	zlm	split
		Standard Malay	zsm	split

The language entries for Pasemah (Besemah), Serawai, Bengkulu, Semendo, Lintang, Ogan and Enim, lects mostly found in the highlands of the Bukit Barisan mountains between South Sumatra, Lampung, and Bengkulu provinces, were recognized as mutually comprehensible dialects and grouped under the overall heading of 'CENTRAL MALAY' [pse], which we now recommend be renamed 'SOUTH BARISAN MALAY'. ('Central Malay' was chosen as a translation of the historic Dutch name for these lects, *Midden Maleisch*, but is less accurate and descriptive than the recently recommended name offered by a long-time linguistic fieldworker (McDonnell 2014).) Anderbeck (2007b) serves as the original ISO change request.

The entries for Musi, Sekayu, Rawas, Palembang, Lematang and Penesak were recognized as mutually comprehensible dialects and grouped under the overall heading of MUSI [mui]. Anderbeck (2007c) covers the change request for all except Sekayu, which had already been merged with MUSI as of Ethnologue 15th edition.

Lembak (alternate name: Col, pronounced 'choal') and Sindang Kelingi were merged as COL [liw] (Anderbeck 2007d).

Muko-Muko was merged with MINANGKABAU [min] as they share clear linguistic and sociolinguistic affinity (Anderbeck 2007e).

KAUR [vkk] and PEKAL [pel] in Bengkulu province were left untouched as distinct languages.

HAJI (Aji), previously identified as a dialect of Malay [mly], was demonstrated in Anderbeck (2007f) to be better described as a distinct language, chiefly on the basis of a strikingly different lexicon from its linguistic neighbors. Mulak, also on the list of MALAY [mly] dialects, was removed as it is actually an alternate name for Kaur [vkk].

A bit further afield but related to SSML, the entry for Malay [mly] was retired and split into (VERNACULAR) MALAY [zlm] and STANDARD MALAY [zsm], the official language of Malaysia and Brunei (Anderbeck 2007g). While STANDARD MALAY is of narrow denotation as the single prescribed official standard, the former is very broad, encompassing vernacular Malay lects in Sumatra, Peninsular Malaysia and Borneo. The two differ

grammatically and phonologically (Asmah Haji Omar 1975; Benjamin 1993) as well as sociolinguistically (Adelaar & Prentice 1996). The relevance of (VERNACULAR) MALAY [zlm] to this study is that this ISO code encompasses a number of Malay dialects which differ little from the canonical vernacular Malay of Riau-Johor (see §1.2.4 above). In previous Ethnologue editions, this included named SSML dialects, but those dialects are now all subsumed under other SSML-specific language entries such as MUSI or SOUTH BARISAN MALAY. Said in another way, all known Malay dialects in southern Sumatra show greater affinity to each other than to the Malay of, say, Johor or Riau.

Finally, it is worth mentioning that, in the 16th edition (Lewis 2009) the Ethnologue added the concept of *macrolanguage*. A macrolanguage, according to their definition, is "multiple, closely related individual languages that are deemed in some usage contexts to be a single language… individual languages that comprise a macrolanguage must be closely related, and there must be some domain in which they are commonly viewed as comprising a single language." Given the unitary perception of 'Malay' in some contexts, a number of ISO listings were attributed to a single MALAY macrolanguage [msa]. Besides INDONESIAN and STANDARD MALAY, all SSML languages were deemed part of the MALAY macrolanguage (Anderbeck and Ethnologue personal correspondence).

Needless to say, little of the above was clear to us when beginning background research and fieldwork. In the following subsection we discuss our terminological choices, designed to make this journey through the jungle of Malay lects a little more navigable.

1.6 Terminology and clustering

The SSML dialect network is so complex that we have required five levels of terminology at the language level and below (contrast that with typical treatments that have at most the level of 'language' and 'dialect'). The terminology we have settled on is:

0. MACROLANGUAGE (designated by all caps and/or the ISO code)
1. LANGUAGE (designated by all caps and/or the ISO code)[10]
2. DIALECT CLUSTER (designated by small caps)
   (2.5 DIALECT SUBCLUSTER – used in only 1 of the 4 dialect clusters)
3. Dialect – this is the closest level to the indigenous usage of ethnonyms
4. Subdialect – this is given, when appropriate, after the dialect name
5. Subsubdialect (very rare –only used a few times)
6. Research site

So, for example, in this framework the research site Sukaraja (6) is a representative of the Besemah (subdialect: Kikim) dialect (4 and 5), which is part of the HIGHLAND dialect cluster (2) which is part of the SOUTH BARISAN MALAY[11] 'language' (1) of MALAY (0).

In following the Ethnologue definition of 'language' reviewed above and focusing on comprehension (whether reported or tested), cross-checked with linguistic similarity, we are making the prediction that speakers of dialects within, say, MUSI [mui], will understand each others' speech better than they would a dialect within, say, SOUTH BARISAN MALAY. As will become obvious, this claim frays at the borders we've drawn (this is a dialect network, after

[10] Section 0 below explores the concept of 'language' in some detail. For now, whenever you read "X language", substitute the phrase "cluster of related dialects whose outer boundaries are boundaries of comprehension with speakers outside the cluster". Aren't you glad we just used 'language'?
Given the ambiguous language/dialect status of Col, Pekal, Kaur and Kubu, for the remainder of this report we will use regular (not all-caps) font to identify these lects, which for the first three emphasizes their respective membership in the **UPPER MUSI** and HIGHLAND clusters.

[11] The maps use some now-abandoned terminology like "Central Malay" for SOUTH BARISAN MALAY, "Ogan" for the **OGANIC** cluster, "Musi Malay" for the MUSI language, "Musi" for the **UPPER MUSI** cluster, "Palembang" for the **PALEMBANG-LOWLAND** cluster and "Coastal Malay" for the Pesisir dialect.

all). Nevertheless, we believe the data largely bear out this system of clustering. The following set of tables lays out the SSML lects researched as part of this study, following our system of clustering.[12] See Table 1.4 for a summary of the languages treated here.

Table 1.4 SSML lects in this report, clustered

LANGUAGE						
MUSI		S. BARISAN MALAY		MINANGKABAU	HAJI	KUBU
DIALECT CLUSTER						
PALEMBANG -LOWLAND	UPPER MUSI	OGANIC	HIGHLAND	(Muko-Muko dialect)	(none)	(none)

Since HAJI and KUBU do not have any dialects discussed in this report, and since Muko-Muko is the only MINANGKABAU lect treated here, the following tables (1.5, 1.6, 1.7, and 1.8) only expand upon MUSI and BARISAN.[13]

Table 1.5 PALEMBANG-LOWLAND dialect cluster (MUSI language)

LANGUAGE					
MUSI					
DIALECT CLUSTER					
PALEMBANG-LOWLAND					
Subcluster					
PALEMBANG			LOWLAND		
Dialect					
Pal. Lama	Pal. Pasar	Pesisir	Penesak	Lematang Ilir	Belide

Table 1.6 UPPER MUSI dialect cluster (MUSI language)

LANGUAGE			
MUSI			
DIALECT CLUSTER			
UPPER MUSI			
Dialect			
Musi (Sekayu, Kelingi, Penukal)	Pegagan	Col (Sumsel, Bengkulu)	Rawas (Rupit, Tengah, Ulu)

Table 1.7 OGANIC dialect cluster (BARISAN language)

LANGUAGE		
SOUTH BARISAN MALAY		
DIALECT CLUSTER		
OGANIC		
Dialect		
Rambang	Enim (Ilir, Tengah, Ulu)	Ogan (Ilir, Tengah, Ulu)

[12] Note: we are not claiming any of these divisions as historical linguistic subgroups, rather merely as convenient and supportable clusters. (One will notice, for example, that a number of innovations in dialects like Pekal are shared with unrelated neighboring languages like Rejang.) As mentioned elsewhere, the named dialects correspond quite closely to emic (insider) ethnolinguistic groupings, while the labels above the dialect level are of our own invention.

[13] One will have noticed by now that many lect names employ colored fonts. A name in the green range denotes a MUSI lect (differentiated by DIALECT CLUSTER), while a blue-purple range denotes a SOUTH BARISAN MALAY lect similarly differentiated. The few MINANGKABAU lects are given red.

Table 1.8 HIGHLAND dialect cluster (BARISAN language)

LANGUAGE								
SOUTH BARISAN MALAY								
DIALECT CLUSTER								
HIGHLAND								
Dialect								
Bengkulu	Lintang	Besemah (Kikim)	Lematang Ulu	Benakat	Semenda	Serawai	Kaur	Pekal

This approach to 'language identification' is explicitly bottom-up. We agree with Kamusella (2012:68) that, as we consider various kinds of communalects, the most basic level of abstraction is the dialect, formed by a cohesive speech community. Over time, as original speech communities have grown, split, migrated, or diverged in other ways, additional dialects are birthed. At some point, due to lack of contact or social disunity, certain lects become unintelligible to their related dialects through separate linguistic evolution. (This is where Mufwene's (2001:171) metaphor of language as a parasitical species provides insight, as we consider the development of related species which are no longer able to interbreed.) At this point we (typically) label them 'languages'.

In the final section of the introduction, we give some more details on the main SSML dialects, tipping our hand in the process by revealing some of our analysis.

1.7 Dialect groups of southern Sumatra

The dialects briefly described below are the ones that were investigated as a part of the series of surveys of southern Sumatra that this report gives the results of. For the purpose of helping the reader who may run into these lects' names in other publications or in other parts of this report, the names, alternate names, previous ISO codes, location, and numbers of speakers are given in this section. All populations are estimates based off of village populations in the speech area using the 2003 Indonesian Agricultural Survey (PODES).

1.7.1 Dialects of the MUSI language

We begin our presentation of SSML dialects with a look at those we have clustered and collectively labeled the MUSI language. Within MUSI, the first entity we examine is not a dialect but a cluster of dialects which all find their home in the city of Palembang.

Palembang, [previously plm] like other large cities around the world, is more cosmopolitan than the surrounding countryside, with a large number of different ethnicities represented there. As such, the older Palembang language, which itself was influenced heavily by Javanese, has been adapted and has innovations to fit the needs of the current inhabitants.

The language situation in the city of Palembang is similar to that described in its sister city of Jambi to the north in Jambi Province as described by Kristen Anderbeck (2010:4–10) and in Riau as described by David Gil (1994:179–200). Both describe a bifurcation between a vehicular urban lect labeled 'Jambi Indonesian'/'Riau Indonesian' and basilectal or traditional forms of Malay. The best-known and –documented case, however, is that of Jakarta. In the Indonesian capital and centuries-old entrepot, the old vehicular Malay lect, called Betawi (or Jakarta Malay), is spoken by ethnic inhabitants of the same name (Ikranegara 1980; Adelaar 1992). Jakarta Malay shows strong influence from Javanese, Sundanese and Balinese. Betawi, in turn, has served as the source of numerous features within a new vehicular lect based on Indonesian (Bahasa Indonesia), increasingly seen as distinct from both Betawi and Indonesian and often labeled 'Jakarta Indonesian' (Ewing 2005; Wouk 2004; Cole, Hermon & Tjung 2006). Both varieties are marked by the use of the *-in* verbal transitive prefix and frequent occurrence of schwa in ultimate closed syllables where, in Indonesian (and nearly all other Malay lects), only *a* is found, e.g., *bənər* 'true' vs. *bənar*, *daləm* 'inside' vs. *dalam*.

The PBh publication *Struktur Bahasa Melayu Palembang* (Dunggio et al. 1983) states that there are two varieties of Palembang: refined Palembang (*baso Pelembang alus*) and everyday

Palembang (*baso P'lembang sari-sari*). Although they treat only the latter, the authors say the refined language was used during the reign of the kingdom established in Palembang by the Majapahit empire, and therefore has many words from Javanese. They say it is still used by old people, important leaders, and between respected people, particularly in solemn ceremonies. The authors say that the everyday version of Palembang is used by younger people both in ceremonies and in everyday life. Our analysis has largely confirmed this, but we would add that *traditional Palembang* (which we refer to as *Palembang Lama*; old Palembang) is also spoken in villages outside Palembang (relic areas).[14] We would also add that the most important function of *everyday Palembang* is not its everyday-ness but rather its roles as the primary language of wider communication (LWC) in South Sumatra. The latter is used as a language between people of different ethnic backgrounds, in commercial transactions, and in local media, hence it is also given the name *Palembang Pasar* (marketplace Palembang).[15]

The number of respondents who identified themselves as ethnically Palembang during the 2000 census was almost one million. When combined with the figures for those identifying themselves as Musi Banyuasin or identifying themselves simply by their village name or as *Melayu* who live near the coast in the greater Palembang speaking area, the total figure arrived at from the census and other sources is 1,600,000 first-language Palembang speakers.[16]

The large number of speakers and its strategic position of a language of wider communication between people from different ethnic backgrounds in South Sumatra make *Bahasa Palembang* an important subject of study, particularly with regards to language planning. When one counts second-language speakers, the number of speakers rises considerably, as most inhabitants of South Sumatra have some familiarity with a variant of Palembang and use it as a LWC.

Structurally, what speakers collectively refer to as *baso P'lembang* (Palembang language) is very similar to Pesisir and the LOWLAND subcluster—the most significant difference being that language varieties carrying the label 'Palembang' usually (and prominently) reflect PM final **a* as *o*.[17] A few linguistic differences can be seen, however, between Palembang Lama and Palembang Pasar. Palembang Lama retains more of the earlier Javanese lexicon (Tadmor 2001) whereas Palembang Pasar demonstrates a closer lexical affinity with Indonesian (Standard or Jakarta): the varieties of Palembang Pasar have between 10-15% greater lexical cognacy with Standard Indonesian than the Palembang Lama varieties. Palembang Pasar contains the cosmopolitan phonological innovation of the apical trill *r* rather than the voiced/voiceless velar fricative *ɣ*/*x* often seen in Palembang Lama. The *–in* verbal transitive suffix, seen as coming from Jakarta Indonesian, can be found in Palembang Pasar (particularly more urban varieties) but not, to our knowledge, in Palembang Lama. Perhaps a related fact is that one Palembang Lama site has a *–kən* (rather than the expected *–kan*) suffix. Finally, what we call 'Java schwa' (see §5.10) can be found in Palembang Lama but not Palembang Pasar.

The differences between Palembang Lama and Palembang Pasar are summarized in Table 1.9.

---

[14] Respondents in the city of Palembang called that spoken in the relic areas *Palembang Kasar* (vernacular or 'coarse' Palembang) or *Palembang Dusun* (village Palembang). The speakers themselves say they speak *Palembang Asli* (original/traditional Palembang).

[15] Newsreport stories and columns are written in Indonesian, but when criminals or eyewitnesses to a crime are quoted, Palembang Pasar (the basilect form) is used. Newspaper columns with humorous or off-color stories also use the mesolect form. The popular TV show about crime also uses Palembang Pasar. Most radio shows and advertisements use Palembang Indonesian, a form closer to but still distinguishable from the acrolect, Standard Indonesian.

[16] An example of how individual population names need to be added can be seen with the census data containing the name 'Teloko' for a language group, when this is a Palembang-speaking village near Kayu Agung.

[17] One site, Air Itam (the closest to the Pegagan language cluster), has the characteristic word final *e* for PM **-a* like the other Pegagan sites, but lacks the other distinctive innovations found in Pegagan. Additionally, two traditional Palembang sites in rural areas reflect *i* in the same environment.

Table 1.9 Varieties of Palembang with characteristics

Variety/Feature	**Palembang Pasar**	**Palembang Lama**
-final *a	*-o*	some *-i*, some *-o*
***r**	mixed depending on basilect, *r* in city	some sites *ɣ*, others *r*
'Java schwa'	no	yes
***-kən* suffix**	no	yes (one site)
Java. vocab.	less	more
***-in* suffix**	yes	no
Sociolinguistic function	vehicular language, increasing urban first-language	mother tongue in traditional and relic areas

The fourteen sites visited as samples of Palembang as listed in Table 1.13 are also listed below, with some additional information about the type of Palembang spoken.

VILLAGE	**DIALECT**	**CODE**
Teloko	Palembang Lama	PL-TEL
Paku	Palembang Lama	PL-PAK
Palembang Lama	Palembang Lama	PL-PL
Sarang Lang	Palembang Lama	PB-SL
Pemulutan	Palembang Lama	PB-PM2
Pelabuhan Dalam	Palembang Lama	PB-PM
Palembang City 1	Palembang Pasar (more acrolectal, towards Indonesian)	PB-SH1
Pulau Betung	Palembang Pasar (more basilectal, everyday)	PB-BET
Gasing Laut	Palembang Pasar (more basilectal, everyday)	PB-GL
Duren Gadis	Palembang Pasar (more basilectal, everyday)	PB-DG
Siju	Palembang Pasar (more basilectal, everyday)	PB-SJU
Air Itam	Palembang Pasar (more basilectal, everyday)	PB-AH
Palembang City 2	Palembang Pasar (more urban)	PB-SH2
Tangan Buntung (within Palembang)	Palembang Pasar (mesolectal, spoken as L2)	PB-SH3

Pesisir: The lects collectively labeled **Pesisir** are listed as Malay [mly] in the 15th edition of the Ethnologue and mapped as part of the Malay dialect chain going from Riau in the north of Sumatra to the border of Lampung in the south of Sumatra. In the surveys, it was found that the variety of Malay spoken to the north, east, and west of Palembang shared a common history, lexicon, and comprehension with speakers of Palembang. Unlike the Palembang sites listed above, which uniformly exhibit *o* for word final **a*, these sites have much more variation, with four exhibiting the high or mid central unrounded vowel *ɨ*/*ɘ*. Two sites, Simpang Bayat and Supat, have the word final *o* like Palembang, but are sufficiently removed geographically that they see themselves as speakers of *Melayu*, not as speakers of Palembang. The population for this area is not enumerated in the census as a distinct ethnolinguistic group in the census, but its speakers are most likely captured in the enumeration for Malay, Musi Banyuasin, and for Palembang speakers in the area. If the Palembang proper portion of the population grouping is subtracted out of the 1,600,000 figure for the group, this would leave approximately 600,000 speakers of Pesisir, though the distinction between the two is tenuous, explaining why these are counted together as one population group. The Musi Banyuasin speakers are estimated at around 220,000, with the remaining 400,000 coming from the other Pesisir communities. The Pesisir areas are sparsely settled with poor soil for farming, but have witnessed the development of some large transmigrant communities of Javanese near the Jambi and Lampung borders. Pesisir has a very large geographical spread, and therefore it is not surprising to see its lects

possess words from neighboring non-Pesisir varieties in their lexicon.[18] This dialect grouping is considered to be closely related to the Palembang dialects, part of the proposed PALEMBANG-LOWLAND cluster of the MUSI language.

The five sites visited as samples of Pesisir as listed in Table 1.13 are recapitulated below.

VILLAGE	CODE
Lebung Gajah (Tulung Selapan)	CST-LG
Pangkalan Balai	CST-DB
Supat	CST-SP
Simpang Bayat	CST-SB
Sungai Menang	CST-SM

The **Penesak** dialect [previously pen], listed as a language in the 15th edition of the Ethnologue, is spoken in Ogan Ilir and Ogan Komering Ilir. The Penesak number is estimated to be approximately 130,000 speakers. The Penesak dialect is split by the two main rivers in southeast of Palembang, the Ogan and the Komering. The main Penesak speech area is west of the Ogan, located around Tanjung Batu, while the Pedamaran dialect is located to the east of the Komering River and south of Kayu Agung. Penesak has the subdialects of Burai, Penesak, Meranjat, and Pedamaran. The Burai dialect reflects word final **a* as *i*, while Tanjung Batu shows *o*. The Pedamaran informant said that Pedamaran was almost identical to the Penesak spoken in Tanjung Batu and the inhabitants of Pedamaran came from Tanjung Batu as immigrants at some time in the past to settle the area on the other side of the Komering River.[19] This differs from Mitani, who considered Pedamaran to be part of the Ogan language along with "Pegagan Ulu", which appears to be Ogan Ilir, around Rantau Alai. Meranjat is referred to as "high Penesak" by some informants during the survey in the area and is spoken south of Tanjung Batu and north of the Rambang language variety. Penesak is proposed as a dialect in the LOWLAND subcluster (PALEMBANG-LOWLAND cluster) of the MUSI language. The three Penesak sites visited are listed below.

VILLAGE	SUBDIALECT	CODE
Tanjung Batu		PB-TB
Burai	Burai	PB-BR
Pedamaran		PB-PDR

Lematang Ilir [previously lmt] is a dialect spoken in Muara Enim along the lower reaches of the Lematang River as it approaches the Musi. In the 15th edition of the Ethnologue it appears that the language called Lematang is a conflation of Lematang Ilir with Belide. The geographic area for Lematang Ilir was found to be quite small, with few phonological innovations setting it apart from the other nearby LOWLAND subcluster varieties of Penesak and Belide. The reflex for word final **a* is *e* in Tanah Abang Selatan, the site nearest to the UPPER MUSI cluster having the same reflex, and is *i* in Danau Rata, similar to the nearby Burai subdialect of Penesak. Lematang Ilir is proposed as a dialect of the LOWLAND subcluster varieties in the PALEMBANG-LOWLAND cluster of the MUSI language. The two sites are listed below.

VILLAGE	CODE
Tanah Abang Selatan	PB-TA
Danau Rata	PB-SR

[18] Mitani noted that Tulung Selapan (cf. Lebung Gajah) had some HIGHLAND Malay words.

[19] Pedamaran had the highest percentage shared cognate (PSC) with Standard Indonesian of all 83 sites visited with a 91% PSC.

The PBh publication on Lematang, *Fonologi dan Morfologi Bahasa Lematang (FMBL)*, states that Lematang refers to the language spoken along the Lematang River from district Merapi in Lahat regency until the districts of Gunung Megang and Rambang Dangku in Muara Enim regency. The authors also describe two dialects of Lematang Ilir, one centered in the village of Gunung Megang and one in the village of Tanjung, which are both in the Gunung Megang regency (Ihsan et al. 1996:7–8). The authors explain that the variety spoken in Lahat regency (**Lematang Ulu**) is in fact very little different than Besemah, and therefore when they use the term Lematang they are referring to Lematang Ilir, the variety spoken in Muara Enim. This might be better called **Lematang Tengah**, or what the local residents call **Lematang Ilir Ogan Tengah** (an older administrative regency name). The survey visits to the villages along the upper, middle, and lower part of the Lematang River found three different varieties of what people would call Lematang: that spoken in the upstream area (Ulu) and part of the HIGHLAND cluster of BARISAN (Tinggi Hari, Arahan), that spoken in the northern part of Muara Enim and part of the OGANIC group of BARISAN (Penanggiran), and that spoken in the downstream area which is part of the PALEMBANG-LOWLAND cluster of MUSI Malay (Tanah Abang Selatan, Danau Rata).

Table 1.10 Ultimate *i *lowering in Lematang Ilir*

Ultimate High Vowel Lowering	Gunung Megang (PBh)	Tanjung (PBh)	Danau Rata	Tanah Abang Selatan
air	ayeʔ	ayiʔ	ayeʔ	ayeʔ
alir	aler	alir	--	alir
aɲin	aɲen	aɲin	aɲen	--
anjiŋ	anjeŋ	anjiŋ	anjiŋ	anjiŋ
gigit	giget	gigit	--	gigit
lain	laen	lain	lain	lain
Other examples				
kuniŋ	--	--	kuniŋ	kuneŋ
diɲin	--	--	diɲin	diɲen
balik	baleʔ	--	baleʔ	baleʔ

The WIST sites surveyed with complete wordlists and sociolinguistic questionnaires (SLQs) are located far downstream and upstream from the *FMBL* sites, while the one midstream site WIST visited (Penanggiran) located close to the *FMBL* sites was the focus of bilingualism investigation rather than a dialectology study. The WIST downstream sites are both spatially closer to Belide and Purun (Musi | Musi Proper | Penukal) than to the Gunung Megang and Tanjung sites. Both *FMBL* sites are close together at midstream, with the Tanjung site near the Rambang site surveyed, and the Gunung Megang site near the reported language variety "Tamblang" reported on the map for *FMBL* as being in Muara Enim district, on the border with the Musi Rawas regency (Ihsan et al. 1996:130). The two *FMBL* sites have a lexical variation of 12.5% in a 200 item wordlist (conversely a lexical similarity of 87.5%), with the Gunung Megang items exhibiting a preference for lexical items found in the upstream varieties of Malay. Another difference is that Gunung Megang has a diphthong for final *–*ar*, –*aw* (*akaw*), while Tanjung has the apical trill –*ar* (*akar*). Another difference in **r* is that in the medial position, Gunung Megang realizes it as *h (bəhat)* while in Tanjung it is the apical trill *r (bərat*). The two WIST wordlists, from Tanah Abang Selatan and Danau Rata, both have *bərat* and *akar*, which would seem to indicate that the Tanjung dialect was the one sampled based on the reflexes of **r*. However, the book also gives two differences based on vowels that do not seem to line up as well with the evidence gathered in the wordlists. First, Gunung Megang is said to have ultimate high vowel lowering, while Tanjung maintains the high vowel. Secondly Gunung Megang has a central vowel *ə* as the reflex for final **a* while Tanjung shows the high back rounded vowel *u*. The word examples given from the book are shown below in tabular format beside the data obtained. It is clear that the linguistic features which may separate the Tanjung and Gunung Megang dialects are not tightly bundled isoglosses, but appear to have some

diffusion, with the WIST wordlists sites agreeing with Tanjung in the **r* reflexes but show more dissonance as far as vowel qualities. The table below (1.10) shows the words used as evidence in the book for high vowel lowering as well as the WIST sites for comparison's sake.

However, the examples given are for the front unrounded high vowel *i* but not for the back rounded vowel *u*. Looking for some examples with this, it becomes clear that the pattern of high vowel lowering in Gunung Megang is not as clear as it first looked. If a word was not available in the WIST list for a comparison, a comparable word shape was sought (e.g., *buruk*/*busuk*). See Table 1.11.

*Table 1.11 Ultimate *u lowering in Lematang Ilir*

Ultimate High Vowel Lowering	Gunung Megang	Tanjung	Danau Rata	Tanah Abang Selatan
buruk	buhoʔ	buroʔ	busuʔ	busoʔ
garut	gahut	garut	gaot	gaut
perut	pəhut	pərut	pərut	pərut
tumpul	tumpul	--	tompol	tumpul

For the reflexes of word final **a* as seen in Table 1.12, it looks like Tanah Abang Selatan is very close to Gunung Megang with a central vowel. Tanjung and Danau Rata both veer away from this, Danau Rata with the reflex *–e*, found throughout the Musi Proper and Pegagan dialect areas. But if Tanjung's reflex is indeed *–u*, this would mark it as unique among the dialects surveyed. It is more likely that *ɨ* is indicated by the symbol *u*. See Table 6.10 for a list of reflexes of word final **–a*.

*Table 1.12 Final *a in Lematang Ilir*

final *-a	Gunung Megang	Tanjung	Danau Rata	Tanah Abang Selatan
buŋa	buŋə	buŋu	buŋe	buŋɨ
nama	namə	namu	name	namɨ
ɲawa	ɲawə	ɲawu	--	--

The book FMBL appears to use the Gunung Megang dialect as its source, as the examples given through out the book after the initial introduction of Tanjung follow the Gunung Megang patterns. Based on the reports on a visit to Penanggiran, this dialect would fall within the Enim variety of the OGANIC group of BARISAN. Both have the voiceless velar fricative for the reflex of **r* and ultimate high vowel, traits that also characterize Ogan, Enim, and Rambang.

Belide is listed as dialect of Malay [mly] in the 15th edition of the Ethnologue and is spoken to the north of Prabumulih and in a pocket to the north of the MUSI language area in Musi Banyuasin (listed as Lematang in the Ethnologue and shown as Lematang in the Wurm and Hattori map of southern Sumatra). The number of Belide speakers is estimated at 75,000. Belide as a dialect is unique among the SSML lects gathered (except for the peripheral HAJI) in that it retains *a* for word final **a*, with the exception of one site near the UPPER MUSI cluster (Tebing Abang), which has *e* (see §5.6 for a discussion of this innovation). The people in the community of Talang Leban, one of the Belide villages to the north of the Musi Proper area, said that the villages had been established by Belide speakers immigrating from Prabumulih to settle the area. Belide is proposed as a dialect in the LOWLAND subcluster varieties of the PALEMBANG-LOWLAND cluster of the MUSI language. The four Belide research sites are listed below.

VILLAGE	**CODE**
Modong	PB-MD
Cambai	PB-CB
Talang Leban	PB-TL
Tebing Abang	BEL-TA

Rawas [previously rws] is a dialect found in Musi Rawas province. Rawas speakers are estimated at 200,000. The literature available on Rawas stated it was spoken in the Rawas Ilir, Rupit, and Rawas Ulu areas of Musi Rawas. The survey found that what was spoken in Rawas Ilir is actually Musi Proper-speaking, so the actual population is likely somewhat smaller. Rawas as a dialect is spoken upstream from Embacang. The Rawas dialect shares the **ar* > *o* merger found in MUSI, but not the other phonological innovations. Some varieties of Rawas also have deletion of the nasal in nasal voiceless stop consonant clusters. Three Rawas locations were visited and are listed below. Three other locations in Rawas Ilir were investigated to find the reported Rawas variety there but all three were found to be varieties of Musi Proper. The Rawas dialect is proposed as member of the UPPER MUSI cluster of the MUSI language.

VILLAGE	SUBDIALECT	CODE
Muara Rupit	Rupit	RAW-RU
Pasar Surulangun (Sarolangun)	Tengah	RAW-PS
Pangkalan	Ulu	RAW-PN

Musi Proper (labeled as such to distinguish this Musi-named local dialect from the larger UPPER MUSI cluster of the even larger MUSI language) is spoken along the Musi River and its tributaries, starting in Lais and going to the west as far as Rawas Ilir (Embacang), and as far south at BTS Ulu (Cacar). The estimated population of Musi Proper speakers is 600,000. The survey found out what was being called *Musi Banyuasin* (following the geographic label) would fall into the category of Pesisir or LOWLAND subcluster, as it lacked the phonological innovation found in the Musi Proper spoken in Sekayu. The Sekayu subdialect contains the phonological innovations that are found in the other lects of the Musi Proper dialect. The nine Musi Proper sites visited are listed in the table below. One of the subdialects—**Penukal**—is spoken on the border of the Musi Proper and LOWLAND subcluster area and has many of the innovations of Musi Proper, but lacks the **-ri* > *ray* sound change. Another Musi Proper variety in contact with a different dialect is **Kelingi** spoken in Petunang and bordering on the Col language area. This should not be confused with **Sindang Kelingi**, the nearby subdialect of Col. The third MUSI language research point showing influence from contact with a neighboring SSML lect is Embacang, which is close to the **Rawas Rupit** subdialect. The Musi Proper dialect is proposed as a part of the UPPER MUSI cluster of the MUSI language.

VILLAGE	Subdialect	CODE
Petunang	Kelingi	MU-KL
Sadu		MU-SA
Prabumulih Satu		MU-P1
Embacang (Mandi Angin)		MU-EM
Pauh		MU-PH
Bingin Teluk		MU-BT
Sekayu	Sekayu	MU-KY
Purun	Penukal	MU-PEN
Balai Agung	Sekayu	MU-BA

Pegagan is a dialect of UPPER MUSI that is located to the south of Palembang in Ogan Ilir and Ogan Komering Ilir. The number of Pegagan speakers is estimated at 225,000. The Pegagan appear to be a group of UPPER MUSI speakers that migrated further downstream along the Musi River, being separated by a considerable distance from the main UPPER MUSI cluster that ends in Lais by the Lematang Ilir, Belide, and Penesak groups –all composing the LOWLAND subcluster—and bordered by Palembang to the north and east and Ogan to the south. Mitani (1980:14) mentions "Pegagan Ulu" as being part of the Ogan language grouping he proposes along with Pedamaran. This apparently was a reference to the Ogan Ilir language area centered

in Rantau Alai. For the purposes of this report, what Mitani calls "Pegagan Ilir" is what is recognized as Pegagan, based upon the phonological innovations it shares with Musi Proper. Pegagan is proposed as a part of the UPPER MUSI cluster, in the MUSI language. The two Pegagan sites visited are listed below.

VILLAGE	CODE
Pegagan (village not specified)	MU-PG2
Muara Penimbung	MU-PG

Col [previously liw and sdi] was listed under two separate entries in the 15th edition of the Ethnologue. These listings apparently followed the Wurm and Hattori listing of dialects (Lembak and Sindang Kelingi), but promoted these to language status. Their total population is estimated to be approximately 110,000. This is spoken in primarily in Musi Rawas and across the border in Bengkulu in the Bukit Barisan mountain range around Curup and Guru Agung, but there is also a small pocket called Lembak Delapan that is northeast of Bengkulu City, with an estimated population of 25,000 speakers. The listing in Wurm and Hattori is confusing as it conflates **Sindang Kelingi**, a Col dialect, with **Kelingi**, a Musi Proper dialect, and shows it divided from the main body of the Col speaking area, while in fact it is contiguous, with no Musi Proper speaking areas intervening. The Col lect is referred to as **Lembak** in the small area northeast of Bengkulu, but this name is not recognized by most of the other Col speakers, though some have heard of Saling, another ethnonym derived from the eponymous river. The PBh material on them uses the name Sindang Kelingi, with others referring to the language by place names such as Lubuk Linggau and Muara Beliti.[20] The autoglottonym for this group is most commonly *Col*/*Cul*, their word for 'none'. Col shares many of the phonological innovations of Musi Proper, but the comprehension of Col by Musi Proper speakers is reportedly low.

A test of a Musi Proper recorded text in a Col area showed extremely high comprehension by Col speakers who had not come into contact with Musi Proper before, indicating high inherent (though asymmetric) comprehension. Therefore, Col is proposed as a member of the UPPER MUSI cluster of the MUSI language. The six Col research sites are listed below.

VILLAGE	SUBDIALECT	CODE
Jukung		COL-LL
Taba Dendang		COL-TT
Terawas		COL-BKL
Guru Agung		COL-PUT
Pondok Kubang	Lembak	COL-L8
Pelajau	Lembak	COL-PLJ

1.7.2 South Barisan Malay language

Rambang is a variety of OGANIC spoken between the Ogan and Enim rivers, and is located between the dialects of the same names. There are an estimated 140,000 speakers of Rambang. It is part of the OGANIC cluster of the BARISAN language. Pak Iskandar, village head of the Rambang town of Penyandingan reported the following subdialects of the Rambang dialect, including Rambang Dangku (Niru, Perbuli, Lahat), Rambang Niru (Prabumulih, Baturaja), Rambang Kapak Tengah (Pagar Alam) and Anak Rambang Senuling (Pegagan Ulu Suku I in OKI). The PBh publication on Ogan gives the Rambang subdialects as Bingin, Rambang, Lubai, and Tanjung Rambang (Wahab et al. 1990:18). The four Rambang sites researched are listed below, and showed high degree of similarity lexically and phonologically, with

[20] The PBh information on "Lembak" (Aliana, Nursato, Siti Salamah Arifin, Efendi, et al. 1985) refers to one of the dialects of LOWLAND subcluster within the MUSI language.

Penyandingan being the most different and also the furthest removed from the main Rambang area.

VILLAGE	CODE
Penyandingan	RAM-PNY
Tambangan Rambang	RAM-TR
Karangan Bindu	RAM-RK
Jemenang	RAM-RD

Enim [previously eni] is spoken on the Enim River from north of Karya Nyata until Muara Enim. Enim is estimated to have 100,000 speakers. The PBh *Geografi Linguistic Bahasa Enim* listed four different dialect groupings for Enim (Naning et al. 1998), but also divided them into an upstream (Sugih Waras) and a downstream dialect (Lawang Kidul). Lawang Kidul is spoken by the communities of Lingga, Tanjung Enim, Kebon Agung, Darmo, and Seleman. The Sugih Waras dialect reportedly is used in Tanjung Karangan, Tanjung Agung, Muara Emil, Matas, Lesung Batu, Embawang, Padu Raksa, Lebak Budi, Sukaraja, Pandan Dulang, Sugih Waras, Padang Bindu (different than the Padang Bindu for Benakat), Lubuk Nipis, Indramayu, and Bedegung. Enim is proposed as part of the OGANIC cluster in the BARISAN language. The three Enim sites visited are listed below.

VILLAGE	SUBDIALECT	CODE
Tanjung Raja	Ilir	EN-ME
Muara Emil	Tengah	EN-TAS
Indramayu	Ulu	EN-TAN

Ogan [previously ogn] is one of the language varieties in South Sumatra Province, with Ogan communities along the Ogan River from Rantau Alai to as far upstream as the Bukit Barisan mountains. Their estimated population is 150,000. The Ogan language variety has many books published about it by PBh, but these are a bit misleading in that they often place Ogan with other varieties near it. The PBh dictionary for Ogan places Pegagan and Ogan together in one group (perhaps because of geographic proximity) but admits these have different pronunciation. The preface says that due to limited time and the fact that most of the researchers spoke Ogan, the data is all from Ogan, not Pegagan (Aliana, Arif, et al. 1985:ix–x). The PBh volume on dialects of Ogan places Ogan with Rambang, Pegagan, and Penesak (Wahab et al. 1990:16–18). The placement of Ogan with Rambang makes sense as some of the Rambang and Ogan that were surveyed saw very little difference between their culture and language and felt they were fairly similar. However, Penesak and Pegagan both have differences in language and culture with Ogan. The Ogan dialect is reportedly one of the easiest for speakers of other varieties to understand and use, pointing to its possible centrality. One reason it may be understood so easily is because it does not have many phonological innovations and its lexicon contains much of the core vocabulary used by other groups in South Sumatra. The Ogan dialect is proposed as a part of the OGANIC cluster in the BARISAN language. The four Ogan sites visited are listed below.

VILLAGE	SUBDIALECT	CODE
Rantau Alai	Ilir	OG-RA
Pengaringan	Tengah	OG-BR
Damar Pura	Ulu	OG-DP
Belandang	Ulu	OG-UO

Benakat is a dialect found in the Gunung Megang district of Muara Enim, spoken in five villages along the Benakat River (Arifin et al. 2001). Population as given in the PBh publication *Struktur Bahasa Benakat* (Arifin et al. 2001) is approximately 15,000, with around 40% living

outside of the area. Benakat is located where the LOWLAND subcluster and UPPER MUSI cluster of the MUSI language meet BARISAN. It is proposed as a member of the HIGHLAND cluster of the BARISAN language most closely linguistically related to Semenda. It also seems to be in a process of adding many of the innovations from Musi Proper. Lexically it is most similar to the nearby Rambang variety in Jemenang (87%) and Enim's Tanjung Raja (88%). The PBh book reports that Benakat is dropping out of use and being replaced by Indonesian and Palembang Pasar. The survey team did not visit all five villages, but the village visited was given a SLQ in which respondents stated that Benakat was still used in the domestic sphere, and that there was not a change underway where young people used a different language than their parents. It is interesting to note that this is one of the smaller populations of speakers of a dialect, and a majority of them work outside of the area (*merantau*) at some point in their lives. This would indicate a greater likelihood of rapid change in their linguistic situation.

VILLAGE	CODE
Padang Bindu	PB-BN

Bengkulu [previously bke] is the dialect spoken in and around Bengkulu City. The estimated number of speakers of Bengkulu is 60,000. There is a reported variance between the variety spoken in the city and that which is spoken in the villages which are said to be more original (*asli*). The authors of the PBh publication *Pemetaan bahasa daerah di Sumatra Barat dan Bengkulu (PBDSBB)* (Kasim et al. 1987) included two wordlists with the Bengkulu dialect. The survey also took a wordlist from a college educated male student from Bengkulu who was a native speaker of this variety. Variation was noted between all three lists in the amount to which MINANGKABAU innovations were evident and realization of **r*. The variation appears to be due to which variety of Bengkulu is being used; with more urban Bengkulu as a mesolect, the use of the apical trill and a more typical Indonesian phonology is present, where as the basilectal form contains more of the MINANGKABAU innovations and retains the voiced uvular fricative *ʁ*. Bengkulu is proposed as member of the HIGHLAND cluster of the BARISAN language. Lexically it is very similar to Padang (89%).

VILLAGE	SUBDIALECT	CODE
Bengkulu City	Bengkulu Indonesian	BNGKL

Lintang [previously lnt] is a dialect spoken in the highlands of the Bukit Barisan mountains of South Sumatra, around Lahat. The number of Lintang speakers is estimated at around 190,000. This dialect is very similar to **Besemah** and Besemah's Kisam dialect,[21] two neighboring lects, but markedly different than the **Col** language spoken to the north. It is proposed as a member of the HIGHLAND cluster of the BARISAN language. The three Lintang towns visited are listed below.

VILLAGE	CODE
Batu Galang	BES-MP
Lubuk Puding Baru	BES-UM
Terusan Baru	BES-TT

Besemah [previously pse in a narrower denotation] is proposed as the reference dialect for the HIGHLAND cluster of the BARISAN language. Within Besemah are the subdialects of Pasemah, Kisam, and Kikim.[22] The speakers of Besemah are estimated to be approximately

[21] One Lintang village head mentioned the acronym LiKiPaLe (Lintang, Kisam, Pasemah, Lematang) in relation to the varieties that were related.

[22] Jaspan, in his 'Pasemah' article (Lebar 1972:34–35) lists two additional "ethnolinguistically kindred groups", namely Gumai (Gumay) and Mekakau (Makakau). We are not aware of linguistic descriptions of either group specifically, but we would be quite surprised if either stood out

330,000, while the speakers of the Kikim variety, which includes Kisam, are estimated to be 80,000. This gives an estimated total of 410,000. The surrounding lects of Serawai, Semenda, and Lintang see their language as having originated in Besemah. For more about the use of the name Besemah rather than Pasemah, see William A. Collin's doctoral dissertation, *Besemah concepts: a study of the culture of a people of South Sumatra* (1979). The Besemah have a very strong sense of ethnolinguistic self identity and reportedly have a harder time 'hiding' their accent than other HIGHLAND speakers such as Serawai. The five Besemah sites visited are listed below.

VILLAGE	SUBDIALECT	CODE
Sukaraja Kisam	Kisam	BES-MDK
Sukaraja	Kikim	BES-KT
Pematang Bango		BES-PA
Muara Sindang		BES-AK
Lawang Agung		BES-BK

Lematang Ulu is not the variety referred to in earlier editions of the Ethnologue as Lematang [previously lmt], which refers to **Lematang Ilir** and **Belide**, which are in the LOWLAND subcluster of the PALEMBANG-LOWLAND cluster of the MUSI language. This variety is spoken along the upper reaches of the Lematang River in the Bukit Barisan Mountains. There are an estimated 275,000 speakers of Lematang Ulu. Lematang Ulu is proposed as part of the HIGHLAND cluster of the BARISAN language on the basis of high comprehension and a shared ethnic identity. The two Lematang Ulu villages visited are listed below.

VILLAGE	CODE
Tinggi Hari	LT-PP
Arahan	LT-MR

Semenda [previously sdd] is a dialect spoken in the Bukit Barisan Mountains of South Sumatra, Bengkulu, and Lampung. The Semenda are estimated to have 230,000 speakers, with the majority in South Sumatra Province. The heartland of the Semenda is southwestern corner of Muara Enim stretching down to western Ogan Komering Selatan. This area used to be called LIOT (Lematang Ilir Ogan Tengah) but has since been divided to form new regencies. The Semenda people have also moved to settle areas further south in Bengkulu and in Lampung (Saleh et al. 1979:7–10, 16). Semenda has varieties that have reflexes of final **a* as the central vowel *ɨ* and *ə*, which can be seen in how some varieties of Semenda are spelled Semende. Perhaps from speakers of Palembang Pasar they have also been called Semendo, though the WIST wordlists for Semenda, as well as other wordlists, did not show any Semenda villages that have *o* as a reflex of final **a*. Below are the six Semenda sites visited.

VILLAGE	SUBDIALECT	CODE
Muara Sindang Tengah	Semenda	SEM-PB
Karya Nyata	Semenda	SEM-SDL
Penyandingan	Semenda	SEM-PNY
Muara Dua (Pulau Panggung)	Semenda	SEM-MD
Bandar Agung	Semenda	SEM-BA
Tanggamus (map: Muara Dua village in Lampung)	Semenda	SEM-LP

significantly from the HIGHLAND cluster. Sakai (2006:39) estimates Gumay population as around 10,000.

Semenda is proposed as a dialect of the HIGHLAND cluster of the BARISAN language on the basis of high reported comprehension with other lects that composed the group. Semenda is the HIGHLAND variety that borders both Enim and Ogan and both of these OGANIC group varieties report good comprehension of Semenda. The PBh publication *Bahasa Semende* helpfully notes that there are two dialects of Semenda, Pulau Panggung and Ulu Enim. Pulau Panggung is futher subdivided into the subdialects Pulau Panggung City, Tanjung Laut, and Perapau (Saleh et al. 1979:9). Of the sites visited, it appears that those around Muara Dua and Karya Nyata would fall into the Pulau Panggung regional dialect, while those that are part of Muara Sindang Tengah, Bandar Agung, and Tanggamus would fall into the Ulu Enim dialect area. According to *Bahasa Semende*, the Ulu Enim dialect is said to be almost identical to Besemah. Checking this against the lexical evidence (see §4.1.2), it is, in fact, the Pulau Panggung site (SEM-MD) that scores a higher cognate rate with all varieties of Besemah than all other varieties of Semenda. Another PBh publication *Morfologi dan sintaksis bahasa Semende* (Saleh et al. 1985:3–4) gives two different dialects with different names: Semenda Darat and Semenda Lembak. Speakers of Semenda Darat are found in Muara Enim, while speakers of Semenda Lembak are found in Ogan Komering Ulu (now Ogan Komering Ulu and Ogan Komering Ulu Selatan). There is supposedly little difference in these dialects. Again, the wordlists taken by WIST would break out in the same way, with Tanggamus, Bandar Agung, and Muara Sindang Tengah being part of the Semenda Lembak dialect area, with Penyandingan and Muara Dua as part of the Semenda Darat area.

Serawai [previously srj] is a dialect spoken in Bengkulu Province, a proposed member of the HIGHLAND group in the BARISAN language. The number of Serawai (also spelled Seraway) speakers is estimated at 290,000. There is reportedly high comprehension between Serawai and the neighboring **Besemah** and **Bengkulu** lects, and lower comprehension between Serawai and the **Kaur** lect. Locals delineate two subdialects of Serawai: a northern one (Talo) marked by final **a* going to *o* and a southern one marked by the final **a* going to the diphthong *aw* (Manna). The break between the two is the border of the regency of Semidang Alas to the northwest (*o*) and the regencies of Pinoraya and Pino to the southeast (*aw*). The Serawai village was visited that spoke the Manna subdialect.[23] In addition, four Serawai wordlists from *Pemetaan basaha daerah di Sumatra Barat dan Bengkulu* were used in analysis.[24] The Serawai village visited is listed below.

VILLAGE	SUBDIALECT	CODE
Napal Melintang	Manna	SRW-NM

Pekal [pel] is a dialect/language spoken in northern Bengkulu, in the Ketahun district of Bengkulu Utara regency. It is bordered by the Muko-Muko dialect to the northwest, the Rejang language to the southeast, and the Jambi, Col, and Rawas lects to the north. There are an estimated 30,000 speakers of Pekal. It is listed as a dialect of MINANGKABAU in Wurm and Hattori (1983) and as a separate language in the 15th edition of the Ethnologue. According to the PBh publication *Struktur Bahasa Pekal*.[25] Pekal is an ethnolinguistic group that is reportedly an assimilation of the Muko-Muko with the Rejang. There are some linguistic items from Rejang and phonological innovations from MINANGKABAU that give some credence to this assertion, but Pekal sites visited viewed themselves as ethnically Malay. Speakers of the

[23] As with many of the 'dialects' listed here, it probably could not be proven that there is 'a Serawai dialect' from a linguistic perspective. We are blessed with eight 'Serawai' wordlists; their (slim) differences, like those of the wider dialect network beyond Serawai, are marked by cross-cutting innovations. Innovations in **h* and **r*, for example (see §5.1 and §5.3), do not track with reflexes of **a*.

[24] The list used by Adelaar (1992) is from the *o* subdialect. That list is exceptional among the eight Serawai lists we have seen in that (as far as the non-IPA orthography can be interpreted) it is the only list with devoiced **r*.

[25] Which is citing *Monografi Propinsi Bengkulu* (Dinas Pertanian Rakyat 1975).

neighboring MINANGKABAU dialect Muko-Muko and the BARISAN dialect of Bengkulu said they have a hard time understanding Pekal, but Pekal reportedly can understand Bengkulu. Based on this identity and the ability to comprehend a Besemah recorded text, the Pekal are proposed as part of the HIGHLAND cluster of BARISAN. The Pekal village visited is listed below. This data was compared against the two wordlists for Pekal (data points 24 and 26) found in *Pemetaan basaha daerah di Sumatra Barat dan Bengkulu* (Kasim et al. 1987).

VILLAGE	CODE
Pulau Baru	PKL-PL
Napal Putih	PKL-NP

Kaur [vkk] is a Malayic dialect/language spoken in southern Bengkulu (referred to as Mulak in Wurm and Hattori). There are an estimated 40,000 Kaur speakers. The Kaur speaking area lies to the south of the Besemah speaking area in Bengkulu and north of the Nasal and Lampung speaking areas. Two dialects are reported by Indonesia Pelangi Nusantara: the southern Bintuhan, which is influenced more by Lampung, and North Kaur, which exhibits more Besemah influence (2010:45). Kaur fits into the HIGHLAND cluster of the BARISAN language, with speakers of Kaur reporting good comprehension of Serawai and demonstrating good comprehension of the Besemah recorded text. Two Kaur sites were visited where a full and a partial wordlist were taken. For analysis these were used in addition to the Kaur Holle list and three other wordlists obtained from *Pemetaan basaha daerah di Sumatra Barat dan Bengkulu* (data points 41, 43, 44).

VILLAGE	SUBDIALECT	CODE
Bintuhan	Bintuhan	KAU-BN
Jembatan Dua	Bintuhan	KAU-J2

1.7.3 Other groups

Muko-Muko [previously vmo] is a dialect of MINANGKABAU spoken in northern Bengkulu. There are an estimated 50,000 Muko-Muko speakers. The Muko-Muko area lies south of Sumatra Barat and the MINANGKABAU area, to the east of Kerinci and Rejang speaking areas, and to the north of Pekal and Bengkulu speaking areas. During the survey it was found that the Muko-Muko speakers do not understand Pekal speakers and that the Muko-Muko view themselves as being part of the MINANGKABAU. This language variety was proposed as a merger into the MINANGKABAU language. The Muko-Muko language data was obtained from *Pemetaan basaha daerah di Sumatra Barat dan Bengkulu* with five data points analyzed (data points19-23) and from the two sites visited.

VILLAGE	CODE
Pondok Lunang	MUK-PL
Suka Pindah	MUK-SP

Kubu [kvb] is a Malayic dialect/language spoken in southern and central Sumatra. There are an estimated 10,000 Kubu speakers. The Kubu (Anak Dalam, Rimba) have traditionally been a nomadic, forest dwelling people, making their enumeration for any time of census particularly difficult. Another difficulty is that this name is often used as a cover term by outsiders for people who may have different ethnic identities who have a nomadic scavenging lifestyle. According to the PBh study on the Kubu language, the Kubu formerly lived as "vagabonds and nomads",[26] living very simply, but now are becoming more influenced by surrounding populations and becoming more sedentary (Dunggio et al. 1985). According to

[26] Here and elsewhere, English-language quotations of Indonesian works reflect our own translation.

different ethnologists, they still rely heavily on what they are able to obtain from the forests and rivers of the natural environment (Weintré 2003; Sandbukt 1995). The increasing number of transmigrants and other local Malay groups competing for the ever-decreasing forest resources such as rattan, turtles, and other rare animals that can be sold has increased the pressures on the traditional Kubu way of life. The Kubu, according to the Ethnologue (15), do not like being called the *Kerinici*. It is not clear why someone would use this appellation, as the *Kerinci* is a separate Malayic group. According to the PPPB PBs study on Kubu dialectology (Maryono, Setyonegoro & Kusmana 1997), they do not like being called the Kubu, but prefer the name given them by the government, *Suku Anak Dalam*. Weintre and the Wikipedia site give the name *orang Rimba* (forest people) as another name for the Kubu. In the village of Sungai Kijang that the survey team visited, the Kubu speakers said that they preferred the term *Suku Anak Dalam* (the government given ethnonym) or called themselves *Anak Perdalaman*. Others have called them *orang Hutan* (forest people). The largest portion of the Kubu population lives in Jambi. Most surrounding groups don't see a relation between themselves and the Kubu. Likewise, the Kubu see themselves as having different cultural traditions than other Malay groups. Therefore, Kubu was recommended to be retained as a separate language listing.

VILLAGE	**CODE**
Sungai Kijang	KUBU

The next section (philosophy and goals) lays out the goals, questions and philosophy which guided our research and our write-up, while the section which follows (methodology) explains what we did and when we did it.

2 Philosophy and goals

This report is the result of background research and field survey focusing on the Malayic lects of southern Sumatra. The primary sources for background research on these Malayic lects are the PBh monographs, the Holle lists, wordlists collected and shared by other linguists, a limited number of linguistic reports on language varieties in southern Sumatra, and interviews of those knowledgeable about the language situation in southern Sumatra. Those areas highlighted as possessing a unique lect were then visited and investigated in the field research phase. SIL's West Indonesian Survey Team (WIST) conducted survey in both urban locations and village settings in Lampung, South Sumatra, and Bengkulu provinces in order to measure similarity and record reported comprehension between the Malayic dialects of southern Sumatra. Research in investigating the dialects focused on gathering linguistic, sociolinguistic, and development-related data related to the team's goals. This report summarizes the findings of the five language survey trips which took place from 2003-2007. The trips consisted of rapid appraisal survey, with an emphasis on dialectology and gaining a better initial understanding of the Malayic lects of southern Sumatra. The research trips were conducted under the auspices of *Lembaga Ilmu Pengetahuan Indonesia* (LIPI) from 2003-2006 and Badan *Pemberdayan Masyarakat dan Desa* (BPMD) from 2006-2008. (See Figure 2.3 and Table 2.13).

2.1 Philosophy of Rapid Appraisal research

The methodology and fieldwork of these surveys were based upon a model of Rapid Appraisal research developed specifically for language assessment. " 'Rapid appraisal', often called a 'level one' survey, is an initial overview of an area, language, ethnolinguistic group, or multilingual society. It develops the broad strokes of understanding of the situation" (Bergman 1991). "The key objective of this method is to formulate hypotheses to be tested in a more in-depth survey or language assessment" (Wetherill 1995).

Some may ask why Rapid Appraisal research in this area is necessary, given that previous research has been done. Most of the previous studies focused on individual lects rather than taking a broad look at area as a whole. Mitani (1980) is one of the few exceptions, but that study

(mentioned above) only considered lexicostatistic evidence, and that without a comparative foundation or published data. Our team felt that it was important to consider other factors, such as comparative, sociolinguistic and comprehension evidence. Some may press further and ask why our research stops at the admittedly sketchy level of Rapid Appraisal and does not go deeper into more conclusive research. Rapid Appraisal research is limited in scope and depth, mainly because it is limited in time. The goal is to gain broad and basic understanding of large areas in short amounts of time. This research is foundational and is meant to be followed by more focused, indepth research and analysis. Suggestions for further research are given in §8.2.

2.2 Language identification and reality

"All models are wrong but some are useful" (Box & Draper 2007)

Given SIL's historic stewardship of the language register found in the Ethnologue (see §1.5.2), one of the typical tasks of an SIL survey team could be labeled *language identification* - which languages are spoken where, by how many people? Which dialects belong to the language? The result is a listing in the Ethnologue (or elsewhere) of languages in a given region: 112 in Sulawesi, 14 in Kazakhstan. Language maps show in pretty colors where one language stops and another starts.

It has long been recognized, however, that reality is not that simple. Despite Bloomfield's (1935) confident explanation that that languages are mutually incomprehensible lects while dialects are mutually comprehensible lects, drawing language boundaries in real-life situations poses a number of problems. Hymes (1974:123 as quoted by Kluge 2007) observes that "what counts as a language boundary cannot be defined by any purely linguistic measure. Attitudes and social meanings enter in as well." Political realities intervene. Sometimes intelligibility is one-way. Different (or no) language names are provided by the people (Mühlhäusler 1996:6). Dialect continua like SSML allow multiple options for slicing and dicing. Haugen (1972a:237) noted that the terms 'language' and 'dialect' "represent a simple dichotomy in a situation that is almost infinitely complex".

As mentioned above, whenever we write 'X language', we intend to signify 'a cluster of related dialects whose outer boundaries are boundaries of comprehension with speakers outside the cluster'. Although we by necessity use a rather rigid system for labeling 'dialects', 'clusters' and 'languages', we acknowledge that reality is significantly more complex (in a number of dimensions) than whatever system we might use. But our hope is that our report and the classification system within will at least provide a relatively useful model for understanding some aspects of southern Sumatran Malay.

2.3 Area of study

Southern Sumatran Malay (SSML) is, for this report, defined as the vernacular Malay lects spoken in the provinces of South Sumatra, Lampung, and Bengkulu. The rationale for restricting the scope of the report to these areas and not including other surrounding language varieties is the significant differences between SSML and the neighboring varieties of Jambi Malay and MINANGKABAU.[27] The SSML lect HAJI is significantly different and treated elsewhere (Anderbeck 2007f). The islands of Bangka and Belitung and their languages are similarly not included in the dialects treated here (cf. Nothofer 1997 for an indepth study of Bangka Malay). Finally, the lects spoken by the Anak Dalam tribe, often labeled 'Kubu', are only briefly treated in this report, as the majority of speakers resides in Jambi Province (see Maryono et al. (1997) and Anderbeck (2008) for discussions on Kubu dialectology).

[27] For differences and relationships between Jambi and South Sumatran Malay, see Anderbeck (2008:160). MINANGKABAU exerts an influence on the lects of northern Bengkulu, but the differences are discussed in the phonological and sociolinguistic sections of this report. Muko-Muko was also visited but found to be part of the MINANGKABAU dialect cluster.

2.4 Goals

2.4.1 Broad questions

The broad questions of this study are:

- What clusters (and subclusters) of SSML dialects emerge from the analysis of data from lexicostatistics, historical-comparative, questionnaires and Recorded Text Tests?
- What are the sociolinguistic dynamics of the speech communities (language attitudes, use, stability, change, and death)?
- Which variety or varieties of SSML might be most amenable for development?

In order to answer these questions, several goals were set. They are listed below.

2.4.2 Identify and cluster SSML lects

This goal is to identify all significant SSML lects and determine optimal clustering into higher-level groupings (whether Mitani's Highland and Lowland groups or something different). The research questions are:

- What is the estimated level of lexical and phonological similarity between SSML lects?
- What is the approximate degree of comprehension (whether inherent or acquired) between lects?
- How can one describe attitudes and self-perception regarding ethnic identity and inter-dialectal comprehension and receptivity?

2.4.3 Language vitality and multilingualism goals

The goals in this area are:

- Investigate patterns of language use within SSML, for the purpose of determining language shift and/or stability, toward predicting language vitality.
- Investigate interaction between ethnic/dialect groups (that is, the interaction of SSML peoples with other SSML peoples and with non-SSML peoples) to determine levels of multilingualism and language vitality.
- Observe aspects of local culture, economy and lifestyle that influence language use and vitality.

2.4.4 Language planning goals

The goals in the area of language planning are:

- Discover the linguistic factors that could influence language development and language program planning within the region.
- Investigate language attitudes in the area in order to determine the perceived need for language planning. Investigation of language attitudes should also give indications as to which lect(s) could serve as a reference dialect for a language development program.
- Discover other sociolinguistic factors (such as literacy levels and attitudes toward literature development) that could influence language planning.

2.4.5 Training goals

As our journeys nearly universally included researchers for whom this was their first field experience, we also had some training goals (not discussed further in this report):

- Orient trainees to Sumatra, in the areas of culture and language, specifically the multilingual and sociolinguistic dimension.
- Train trainees in wordlist collection.
- Provide experience for trainees in gathering sociolinguistic questionnaires and making and recording observations.

- Train trainees to develop and use a Rapid Appraisal Recorded Text Test (RA-RTT) to determine mutual comprehension between language varieties.

3 Methodology

Several different linguistic and sociolinguistic tools were used to fulfill the research goals stated above. The initial field research consisted of four parts: transcribing and recording a wordlist, administering a sociolinguistic questionnaire (SLQ), collecting an inventory of sentences, and interviewing residents about their language and culture. The follow-up field research consisted of testing dialect groupings through the use of the Rapid Appraisal Recorded Text Test (RA-RTT). Full 358-item wordlists (or in some cases partial wordlists) were collected, sociolinguistic questionnaires (SLQs) were administered, and an inventory of 23 sentences was gathered. The interviews were used to check and reject hypotheses of language groupings and learn more about the cultural history of the speech community being visited. The SLQ contained questions concerning use of language in different domains, language maintenance, dialects, comprehension of other Malayic lects, and attitudes towards their own lect. The results from the SLQ were used in grouping speech communities according to the identity and response of the speakers. The 358 item wordlists were collected from adult speakers who were native to their respective language areas with limited time outside of their areas. The basic Austronesian (BAN) portion of the list (Blust 1981) was used in the lexicostatistical comparison of lects (see §4), and the entire list was used in the comparison of phonological phenomena (see §5). The sentence inventory has the potential to be used to explore grammatical differences between the lects but no extensive analysis has been done at this time. A cursory analysis of the same 23 sentences elicited in numerous locations shows a high degree of uniformity. The RA-RTT was conducted at ten locations using three recorded stories in dialects of the three Malayic languages under investigation to observe how well speakers of other dialects understood these dialects and test hypotheses regarding grouping of dialects (see §6).

More detailed explanations of these tools, methodologies and data gathered with them are given in their respective sections below.

3.1 Research sites

The total sites visited in Lampung, Bengkulu, and South Sumatra number ninety. Sites where at least a 200-item BAN wordlist were taken by WIST total seventy-six, although partial wordlists were taken in most of the remaining sites.[28] An additional four wordlists from the Palembang area (Tadmor 2001) and one Holle wordlist from Kaur (Stokhof 1987) were used in the analysis, giving a total of eighty-one BAN wordlists. Sentences were elicited in sixty-nine sites and sociolinguistic questionnaires (SLQs) in seventy-two. Additional wordlists from Bengkulu and West Sumatra,[29] Jambi, Bangka, Belitung, and other Malayic speaking regions were also used for comparison. The information obtained from these research sites was used in the lexicostatistic, phonological, and sociolinguistic analysis. See Table 3.1 for a complete chronological listing of the research sites and the activities performed at these sites and Figure 3.1 for a graphic representation of research sites.

[28] The full (2005 revision) WIST wordlist template is provided as a worksheet in the linked wordlist materials (McDowell and Anderbeck 2020), as are the full wordlists including metadata.

[29] From Pemetaan Bahasa Daerah di Sumatra Barat dan Bengkulu (Kasim et al. 1987) came 44 word lists, some of which were used in various parts of the analysis.

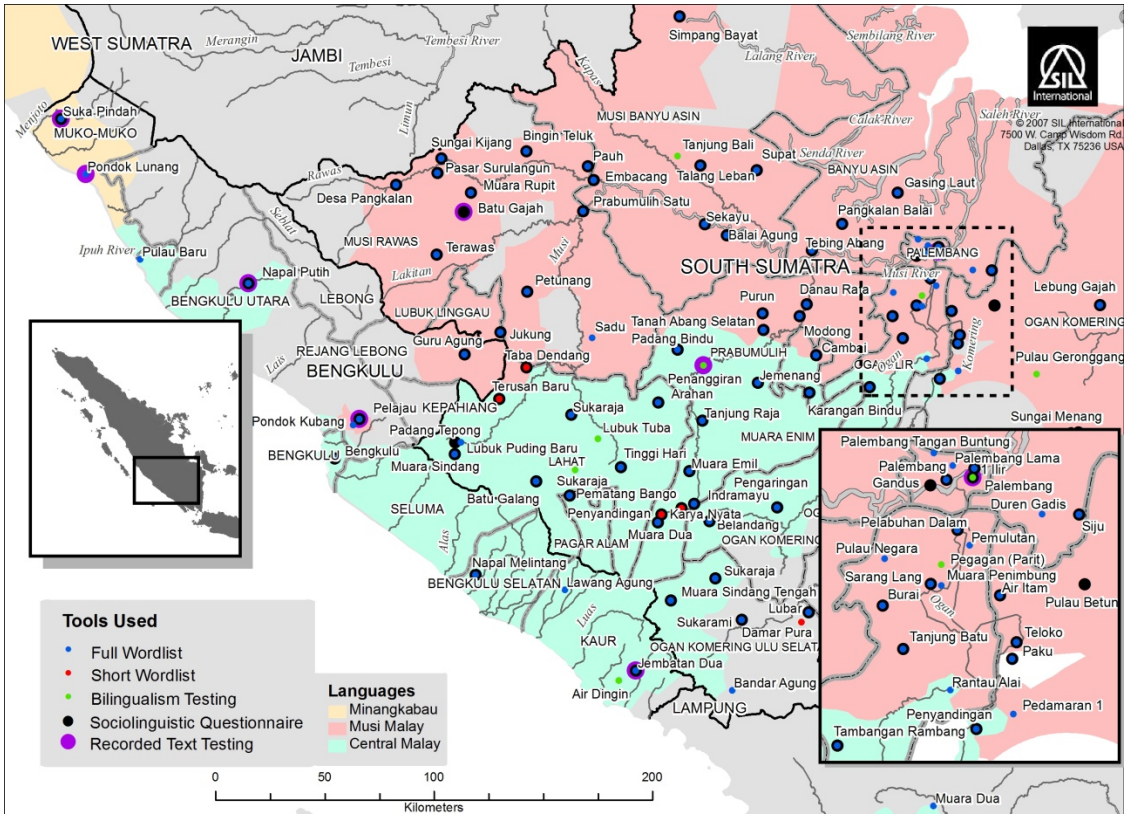
Table 3.1 List of research sites visited in southern Sumatra, with instruments used

Date	Village	Dialect	Code	SLQ	Wordlist	Sentences
04-Sep-03	Bandar Agung	Semenda	SEM-BA	no	full	yes
24-Sep-04	Paku	Palembang Lama	PL-PAK	yes?	partial	yes
27-Sep-04	Sukarami	HAJI	HAJ	yes	full	yes
28-Sep-04	Lubar	HAJI	--	yes	no	no
26-Sep-04	Damar Pura	Ogan Ulu	OG-DP	yes?	1-178	yes
29-Sep-04	Penyandingan	Rambang	RAM-PNY	no	full	yes
30-Sep-04	Teloko	Palembang Lama	PL-TEL	no	full	yes
25-Apr-04	Palas Aji	HAJI	--	yes	no	no
03-May-05	Tanggamus	Semenda	SEM-LP	yes	full	no
26-Jun-06	Pondok Kubang	Col Lembak	COL-L8	no	full	yes
21-Sep-06	Tanjung Batu	Penesak	PB-TB	yes	full	yes
21-Sep-06	Saka Tiga	Penesak	-	yes	no	no
21-Sep-06	Muara Penimbung	Pegagan	MU-PG	yes	full	yes
22-Sep-06	Burai	Penesak Burai	PB-BR	yes	full	yes
22-Sep-06	Tambangan Rambang	Rambang	RAM-TR	yes	full	yes
22-Sep-06	Karangan Bindu	Rambang	RAM-RK	yes	full	no
23-Sep-06	Padang Bindu	Benakat	PB-BN	yes	full	yes
23-Sep-06	Jemenang	Rambang	RAM-RD	yes	full	yes
23-Sep-06	Pengaringan	Ogan Tengah	OG-BR	yes	full	yes
24-Sep-06	Batu Galang	Lintang	BES-MP	yes	full	yes
25-Sep-06	Belandang	Ogan Ulu	OG-UO	yes	full	yes
25-Sep-06	Tinggi Hari	Lematang Ulu	LT-PP	yes	full	yes
25-Sep-06	Lubuk Puding Baru	Lintang	BES-UM	yes	222	yes
25-Sep-06	Sukaraja Kisam	Besemah	BES-MDK	yes	full	yes
26-Sep-06	Indramayu	Enim Ulu	EN-TAN	yes	full	yes
26-Sep-06	Arahan	Lematang Ulu	LT-MR	yes	full	yes
26-Sep-06	Muara Sindang Tg.	Semenda	SEM-PB	yes	full	yes
27-Sep-06	Purun	Musi Penukal	MU-PEN	yes	full	yes
27-Sep-06	Tanah Abang S.	Lematang Ilir	PB-TA	yes	full	yes
27-Sep-06	Muara Emil	Enim Tengah	EN-TAS	yes	full	yes
27-Sep-06	Karya Nyata	Semenda	SEM-SDL	yes	50	no
27-Sep-06	Penyandingan	Semenda	SEM-PNY	yes	75	yes
28-Sep-06	Modong	Belide	PB-MD	yes	full	yes
28-Sep-06	Jukung	Col	COL-LL	yes	full	yes
28-Sep-06	Danau Rata	Lematang Ilir	PB-SR	yes	full	yes
28-Sep-06	Tanjung Raja	Enim Ilir	EN-ME	yes	full	yes
28-Sep-06	Sukaraja	Besemah (Kikim)	BES-KT	yes	full	yes
29-Sep-06	Cambai	Belide	PB-CB	yes	full	yes

Date	Village	Dialect	Code	SLQ	Wordlist	Sentences
29-Sep-06	Taba Dendang	Col	COL-TT	yes	100	yes
29-Sep-06	Rantau Alai	Ogan Ilir	OG-RA	yes	1-270	yes
29-Sep-06	Terusan Baru	Lintang	BES-TT	yes	100	yes
29-Sep-06	Pematang Bango	Besemah	BES-PA	yes	full	yes
29-Sep-06	Muara Dua	Semenda	SEM-MD	no	full	no
30-Sep-06	Terawas	Col	COL-BKL	yes	full	yes
30-Sep-06	Pelabuhan Dalam	Palembang Lama	PB-PM	yes	full	yes
30-Sep-06	Muara Sindang	Besemah	BES-AK	yes	full	yes
1-Oct-06	Guru Agung	Col	COL-PUT	yes	210	yes
22-Mar-07	Pulau Betung	Palembang Pasar	PB-BET	yes	1-122	no
22-Mar-07	Ulak Depati	Pesisir	--	yes	no	no
22-Mar-07	Air Itam	Palembang Pasar	PB-AH	yes	full	no
23-Mar-07	Gasing Laut	Palembang Pasar	PB-GL	yes	full	no
23-Mar-07	Gandus	Palembang Pasar	--	no	no	yes
24-Mar-07	Palembang city	Palembang Pasar	--	no	no	yes
26-Mar-07	Palembang City 1	Palembang Pasar	PB-SH1	yes	full	yes
27-Mar-07	Lebung Gajah	Pesisir	CST-LG	yes	full	yes
27-Mar-07	Prabumulih Satu	Musi Proper	MU-P1	yes	full	yes
28-Mar-07	Sekayu	Musi Sekayu	MU-KY	yes	full	yes
28-Mar-07	Palembang City 2	Palembang Pasar	PB-SH2	yes	full	yes
29-Mar-07	Muara Rupit	Rawas Rupit	RAW-RU	yes	full	no
21-May-07	Pangkalan Balai	Pesisir	CST-DB	yes	full	yes
22 May 07	Tangan Buntung	Palembang Pasar	PB-SH3	no	full	no
23-May-07	Supat	Pesisir	CST-SP	yes	full	yes
23-May-07	Balai Agung	Musi Sekayu	MU-BA	yes	full	yes
24-May-07	Simpang Bayat	Pesisir	CST-SB	yes	full	yes
25-May-07	Talang Leban	Belide	PB-TL	yes	full	yes
26-May-07	Duren Gadis	Palembang Pasar	PB-DG	no	full	no
26-May-07	Siju	Palembang Pasar	PB-SJU	yes	full	yes
28-May-07	Tebing Abang	Belide	BEL-TA	yes	full	yes
29-May-07	Sungai Menang	Pesisir	CST-SM	yes	full	yes
31-May-07	Sungai Kijang	Kubu	KUBU	yes	full	yes
31-May-07	Pangkalan	Rawas Ulu	RAW-PN	yes	full	yes
1-Jun-07	Petunang	Musi Kelingi	MU-KL	yes	full	yes
1-Jun-07	Sadu	Musi Proper	MU-SA	no	full	no
2-Jun-07	Pauh	Rawas Ilir	MU-PH	yes	full	yes
2-Jun-07	Embacang	Musi Proper	MU-EM	yes	full	yes
3-Jun-07	Pasar Surulangun	Rawas Tengah	RAW-PS	yes	full	yes
4-Jun-07	Bingin Teluk	Musi Proper	MU-BT	yes	full	yes
6-Jun-07	Pulau Baru	Pekal	PKL-PL	no	full	yes

Date	Village	Dialect	Code	SLQ	Wordlist	Sentences
8-Jun-07	Bengkulu City	Bengkulu Ind.	BNGKL	yes	full	yes
9-Jun-07	Lawang Agung	Besemah	BES-BK	no	full	no
9-Jun-07	Napal Melintang	Serawai Manna	SRW-NM	yes	full	yes
10-Jun-07	Bintuhan	Kaur	--	no	partial	no
Tadmor	Palembang Lama	Palembang Lama	PL-PL	no	200	no
Tadmor	Sarang Lang	Palembang Lama	PB-SL	no	200	no
Tadmor	Pemulutan	Palembang Lama	PB-PM2	no	200	no
Tadmor	Pegagan	Pegagan	MU-PG2	no	200	no
Holle	Kaur/Mulak	Kaur	KAUR	no	263	no
27-Mar-08	Batu Gajah	Rawas		RTT only		
26-Mar-08	Penanggiran	Enim/Rambang		RTT only		
25-Mar-08	Ilir Barat Satu	Palembang Pasar		RTT only		
24-Mar-08	Satu Ilir	Palembang Pasar		RTT only		
29-Mar-08	Pelajau	Col Lembak	COL-PLJ	yes	full	yes
31-Mar-08	Jembatan Dua	Kaur	KAU-J2	yes	full	yes
1-Apr-08	Napal Putih	Pekal	PKL-NP	yes	full	yes
2-Apr-08	Suka Pindah	Muko-Muko	MUK-SP	yes	full	yes
3-Apr-08	Pondok Lunang	Muko-Muko	MUK-PL	no	full	yes
3-Apr-08	Pedamaran (Satu)	Penesak	PB-PDR	no?	full	no?

Figure 3.1 Research locations



3.2 Lexicostatistic theory and methodology

Lexicostatistical analysis in this study is regarded as a good first step in helping us determine the relative *synchronic* similarities of two different lects (cf. Boone 1994). Since its invention as a means of comparing lects, some of the applications of lexicostatistics have come under heavy criticism from within the linguistic community. This criticism was largely based upon the fact that many early proponents of lexicostatistics claimed that the method could provide a quantitative means of genetically classifying lects as one or multiple languages based upon their lexical similarity, as well as estimating when these lects diverged in time (glottochronology). Such precision in determining the genetic relationship between two varieties overlooked many important factors, including sociolinguistic and language identity concerns, as well as inherent weaknesses and variables within the application of lexicostatistics itself.

Our interests in using lexicostatistics, in contrast, are strongly oriented toward synchronic questions like comprehension and barriers to communication. In researching SSML, the authors have used lexicostatistics as an initial guide in looking at the relative lexical distance between the varieties being studied. Some aspects of the lexicostatistical analysis also serve to confirm locally perceived similarity within SSML. Cognate decisions were made using the comparative method, using Adelaar's reconstructions of Proto-Malayic (1992).[30]

[30] In theory, the comparative method allows one to distinguish between native and borrowed forms. However, it was not possible to consistently distinguish between the two given the inherent closeness of the varieties studied.

For performing this lexicostatistical analysis, a program called WordSurv 6.0/6.0.2 was used. WordSurv[31], developed by SIL International and Taylor University, allows the researcher to enter or import multiple wordlists for comparison. After the researcher determines the cognate sets for each lexical item, WordSurv then calculates the percentage of lexical similarity between each lect for which is has been given data.

Probst (1992) is an evaluation of different ways to handle difficult issues in lexicostatisics, particularly polymorphemic words in agglutinative and inflecting languages. His general advice, which we tried to follow, is to be consistent and document one's decisions.

The cognate decision is a simple yes/no dichotomy; either a word is considered part of a certain cognate group or it is not. In many cases, particularly since the comparative work has already been completed for PM, the cognate decisions were clear. However, there were many cases where the decision was not so clear-cut. Our methodological issues (and decisions) will now be discussed.

In contrast to standard lexicostatistic practice, WordSurv allows for the calculation of synonyms. For example, if Language 1 has Form A, Language 2 has Form B, but the list representing Language 3 has both Forms A and B as synonyms, the default statistical option for WordSurv would be to group Language 3 with Language 2 (on the basis of Form B) and also with Language 1 (on the basis of Form A). This leads to somewhat higher cognate percentages than would be the case if all synonyms were disallowed. However, disallowing synonyms is very difficult when one's wordlist informant is not able to tell the elicitor which form is primary, and which is secondary. Additionally, our synchronic goals did not militate against the inclusion of synonyms, therefore our team included them in the calculations. This also better captures the lexicon of SSML, where a single gloss in Indonesian may have two synonyms that are both commonly used within one speech community.

The question arose on how to count polymorphemic words, specifically:

- words with a single stem plus affixes;
- reduplication, both whole and partial;
- compound words, like 'tree skin' for bark; and
- stems with clarifying words, like '(bird) feathers', versus just 'feathers'

For affixed words, we disregarded affixes and particles such as *ŋə-*, *N-*, *mə-*, *ma-*, *pə-*, *ku-*, *di-*, *kə-*, *sə-*, *-əm-*, *-ar-*, *-kə(n)*, *-an*, and *-i*. Although the presence or absence of some of these affixes could easily have an effect on comprehension, we had no principled way of deciding which would and which would not. We also decided to disregard full and partially-reduplicated particles and group the words as cognate as long as the stem matched. We decided that for compound words (i.e., where both words were integral to the meaning), both words needed to be present to form the same cognate group. So 'tree skin' and 'wood skin' would not be grouped together. But in the case of stems with clarifying words, we decided to match the stem only and discard the clarifying words, thus 'bird feathers' was grouped with 'feathers'.

As a very rough indicator, above 90% cognate rate is considered to have low difficulty, while below 70% cognate rate is considered to be an indicator that difficulty in comprehension may be posed based on a significant portion of the lexicon being different. So while percentages are a poor measurement of linguistic relationship, low percentages can be an indicator that inherent intelligibility is low (Boone 2007).

The use of the BAN list in our lexicostatistics facilitated comparison of the data gathered during the six survey trips with data collected by other linguists.

[31] <http://www.sil.org/computing/survey/wordsurv.htm>

1.1 Phonological analysis

After the wordlist data were checked, correspondences were established for the consonant and vowel phonemes based on the PM phoneme inventory as described by Adelaar (1992). To illustrate the method, we use the PM phoneme **s*. The occurrences of this phoneme were separated by potentially significant environment into word-initial, word-medial and word-final. In SSML, PM **s* turns out to be quite conservative, rarely changing to (say) *h* in final position, in contrast with Jambi Malay or MINANGKABAU to the north. However, we discovered that, in a number of Col lects, initial **s* > *h*, e.g., **sakit* 'sick, painful' is reflected as *haket*.

Results of the analysis for this phoneme and others are discussed in §5.

1.2 RTT rationale, procedures used, sampling

The Recorded Text Test (RTT) (Casad 1974) is based upon the assumption that a person's ability to retell a story heard in another lect corresponds to his or her ability to comprehend that lect. In practice, this tool can differentiate between very low levels of comprehension in the second language (L2) and moderate/high levels of comprehension. It cannot reliably distinguish between moderate and high levels of comprehension of the L2.

A detailed description of standard procedures for the Rapid Appraisal (RA) RTT (Stalder 1996), procedures our team followed in developing and administering the RA-RTT, site selection, and analysis methods is given in Appendix 1, along with the three RA-RTT stories and scores from each village. In brief, the RA-RTT retelling method requires a group of subjects to listen to a recorded story in another lect and retell it segment by segment, paraphrasing it in their mother tongue or into a language of wider communication (LWC).

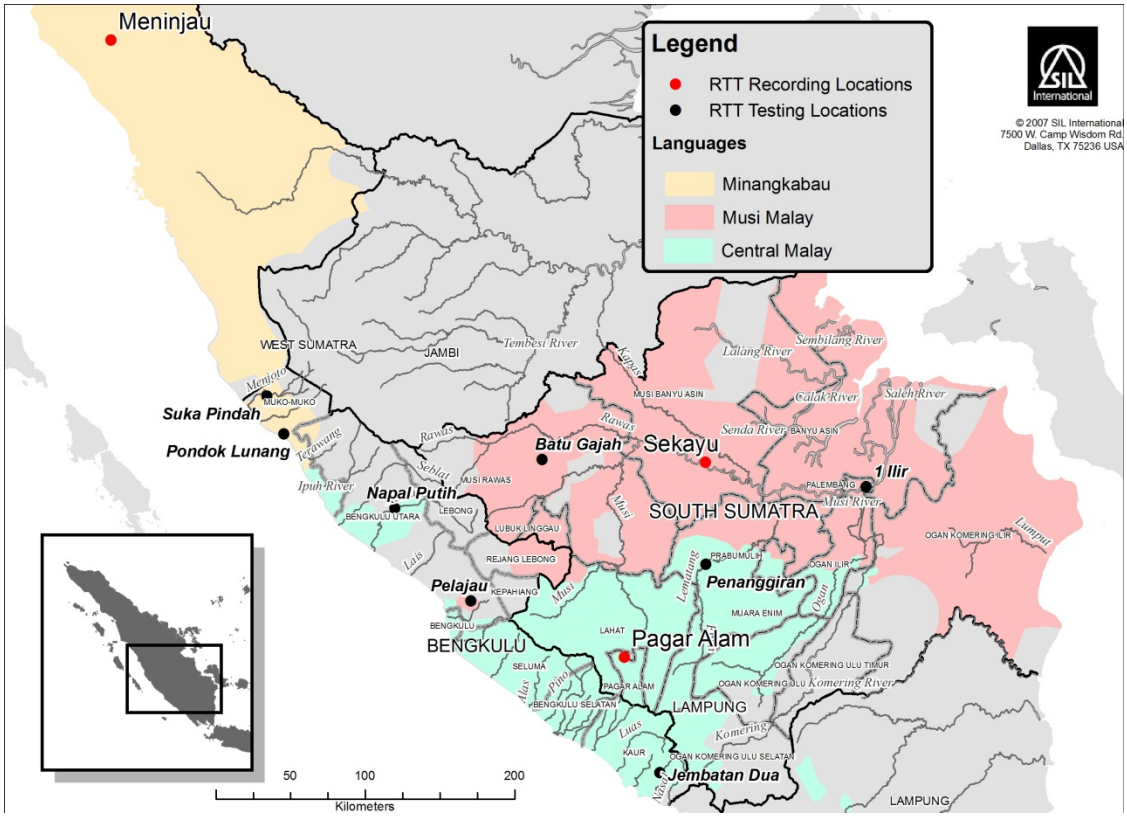
Three stories were tested. One story was told by a speaker from the Sekayu area in the local subdialect of what we call the Musi Proper dialect. One story was told by a speaker from Sukaraja Kikim, a village half an hour from Pagar Alam in the local dialect of Besemah (part of the SOUTH BARISAN MALAY cluster). One story was in the Maninjau dialect of MINANGKABAU (not far from Bukit Tinggi). All stories were told in the everyday register. These stories were chosen from areas in what seemed to be the 'dialect centers' of their respective languages, though for MINANGKABAU Bukit Tinggi would have been the preferred dialect over Maninjau.

These stories were then tested in eight villages and two urban setting: three Musi Proper villages and two locations in Palembang, three BARISAN villages, and two MINANGKABAU villages. The results from the analysis of the SLQs and from the phonological innovations allowed researchers to predict which varieties were of the MINANGKABAU, BARISAN and MUSI languages and to test these predictions using the RA-RTT. See Figure 3.2 for sources of the RTT texts and where they were tested.

All three stories were not tested in each village as the analysis had already produced language groupings; rather, the variety (ies) that was projected to match with the tested village was used. This testing confirmed statements about comprehension that had been made in the SLQ and gave researchers a sense of the quality of that comprehension.

Respondents heard the stories in the vernacular and retold the stories in Indonesian or in their vernacular, at which point it was translated into Indonesian by a local translator. The respondents were then scored based upon the completeness of the core elements (items of the story that all hometown listeners recounted in retelling the story). Minor details (items the majority but not all hometown respondents had retold) were also noted if supplied by the test respondents.

Figure 3.2 RA-RTT sources and testing locations



3.5 Sociolinguistic questionnaires

Sociolinguistic questionnaires help answer questions regarding language use and vitality, language shift, dialectology, and language attitudes.

3.5.1 Sociolinguistic methodology

Sociolinguistic questionnaires (SLQs) were administered at most sites where language data was taken (73 sites in toto; see Table 3.1 above for a listing of SLQ-full and SLQ-less sites), as well as some areas where the researchers already had adequate linguistic data but wished to know more about language attitudes, language use, and dialectology. The sites are broken down here in Table 3.2 by dialect and number of SLQ sites:

Table 3.2 SLQs administered by dialect

Dialect	# of SLQs
Belide	4
Lematang Ilir	2
Palembang Lama	2
Palembang Pasar	6
Penesak	3
Pesisir	6
Col	5
Musi Proper	7
Pegagan	1
Rawas	4
Enim	3
Ogan	4
Rambang	3
Benakat	1
Bengkulu	1
Besemah	4
Kaur	1
Lematang Ulu	2
Lintang	3
Pekal	1
Semenda	4
Serawai	1
HAJI	3
Kubu	1
Muko-Muko	1

The SLQs were typically administered to a group of people from the area, though in a few cases it was to a single individual from the area such as the village leader. The village leader was asked to supply a mixed group of men and women native to the area, both older and younger, to answer questions regarding language use. As these people were chosen by the village leader (in most cases) rather than a random selection, this introduces the possibility of skewing the answers – as the residents may have been selected for their ability to speak Indonesian. However, most of the Malayic speaking communities we visited reported a very high percentage of residents could use Standard Indonesian (SI) for communicating. A test of bilingualism for Indonesian among the UPPER MUSI and BARISAN dialect clusters backed this up, with the average resident being capable of communicating about their daily lives using Indonesian. Based on this, these researchers believe that the nonrandom nature of the group interviews did not significantly alter the results. This had the advantage of requiring less screening and not requiring extensive sampling. The answers recorded reflect the group consensus—though if any individual objected strongly to the group consensus it was noted as well.

The SLQ sites chosen do not necessarily represent a *representative* snapshot of the language situation throughout the region, as for the most part rural villages were chosen for dialectology purposes. This means that towns with large amounts of immigrants or with no dominant ethnic group were generally avoided (with the notable exception of Palembang, Bengkulu, and a few other locations) and more remote villages were favored. Additionally, villages that could not be reached using a SUV were not visited as well. (This was not a major limiting factor as the majority of villages in southern Sumatra can now be reached with a motorcycle or SUV). It is anticipated that

both domains of L1 and maintenance of L1 would decrease in the towns with larger mixed populations and that language vitality would be stronger in the remote villages not surveyed due to inaccessibility.

Table 3.3 below is a list of the nine communities—outside of the cities of Bengkulu and Palembang—where respondents reported that their town had 40% or more immigrants. It is not clear that all of these 'immigrants' are from other ethnic groups, but rather that they are considered not to be original inhabitants by the respondents. The 'dialect' and 'code' of the table pertain to the main autochthonous SSML group of the village.

Table 3.3 Villages reporting high percentage of immigrants

REPORTED % IMMIGRANT	VILLAGE	Dialect	CODE
75%	Sungai Kijang	Kubu	KUBU
70%	Simpang Bayat	Pesisir	CST-SB
60%	Sukaraja Kisam	Besemah	BES-MDK
50%	Karya Nyata	Semenda	SEM-SDL
50%	Balai Agung	Musi Proper Sekayu	MU-BA
50%	Air Itam	Palembang Pasar	PB-AH
50%	Pasar Surulangun	Rawas Tengah	RAW-PS
40%	Prabumulih Satu	Musi Proper	MU-P1
40%	Tinggi Hari	Lematang Ulu	LT-PP

The sociolinguistic questionnaires dealt with attitude toward language use including use in different domains, language shift, language maintenance, language planning, and dialectology. The dialectology section of the SLQ provided respondents an opportunity to give their emic perspective on how their lect and those surrounding it were related, including questions on comprehension, common ancestry, and languages used when speaking to people from different areas. Not all parts of the SLQ were administered in every area;[32] the data presented in §7 represents the answers obtained from the places where the SLQ was administered in full. The SLQ used in the different surveys was also slightly different depending on the area being surveyed.[33]

Another tool used in the SLQ was to show the group a map showing the different dialect varieties bordering their area and asking questions about language similarity and differences. There responses were used to correct and verify the language maps WIST had for southern Sumatra.

3.5.2 Presentation of Data

The research site and the dialect group of the spoken variety from the site are noted in the various SLQ data tables. The relations of this lect to other lects were variously estimated by the respondents to be the same, similar, or different.

Sociolinguistic data and analysis are presented in §7.

*** FINDINGS ***

The following four chapters present the findings of our research, integrated with the findings of previous researchers. The first chapter (§4) looks at lexicostatistical analysis of the wordlists collected, the second chapter (§5), by far the longest, examines innovations in

[32] For example, if complete SLQs had already been done for several UPPER MUSI and PALEMBANG sites, and the researchers wanted to find out the language spoken in a community between an UPPER MUSI and a PALEMBANG area, only the dialectology portion of the SLQ would be administered.

[33] For example, the dialectology section used in the Bengkulu SLQ included languages of Bengkulu, western South Sumatra and Jambi, and West Sumatra, while the dialectology section used in Palembang had languages of South Sumatra without mentioning languages of Jambi, West Sumatra, or Bengkulu.

the phonology of the various lects, the third chapter (§6) presents data from recorded text testing, while the fourth (§7) provides the results from sociolinguistic questionnaires.

4 Lexicostatistic and lexical analysis

With 95 wordlists compared for lexical similarity, the sheer size of the results is such as to be prohibitively large to place on one sheet for comparison. Therefore, the results are broken into separate tables which provide comparisons of individual sites within a dialect cluster (dialect) and between clusters. Our analysis is therefore bottom-up, without allowing previous linguistic classifications to unduly influence the cluster decisions. The location of the research site is given in a code in the table, and the full information on the research site can be found in Table 3.1.

4.1.1 OGANIC cluster

The OGANIC cluster, comprised of Ogan, Rambang, and Enim, consisted of nine locations compared, along with three other wordlists: Proto-Malayic (PM), Standard Indonesian (SI), and Standard Malay (SM). Within the OGANIC cluster (Table 4.1), the highest percentage of shared cognates (PSC) is 92% between two Rambang varieties and the lowest cognate count 79% between two Ogan and two Rambang sites. The nine varieties have a high internal PSC of 85% and an average PSC of 78% with PM and 74% with SI. While we can not predict high comprehension based upon lexical similarity, we can rule out that lexical dissimilarity would be a major factor inhibiting comprehension. If there is a problem in comprehension within the OGANIC cluster, it is likely to occur as a result of different phonological changes and phonetic realizations that transform the same lexical items. With the exception of some different phonological phenomena in the diphthongs and vowel epenthesis, there is little variation phonologically. Moreover, according to the SLQ responses, there is high intercomprehension between members of the OGANIC cluster, though the upstream varieties may be considered 'odd' in the way the words are realized due to the vowel changes.

Table 4.1 OGANIC cluster lexicostatistics results

Variety	PM	SI	SM	OG-RA	OG-BR	OG-UO	EN-TAN	EN-TAS	EN-ME	RAM-RD	RAM-RK	RAM-TR	RAM-PNY
PM	1	.91	.89	.79	.76	.75	.76	.78	.78	.79	.82	.76	.79
SI	.91	1	.98	.76	.7	.71	.7	.73	.72	.74	.8	.74	.76
SM	.89	.98	1	.76	.7	.71	.7	.72	.72	.73	.8	.74	.76
OG-RA	.79	.76	.76	1	.84	.8	.84	.83	.82	.89	.89	.87	.9
OG-BR	.76	.7	.7	.84	1	.89	.88	.86	.84	.82	.81	.84	.79
OG-UO	.75	.71	.71	.8	.89	1	.89	.87	.85	.81	.79	.81	.8
EN-TAN	.76	.7	.7	.84	.88	.89	1	.9	.88	.86	.83	.83	.81
EN-TAS	.78	.73	.72	.83	.86	.87	.9	1	.88	.84	.83	.85	.81
EN-ME	.78	.72	.72	.82	.84	.85	.88	.88	1	.85	.83	.82	.8
RAM-RD	.79	.74	.73	.89	.82	.81	.86	.84	.85	1	.92	.89	.87
RAM-RK	.82	.8	.8	.89	.81	.79	.83	.83	.83	.92	1	.89	.86
RAM-TR	.76	.74	.74	.87	.84	.81	.83	.85	.82	.89	.89	1	.84
RAM-PNY	.79	.76	.76	.9	.79	.8	.81	.81	.8	.87	.86	.84	1

4.1.2 HIGHLAND cluster

The HIGHLAND cluster, comprised of Besemah, Kikim, Semenda, Benakat, Lintang, Kaur, Lematang Ulu, Bengkulu, Serawai, and Pekal, consists of 20 locations compared, along with PM, SI, and SM. Due to the large number of sites, not all of them were able to be represented on the horizontal axis, but all are retained in the vertical axis of Table 4.2.[34] Please see the linked materials for a complete listing (McDowell and Anderbeck 2020). Within the HIGHLAND cluster, the highest cognate count is 97% between two Lintang sites and the lowest cognate count is 64% between two Kaur sites in southern Bengkulu and two Pekal sites located in northern Bengkulu. The average PSC internal to the cluster is 78%, with the PSC between the cluster members and PM an average of 76% and SI of 72%. The central dialect from a lexicostatistic perspective is the Lintang site BES-UM, with an average PSC with other cluster members of 85%.

The HIGHLAND cluster shows high cognate counts within the dialects (i.e., the three Lintang sites, the four Semenda sites, etc.) and fairly high cognate counts in adjoining dialects. As the distances grow quite far or if one group is isolated by other intervening language groups, the number of shared lexical items decreases. This agrees with the SLQs, which show that comprehension is reportedly high among adjoining HIGHLAND varieties but decreases with distance. Some difficulty in understanding is predicted for the peripheral HIGHLAND cluster varieties (those at the far ends of the 'chain') based on lexical difference. This proved to be the case according to the SLQs, with Kaur and Pekal being harder for other varieties to understand.

[34] The site KAU-BN representing Bintuhan Kaur was a short word list and is in the linked materials but is not included in this section.

Table 4.2 HIGHLAND cluster lexicostatistics results

Variety	PM	SI	SM	LT-PP	PB-BN	SEM-MD	SEM-LP	SEM-PB	BES-UM	BES-PA	BES-MDK	BES-KT	BES-BK	SRW-NM	KAU-J2	BNGKL	PKL-PB	PKL-NP
PM	1	.91	.89	.74	.76	.78	.77	.71	.84	.77	.75	.81	.77	.74	.69	.87	.81	.74
SI	.91	1	.98	.7	.73	.72	.7	.66	.8	.72	.71	.76	.7	.7	.65	.9	.83	.73
SM	.89	.98	1	.71	.73	.72	.69	.67	.79	.72	.72	.76	.71	.7	.64	.89	.81	.72
LT-MR	.72	.68	.68	.88	.81	.86	.81	.81	.9	.86	.82	.86	.84	.83	.69	.7	.66	.67
LT-PP	.74	.7	.71	1	.79	.88	.85	.83	.89	.9	.89	.89	.84	.82	.72	.74	.7	.69
PB-BN	.76	.73	.73	.79	1	.81	.79	.75	.85	.82	.79	.84	.8	.79	.72	.74	.71	.7
SEM-MD	.78	.72	.72	.88	.81	1	.93	.88	.91	.91	.86	.9	.88	.84	.76	.74	.71	.69
SEM-BA	.79	.73	.73	.86	.8	.91	.91	.85	.89	.88	.84	.87	.84	.8	.73	.73	.72	.68
SEM-LP	.77	.7	.69	.85	.79	.93	1	.89	.89	.87	.87	.85	.85	.84	.74	.7	.68	.67
SEM-PB	.71	.66	.67	.83	.75	.88	.89	1	.84	.84	.82	.82	.81	.79	.7	.67	.65	.65
BES-UM	.84	.8	.79	.89	.85	.91	.89	.84	1	.9	.88	.93	.9	.88	.76	.82	.79	.75
BES-MP	.78	.73	.72	.84	.79	.88	.84	.8	.97	.88	.83	.86	.86	.83	.71	.75	.73	.69
BES-PA	.77	.72	.72	.9	.82	.91	.87	.84	.9	1	.9	.91	.88	.84	.75	.77	.72	.68
BES-MDK	.75	.71	.72	.89	.79	.86	.87	.82	.88	.9	1	.87	.84	.82	.7	.72	.7	.67
BES-AK	.76	.72	.72	.92	.81	.9	.86	.84	.92	.92	.88	.92	.88	.86	.75	.76	.73	.7
BES-KT	.81	.76	.76	.89	.84	.9	.85	.82	.93	.91	.87	1	.88	.86	.75	.8	.76	.72
BES-BK	.77	.7	.71	.84	.8	.88	.85	.81	.9	.88	.84	.88	1	.88	.75	.74	.71	.68
SRW-NM	.74	.7	.7	.82	.79	.84	.84	.79	.88	.84	.82	.86	.88	1	.78	.76	.7	.69
KAU-J2	.69	.65	.64	.72	.72	.76	.74	.7	.76	.75	.7	.75	.75	.78	1	.7	.68	.65
Kaur	.72	.7	.7	.71	.73	.74	.71	.73	.74	.76	.69	.74	.73	.74	.87	.69	.71	.64
BNGKL	.87	.9	.89	.74	.74	.74	.7	.67	.82	.77	.72	.8	.74	.76	.7	1	.87	.79
PKL-PB	.81	.83	.81	.7	.71	.71	.68	.65	.79	.72	.7	.76	.71	.7	.68	.87	1	.82
PKL-NP	.74	.73	.72	.69	.7	.69	.67	.65	.75	.68	.67	.72	.68	.69	.65	.79	.82	1

4.1.3 South Barisan Malay dialect cluster

The **SOUTH BARISAN MALAY** language had nine OGANIC cluster sites analyzed and twenty HIGHLAND locations compared. See Table 4.3. Due to the large number of sites, not all are displayed, but a complete table can be found in the linked materials. The average PSC for the entire language was 79%, with an average PSC with PM of 76% and 73% for SI. A clear break in lexical similarity between the OGANIC and HIGHLAND clusters is not evinced, with some HIGHLAND sites having a high PSC with an OGANIC site (e.g., the Lintang site BES-UM having a PSC of 89% with the Enim site EN-ME). The overall internal PSC figures for the BARISAN language are: BARISAN language 79%, HIGHLAND cluster 78%, and OGANIC cluster 85%. The dialect appearing to be geographically and lexicostatistically central for the BARISAN language is the Lintang site Lubuk Puding Baru (BES-UM) with PSC above 80% with all sites excluding the outliers of Pekal and Kaur, and an eye-popping average PSC of 85% with all other 24 varieties in the cluster. Even with the lower PSC with Kaur and Pekal, it still scores a PSC in the mid-70s with the four complete Pekal and Kaur wordlists. The varieties that appear more peripheral to BARISAN based upon PSC are found in Bengkulu: Pekal in the north and Kaur in the south.

Table 4.3 SOUTH BARISAN MALAY lexicostatistics results

Variety	PM	OG-RA	OG-UO	EN-TAN	EN-ME	RAM-RD	RAM-PNY	LT-PP	PB-BN	SEM-PB	BES-UM	BES-PA	BES-MDK	BES-BK	SRW-NM	KAU-J2	BNGKL	PKL-PB
PM	1	.79	.75	.76	.78	.79	.79	.74	.76	.71	.84	.77	.75	.77	.74	.69	.87	.81
OG-RA	.79	1	.8	.84	.82	.89	.9	.78	.84	.75	.88	.8	.79	.76	.77	.73	.8	.76
OG-UO	.75	.8	1	.89	.85	.81	.8	.79	.81	.75	.83	.81	.78	.78	.76	.75	.72	.71
EN-TAN	.76	.84	.89	1	.88	.86	.81	.82	.82	.78	.86	.83	.8	.8	.78	.75	.72	.72
EN-ME	.78	.82	.85	.88	1	.85	.8	.87	.84	.82	.9	.86	.83	.85	.84	.76	.76	.74
RAM-RD	.79	.89	.81	.86	.85	1	.87	.79	.87	.74	.86	.82	.8	.78	.76	.73	.77	.73
RAM-PNY	.79	.9	.8	.81	.8	.87	1	.75	.84	.71	.83	.78	.77	.75	.73	.71	.79	.75
LT-PP	.74	.78	.79	.82	.87	.79	.75	1	.79	.83	.89	.9	.89	.84	.82	.72	.74	.7
PB-BN	.76	.84	.81	.82	.84	.87	.84	.79	1	.75	.85	.82	.79	.8	.79	.72	.74	.71
SEM-PB	.71	.75	.75	.78	.82	.74	.71	.83	.75	1	.84	.84	.82	.81	.79	.7	.67	.65
BES-UM	.84	.88	.83	.86	.9	.86	.83	.89	.85	.84	1	.9	.88	.9	.88	.76	.82	.79
BES-PA	.77	.8	.81	.83	.86	.82	.78	.9	.82	.84	.9	1	.9	.88	.84	.75	.77	.72
BES-MDK	.75	.79	.78	.8	.83	.8	.77	.89	.79	.82	.88	.9	1	.84	.82	.7	.72	.7
BES-BK	.77	.76	.78	.8	.85	.78	.75	.84	.8	.81	.9	.88	.84	1	.88	.75	.74	.71
SRW-NM	.74	.77	.76	.78	.84	.76	.73	.82	.79	.79	.88	.84	.82	.88	1	.78	.76	.7
KAU-J2	.69	.73	.75	.75	.76	.73	.71	.72	.72	.7	.76	.75	.7	.75	.78	1	.7	.68
BNGKL	.87	.8	.72	.72	.76	.77	.79	.74	.74	.67	.82	.77	.72	.74	.76	.7	1	.87
PKL-PB	.81	.76	.71	.72	.74	.73	.75	.7	.71	.65	.79	.72	.7	.71	.7	.68	.87	1

4.1.4 Upper Musi cluster

The UPPER MUSI cluster, consisting of the Musi Proper, Rawas, Pegagan and Col language varieties had 17 wordlists compared. See Table 4.4. The wordlist locations stretch the length of the Musi River basin with two additional sites on the other side of the Bukit Barisan Mountains in Bengkulu, far from the Musi River. The rate of lexical similarity between the sites ranged between a high of 97% for two Col sites and a low of 56% for the furthest upstream Rawas site Pangkalan (RAW-PN) and a Col Lembak site of Pondok Kubang (COL-L8). The average internal PSC for the UPPER MUSI cluster is the high figure of 85%, with an average PSC with PM of 80% and of 78% with SI. The central dialects from a lexical standpoint are MU-BA (Sekayu) and RAW-PS (Pasar Surulangun) with average PSCs of 88%. The high rates of cognacy that are evident overall in the UPPER MUSI cluster predict that the lexicon would not be a factor in lack of comprehension if there is one, but rather sociolinguistic factors or phonological phenomena. Due to the number of sites, not all are displayed on the horizontal axis but all are present on the vertical axis.

Table 4.4 UPPER MUSI cluster lexicostatistics results

Variety	PM	RAW-PS	RAW-RU	RAW-PN	MU-KL	MU-SA	MU-BT	MU-PH	MU-P1	MU-BA	MU-KY	MU-EM	MU-PEN	MU-PG2	MU-PG	COL-PUT	COL-LL	COL-BKL	COL-PLJ
PM	1	.83	.83	.85	.83	.8	.8	.79	.78	.81	.8	.8	.78	.78	.78	.77	.79	.8	.78
SI	.91	.84	.84	.87	.8	.77	.8	.8	.74	.78	.79	.78	.78	.74	.74	.76	.78	.78	.74
RAW-PS	.83	1	.96	.94	.88	.89	.94	.94	.84	.9	.9	.92	.86	.84	.84	.84	.86	.88	.84
RAW-RU	.83	.96	1	.96	.87	.88	.93	.92	.85	.88	.86	.9	.85	.83	.82	.82	.84	.86	.82
RAW-PN	.85	.94	.96	1	.86	.86	.9	.9	.83	.87	.86	.88	.83	.82	.81	.82	.83	.85	.81
MU-KL	.83	.88	.87	.86	1	.91	.86	.83	.91	.9	.87	.87	.84	.81	.82	.9	.91	.92	.88
MU-SA	.8	.89	.88	.86	.91	1	.9	.86	.89	.92	.88	.9	.85	.83	.83	.86	.88	.9	.88
MU-BT	.8	.94	.93	.9	.86	.9	1	.92	.84	.88	.87	.92	.83	.82	.8	.82	.84	.87	.82
MU-PH	.79	.94	.92	.9	.83	.86	.92	1	.82	.85	.82	.95	.83	.82	.78	.81	.8	.84	.81
MU-P1	.78	.84	.85	.83	.91	.89	.84	.82	1	.9	.84	.86	.82	.83	.84	.84	.84	.86	.84
MU-BA	.81	.9	.88	.87	.9	.92	.88	.85	.9	1	.94	.9	.86	.87	.88	.86	.88	.91	.86
MU-KY	.8	.9	.86	.86	.87	.88	.87	.82	.84	.94	1	.86	.82	.83	.84	.84	.84	.88	.82
MU-EM	.8	.92	.9	.88	.87	.9	.92	.95	.86	.9	.86	1	.84	.83	.8	.84	.84	.87	.85
MU-PEN	.78	.86	.85	.83	.84	.85	.83	.83	.82	.86	.82	.84	1	.85	.83	.8	.8	.82	.77
MU-PG2	.78	.84	.83	.82	.81	.83	.82	.82	.83	.87	.83	.83	.85	1	.89	.78	.79	.8	.8
MU-PG	.78	.84	.82	.81	.82	.83	.8	.78	.84	.88	.84	.8	.83	.89	1	.8	.8	.8	.81
COL-PUT	.77	.84	.82	.82	.9	.86	.82	.81	.84	.86	.84	.84	.8	.78	.8	1	.97	.91	.91
COL-LL	.79	.86	.84	.83	.91	.88	.84	.8	.84	.88	.84	.84	.8	.79	.8	.97	1	.92	.89
COL-BKL	.8	.88	.86	.85	.92	.9	.87	.84	.86	.91	.88	.87	.82	.8	.8	.91	.92	1	.89
COL-PLJ	.78	.84	.82	.81	.88	.88	.82	.81	.84	.86	.82	.85	.77	.8	.81	.91	.89	.89	1
COL-L8	.77	.82	.81	.8	.85	.86	.81	.78	.81	.83	.8	.82	.75	.76	.79	.87	.88	.86	.95

4.1.5 Palembang-Lowland cluster

The PALEMBANG-LOWLAND cluster of the MUSI language, consisting of the Palembang, Pesisir, and LOWLAND subcluster language varieties had 27 wordlists analyzed: 9 from the LOWLAND subcluster varieties and 18 from the Pesisir and Palembang varieties. These are presented in three separate tables: one LOWLAND subcluster varieties (Table 4.5), one for the Palembang and Pesisir varieties (Table 4.6), and one for the PALEMBANG-LOWLAND cluster (Table 4.7). To fit the table to the page, the overall PALEMBANG-LOWLAND cluster table displays only a selection of varieties. Please reference the linked materials for complete cross-listing comparing PALEMBANG-LOWLAND cluster varieties. The central dialect for the PALEMBANG-LOWLAND cluster lexicostatistically is the acrolectal Everyday Palembang list (PB-SH1), taken in the city of Palembang from a life-long resident of Palembang. Centrality of this site makes sense due to the large number of wordlists taken in and around Palembang, and it is geographically central, located between the wordlist sites to the south, west, and north. It had an average PSC of 94% with other members of the PALEMBANG-LOWLAND cluster, as compared to the high PALEMBANG-LOWLAND cluster average PSC of 83%. The PSC for the PALEMBANG-LOWLAND cluster with PM is 78% and with SI is 77%. The 18 Pesisir and Palembang varieties show an internally high rate of lexical similarity comparable to that in LOWLAND subcluster. The older varieties of Palembang such PL-PL, have high PSCs with the villages outside of Palembang that use a more rural or archaic form of Palembang such as PB-DG or PB-TEL. The three Palembang Pasar sehari-hari (everyday) varieties showed the variation of lexicon that can occur in the same city with a range of 92 - 94%. This puts in perspective how high the similarity is between these Palembang and Pesisir sites, with an average internal PSC of 87%, and a PSC with PM of 78%. The more contemporary forms of Palembang Pasar are also closer to PM, SI, and SM than the older form, which retains more Javanese loan words. The overall PALEMBANG-LOWLAND cluster shows the high lexical similarity between different varieties within it, in spite of the perceived differences reported in the SLQ.

Table 4.5 PALEMBANG-LOWLAND cluster (LOWLAND subcluster) lexicostatistical results

Variety	PM	SI	PB-PDR	PB-TB	PB-BR	PB-SR	PB-MD	PB-CB	PB-TA	PB-TL	BEL-TA
PM	1	.91	.86	.85	.84	.73	.8	.76	.77	.83	.81
SI	.91	1	.91	.83	.84	.7	.8	.73	.76	.8	.8
PB-PDR	.86	.91	1	.86	.86	.74	.81	.77	.8	.8	.82
PB-TB	.85	.83	.86	1	.9	.84	.85	.86	.84	.86	.87
PB-BR	.84	.84	.86	.9	1	.8	.83	.83	.8	.84	.86
PB-SR	.73	.7	.74	.84	.8	1	.78	.82	.87	.82	.87
PB-MD	.8	.8	.81	.85	.83	.78	1	.9	.8	.89	.8
PB-CB	.76	.73	.77	.86	.83	.82	.9	1	.82	.88	.82
PB-TA	.77	.76	.8	.84	.8	.87	.8	.82	1	.83	.87
PB-TL	.83	.8	.8	.86	.84	.82	.89	.88	.83	1	.87
BEL-TA	.81	.8	.82	.87	.86	.87	.8	.82	.87	.87	1

Table 4.6 PALEMBANG-LOWLAND cluster (Palembang and Pesisir) lexicostatistical results

Variety	PM	SI	CST-SB	CST-DB	CST-SP	CST-LG	CST-SM	PB-AH	PB-SJU	PB-DG	PB-BET	PB-GL	PB-SH2	PB-SH3	PB-SH1	PL-PL	PL-TEL	PB-SL	PB-PM
PM	1	.91	.85	.83	.85	.82	.78	.8	.79	.76	.57	.72	.84	.8	.81	.71	.73	.74	.74
SI	.91	1	.85	.82	.82	.8	.76	.79	.78	.74	.58	.71	.86	.8	.82	.69	.71	.7	.72
CST-SB	.85	.85	1	.87	.9	.86	.86	.86	.86	.81	.68	.8	.86	.86	.86	.77	.81	.78	.77
CST-DB	.83	.82	.87	1	.9	.86	.86	.88	.87	.84	.73	.82	.85	.85	.87	.8	.84	.81	.82
CST-SP	.85	.82	.9	.9	1	.87	.9	.88	.9	.86	.72	.82	.88	.89	.88	.81	.85	.84	.8
CST-LG	.82	.8	.86	.86	.87	1	.92	.92	.88	.85	.77	.82	.87	.86	.88	.81	.86	.82	.83
CST-SM	.78	.76	.86	.86	.9	.92	1	.92	.92	.87	.78	.84	.86	.86	.88	.85	.89	.86	.85
PB-AH	.8	.79	.86	.88	.88	.92	.92	1	.9	.86	.78	.85	.86	.87	.9	.82	.89	.84	.86
PB-SJU	.79	.78	.86	.87	.9	.88	.92	.9	1	.9	.82	.84	.87	.86	.9	.85	.89	.86	.87
PB-DG	.76	.74	.81	.84	.86	.85	.87	.86	.9	1	.84	.9	.88	.88	.92	.89	.93	.89	.87
PB-BET[35]	.57	.58	.68	.73	.72	.77	.78	.78	.82	.84	1	.82	.77	.78	.81	.82	.86	.81	.78
PB-GL	.72	.71	.8	.82	.82	.82	.84	.85	.84	.9	.82	1	.83	.84	.89	.89	.93	.84	.82
PB-SH2	.84	.86	.86	.85	.88	.87	.86	.86	.87	.88	.77	.83	1	.92	.94	.82	.85	.84	.83
PB-SH3	.8	.8	.86	.85	.89	.86	.86	.87	.86	.88	.78	.84	.92	1	.93	.85	.87	.87	.81
PB-SH1	.81	.82	.86	.87	.88	.88	.88	.9	.9	.92	.81	.89	.94	.93	1	.89	.92	.88	.85
PL-PL	.71	.69	.77	.8	.81	.81	.85	.82	.85	.89	.82	.89	.82	.85	.89	1	.92	.89	.83
PL-TEL	.73	.71	.81	.84	.85	.86	.89	.89	.89	.93	.86	.93	.85	.87	.92	.92	1	.9	.87
PB-SL	.74	.7	.78	.81	.84	.82	.86	.84	.86	.89	.81	.84	.84	.87	.88	.89	.9	1	.86
PB-PM	.74	.72	.77	.82	.8	.83	.85	.86	.87	.87	.78	.82	.83	.81	.85	.83	.87	.86	1
PB-PM2	.76	.73	.79	.83	.83	.83	.84	.85	.86	.86	.76	.8	.84	.85	.86	.83	.85	.89	.91

[35] Not a complete wordlist; PSC would be expected to rise with a longer WL.

Table 4.7 PALEMBANG-LOWLAND cluster (selection from Palembang, Pesisir, and LOWLAND subcluster) lexicostatistical results

Variety	PM	SI	CST-SP	CST-SM	PB-AH	PB-SJU	PB-DG	PB-SH3	PB-SH1	PL-PL	PL-TEL	PB-SL	PB-PM2	PB-TB	PB-BR	PB-SR	PB-MD	PB-TA	PB-TL	BEL-TA
PM	1	.91	.85	.78	.8	.79	.76	.8	.81	.71	.73	.74	.76	.85	.84	.73	.8	.77	.83	.81
SI	.91	1	.82	.76	.79	.78	.74	.8	.82	.69	.71	.7	.73	.83	.84	.7	.8	.76	.8	.8
CST-SP	.85	.82	1	.9	.88	.9	.86	.89	.88	.81	.85	.84	.83	.88	.88	.82	.83	.83	.9	.93
CST-SM	.78	.76	.9	1	.92	.92	.87	.86	.88	.85	.89	.86	.84	.87	.85	.86	.8	.87	.84	.91
PB-AH	.8	.79	.88	.92	1	.9	.86	.87	.9	.82	.89	.84	.85	.86	.84	.86	.81	.87	.84	.91
PB-SJU	.79	.78	.9	.92	.9	1	.9	.86	.9	.85	.89	.86	.86	.86	.87	.83	.8	.83	.86	.91
PB-DG	.76	.74	.86	.87	.86	.9	1	.88	.92	.89	.93	.89	.86	.82	.84	.78	.78	.79	.82	.85
PB-SH3	.8	.8	.89	.86	.87	.86	.88	1	.93	.85	.87	.87	.85	.83	.83	.79	.81	.81	.84	.9
PB-SH1	.81	.82	.88	.88	.9	.9	.92	.93	1	.89	.92	.88	.86	.84	.86	.8	.81	.82	.83	.89
PL-PL	.71	.69	.81	.85	.82	.85	.89	.85	.89	1	.92	.89	.83	.79	.77	.77	.76	.78	.78	.82
PL-TEL	.73	.71	.85	.89	.89	.89	.93	.87	.92	.92	1	.9	.85	.82	.8	.81	.79	.79	.81	.86
PB-SL	.74	.7	.84	.86	.84	.86	.89	.87	.88	.89	.9	1	.89	.83	.78	.8	.75	.78	.79	.84
PB-PM2	.76	.73	.83	.84	.85	.86	.86	.85	.86	.83	.85	.89	1	.81	.79	.78	.74	.77	.79	.85
PB-TB	.85	.83	.88	.87	.86	.86	.82	.83	.84	.79	.82	.83	.81	1	.9	.84	.85	.84	.86	.87
PB-BR	.84	.84	.88	.85	.84	.87	.84	.83	.86	.77	.8	.78	.79	.9	1	.8	.83	.8	.84	.86
PB-SR	.73	.7	.82	.86	.86	.83	.78	.79	.8	.77	.81	.8	.78	.84	.8	1	.78	.87	.82	.87
PB-MD	.8	.8	.83	.8	.81	.8	.78	.81	.81	.76	.79	.75	.74	.85	.83	.78	1	.8	.89	.8
PB-TA	.77	.76	.83	.87	.87	.83	.79	.81	.82	.78	.79	.78	.77	.84	.8	.87	.8	1	.83	.87
PB-TL	.83	.8	.9	.84	.84	.86	.82	.84	.83	.78	.81	.79	.79	.86	.84	.82	.89	.83	1	.87
BEL-TA	.81	.8	.93	.91	.91	.91	.85	.9	.89	.82	.86	.84	.85	.87	.86	.87	.8	.87	.87	1

4.1.6 MUSI language

The MUSI language lexicostatistical analysis was done on the previously discussed 27 PALEMBANG-LOWLAND cluster varieties and 18 UPPER MUSI cluster varieties. Including all varieties within its two clusters, the MUSI language has an average internal PSC of 82% and an average PSC with PM of 79% and with SI of 78%. Differences between these two clusters lie more in the different phonological innovations than in a different lexicon, though the Palembang sites (particularly the archaic form of Palembang) have a higher percentage of Javanese loan words and thus, a lower lexical similarity (usually about 10% lower) when compared with UPPER MUSI cluster sites. Table 4.8 is a selection of MUSI sites.

Table 4.8 Selected varieties from MUSI language

Variety	PM	CST-SP	CST-SM	PB-AH	PB-DG	PB-SH1	PL-PL	PB-PM2	PB-TB	PB-BR	PB-MD	BEL-TA	RAW-PS	RAW-PN	MU-KL	MU-BA	MU-PEN	MU-PG2	COL-LL	COL-PLJ
PM	1	.85	.78	.8	.76	.81	.71	.76	.85	.84	.8	.81	.83	.85	.83	.81	.78	.78	.79	.78
CST-SP	.85	1	.9	.88	.86	.88	.81	.83	.88	.88	.83	.93	.9	.9	.86	.9	.83	.84	.86	.84
CST-SM	.78	.9	1	.92	.87	.88	.85	.84	.87	.85	.8	.91	.86	.84	.82	.85	.85	.83	.8	.77
PB-AH	.8	.88	.92	1	.86	.9	.82	.85	.86	.84	.81	.91	.86	.84	.84	.86	.84	.82	.82	.8
PB-DG	.76	.86	.87	.86	1	.92	.89	.86	.82	.84	.78	.85	.85	.82	.78	.81	.78	.77	.75	.74
PB-SH1	.81	.88	.88	.9	.92	1	.89	.86	.84	.86	.81	.89	.88	.88	.82	.83	.8	.78	.79	.77
PL-PL	.71	.81	.85	.82	.89	.89	1	.83	.79	.77	.76	.82	.79	.76	.75	.8	.76	.78	.7	.69
PB-PM2	.76	.83	.84	.85	.86	.86	.83	1	.81	.79	.74	.85	.82	.79	.77	.79	.76	.78	.73	.71
PB-TB	.85	.88	.87	.86	.82	.84	.79	.81	1	.9	.85	.87	.88	.86	.84	.88	.84	.84	.8	.8
PB-BR	.84	.88	.85	.84	.84	.86	.77	.79	.9	1	.83	.86	.88	.87	.82	.82	.82	.81	.8	.77
PB-MD	.8	.83	.8	.81	.78	.81	.76	.74	.85	.83	1	.8	.86	.84	.8	.8	.8	.78	.75	.75
BEL-TA	.81	.93	.91	.91	.85	.89	.82	.85	.87	.86	.8	1	.89	.87	.87	.89	.86	.85	.82	.81
RAW-PS	.83	.9	.86	.86	.85	.88	.79	.82	.88	.88	.86	.89	1	.94	.88	.9	.86	.84	.86	.84
RAW-PN	.85	.9	.84	.84	.82	.88	.76	.79	.86	.87	.84	.87	.94	1	.86	.87	.83	.82	.83	.81
MU-KL	.83	.86	.82	.84	.78	.82	.75	.77	.84	.82	.8	.87	.88	.86	1	.9	.84	.81	.91	.88
MU-BA	.81	.9	.85	.86	.81	.83	.8	.79	.88	.82	.8	.89	.9	.87	.9	1	.86	.87	.88	.86
MU-PEN	.78	.83	.85	.84	.78	.8	.76	.76	.84	.82	.8	.86	.86	.83	.84	.86	1	.85	.8	.77
MU-PG2	.78	.84	.83	.82	.77	.78	.78	.78	.84	.81	.78	.85	.84	.82	.81	.87	.85	1	.79	.8
COL-LL	.79	.86	.8	.82	.75	.79	.7	.73	.8	.8	.75	.82	.86	.83	.91	.88	.8	.79	1	.89
COL-PLJ	.78	.84	.77	.8	.74	.77	.69	.71	.8	.77	.75	.81	.84	.81	.88	.86	.77	.8	.89	1

4.1.7 Variety of Malayic languages

Lexicostatistical analysis was also done on other Malayic varieties to compare the languages of southern Sumatra with these other varieties. Due to the limitations of the printed page, not all varieties analyzed can be presented together, but a selection of Malayic varieties used in the lexicostatistical analysis is presented in Table 4.9. Beyond the sites visited by WIST, wordlists are used from published sources and from unpublished sources by permission. Varieties that are included include MINANGKABAU (MIN1) (Adelaar 1992), Jambi Ilir (JI-ML) (Anderbeck 2008), Jambi Ulu (JU-MS) (Anderbeck id.), HAJI (HAJI) (Anderbeck 2007f), Kerinci (KER) (Prentice & Usman 1978), Duano of Riau (DNO-KB) (Seidlitz 2007), Lom of Bangka Island (Smedal 1987), Deli Malay of North Sumatra (Blust 1988), and (Sumsel) Kubu (WIST data). Not all varieties could be fit on the horizontal axis which are found on the vertical axis.

Table 4.9 Selection of Malayic varieties lexicostatistic results

Variety	PM	Deli	LomTT	PB-SH1	BEL-TA	RAW-PS	MU-BA	COL-LL	OG-UO	RAM-PNY	BES-UM	SRW-NM	KAU-J2	PKL-PB	HAJI	JU-MS	KUBU	KER	MUK-PL	MUK-SP	MIN1
PM	1	.86	.8	.81	.81	.83	.81	.79	.75	.79	.84	.74	.69	.81	.63	.78	.78	.88	.82	.77	.87
Deli	.86	1	.72	.76	.75	.8	.76	.75	.68	.72	.77	.69	.63	.79	.57	.73	.73	.83	.82	.78	.85
LomTT	.8	.72	1	.75	.71	.77	.7	.71	.65	.68	.75	.67	.66	.71	.56	.69	.73	.74	.71	.68	.75
DNO-KB	.59	.54	.53	.52	.52	.56	.52	.51	.48	.51	.53	.48	.46	.55	.43	.51	.54	.6	.55	.52	.58
PB-SH1	.81	.76	.75	1	.89	.88	.83	.79	.74	.81	.82	.74	.69	.78	.62	.72	.77	.81	.76	.72	.78
BEL-TA	.81	.75	.71	.89	1	.89	.89	.82	.76	.85	.86	.75	.72	.77	.63	.72	.81	.8	.76	.72	.77
RAW-PS	.83	.8	.77	.88	.89	1	.9	.86	.76	.82	.86	.77	.71	.84	.63	.8	.86	.84	.83	.79	.84
MU-BA	.81	.76	.7	.83	.89	.9	1	.88	.72	.82	.85	.76	.69	.75	.62	.73	.84	.78	.77	.74	.78
COL-LL	.79	.75	.71	.79	.82	.86	.88	1	.71	.77	.83	.76	.68	.77	.61	.71	.78	.76	.76	.73	.78
OG-UO	.75	.68	.65	.74	.76	.76	.72	.71	1	.8	.83	.76	.75	.71	.65	.68	.7	.72	.7	.68	.72
EN-TAN	.76	.7	.65	.76	.78	.76	.76	.72	.89	.81	.86	.78	.75	.72	.7	.68	.74	.72	.71	.68	.7
RAM-PNY	.79	.72	.68	.81	.85	.82	.82	.77	.8	1	.83	.73	.71	.75	.65	.69	.77	.77	.73	.69	.75
BES-UM	.84	.77	.75	.82	.86	.86	.85	.83	.83	.83	1	.88	.76	.79	.69	.74	.8	.83	.79	.76	.79
SRW-NM	.74	.69	.67	.74	.75	.77	.76	.76	.76	.73	.88	1	.78	.7	.63	.68	.72	.72	.71	.69	.72
KAU-J2	.69	.63	.66	.69	.72	.71	.69	.68	.75	.71	.76	.78	1	.68	.73	.66	.69	.68	.67	.67	.68
PKL-PB	.81	.79	.71	.78	.77	.84	.75	.77	.71	.75	.79	.7	.68	1	.6	.78	.74	.85	.84	.8	.84
Aji	.63	.57	.56	.62	.63	.63	.62	.61	.65	.65	.69	.63	.73	.6	1	.56	.61	.61	.61	.58	.61
JI-ML	.79	.78	.73	.81	.8	.83	.78	.76	.71	.78	.76	.7	.67	.78	.59	.77	.79	.8	.83	.78	.83
JU-MS	.78	.73	.69	.72	.72	.8	.73	.71	.68	.69	.74	.68	.66	.78	.56	1	.76	.82	.84	.81	.85
KUBU	.78	.73	.73	.77	.81	.86	.84	.78	.7	.77	.8	.72	.69	.74	.61	.76	1	.79	.78	.75	.79
KJ1	.84	.79	.74	.84	.84	.86	.79	.78	.75	.76	.8	.73	.68	.79	.61	.76	.77	.81	.81	.76	.81
KER	.88	.83	.74	.81	.8	.84	.78	.76	.72	.77	.83	.72	.68	.85	.61	.82	.79	1	.87	.8	.9
MUK-PL	.82	.82	.71	.76	.76	.83	.77	.76	.7	.73	.79	.71	.67	.84	.61	.84	.78	.87	1	.97	.92
MUK-SP	.77	.78	.68	.72	.72	.79	.74	.73	.68	.69	.76	.69	.67	.8	.58	.81	.75	.8	.97	1	.89
MIN1	.87	.85	.75	.78	.77	.84	.78	.78	.72	.75	.79	.72	.68	.84	.61	.85	.79	.9	.92	.89	1

HAJI, one of the other Malayic languages spoken in South Sumatra along with MUSI and BARISAN, showed a remarkably different lexicon than the varieties surrounding it, only scoring above 70% in three varieties out of the 94 compared: 73% with the Kaur site of Jembatan Dua, 71% with the Kaur Holle list, and 70% with the upstream Enim site of Indramayu. While a lexical similarity above 70% is no guarantee of comprehension, usually a figure below the 70% limit indicates comprehension will be poor due to different vocabulary. The BARISAN dialects have an average PSC with HAJI of 65% while the MUSI dialects have one of 60%. This compares with a HAJI-PM PSC of 63% and to SI of 60%. With this in mind, the predicition is that most speakers of different varieties of southern Sumatran Malay would have difficulty in understanding HAJI. (However, based on the responses to the HAJI SLQ, the communication would probably occur in Palembang Pasar as the LWC).

The neighboring varieties of Jambi Ilir and Jambi Ulu evinced a much higher degree of lexical similarity, with high PSCs with MUSI (though not as similar to BARISAN). Not surprisingly, the geographically closer an area is to the Jambi Ilir or Jambi Ulu site, the greater the lexical similarity. Thus the Pesisir site of Simpang Bayat, located on the Trans Sumatran highway within an hour of the Jambi border, has an 85% similarity with Jambi Ilir (JI-ML) on the other side of the Jambi border. The other Pesisir varieties of the PALEMBANG-LOWLAND cluster of the MUSI language also reflect similar rates. Jambi Ilir has an average PSC of 78% with MUSI, 71% with BARISAN, and 79% with PM. Jambi Ulu has an average PSC of 72% with MUSI, 68% with BARISAN, and 77% with PM. In short, Jambi Ilir has a great deal of lexical similarity with the Pesisir and Palembang dialects of MUSI while Jambi Ulu would present some difficulty in comprehension caused by a different lexicon (not to mention its unique phonological innovations) for speakers of BARISAN.

Across the Straits of Bangka, the varieties of Lom and Bangka could be considered bordering the eastern edge of the MUSI language area, facing the swamps of South Sumatra where Pesisir is spoken. The Bangka site BK-Dul (Nothofer 1997) has an average PSC of 72% with MUSI, 68% with BARISAN, and 79% with PM. The Lom site has an average PSC of 71% with MUSI, 68% with BARISAN, and 80% with PM. However, the Pesisir locations have a much higher PSC with Lom and Bangka. For example, Simpang Bayat (CST-SB) has a PSC of 77% with both Bangka and Lom, with other Pesisir varieties having similar PSCs. This points out that the MUSI varieties near the Strait of Bangka have higher lexical similarity to Lom and Bangka, while the MUSI and BARISAN dialects as a whole are at the probable lower end of possible comprehension by Bangka Malay speakers based on lexical similarity.

Moving further up the east coast of Sumatra (Deli) and into the Riau Islands (Duano), we investigate whether other Malayic coastal varieties continue this pattern of sharing a large number of lexical items with the Pesisir varieties, but much lower percentages of shared cognates with the other MUSI and BARISAN varieties. Deli Malay has an average PSC of 74% with MUSI, 70% with BARISAN, and 86% with PM. However, with the Pesisir site of Simpang Bayat, it has a PSC of 81%. So Deli Malay holds to the pattern seen in Lom, Bangka, and Jambi Ilir. Duano has an average PSC of 50% with MUSI, 47% with BARISAN, and 58% with PM. While having a slightly higher PSC with the Pesisir site of Simpang Bayat (58%), the extremely low cognate rates indicate communication in their own dialects would not be possible.

Moving over to the center of southern Sumatra, the Kubu variety at Sungai Kijang (KUBU) and the variety from Bukit Tembesu (KJ1) in Jambi (Maryono, Setyonegoro & Kusmana 1997) share a surprisingly low PSC at first glance of 77%. However, given the small size of most Kubu groups and the extent to which they use other dialects (based on the SLQ from Sungai Kijang where Rawas was used in almost every domain other than in the home), it should be less surprising that there is less lexical similarity in these disbursed communities. These two Kubu sites have an average PSC of 82% with MUSI Malay, 74% with BARISAN, and 81% with PM. So they both have a higher similarity with an aggregate of MUSI sites than with their counterpart Kubu variety—seen

in the fact that they have their highest PSC of 86% with the the Rawas site of Pasar Surulangun (RAW-PS), with the next highest being other nearby Rawas and Musi Proper dialects, rather than the expected Jambi Ulu dialect.

Continuing westward, we find the Malayic varieties of Kerinci and MINANGKABAU. The Kerinci list has an average PSC of 75% with MUSI, 78% with BARISAN, and 88% with PM. The MINANGKABAU wordlist MIN1 has an average PSC of 73% with MUSI, 76% with BARISAN, and 87% with PM. From this we conclude the lexicon would not present a major obstacle for speakers of BARISAN or speakers of the Rawas dialect of MUSI who might come into contact with speakers of either MINANGKABAU or Kerinci. However, the fact that both of these Malayic varieties have a much higher PSC with PM indicates they should not group with MUSI or BARISAN based on the lexical evidence. Also, due to the phonological innovations both Kerinci and MINANGKABAU have, it is unlikely that speakers of BARISAN would understand them without prolonged exposure (which probably explains why many Pekal speakers stated they could understand MINANGKABAU). The Rawas location of Pasar Surulangun (RAW-PS), located near the border with Jambi and to the southwest of the MINANGKABAU and Kerinci speech areas has a fairly high PSC of 84% with MINANGKABAU and Kerinci.

The Muko-Muko dialects of MINANGKABAU showed an average PSC of 91% with MIN1 (Padang), and varied but slightly in their individual percentage, with Pondok Lunang having a PSC of 92%, compared to 89% PSC between MIN1 and Suka Pindah. The MINANGKABAU-influenced sites of Pekal have lower rates of lexical similarity with MIN1. The coastal BARISAN varieties of Bengkulu have absorbed more MINANGKABAU words into their lexicon than those in the mountains; for example in Pekal the coastal town of Pulau Baru (PKL-PL) has a PSC of 84% and the mountain town of Napal Putih has a much lower PSC of 76%. The city of Bengkulu, while further removed from Padang, has had even more vocabulary enter its lexicon due to the large number of MINANGKABAU merchants in the city, showing a PSC with MIN1 of 89%.

While lexicostatistical analysis did not include Bornean Malayic wordlists, we would expect that the percentage of shared cognates would be significantly lower than with other Sumatran Malay lects. Likely the highest percentages would be found between Pesisir/Palembang lects and similarly port-oriented Malays like Pontianak, Brunei or Banjar. In fact, what we do see when we compare Palembang Lama lects to Banjar Hulu (Adelaar 1992) is a striking correlation of Javanese loanwords: *gulu* 'neck', *ilat* 'tongue', *buntut* 'tail', *iwak* 'fish', *uyah* 'salt', *(h)abaŋ* 'red' and *(h)aɲar* 'new'. Banjar is also heavily influenced by Javanese, but it is interesting that they track so closely in terms of *which* lexemes are borrowed. This seems to connote significant contact between the ports of Palembang and Banjar.

4.2 Summary of lexicostatistic analysis

The lexicostatistic cognate counts provide some backing for the claimed comprehension or lack of comprehension by informants discussed in §7 below. These counts provide a *synchronic* look at the language situation in the research sites at the time of the visits. As such no attempt was made to discard non-native forms of the word, an exercise that would have been difficult particularly for the lects around Palembang. Rather the focus is on gauging the difference in lexical items as a possible indicator of difficulty in comprehension.

In all sites sampled, geographic proximity has a large influence on the similarity of cognates. Thus a HIGHLAND site might share a higher cognate count with a nearby UPPER MUSI site than with another HIGHLAND site that is far away. This is not surprising when we remember we are looking at a dialect network with fuzzy boundaries (when boundaries exist at all).

The percentage of shared cognates with in the language, within the cluster (if appropriate), and with PM and SI are recapitulated below in Table 4.10.

Table 4.10 Percentage shared cognates in languages of southern Sumatra

Language	PSC in language	Cluster	PSC in cluster	PSC with PM	PSC with SI
MUSI	83%	UPPER MUSI	85%	80%	78%
		PALEMBANG-LOWLAND	83%	78%	77%
BARISAN	79%	HIGHLAND	78%	76%	72%
		OGANIC	85%	78%	74%
Kubu[36]	77%		77%	81%	81%
MINANGKABAU[37]	93%	Muko-Muko	89%	82%	82%
HAJI	n.a.		n.a.	63%	60%

In the BARISAN language, the sampled lects show a very high correlation in the OGANIC cluster and a lower one in the HIGHLAND cluster, with a high PSC for the overall language. In the HIGHLAND cluster, the central lects centered around Lahat (Semenda, Besemah, Kikim, Lintang, Lematang Ulu) are very closely related lexically, explaining the sociolinguistic reports of high comprehension and similar ethnic identity. The varieties in the HIGHLAND cluster that had lower reported comprehension by other dialects—Kaur and Pekal—also had lower cognate counts that may explain some of the low comprehension. While the Bengkulu dialect had lower cognate rates with the other varieties in the HIGHLAND cluster, it was reportedly understandable. Bengkulu itself shared a higher cognate rate with Padang, perhaps due to the long influence of MINANGKABAU tradespeople and merchants living in the city. Bengkulu speakers felt that the HIGHLAND varieties of Besemah and Serawai were comprehensible. For the OGANIC cluster, all sampled sites had high rates of cognacy, partially explaining the reported high comprehension between lects.

The MUSI language showed high correlation in the individual lects and the clusters. As one would expect in a dialect change following a river, Pegagan (the downstream variety) was closer to Musi Proper (midstream and upstream) than Rawas (upstream) in the UPPER MUSI cluster. The high cognate counts between Col and the dialects of Musi Proper and Rawas (mid 80s), offers support to the results of the RTT and the SLQs which indicated that Col fits in the UPPER MUSI cluster of the MUSI language. This fact points out the importance of using comprehension tests, shared innovations, and sociolinguistic field work to predict comprehension rather than just using lexicostatistics. The PALEMBANG-LOWLAND cluster also had fairly high cognacy rates with the Pesisir, while internally everyday Palembang Pasar and Palembang *asli*/*dusun* had more words in common with each other than the nearby LOWLAND subcluster varieties.

The Kubu lect was not placed with either the UPPER MUSI or HIGHLAND clusters as they very much seem to have their own ethnolinguistic identity. From the standpoint of common vocabulary (as well as geography), they are closer to the MUSI language than the BARISAN cluster.

The Muko-Muko variety inclusion in MINANGKABAU was not contradicted by the lexicostatistical evidence, which also indicated a number of MINANGKABAU words in some varieties of Pekal and Bengkulu.

*4.3 Two reanalyzed final *r words*

Still on the subject of the SSML lexicon but shifting from statistics, we present here a brief discussion on two words, both ending in PM **r*, which have been reanalyzed in SSML or elsewhere: **lihər* and **air* 'water'. Both words' reconstructions are complicated.

Adelaar (1992:91) reconstructs PM **air* 'water'. However, he notes that Standard Malay also has the (odd) form *ayər*. He references further discussion (perhaps including the source?) in section "2.1.3" but section 2.1.3 does not seem to contain any discussion of the lexeme. Outside this cryptic reference, it seems that only in SSML has **air* been reanalyzed as as ⊗*ayar*.

[36] Only for the two Kubu sites compared lexically.
[37] Only for the three MINANGKABAU sites compared lexically.

The reconstruction for 'neck' is a bit more complicated. Adelaar (id. p. 39) reconstructs **lihər* based on a sole witness for the final vowel, that of Serawai (*liax*). All other lects in his sample point to a reconstruction of ⊗*lihir*. As it turns out, a large number of SSML lects evince what can be analyzed as pre-SSML *liar* 'neck'. In SOUTH BARISAN MALAY Malay, these lects include sampled sites from Rambang, Ogan, Pekal, Kaur, Besemah, Semenda, Lematang Ulu, Lintang and Kikim, while in MUSI, they include Musi Proper, Rawas and Pesisir, but not Palembang or Col. Besides SSML, reflexes of **lihər* are also found in Bangka and MINANGKABAU's Paseman lect (Sastra 1994), while in most of MINANGKABAU and indeed most of the rest of Sumatran Malay, one sees reflexes of ⊗*lihir*. Given that *liar* seems to be a retention, we will not attempt to incorporate it in our clustering of dialects.

Occurrences of *ayar* are less common than that of *liar*, but still over twenty sites have it: one Palembang Pasar site, the majority of Musi Proper and Rawas sites, Kaur, Pekal, one Rambang and some Ogan and Enim sites. If there is a possible cluster that links these dialects while excluding those without *ayar*, it escapes us.

We now shift our attention from lexical distribution to the geographical spread of phonological and morphological innovations.

5 Phonological and morphological innovations

The intent here of 'phonological innovations' is sound changes that have taken place systematically in a particular lect. For example, in certain varieties final **s* goes to *h* not just in one lexeme like *taŋis* 'cry'> *taŋih*, but in all instances of final **s*. Due to a largely similar lexicon and common traditions the Malayic language varieties of southern Sumatra can be hard to group. This comparison within similar environments is at the heart of the historical-comparative method, and can be helpful in understanding similarities and differences in dialect networks.

When speaking of phonological innovations from Proto-Malayic, it is helpful to differentiate between a number of phonotactic environments. The vast majority of native Malay words are, of course, disyllabic. Trisyllabic words are under strong pressure to reduce in some way to two syllables, and because of the lack of data will not be given more than passing treatment in this report. CVC structures are most common, with open syllables (VC or CV) and medial nasal-stop clusters also common. When looking at consonant innovations in disyllabic words, four environments will be differentiated: initial, medial, medial as part of a cluster, and final. Medial environments also are helpfully subdivided by whether the surrounding vowels are like or unlike. Final environments are often affected by whether the preceding vowel is high (*u*, *o*, *i*, *e*) or low (*a*).

When looking at vowel innovations, we will need to differentiate between ultimate and penultimate environments.

Here is a summary of common sound changes that have happened in one or more Malay varieties of southern Sumatra. For convenience sake they are separated into consonant and vowel innovations, although in some cases both the vowel and consonant are affected as a package. Their elaboration is somewhat cryptic but readers familiar with Austronesian sound changes, particularly those on Sumatra, will find these familiar, and they will receive more detailed explication below. In brackets are the basic parameters which can occur.

Relevant linguistic factors for grouping dialects:

- medial **h* between same vowels [yes, no, *ʔ*, mixed, unsure] (§5.1)
- final **h* [yes, no, mixed, unsure] (§5.1)
- reflexes of **s* [*s*, *h*] (§5.2)
- initial **r* [VD/VL, ALV/Back, *h*, Ø] (§5.3)
- medial **r* [VD/VL, ALV/Back, *h*, Ø] (§5.3)
- **-ri* sequences [*i*, *ay*] (§5.3)
- final HV-**r* (§5.3)
- final LV-**r* (§5.3)
- medial nasal- voiced stop [one/two segments] (§5.4)
- medial nasal- voiceless stop [one/two segments] (§5.4)

- penultimate high-vowel lowering and/or split (§5.5.1)
- ultimate high-vowel lowering and/or split (§5.5.2)
- final **a* [front, back, middle, low i.e., *e*, *o*, *ə*/*ɘ*/*ɨ*, *a*] (§5.6)
- final *diphthongs [diphthong, monophthong] (§5.7)
- ultimate high-vowel diphthongization (§5.8)
- penultimate closed **a* > *e* (§5.9)
- final closed **a*/**ə* 'Java schwa' [*a*, *ə*] (§5.10)
- initial voiced obstruent nasal assimilation [yes/no/stop but not affricate] (§5.11)
- transitive suffixes [*-kan*, *–ka* , *-an*, *-a*] (§5.12)
- kin terms [*ʔ* vs. *ŋ* e.g., *bapaʔ* vs. *bapaŋ*] (§5.13)
- **air* vs. reanalyzed *⊗ayar* 'water' (§4.3)

Not listed here is the complex of innovations associated with MINANGKABAU, which particularly affect ultimate syllables (e.g., *cəpat* 'fast' > *capeʔ*). While treated only cursorily in this report, it should be noted that these innovations have diffused down the west Sumatran coast down to Bengkulu city, in decreasing degree from north to south. See Adelaar (1995) for a discussion of the common MINANGKABAU sound changes.

We begin by discussing a number of innovations primarily affecting consonants.

*5.1 Realization of PM *h*

PM **h* is one of the phonemes that more easily undergoes deletion or change in SSML. HAJI, Kaur, and Kubu are the most conservative lects in this regard, as they retain **h* in word initial, word medial position between like vowels, and word finally, representing relic areas. See Anderbeck (2008:54–55) for a discussion of **h* in Jambi Kubu, where three of five Kubu sites still show high retention of word initial **h*. No grouping is proposed based on retentions, but it is interesting to note that the varieties that retain **h* in all positions were referred to as hard to understand or as strange by speakers of other varieties. Of course these lects which are more conservative in this regard have their own innovations in other respects, also helping explain this. See Figure 5.1.

The most widespread innovation in SSML (and likely elsewhere in Malayic lects) is the erasure of **h* word initially (e.g., *itam* 'black < **hitam*). Less common are the loss of **h* medially between like vowels (e.g., *tau* 'know' < PM **tahu(ʔ)*) and finally (e.g., *jao* 'far' < **jauh*). As these innovations are less widespread they are therefore more useful in grouping the dialects (refer to Table 5.1, Figure 5.2 and Figure 5.3).

Word medially, **h* deletion occurs in all members of the UPPER MUSI cluster. The PALEMBANG-LOWLAND cluster primarily preserves the **h* between like vowels, with the exception of Sarang Lang and two Belide villages. This may indicate the spread of this innovation from the UPPER MUSI cluster to nearby communities. The BARISAN language shows preservation of the **h* in all varieties with the exception of Benakat, the dialect bordering on the Musi River. It appears that it has adopted the medial **h* deletion innovation. Conversely, the glottal obstruent **h* has strengthened to a glottal stop in Serawai and Lintang (e.g., *jaʔat* 'bad, evil' < **jahət*), indicating a shared innovation in these lects of the HIGHLAND cluster.

In word final position, **h* is preserved in almost all BARISAN lects, while in the MUSI lects there appears a widespread elision of final **h* in the lects in the UPPER MUSI cluster and in sites near Palembang. The dialect center of Musi Proper, Sekayu, was less clear, with both preservation and deletion of final **h*. The PALEMBANG-LOWLAND cluster has deletion of **h* in sites located around Palembang, but the innovation has not spread to all of the varieties, particularly those located at a distance from Palembang. The dialect of Col, which often shares innovations with the Rawas and Musi Proper varieties in this case had one site with almost complete preservation of final **h*, one site with complete deletion of **h*, and other sites showing a mixture. In the HIGHLAND cluster, no WIST research points evinced loss of **h*, but one PBh research point for Bengkulu had loss of final **h*, while the other research point preserved it. The OGANIC cluster was also nearly uniform in preservation of final **h*—the exception being Rantau Alai which is the closest OGANIC site to the nearby MUSI dialect Pegagan. It appears the **h* deletion innovation has been adopted in the

downstream OGANIC region, with even the midstream Pengaringan site evincing significant loss of final **h*. In Pekal the final **h* is consistently strengthened to a glottal stop, e.g., *putia?* 'white' < **putih*.

The innovation of **h* deletion that appears to be the most helpful in grouping the lects is the deletion of **h* between like vowels (e.g., **jahat*, **puhun* 'tree'), showing an innovation operating across a large geographical area in a restricted number of lects that are either members of the UPPER MUSI cluster or border on the UPPER MUSI cluster. For example, the Air Itam (PB-AH) site also is next to the Pegagan speech area and also has many speakers of this variety living in their village. Geographic proximity facilitating the spread of an innovation appears to also be a factor for the two Belide sites (PB-CB and PB-TL) that are located close to the Musi River and UPPER MUSI speech communities.

Figure 5.1 preservation of initial *h

Figure 5.2 innovations of *-h

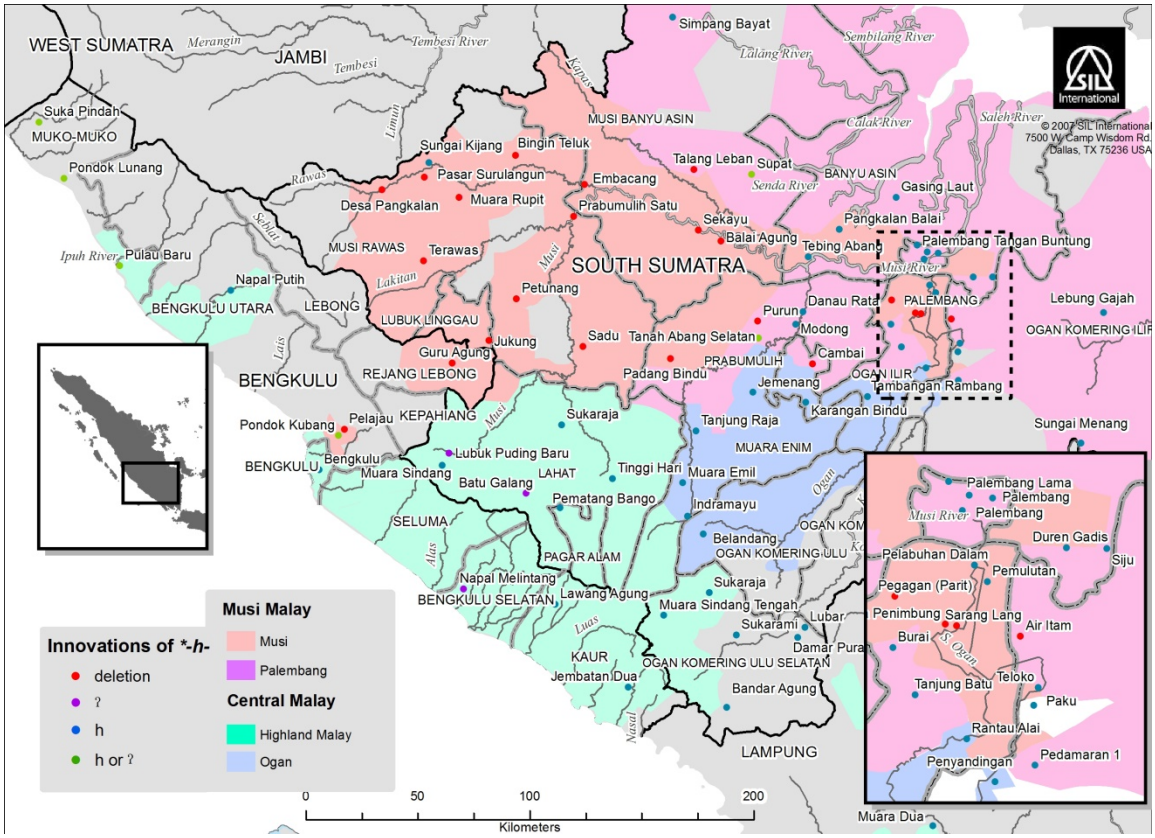


Figure 5.3 innovations of *-h

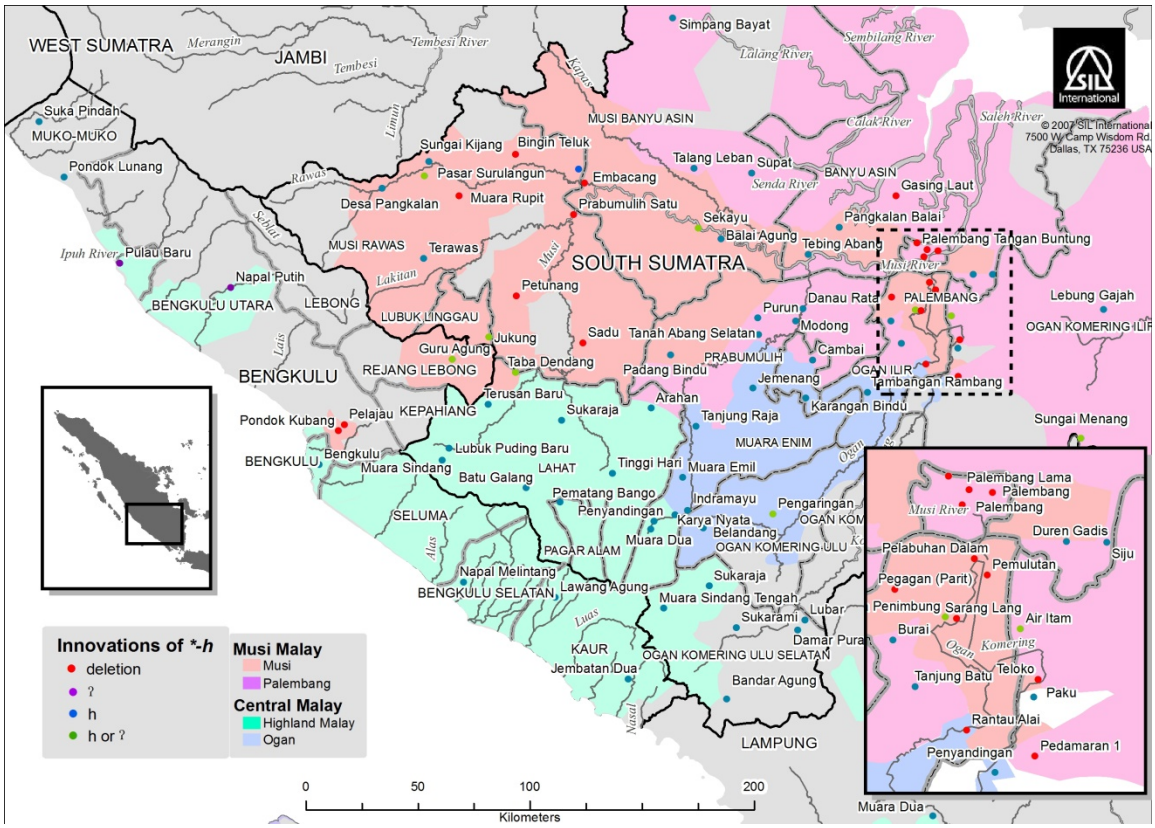


Table 5.1 Summary of *h *(MUSI)*

		Village	Code	#h__	v_h_v	__h#	%h#	%?#
(LOWLAND)	Penesak	Tanjung Batu	PB-TB	--	h	h	93%	0%
		Pedamaran	PB-PDR	--	h	h	82%	0%
		Burai	PB-BR	--	h	h	85%	0%
	Lem. Ilir	Tanah Abang Sel.	PB-TA	--	mixed	h	77%	4%
		Danau Rata	PB-SR	--	h	h	78%	4%
	Belide	Modong	PB-MD	--	h	h	100%	0%
		Cambai	PB-CB	--	—	h	90%	0%
		Talang Leban	PB-TL	--	—	h	100%	0%
		Tebing Abang	BEL-TA	--	h	h	85%	4%
PALEMBANG-LOWLAND (PALEMBANG)	Pal. Lama	Teloko	PL-TEL	--	h	--	9%	0%
		Paku	PL-PAK	mixed	h	h	100%	0%
		Palembang Lama	PL-PL	--	h	--	0%	0%
		Sarang Lang	PB-SL	--	—	--	5%	0%
		Pemulutan	PB-PM2	--	h	--	0%	0%
		Pelabuhan Dalam	PB-PM	--	h	--	27%	0%
	Pal. Pasar	Palembang City 1	PB-SH1	--	h	--	15%	0%
		Palembang City 2	PB-SH2	--	h	--	32%	0%
		Tangan Buntung	PB-SH3	--	h	--	18%	4%
		Gasing Laut	PB-GL	--	h	--	5%	0%
		Duren Gadis	PB-DG	--	h	h	73%	4%
		Air Itam	PB-AH	--	—	mixed	46%	12%
		Siju	PB-SJU	--	h	h	74%	4%
	Pesisir	Lebung Gajah	CST-LG	--	h	h	75%	0%
		Pangkalan Balai	CST-DB	--	h	h	81%	4%
		Supat	CST-SP	--	mixed	h	89%	7%
		Simpang Bayat	CST-SB	--	h	h	89%	7%
		Sungai Menang	CST-SM	--	h	mixed	59%	7%
UPPER MUSI	Rawas	Muara Rupit	RAW-RU	--	—	--	16%	4%
		Pasar Surulangun	RAW-PS	--	—	mixed	56%	3%
		Pangkalan	RAW-PN	--	—	h	81%	3%
	Musi Proper	Petunang	MU-KL	--	—	--	10%	3%
		Sadu	MU-SA	--	—	--	6%	3%
		Prabumulih Satu	MU-P1	--	—	--	15%	4%
		Purun	MU-PEN	--	—	h	77%	3%
		Bingin Teluk	MU-BT	--	—	--	19%	4%
		Sekayu	MU-KY	--	—	mixed	40%	4%
		Balai Agung	MU-BA	--	—	h	82%	4%
	Pegagan	Muara Penimbung	MU-PG	--	—	mixed	40%	4%
		Pegagan	MU-PG2	--	—	--	0%	0%
	Col	Jukung	COL-LL	--	—	mixed	allomorphs	
		Taba Dendang	COL-TT	--	na.	mixed	58%	0%
		Guru Agung	COL-PUT	--	—	mixed	55%	0%
		Terawas	COL-BKL	--	—	h	89%	0%
		Pelajau (L8)	COL-PLJ	--	—	--	3%	3%
		Pondok Kubang	COL-L8	--	mixed	--	0%	4%

*Table 5.2 Summary of *-h (non-MUSI)*

		Village	#h_	v_h_v	_h#	%h#	%ʔ#
BARISAN	OGANIC Rambang	Penyandingan	--	h	h	82%	0%
		Tam. Rambang	--	h	h	93%	0%
		Karangan Bindu	--	h	h	93%	4%
	OGANIC Enim	Jemenang	--	h	h	92%	4%
		Tanjung Raja	--	h	h	100%	0%
		Muara Emil	--	h	h	100%	0%
		Indramayu	--	h	h	100%	0%
	OGANIC Ogan	Rantau Alai	--	h	--	0%	0%
		Pengaringan	--	na.	mixed	42%	0%
		Damar Pura	--	h	h	86%	7%
		Belandang	--	h	h	96%	0%
	HIGHLAND Pekal	PBDSBB 24	--	h	ʔ	18%	82%
		PBDSBB 26	--	h	ʔ	9%	91%
		Pulau Baru	--	h	ʔ	4%	88%
		Napal Putih	--	mixed	ʔ	0%	100%
	HIGHLAND Bengkulu	Bengkulu PBh	--		--	0%	0%
		Bengkulu PBh	--		h	100%	0%
		Bengkulu City	--	h	h	68%	0%
	HIGHLAND Lintang	Batu Galang	--	ʔ	h	93%	4%
		Lubuk P. Baru	--	ʔ	h	76%	0%
		Terusan Baru	--	na.	h	83%	0%
	HIGHLAND Besemah	Sukaraja Kisam	--	h	h	100%	0%
		Sukaraja Kikim	--	h	h	100%	0%
		Pematang Bango	--	h	h	96%	0%
		Muara Sindang	--	h	h	93%	4%
		Lawang Agung	--	h	h	100%	0%
	HIGHLAND Lematang Ulu	Tinggi Hari	--	h	h	96%	0%
		Arahan	--	na.	h	100%	0%
	HIGHLAND Semenda	M. Sindang Tg.	--	h	h	96%	0%
		Semende-Holle	--	h	h	100%	0%
		Karya Nyata	--	na.	h	100%	0%
		Penyandingan	--	na.	h	100%	0%
		Muara Dua	--	na.	h	96%	0%
		Bandar Agung	--	h	h	96%	0%
		Tanggamus	--	h	h	96%	0%
	HIGHLAND Benakat	Padang Bindu	--	—	h	89%	4%
	HIGHLAND Serawai	Napal Melintang	--	ʔ	h	79%	0%
	HIGHLAND Kaur	Jembatan Dua	h	h	h	92%	0%
		PBDSBB 43	h	h	h	90%	0%
		PBDSBB 44	h	h	h	100%	0%
	Muko-Muko	Pondok Lunang	--	mixed	h	100%	0%
		Suka Pindah	--	mixed	h	96%	0%
	HAJI	Sukarami	h	h	h	100%	0%
	Kubu	Sungai Kijang	h	h	h	75%	0%

*5.2 PM *s > h, x*

PM **s* rarely shows innovation in medial or final position in SSML (contra MINANGKABAU, Kerinci and Highland Jambi with regards to final **s*). However, in the Col lects, initial **s* commonly goes to the voiceless glottal or velar fricative. This innovation was also noted in the Jambi Ulu dialect of Lubuk Kepayang (Anderbeck 2008:49). In all the Col varieties other than Lembak Delapan (Pondok Kubang and Pelajau in Bengkulu province)[38] initial **s* is realized as *x* or *h*, e.g., *hakit* 'sick, painful' < **sakit*. See Table 5.3 and Figure 5.4. Other SSML lects maintain **s* in word initial position, and accordingly are not displayed in the table.

Figure 5.4 innovation of *s > h, x

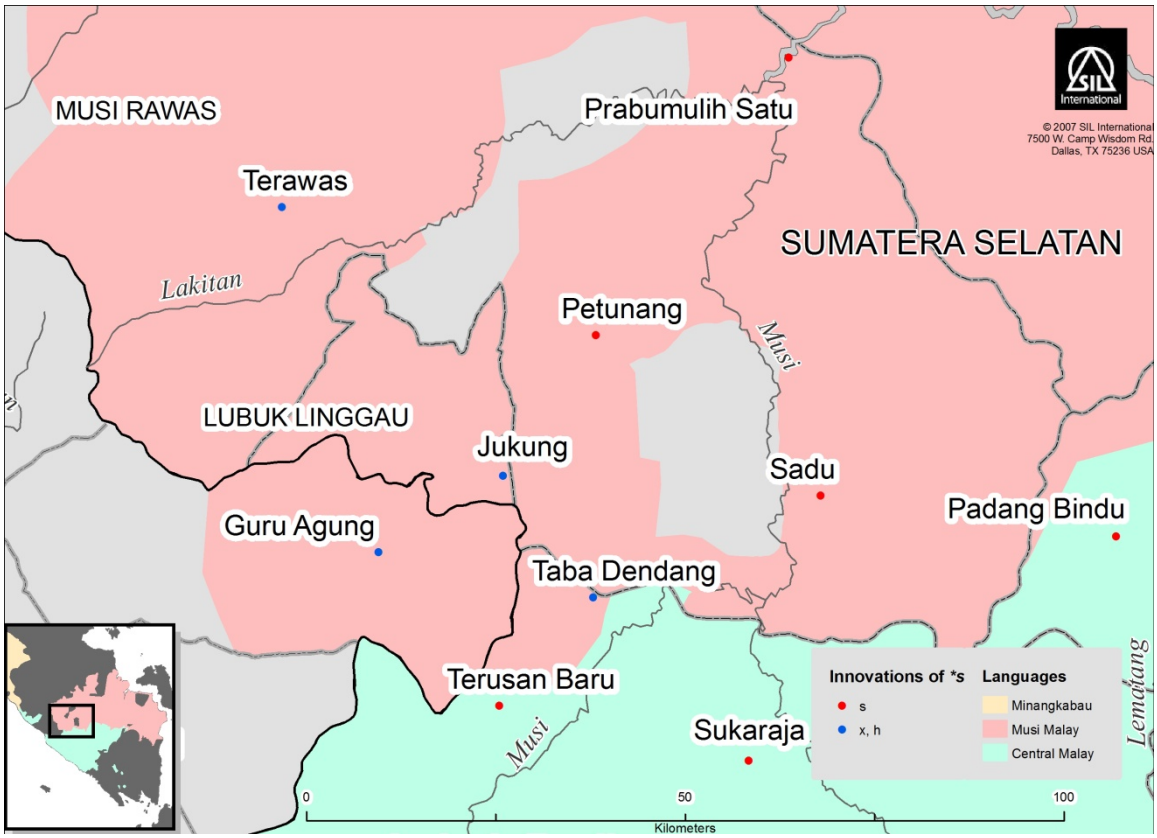


Table 5.3 *s > h, x

Notes	Village	Dialect	Code	s>h, x
	Jukung	Col	COL-LL	x, h
	Taba Dendang	Col	COL-TT	x, h
	Guru Agung	Col	COL-PUT	x, h
	Terawas	Col	COL-BKL	x, h
PBh list has mixed evidence, with 3 *h*/ 4 *s*	Pelajau (Lembak)	Col Lembak	COL-PLJ	s
	Pondok Kubang (Lembak)	Col Lembak	COL-L8	s

*5.3 Reflexes of PM *r in SSML*[39]

Reflexes of Proto-Malayic (PM) **r* in southern Sumatran Malay (SSML) are complex and various. Approximately twelve different phonetic reflexes of PM **r* have been identified, not counting the

[38] The Lembak Delapan list was checked against a PB list for the same area which showed mixed results, with three examples showing evidence for the innovation and four retaining **s*.

[39] The material in this and the high vowel sections was first presented as a conference paper (Anderbeck 2007h).

effects that **r* exerts over the vowels around it. In some cases, native reflexes have been replaced by a borrowed apical trill; in many more there is now a lexically-conditioned phonemic split between native and borrowed reflexes. Although complex, there are patterns that can be identified, in terms of geography and phonotactic environment, and these patterns are often suggestive of shared history.

5.3.1 Background

Proto-Malayo-Polynesian (PMP) has two *r*-like phonemes reconstructed, **r* and **R* (Blust 1999). The existence of the former, **r*, is in dispute, but if it did exist was considered to be an alveolar tap (Blust 1990:235) or trill (Ross 1992:40). The latter proto-phoneme, **R*, is considered to have been an alveolar or uvular trill (Blust id.). Adelaar (1992:83–86) concludes that, if both PMP proto-phonemes did exist, they must have merged in Proto-Malayic, and the phonetic quality of the resulting proto-phoneme (PM **r*) was a velar or uvular fricative. Collins (1986:11) came to basically the same conclusion. Dialectological research in Sumatra, for the most part, bears both of these points out. We believe we can go further and say that this back fricative was voiced (see §5.3.5 for a discussion of this point).

5.3.2 Intrusive trill

Adelaar (1992:83–86) notes a split in the Serawai reflexes of PM **r*, with some words reflecting an apical trill and others a back (read: velar or uvular) fricative. Adelaar attributes this innovation (in Serawai) particularly to the presence of loanwords from Standard Indonesian and more generally to the phonological pressure of Indonesian on Serawai. As we will see, this split is not limited to Serawai but is found in large parts of SSML. Does the larger corpus of data now available support Adelaar's conclusions?

Looking at the dialectological data from the region in focus, we see an entire spectrum with regard to the apical *r* innovation. A few areas have only the back fricative, even assimilating loanwords which originally had an apical trill. For example, Teloko (Palembang Lama) has *kəʁiŋət* 'sweat', from Javanese *kəriŋət*. On the other end of the extreme some varieties have only the apical trill, even in words that are clearly native to southern Sumatra like *liar* 'neck'.

In the rest of the spectrum are the majority of SSML lects, with some having 90% native reflexes and 10% borrowed trills, some the reverse, and some having quite a few of both. What patterns can be seen in the split between back fricatives and the innovative apical trill? Most important is lexical conditioning. Some lexical items always or nearly always have the innovative trill in the sample of 80-odd data points. Examples of words in this category are *rikin* 'count' (< Dutch), *ratus* 'hundred', *baru* 'new', *soroŋ*/*suruŋ* 'push', and *kubur* 'bury' (< Arabic). Conversely, there is another subset of **r* words which rarely innovates to the apical trill, such as *rusa* 'deer', *darah* 'blood', and *jəmur* 'dry (clothes)'.

The question that arises is why some lexemes frequently reflect [r] and others most often exhibit a back fricative. There are probably multiple explanations. Some words, as noted above, are loans. One phonological conditioning environment is word-medial between like vowels, for example, *buru* 'hunt' and *garam* 'salt'. Not all words in this environment innovate to trill, but a high proportion does. Another tendency is nearly as strong, which is for initial position to have the most common shift to [r], and final the least. See Table 5.4.

*Table 5.4 *ʀ to [r] by position*

position	average shift to [r]
initial	55%
medial	34%
final	16%

The most extreme example of this is Col, which averages 80% > [r] in initial position, 50% in medial, and 16% in final, although the whole MUSI language has a similarly pronounced pattern.

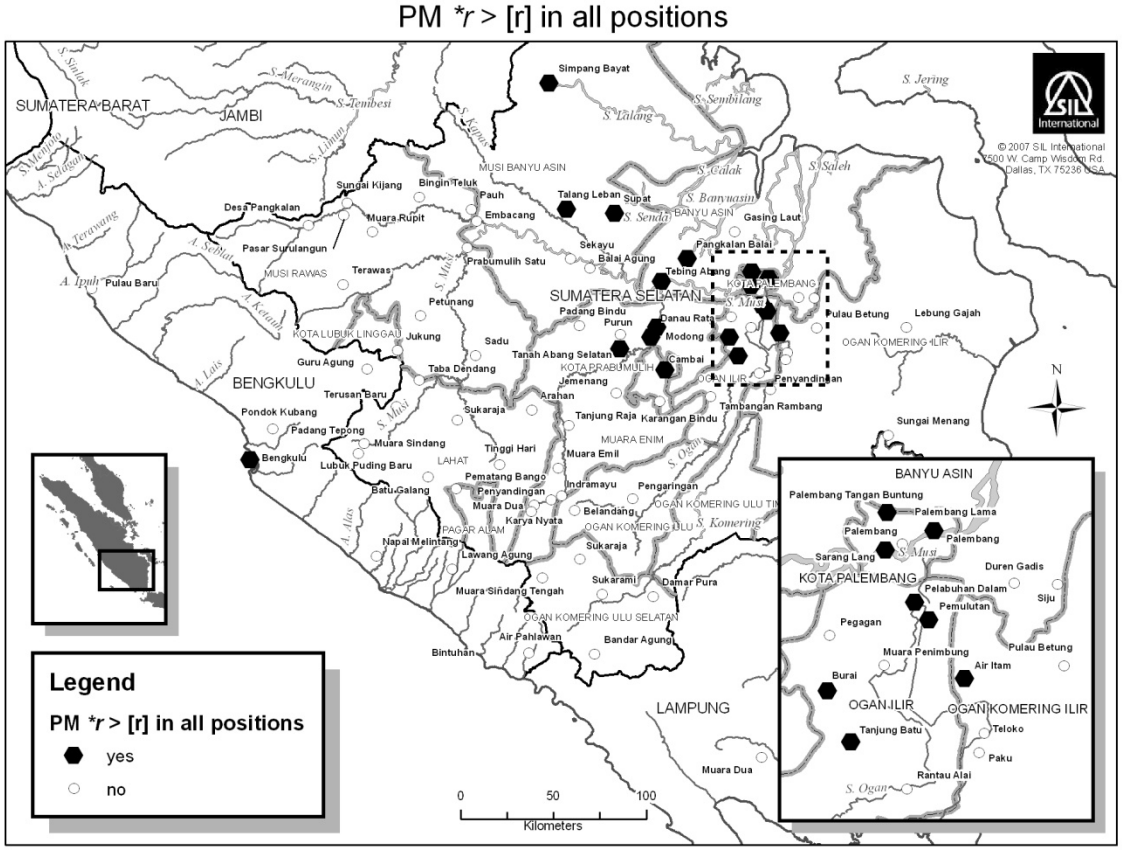
When one looks at the geographical distribution of this innovation, an interesting pattern emerges. For the most part this is a cosmopolitan innovation, resisted by Palembang Lama lects but embraced most vigorously by Palembang Pasar and Pesisir lects north of Palembang (see Figure 5.5). Made clear by this map is that there are no upstream sites which show a total conversion to the apical *r*, only downstream sites.

In several wordlists close to Palembang it appears that the lone exception to the trill innovation is the word for 'snake'. Instead of the expected ular we see *ulo*, *uli* etc. However, this apparent exception can be explained away by noting that the final vowel in this lexeme is identical to the set of words ending in **a*, indicating that the word has been reanalyzed as ⊗*ula*. See Table 5.5.

Table 5.5 Reanalysis of ular 'snake'

Location	'snake'	*-*a*
Pemulutan (both lists)	uli	i
Sarang Lang	ulɔ	ɔ
Tangan Buntung	ulo	o
Palembang City	ulo	o
Duren Gadis	ulo	o
Gasing Laut	ulo	o

Figure 5.5 PM *r *> [r] in all positions*



Lexical Conditioning

In the areas of a split in reflexes of *r*, more important than phonotactic environment is lexical conditioning. Some lexical items always or nearly always have the innovative trill in the sample of 80-odd data points.[40] See Tables 5.6 and 5.7.

Table 5.6 Lexical items unanimous > [r]

word	**gloss**	**comment**
rikin	count	< Dutch
ratus	hundred	
ribu	thousand	
rotan	rattan	
rakit	raft	
rəjuŋ	sing	only found in Semenda
rupuʔ	think	only found in Semenda
rənay	dirty	only found in Col
kəru-an	know	
guruh	thunder	
baru	new	
karut	evil	
pərau	boat	
paraŋ	machete	
marah	angry	
purik	angry	
soroŋ/suruŋ	push	
surum	wear	
kotor	dirty	
lumpur	mud	other words are more common for 'mud'
kubur	bury	< Arabic
aŋar	new	
putar	revolve	

Table 5.7 Lexical items mostly > [r]

word	**gloss**	**comment**
bə-rənaŋ	swim	
rambut	hair	
rai	face	
buru	hunt	
buruŋ	bird	
garam	salt	
paraʔ	close	significant minority retains back fricative
biraʔ, bireʔ	defecate	
kəriŋat	sweat	
pikir	think	< Arabic
pasir	sand	
bayar	pay	significant minority retains back fricative

Conversely, there is another subset of *r* words which rarely innovates to the apical trill. This category is too fuzzy to delineate exactly, but some prominent words in this category (written in their canonical form) are *rusa* 'deer', *darah* 'blood', *jarum* 'needle', *garut* 'scratch', *hari* 'day', *jari* 'finger', *urat* 'vein', *duri* 'thorn', *bəras* 'rice', *duara* 'door', *harum* 'fragrant', *(kə)mari* 'yesterday', *bibir* 'lip', *ilir* 'downstream', *tidur* 'sleep', *təlur* 'egg', *niur* 'coconut', *kapur* 'chalk (lime)', and *jəmur* 'dry (clothes)'.

[40] The data points which are all one way or the other are not included in these calculations.

Due to lack of depth in the dialectological data, we will not revisit the question of whether the split in reflexes traces back to different phonemes in Proto-Malayo-Polynesian. But are there other possible causes? One can notice that, in nearly half of the examples given in the tables above, the *r* environment is word-medial. There are one and a half times as many examples of [r] sandwiched between like vowels, than between different vowels. This seems exceptionally high. As a point of comparison, the first 100 Indonesian lexemes of the Basic Austronesian Wordlist show 57% different vowels in the two syllable positions, or almost opposite the ratio for the southern Sumatra [r] words. As another point of comparison, in the set of words mentioned above which rarely innovate to the apical trill, only one of the eleven word-medial examples is between like vowels. Might there be something about the articulatory environment of *r* between like vowels that has triggered a tendency toward shift to an apical trill?

There are a few words which exist as *d*/*r* doublets, geographically speaking: *bərənaŋ*/*bədənaŋ* 'swim' and *dai*/*rai* 'face'. In both of these examples, the *r* reflexes are nearly unanimously the apical trill, leading one to the likely conclusion that **d* weakened to a trill or tap at the same point of articulation.

Returning, then, to the topic of lexical conditioning, there are a number of words that would seem to have been borrowed from SI or other cosmopolitan sources. Among these candidates are words which are actually descended from PM. For example, the words *ratus* 'hundred' and *ribu* 'thousand'. Even in Jambi, which is more conservative in its phonology, these two words differ from the norm by exhibiting the apical trill (Anderbeck 2008). It seems likely that the above words were lost in some colloquial varieties of Sumatran Malay, then re-borrowed from the standard. Another commercial term, *bayar* 'pay' may have a similar history. Likewise, PM **buru* 'hunt' has been reconstructed, but in southern Sumatra the reflexes mostly have the apical trill. Ogan (Pengaringan) seems to have two different reflexes of PM **buru*: [buru] 'hunt' and [buχu] 'chase away'. It is very interesting that the latter, native-seeming reflex, preserves the PMP definition (while nearly all other ML varieties we are aware of have the innovative definition of 'hunt'). [41] At the same time it seems to indicate that, for some reason, [buru] 'hunt' has been borrowed from another source.

Additionally, there are certain words which, whatever their PM pedigrees, are not common in Sumatran Malay. This is the case with *pasir* 'sand', *putar* 'turn, revolve', *kotor* 'dirty' and, specifically in southern areas, *rambut* 'hair' and *baru* 'new'. As for the rest of the words noted above, we have no explanation at present.

5.3.3 *r *to ʔ*

Another interesting innovation is **r* > *ʔ* which happens only in word-final position. Like the apical trill innovation, **r* > *ʔ* is lexically conditioned, occurring frequently in some lexemes and rarely or never in others. Words which universally innovate are *s(əb)utiʔ* < **sə-butir* 'one', *ikoʔ* < **ikur* 'tail' and *bəsaʔ* < **bəsar* 'tail'. Frequent innovators are *tidoʔ* < **tidur* 'sleep', *təloʔ* < **təlur* 'egg', *aiʔ* < **air* 'water', and, less frequently, *libaʔ* < **libar* 'wide'.

Unlike the apical trill innovation, the glottal stop innovation seems it could be homegrown rather than imported. At the minimum this innovation has spread as a *process* (applied to specific words) rather than as lexemes borrowed en toto. Consider the following evidence: *ʔ* occurs in lexemes that otherwise have typical regional phonological processes acting on them. For example, *səbutiʔ* 'one' is frequently shortened to *sutiʔ*, making the word disyllabic, as happens with other words like *boyo* (< *buaya*). In areas with ultimate closed high vowel lowering (§5.5.2), the vowels in the *ʔ* words are also lowered (e.g., *təloʔ*), whereas in areas without vowel lowering, the vowels in question are high (*təluʔ*). Most tellingly we have cases where the vowel went through some distinctive sound changes (see below about Musi vowel merger before final **r*) before **r* > *ʔ*. [42]

---

[41] Kota Bangun (Kutai Lakes) Malay in East Kalimantan province also has *buru* 'chase after; chase out' (Collins p.c.).

[42] The phonological congruence of the **r* > *ʔ* set to the varieties in question begs the question: if this innovation has spread via the phonology versus the lexicon, how can we say that this innovation is

In terms of geographical distribution, there is not really an identifiable center of the innovation*r > ʔ; it is fairly evenly spread through large parts of SSML. A few clusters have percentages that average 10% *above the median*, namely **LOWLAND subcluster**, some dialects of **Palembang**, and (slightly less frequent but more consistent) the **BARISAN** subclusters of Besemah, Semenda and Lintang.

The remainder of this section will look exclusively at the reflexes of PM *r which did not innovate to either [r] or [ʔ] but rather retained some aspect of back fricative.

5.3.4 UPPER MUSI cluster innovations

The collection of lects often identified (moving downstream) as Lembak/Col, Sindang, Musi Proper/Sekayu and Pegagan/'Ogan Ilir' show a striking cluster of innovations to *r. Here we discuss three innovations by which these lects stand together and stand apart from the rest of SSML.

- Initial *r > Ø. Only in the lects mentioned above do the majority of non-[r] reflexes completely elide in *initial* position, e.g., *umah* 'house' < rumah and *use* 'deer' < **rusa*.
- Final **ri* sequences. Perhaps the most unique innovation of the UPPER MUSI cluster is change of the sequence **-ri* > *ray*, e.g., **jari* 'finger' > *jaray* and **duri* 'thorn' > *duray*.
- Merger of final **ar*, **ur*, **ir*. This final innovation is actually itself a concatenation of a number of related processes and results. In the UPPER MUSI cluster, final **ar*, **ur*, **ir* (when not > [r] or [ʔ]) have all merged to *o(ɰ)*. This can possibly be attributed to coda-driven nasalization of the syllables which then makes the *r* redundant for differentiating such a word from other potential homophones. Examples are **bibir* 'lip' > *bibo*, **hilir* 'downstream' > *ilo*, **tidur* 'sleep' > *tido*, **kapur* 'lime' > *kapo*, **lihər* 'neck' > *lio* and **akar* 'root' > *ako*.

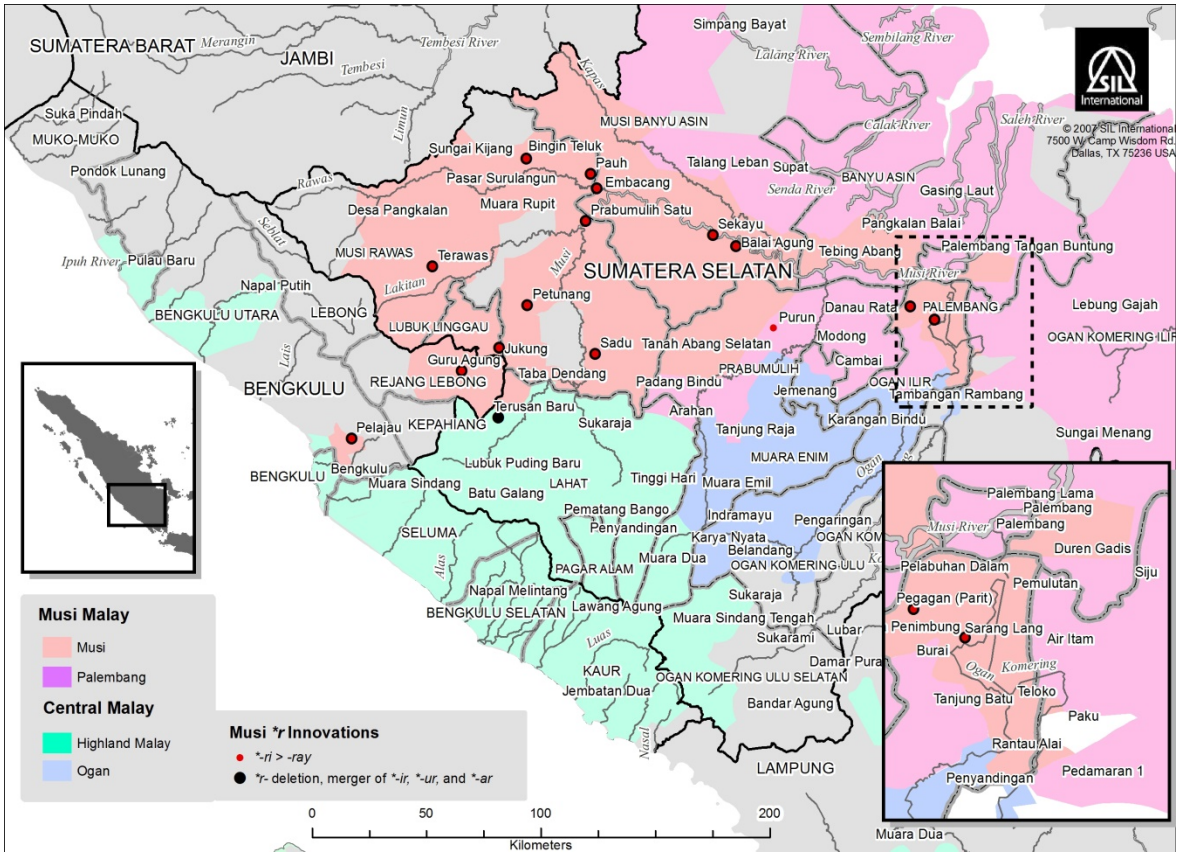
This is a striking innovation and it is worth digging into the question of how this may have developed. Thankfully the dialectological data give us some evidence. This is how it seems to have happened. First, the voiced fricative underwent lenition. Then, to perceptually compensate for the lenition, the final consonant had a secondary articulation applied of lip rounding (cf. Lebung Gajah's [bibiw͡ɰ]). Third, the lip rounding began to have vocalic manifestations, producing sequences like Arahan's [bibiew]. Finally, the lip rounding completely subsumed the previous vowel (and in some cases the original consonant also disappeared entirely). Some of these innovations happened in areas around the UPPER MUSI cluster, particularly Lematang (**LOWLAND**) and Pesisir areas southeast of Palembang. But only in UPPER MUSI itself did all the innovations come together for a full merger of all three *V*r* sequences.

See Figure 5.6 for a visual presentation of the geographical extent of the innovations discussed in this section. Note also that the sites on the Rawas tributary of the Musi are largely exempt from these innovations.[43]

lexically conditioned? Our only response at present is that 'lexical conditioning' is another way of saying that, if there was a phonological conditioning factor, that factor is not apparent to us.

[43] The Rawas sites are limited to the innovation of **ar* > *o*, e.g., **akar* > *ako*.

Figure 5.6 UPPER MUSI cluster innovations



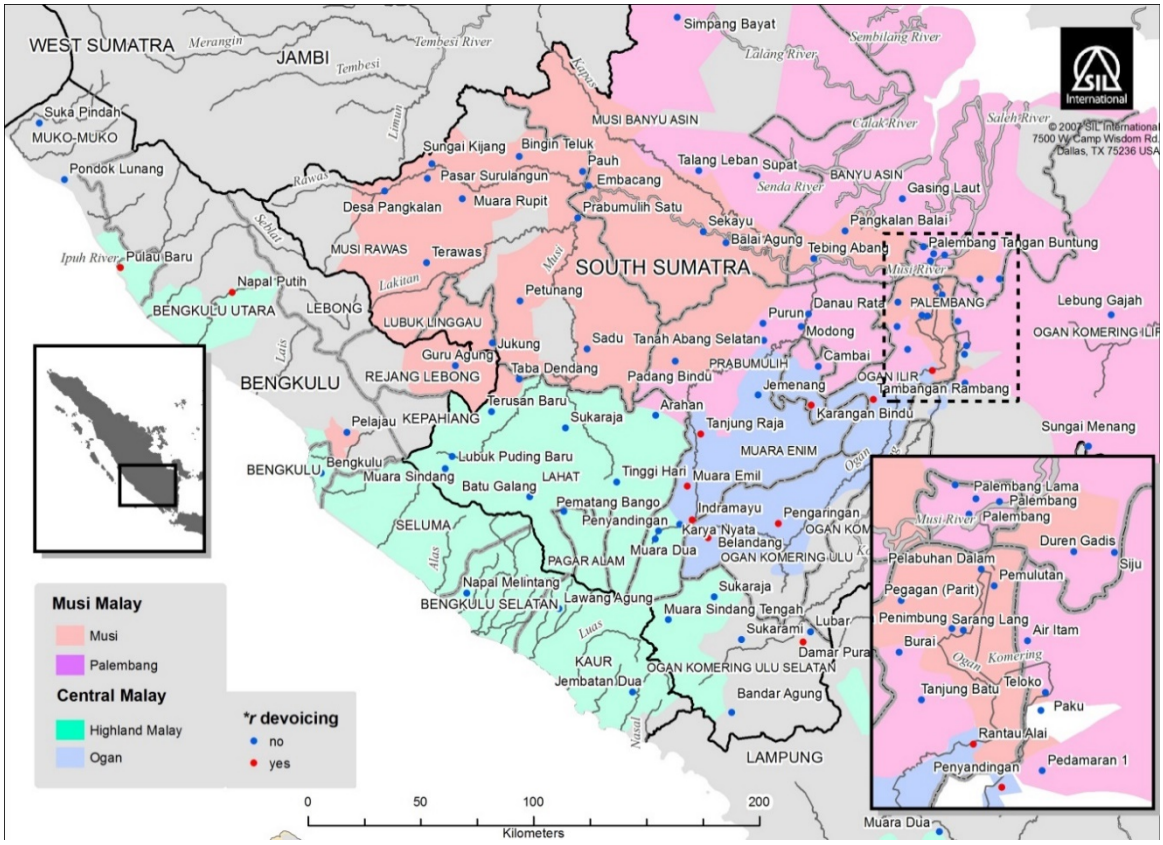
5.3.5 *OGANIC subcluster devoicing of* *r

In the lects of the OGANIC subcluster, native reflexes of **r* are devoiced in all positions. Some examples are given from Belandang in the furthest upstream reaches of the Ogan River: *χumpʊt* 'grass', *kuχʊs* 'skinny' and *nioχ* 'coconut'.

We conclude that the directionality of this change should be seen as from voiced to voiceless for two reasons. First, the innovation occurs in a limited geographical/cluster range, with few exceptions restricted to OGANIC and Pekal. (In Pekal this devoicing is followed by subsequent lenition to *h* word-finally.) Second, other changes involving **r* could only have happened with an earlier voiced fricative. These changes include sporadic fortition of **r* > *g* (e.g. *gumput* 'grass' < **rumput*) in HIGHLAND lects, and voicing-driven nasalization of final *V*r* in UPPER MUSI (§5.3.4).

See Figure 5.7 to see the distribution of this devoicing innovation.

*Figure 5.7 OGANIC subcluster devoicing of *r*



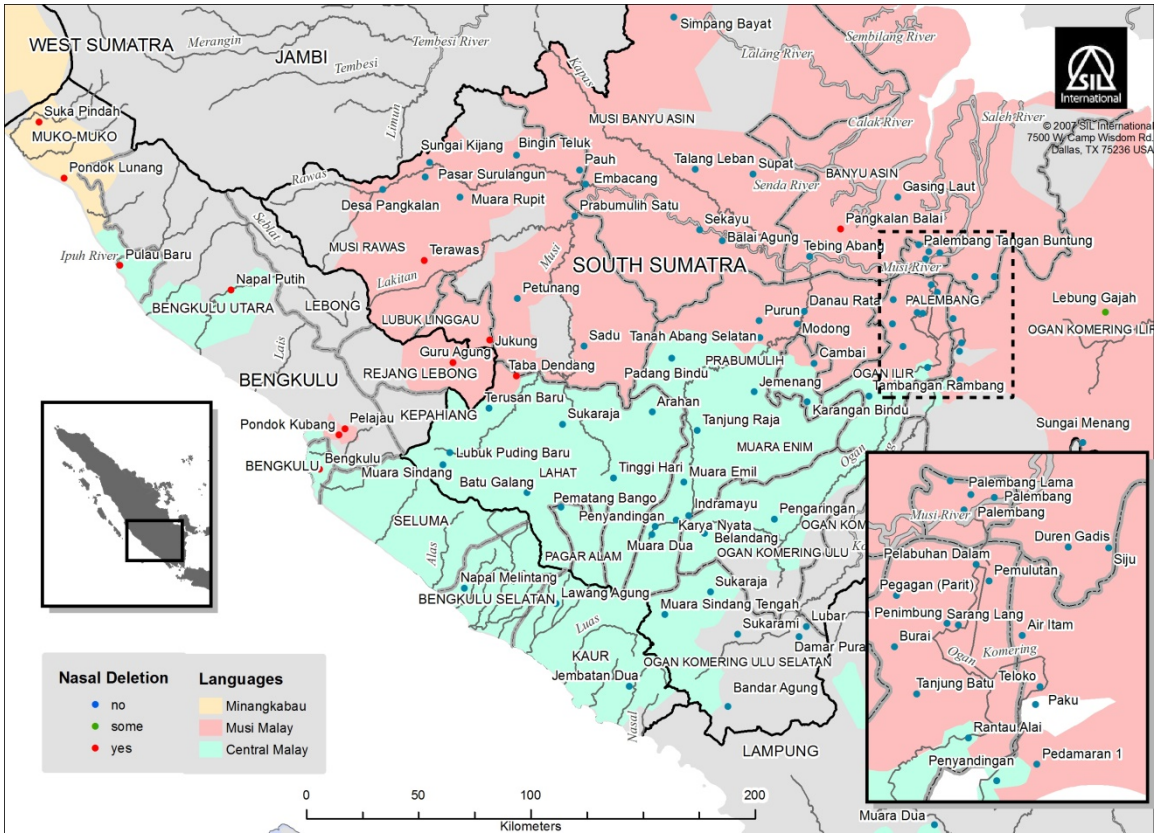
5.3.6 *Miscellaneous* *r *innovations*

- The Kubu (Anak Dalam) site in the Rawas area shows devoicing of **r* in final syllables only, e.g., *tiduχ* 'sleep'.
- A number of sites in the HIGHLAND cluster (Tinggi Hari, Arahan, Kikim Sukaraja, Benakat Padang Bindu, all Semenda sites, and Besemah Sukaraja Kisam) show **-ur* > *ow*, e.g., *təlow* 'egg' < **təlur*.
- Pekal in northern Bengkulu province has **r* > *h* in all positions (and further elision to Ø in initial position), e.g., *imbu* < **rimba* 'forest', *buhu* < **buru* 'hunt, *ulah* < **ulər* 'snake'.
- Bengkulu and Kaur dialects evince epenthesis of low vowel between high vowel and final **r* (but not other consonants), e.g., Kaur *təlʊᵊʕ* 'egg' < **təlur*. This is fairly common for western Sumatran Malay varieties and is most likely an innovation that has spread from MINANGKABAU.
- In some dialects of MINANGKABAU, including the Pancuang Soal/Muko-Muko dialect north of Pekal, final **r* is completely elided, e.g., *ula* 'snake' < **ulər*, *təluᵊ* 'egg' < **təlur*.
- In HAJI, the high vowel is lowered before **r* (but not before other consonants; Anderbeck 2007a). Given the geographical separation from West Sumatra and lack of any other shared innovations, it is doubtful that this innovation is from MINANGKABAU influence.

5.4 Nasal deletion before voiceless stops

A limited number of SSML lects exhibit deletions of nasal before voiceless stops in word medial position (e.g., *jatuŋ* 'heart' < PM **jantuŋ*), or a marked reduction in prominence (*jaⁿtuŋ*). This appears to be an areal feature, occurring in the northwest corner of South Sumatra Province and the adjoining area in Bengkulu. It is also found to a much lesser extent in the large swamp land to the southeast of Palembang. See Table 5.8 and Figure 5.8.

Figure 5.8 occurrence of nasal deletion before voiceless stops



Of the Malay lects examined in the survey, three exhibit loss of nasal before a voiceless stop: Col and Rawas (part of the MUSI language) and Pekal (part of the BARISAN language). To a lesser extent, the Lebung Gajah variety of Pesisir (part of the PALEMBANG-LOWLAND cluster of the MUSI language) shows loss of prominence of nasal before voiceless stops. Two other non-Malayic varieties also showed the same innovation: Muko-Muko (part of the MINANGKABAU language) and Rejang, with the five Rejang sites surveyed by WIST in 2005 also showing loss of nasal before voiceless stops (West Indonesia Survey Team 2006; Blust 1984:428).

The research sites where deletion of nasal before voiceless stops occurs (excluding apparent doublets) is presented in the tables below. No environment seemed to condition the deletion of nasal before voiceless stops in those where it only occurred part of the time. Data collected by WIST was compared against wordlists from PBh publications, in particular *Pemetaan Bahasa Daerah Sumatra Barat dan Bengkulu* (PBDSBB), and similar results were obtained.

The nasal deletion before voiceless stops is seen as an areal feature as it does not extend to all members of one language, but rather extends to the closely located varieties of disparate languages. For example, the nasal deletion seen in Muko-Muko is not found in other MINANGKABAU dialects. Likewise, the Rawas lect Pasar Surulangun only has a very limited amount of the nasal deletion innovation, compared to the Rupit variety. The Pangkalan upstream Rawas variety shows a fairly high occurrence of nasal deletion preceding a voiceless stop (80%) but the informant was inconsistent in this feature of her speech, sometimes pronouncing a word with and without the nasal (e.g., rumpʊt/ ɯɯ^{m}pʊt). Another evidence for this being part of geographic spread is its high occurrence in Petunang (56%), part of the Musi Proper dialect of the MUSI language, whose other dialects do not show this innovation. This is also backed up by the PBh wordlist for Sindang Kelingi, with an even higher percentage of 82%. In the BARISAN language, the nasal deletion before a voiceless stop is found in Pekal, but it is not seen again further south (Bengkulu, Besemah, Kaur, etc.).

The frequency of the innovation decreases as one moves away from the center, another indication of an areal feature. Petunang again is an illustration of this with a lower frequency of

nasal deletion (56%) than the Col languages further upstream from it (100%), but a higher percentage than the Musi Proper dialects downstream (0%). Interestingly, in words where there is an apparent doublet (e.g., *gacaŋ*/*gancaŋ* 'fast'), the sites that use the item without the nasal are located in or nearby the areas where nasal deletion has occurred.

Table 5.8 occurrence of nasal deletion before voiceless stops

Language	MUSI language—PALEMBANG subcluster (Palembang and Pesisir)			
Site	Sarang Lang	Lebung Gajah	Sungai Menang	Siju
nasal	Deletion	reduction	reduction	reduction
occurrence	10%	56%	17%	5%
after high vowel/ low vowel	Low	both	high vowel	neither (one case)

Language	MUSI language—UPPER MUSI cluster (Rawas and adjacent Musi Proper varieties)					
Site	Pasar Surulangun (Rawas Ulu)	Pangkalan (Rawas Ulu)	Rupit (Rawas)	RAWAS (PBh)	Petunang (Musi Proper Kelingi)	Sindang Kelingi (PBh)
nasal	reduction	deletion	deletion	deletion	reduction	deletion
occurrence	7%	80%	88%	100%	56%	82%
after high vowel/ low vowel	low vowel (one case)	both	both	both	both	both

Language	Col varieties of UPPER MUSI				
Site	Terawas, Guru Agung, & Taba Dendang	Jukung	Pondok Kubang (Bengkulu)	Lembak (PBh)	Pelajau (Bengkulu)
nasal	deletion	deletion	deletion	deletion	deletion
occurrence	100%	90%	70%	100%	100%
after high vowel/ low vowel	both	both	both	both	both

Language	Rejang	Muko-Muko varieties of MINANGKABAU		
Site	Five sites surveyed by WIST, 2005	MUKO-MUKO PBDSBB 20-23	Pondok Lunang	Suka Pindah
nasal	deletion	deletion	deletion	deletion
occurrence	100%	85% - 100%	93%	100%
after high vowel/low vowel	both	both	both	both

Language			Pekal varieties of SOUTH BARISAN MALAY
Site	Napal Putih	Pulau Baru	PEKAL PBDSBB 24, 26
nasal	Deletion	deletion	deletion
occurrence	100%	100%	100%
after high vowel/ low vowel	Both	both	both

We now move to a discussion of vowel innovations.

5.5 PM high vowels in southern Sumatra[44]

Proto-Malayic, it is generally agreed, had four plain vowels including two high vowels **u* and **i* (Adelaar 1992:32). In many Malay varieties, the high vowels underwent a split where **u* became *u* or *o*, and **i* became *i* or *e*. For example, **ikur* 'tail' became *ekor*, **libar* became *lebar*, while **pisaŋ* and **urat* kept their original shapes.

While for a time it was thought that all Malay varieties had undergone this split (Asmah Haji Omar 1977), a number of publications since then have documented Malayic lects that did not undergo a split in the high vowels. The clear pattern for penultimate vowel split is that the areas most exposed to 'standard' or 'cosmopolitan' Malay varieties are affected, while more isolated are not. Anderbeck (2008) shows that, for Jambi Malay, the downstream and coastal varieties have the split, while upstream varieties do not but rather retain the 4-vowel system. How about SSML lects?

When the subject of lowering of Proto-Malayic high vowels is discussed, we need to distinguish between occurrences in the ultimate and penultimate syllables, as lowering in those environments differs in terms of triggers and geographic distribution. Ultimate vowel lowering is more frequently non-phonemic, often triggered by closed syllables in general and/or non-oral consonants. Penultimate vowel lowering is phonemic and in some cases lexically-conditioned. Other triggers are more complex. Adelaar (1992) treats this topic in satisfying depth, although his work gives the unfortunate impression that both penultimate and ultimate vowel splits are the same innovation. We hope to demonstrate in greater detail than in Anderbeck (2008) that they are in fact distinct. Table 5.9 gives a summary of the differences.

Table 5.9, Summary of penultimate and ultimate high vowel lowering

penultimate	**ultimate**
triggered primarily by presence of *high vowel in closed ultima	not influenced by penultimate vowel
phonemic	non-phonemic
'cosmopolitan' geographical distribution	largely unrelated to cosmopolitan influences

We begin by focusing on the split in penultimate high vowels.

5.5.1 Split in penultimate high vowels

While an absolute rule cannot be formulated to predict which SSML lexemes will exhibit high vowel lowering in the penult, the tendency could be given like this:[45]

> Penultimate high vowel lowering frequently occurs in lexemes with closed ultima which themselves contain lowered high vowels.

In phonological notation, this tendency can be expressed by: hV > mid-vowel/ _C(C)hVC#. Examples of this phenomena are *pendek* 'short' < **pindik*, *bohoŋ* 'lie' < *buhuŋ*, *gosoʔ*/*kosoʔ* 'rub' < **kusuk* and *beloʔ* 'turn' < **biluk*. This rule is not absolute: many lexemes with the same conditions do not undergo penultimate high vowel lowering, such as *guruh* 'thunder', *busuk* 'rotten', *idup* 'live', *diŋin* 'cold', etc.

Lowering does not occur in lexemes with open final syllables.[46] SSML examples of penultimate high vowel lowering in lexemes with ultimate low vowels are so rare as to be noteworthy: *tembaʔ* 'shoot', *rotan* 'rattan', *ŋ-oap* 'yawn', *otaʔ* 'brain', and *obat* 'medicine'. Not

[44] Due to the exigencies of time, this section unfortunately does not contain data from later visits to Pesisir, additional Musi, Rawas and Bengkulu sites, as well as Kubu.

[45] The factors discussed here do not all apply to Pekal, the most phonologically aberrant SSML lect.

[46] There is occasional slight penultimate high vowel lowering in CVCCV lexemes like *mimpi* 'dream'; the vowels in these syllables act in many ways like ultimate closed high vowel lowering.

only are these examples rare, they also are lowered much less frequently in the various lects than those with high ultimate vowels.

It would be tempting to claim, therefore, that penultimate high vowel lowering is triggered by ultimate high vowel lowering. Besides the rather rare counterexamples above where lowering in the penult precedes ultimate low vowels (e.g., *ota?* 'brain'), there is also dialectal counterevidence. In certain UPPER MUSI lects, particularly Col and Pegagan, the penultimate high vowel is frequently lowered while the ultimate vowel remains high, e.g., *tojuh* 'seven', *boliŋ* 'turn' and *tepis* 'thin'. We will thus stay for now with the weaker statement that these phenomena frequently co-occur.

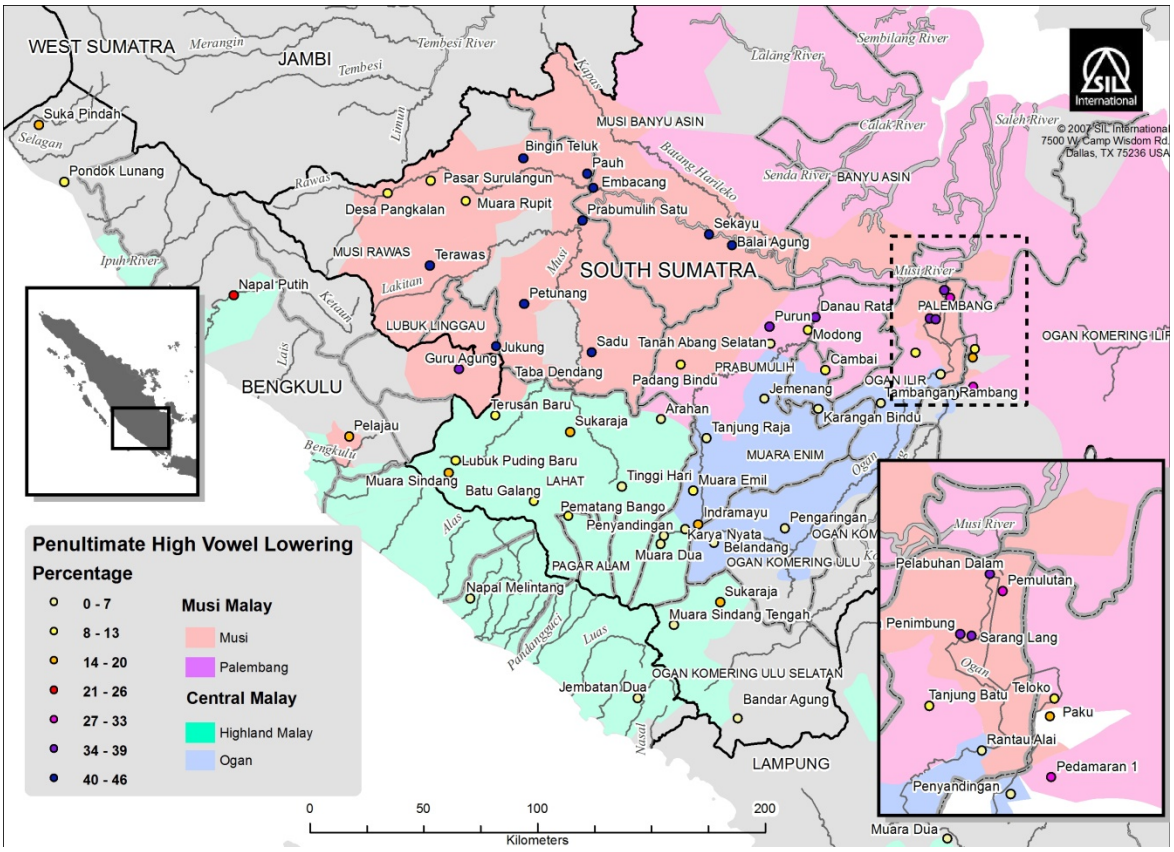
In terms of geographical distribution, penultimate vowel lowering correlates to more 'cosmopolitan' influences, bearing a striking resemblance to the situation in Jambi province to the north (Anderbeck 2008). As in Jambi, the areas of greatest percentage of penultimate high-vowel lowering are around the capital (and port) city, and following up the main river, in this case the Musi. Table 5.10 gives an average of the various clusters.

Table 5.10 Penultimate vowel lowering

Group	**% lowered**
PALEMBANG	27%
LOWLAND	18%
Musi Proper/Pegagan/Col	37%
Rawas	8%
OGANIC	5%
Semenda	5%
Lintang	9%
Besemah	13%
Kaur	4%
Pekal	26%
Muko-Muko (MINANGKABAU)	12%
Indonesian	18%
Jambi Ilir	14%
Jambi Ulu	3%

See Figure 5.9 for a map showing the percentage of penultimate high vowel lowering in the sample.

Figure 5.9 Geographical distribution of penultimate high vowel lowering



A word should be said about lexical conditioning. Particularly for the groups whose average percentage of high-vowel lowering is below 10%, the question could be asked whether for them the split of high vowels is phonemic. While this study does not directly address that question, Table 5.11 gives a listing from our database of all the words which exhibit vowel lowering in 50% or more of the isolects sampled.

Table 5.11 Words with > 50% lowering

word	**meaning**	**comment**
pendeʔ	short	same as Indonesian
tembaʔ	shoot	same as Indonesian
ŋ-omoŋ	talk	same as Indonesian
teteʔ	breast	same as Indonesian
rotan	rattan	same as Indonesian; originally derived from **raut-an*
belok	turn	same as Indonesian
gosok	rub	same as Indonesian
bohong	lie	same as Indonesian
neneʔ	grandmother	same as Indonesian
tokoʔ	neck	not in Indonesian

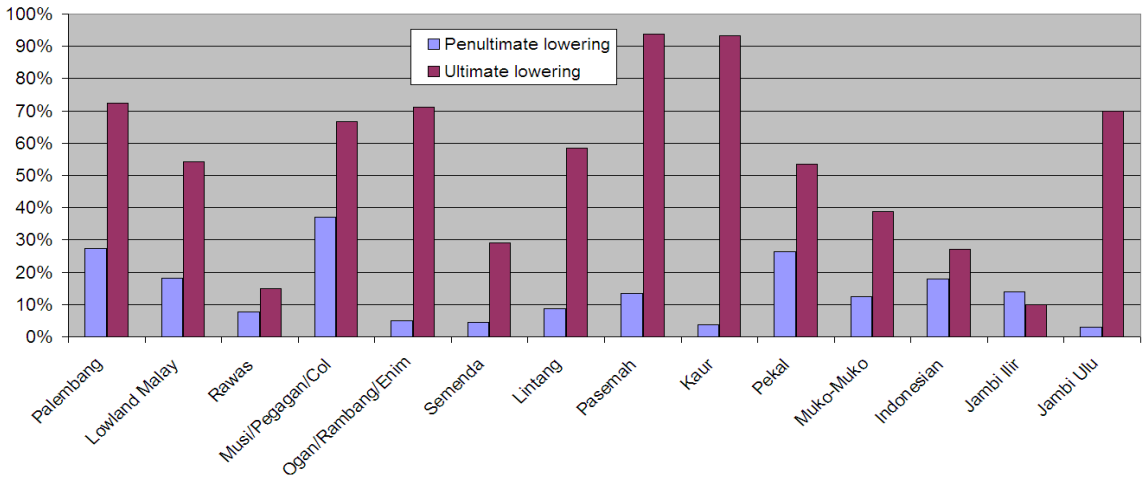
With the exception of one word, all the high-frequency words are identical to Indonesian. While this is not proof that they are Indonesian loans, at the least Indonesian has served to reinforce these

forms, and their presence in areas with few if any other lowered forms would strongly indicate that they indeed are loans.[47]

5.5.2 Lowering of ultimate high vowels

Ultimate high-vowel lowering in closed syllables is a nearly universal feature of SSML, e.g., *minom* 'drink' < **inum*, *tʊmpʊl* 'dull, blunt' < **tumpul*, and *busʊʔ* 'rotten' < **busuk*. In SSML proper, the lowest calculated percentage of final high vowel lowering is Rawas with 15%, coming from lexical items which are lowered with a high frequency across the corpus. The lowest after that is Semenda, which doubles Rawas' percentage to 29%.

Figure 5.10 Penultimate and ultimate high vowel lowering by dialect/subcluster



As can be seen in Figure 5.10, if anything there is an inverse relationship between penultimate and ultimate vowel lowering. The two most active ultimate-vowel-lowerers are Kaur and Besemah, who are most conservative in the penultimate syllable.

Aside from Kaur and Besemah, where ultimate closed high vowel lowering is nearly universal and thus clearly non-phonemic, we need to ask what triggers or conditions vowel lowering in some lexemes and not another.

- The obvious fact is that the ultimate syllable must be closed by a consonant of some sort. Ultimate high vowels are almost never lowered in open syllables.
- Nasalized ultimate vowels are lowered significantly more frequently (78%) than non-nasalized vowels (63%).[48]
- Although lowering occurs preceding all final consonants, the strongest correlation with vowel lowering occurs with final glottal stop.

One factor that is *not* a conditioning factor is the presence or absence of a high vowel in the penult: in a comparison of CVCVC words from all wordlist sites, there was a negligible difference in ultimate lowering when comparing penultimate-high-vowel words (69%) with penultimate-low-vowel words (65%).

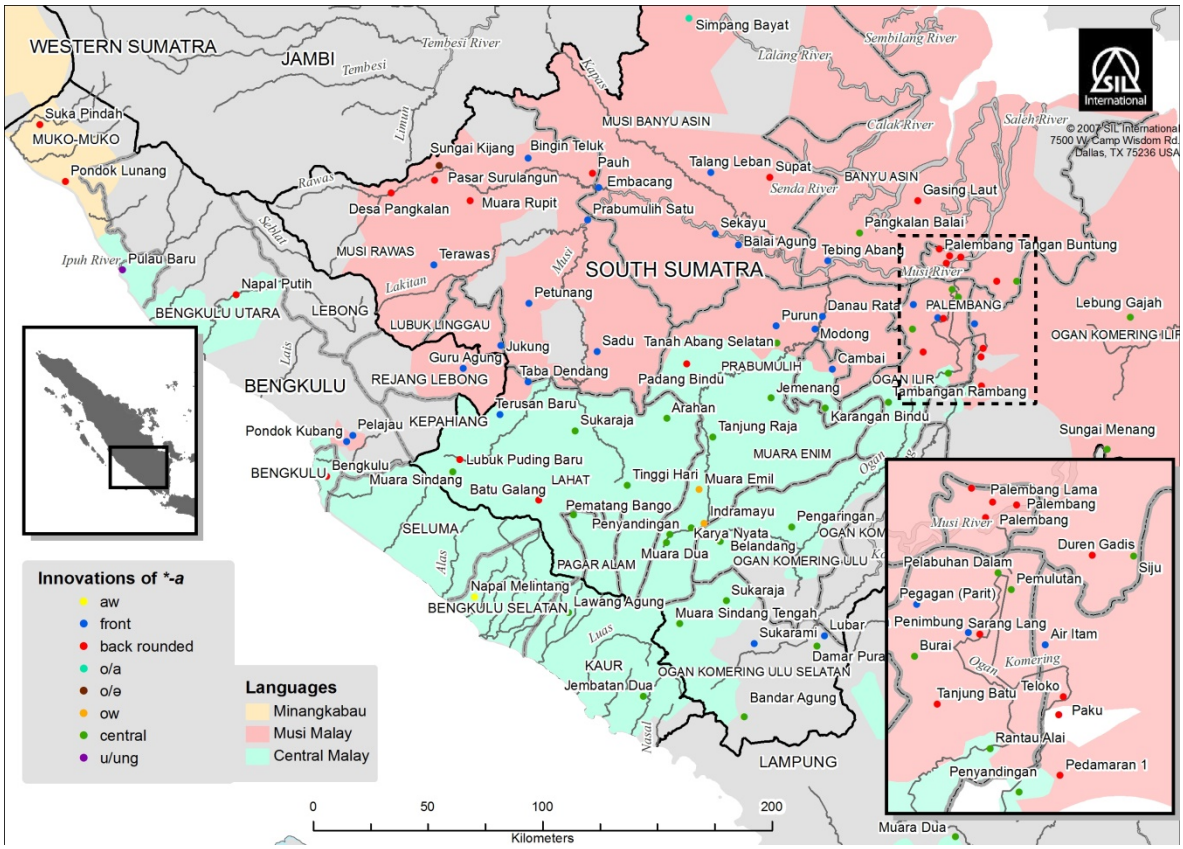
[47] The main challenge to this interpretation is the fairly consistent pattern that these same lects have *utaʔ*, *ubat*, etc. – why did they borrow the hV-hV words but not the hV-lV words?

[48] Given the assumption that all these dialects have onset-driven nasal harmony (cf. Blust 1997), we can predict which ultimate vowels will be nasalized even though nasality is not marked on the wordlists.

*5.6 Reflex of word final *-a*

To the speakers of the various Malay varieties spoken in southern Sumatra, the most salient distinguishing feature used to identify and differentiate their variety from other varieties is the variety's reflex for final *–*a* (e.g., *buŋo*, *buŋe* 'flower' < PM **buŋa*) and the local word for *tidak ada* 'there is not'.[49] Regarding the former innovation, individual sites within a dialect often agreed on the reflex, but the larger level of the cluster and language usually did not present a unified testimony. While the reflex of word final **a* was salient in the speakers' minds, it was less useful in making groupings as just one district on the coast such as Lebung Gajah could have the reflexes *ɨ*, *ə*, *e*, and *o*. However, these lects would have almost identical morphology, vocabulary, and were mutually comprehensible but were given a different name as they had the different reflex of *–*a*. See Figure 5.11.

Figure 5.11 innovations of *-a



5.6.1 South Barisan Malay

The reflex for BARISAN tends to the close to mid central unrounded vowel *ɨ*/*ɘ*/*ə* with the close-mid back rounded vowel *o* as the secondary preference.

5.6.1.1 Oganic

For the OGANIC cluster, the predominate reflex is the mid close central unrounded vowel *ɘ*. The four Rambang sites and the four Ogan sites all have this reflex. The three Enim sites show the variety predicated in *Geografi Linguistik Bahasa Enim* in the wordlists presented there for twenty Enim speaking towns on the Enim River. These are divided into a downstream dialect having *ə*, a midstream dialect generally showing ow, and an upstream section of villages which do not have a clear pattern

[49] See Tadmor (2003) for a discussion of the geographical spread and possible origin of this innovation.

in how they were marked. The Enim sites visited by WIST: Tanjung Raja (downstream from point 1), Muara Emil (point 8), and Indramayu (point 19) are compared with the points below.

Table 5.12 Reflexes of *-a *in Enim (Naning et al. 1998)*

	1-5 Downstream	6	7	8-15	16	17-20 Upstream
*-a	ə	ow	ɵw	ow	o	o/ə/e
	ɔ	ow				ɵw

From looking at Table 5.12 above one can see that the information by WIST showed the same change as one went further upstream as predicted in *GLBE*. One reason there may have been so much variation in reporting the final vowel for the upstream dialect is due to the rare nature of the close-mid central rounded vowel glide *ɵw*, causing it to be transcribed variously as *o* or *ə*. The transcriptions of *e* are predicted to be clerical errors as the font used to differentiate between the *e* and the *ɘ* is not automated but hand marked.

5.6.1.2 Highland

HIGHLAND cluster is similar to the OGANIC cluster in that the word final reflex of **a* is realized as the close to close mid central unrounded vowel (*ɨ*----*ɘ*), with some lects exhibiting close-mid central and back round vowels (*ɵ*, *o*). The two Lintang locations (Batu Galang and Lubuk Puding Baru) and the Bengkulu location that have *o* as the reflex are interestingly the same locations that diverge from the rest of the HIGHLAND cluster and have the *–kan* suffix rather than *–ka(h)* (§5.12.2). The other Lintang site Terusan Baru has the reflex *e*, apparently an influence from the Col and Musi Proper speech areas further downstream on the Musi River. The Benakat site Padang Bindu also has the *o* reflex, though it is unclear if this is due to influence from Palembang, from the Rawas sites, or from the *o* reflex area in the Bukit Barisan mountains. The thirteen Besemah, Lematang Ulu, and Semenda sites show consistency with the central reflexes *ɨ*, *ɘ*, and *ə*. The two HIGHLAND lects which differ from the rest are Serawai and Pekal. The northern variety of Serawai retains *a* while the southern variety — represented by WIST site Napal Melintang — has the diphthong aw. Pekal, ever the innovator, has the highly unusual *u* and the even rarer ʊŋ in the two locations surveyed. This reflex of final **a* was not found in any other surveyed Malay varieties. The Holle wordlist for Kaur from Bintuhan had the close central unrounded vowel *ɨ* while the wordlist collected by WIST from nearby Jembatan Dua has the similar *ɘ*. See map for more detailed information of distribution of the reflex of final **a*.

5.6.2 Musi

MUSI continues the pattern of being different from BARISAN in many phonological aspects. For word final reflexes of *a* MUSI typically reflects *e* or *o*.

5.6.2.1 Upper Musi cluster

The UPPER MUSI cluster evinces *e* along the length of the Musi River, and *o* along the length of the Rawas River. The three Rawas sites and one Musi Proper site on the Rawas River (Pauh) have *o*, while the other eight Musi Proper sites have *e*. The two Pegagan sites also have *e*. The six Col sites have either the close-mid front unrounded *e* or the open-mid front unrounded *ɛ*.

5.6.2.2 Palembang-Lowland cluster

The PALEMBANG-LOWLAND cluster shows more variety than the UPPER MUSI cluster, not surprisingly. The 'old' Palembang Lama dialects and the Pesisir dialects tend to have *ɨ*/*ɘ*; this can be found in six sites, with an additional two sites from LOWLAND subcluster (Tanah Abang Selatan and Burai) having *ɨ* as well. The 'newer' Palembang Pasar lects have *o* (eight sites), which may have come from *ɔ*, which is found in two 'older' Palembang Lama speaking areas, Palembang Lama and Sarang

Lang. The PALEMBANG varieties that are in contact with Musi Proper dialect areas seem to have adopted the *e* reflex, with the LOWLAND subcluster site Danau Rata bordering on the Pegagan area and the Palembang Pasar site Air Itam bordering on the Pegagan area both having *e*. The Belide villages are interesting, with three of the four showing an alternation between *e*/*a* in the word final environment, with *e* being manifested after the voiced stops *g*, *d*, and *ŋ*. Hence in Belide we see *apa* 'what' (< **apa*) but *buŋe* 'flower (< **buŋa*).

5.6.3 Other Languages

From the Kubu wordlists found in Struktur Bahasa Kubu (Dunggio et al. 1985), it appears that *o* is the tendency. However, if we compare these against Dialektologi Bahasa Kubu (Maryono, Setyonegoro & Kusmana 1997), we can find *e* and *a*. This seems to show that Kubu tends to take the reflex of the nearest variety, as the SBK took data points near the Rawas River, which has the *o* reflex (see section 6.2.3.1 for more discussion about where the various books' data points are). For the site WIST visited, Kubu showed an alternation between *a*/*o*. HAJI, ever the conservative, retains *a*.

5.6.4 Summary

The reflexes of word final **a* sound in southern Malay varieties reflect in a small way the variety of reflexes found for this in western Malay. While it is very salient in speakers' minds, it is more useful for distinguishing subdialects of one dialect than for the categories of cluster and language. These can be seen in Figure 5.11. However, the clusters and language do show tendencies, which are reflected in Tables 5.13 and 5.14.

Table 5.13 Reflexes of word final *a *(MUSI)*

		Village	Lect	Code	reflexes of-a	conditioning environment
PALEMBANG-LOWLAND	LOWLAND	Pedamaran	Penesak	PB-PDR	o	
		Tanjung Batu	Penesak	PB-TB	o	
		Burai	Penesak	PB-BR	ɨ	
		Tanah Abang S.	Lem. Ilir	PB-TA	ɨ	
		Danau Rata	Lem. Ilir	PB-SR	e	
		Modong	Belide	PB-MD	a/e	e after g, d, ŋ
		Cambai	Belide	PB-CB	a/e	e after g, d, ŋ
		Talang Leban	Belide	PB-TL	a/e	e after V, g, d, ŋ
		Tebing Abang	Belide	BEL-TA	e	
	PALEMBANG	Teloko	Pal Lama	PL-TEL	o	
		Paku	Pal Lama	PL-PAK	o	
		Palembang Lama	Pal Lama	PL-PL	ɔ	
		Sarang Lang	Pal Lama	PB-SL	ɔ	
		Pemulutan	Pal Lama	PB-PM2	ɨ	
		Pelabuhan Dlm	Pal Lama	PB-PM	ɨ	
		Palembang City 1	Pal Pasar	PB-SH1	o	
		Palembang City 2	Pal Pasar	PB-SH2	o	
		Tangan Buntung	Pal Pasar	PB-SH3	o	
		Gasing Laut	Pal Pasar	PB-GL	o	
		Duren Gadis	Pal Pasar	PB-DG	o	
		Air Itam	Pal Pasar	PB-AH	e	
		Siju	Pal Pasar	PB-SJU	ɨ	
		Lebung Gajah	Pesisir	CST-LG	ɨ	
		Pangkalan Balai	Pesisir	CST-DB	ɘ	
		Supat	Pesisir	CST-SP	o	
		Simpang Bayat	Pesisir	CST-SB	o/a	
		Sungai Menang	Pesisir	CST-SM	ɨ	
UPPER MUSI	Rawas	Muara Rupit	Raw. Rupit	RAW-RU	o	
		Psr Surulangun	Rawas	RAW-PS	o	
		Pangkalan	Rawas Ulu	RAW-PN	o	
	Musi Proper	Petunang	Kelingi	MU-KL	e	
		Sadu	Musi Proper	MU-SA	e	
		Prabumulih Satu	Musi Proper	MU-P1	e	
		Purun	Musi Penukal	MU-PEN	e	
		Embacang	Musi Proper	MU-EM	e	
		Pauh	Musi Proper	MU-PH	o	
		Bingin Teluk	Musi Proper	MU-BT	e	
		Sekayu	Musi Proper	MU-KY	e	
			Musi Sekayu			
	Pegagan	Balai Agung	Musi Sekayu	MU-BA	e	
		MrPenimbung	Pegagan	MU-PG	e	
		Pegagan	Pegagan	MU-PG2	e	
	Col	Jukung	Col	COL-LL	e	
		Taba Dendang	Col	COL-TT	ɛ	
		Guru Agung	Col	COL-PUT	ɛ/eʸ	
		Terawas	Col Lembak	COL-BKL	ɛ	
		Pelajau	Col Lembak	COL-PLJ	e	
		Pondok Kubang	Col Lembak	COL-L8	e	

Table 5.13 Reflexes of word final *a *continued*

Language	Cluster	Dialect	Village	Lect	Code	reflexes of -a
BARISAN	OGANIC	Rambang	Penyandingan	Rambang	RAM-PNY	ɘ
			Tambangan Rambang	Rambang	RAM-TR	ɘ
			Karangan Bindu	Rambang	RAM-RK	ɘ
			Jemenang	Rambang	RAM-RD	ɘ
		Enim	Tanjung Raja	Enim Ilir	EN-ME	ɘ
			Muara Emil	Enim Tengah	EN-TAS	ow
			Indramayu	Enim Ulu	EN-TAN	ɵw
		Ogan	Rantau Alai	Ogan Ilir	OG-RA	ɘ
			Pengaringan	Ogan Tengah	OG-BR	ɘ
			Damar Pura	Ogan Ulu	OG-DP	ɘ
			Belandang	Ogan Ulu	OG-UO	ɘ
	HIGHLAND	Bengkulu	Bengkulu City	Bengkulu Ind.	BNGKL	o
		Lintang	Batu Galang	Lintang	BES-MP	o
			Lubuk Puding Baru	Lintang	BES-UM	o
			Terusan Baru	Lintang	BES-TT	e
		Besemah	Sukaraja Kisam	Besemah	BES-MDK	ɘ
			Sukaraja	Kikim	BES-KT	ɘ
			Pematang Bango	Besemah	BES-PA	ɘ
			Muara Sindang	Besemah	BES-AK	ɘ
			Lawang Agung	Besemah	BES-BK	ɨ
		Lematang Ulu	Tinggi Hari	Lematang Ulu	LT-PP	ɨ
			Arahan	Lematang Ulu	LT-MR	ɨ
		Semenda	Muara Sindang Tengah	Semenda	SEM-PB	ɘ
			Karya Nyata	Semenda	SEM-SDL	ɘ
			Penyandingan	Semenda	SEM-PNY	ɘ
			Muara Dua	Semenda	SEM-MD	ɘ
			Bandar Agung	Semenda	SEM-BA	ə
			Tanggamus	Semenda	SEM-LP	ɨ
		Benakat	Padang Bindu	Benakat	PB-BN	o
		Serawai	Napal Melintang	Serawai (S.)	SRW-NM	aw
		Pekal	Napal Putih	Pekal	PKL-NP	u/ʊŋ
			Pulau Baru	Pekal	PKL-PL	u
		Kaur	Jembatan Dua	Kaur	KAU-J2	ɘ
			Bintuhan	Kaur	KAUR	ɨ
MINANGKABAU		Muko-Muko	Pondok Lunang	Muko-Muko	MUK-PL	o
			Suka Pindah	Muko-Muko	MUK-SP	o
		HAJI	Sukarami	HAJI	HAJ	a
		Kubu	Sungai Kijang	Kubu	KUBU	o/a

Table 5.14 Summary of reflexes of final *a

Language	Decision	Cluster	Decision	Variety	Decision
MUSI	o/ɨ/e	PALEMBANG	o/ɨ	Penesak	o/ ɨ
				Lematang Ilir	e/ ɨ
				Belide	e
				Palembang	o/ɨ
		UPPER MUSI	e	Musi Proper	e
				Rawas	o
				Pegagan	e
				Col	e/ɛ
BARISAN	ə/ɨ	OGANIC	ə	Rambang	ə
				Enim	ə/ow
				Ogan	ə
		HIGHLAND	ə/ɨ/o	Besemah	ə/ɨ
				Semenda	ə
				Serawai	a/aw
				Lintang	o
				Lematang Ulu	ɨ
				Bengkulu	o
				Pekal	u/ʊŋ
				Kaur	ɨ/ə
Kubu	o/ɨ				
HAJI	a				

5.7 Word-final diphthongs and monophthongs

Adelaar reconstructs two final PM diphthongs, both of which occur only at the end of words: **-ay* and **-aw*. Five of the six lects in his sample retained the diphthong, while Jakarta Malay reflected **-ay* as *–e*, and **-aw* as *o*. Some varieties of SSML frequently simplify word final diphthongs to monophthongs. Probably not coincidentally, the only SSML lects which do *not* preserve the diphthong are those in the PALEMBANG subcluster. Within the PALEMBANG subcluster, the varieties further from Palembang City still preserve the diphthong while those closer to the city evince the cosmopolitan innovation of monophthongization of final diphthongs in all or most lexical items. In PALEMBANG versus in Jakarta Malay, however, reflexes of final diphthongs are as likely to be *–i* and *–u* as *–e* and *–o*, cf. Teloko *iju* 'green' < PM **hijaw* and *suŋi* 'river' < **sungay*.

This innovation distinguishes the PALEMBANG varieties from the nearby LOWLAND subcluster and Pegagan varieties. (One Palembang Pasar variety does preserve the diphthong (Tangan Buntung).[50]) Palembang City 2 list represents a lifetime resident from Palembang, and has a mixture of diphthong preservation and simplification. Palembang City 1 list has the Palembang Pasar variety of someone who moved to Palembang as a teenager and shows total monophthongization. This agrees with some of the villages around Palembang which have the same innovation. All varieties outside of the Palembang and Pesisir varieties retain the diphthong. Other non-MUSI lects spoken in southern Sumatra also preserve the diphthong in word final position. See Table 5.15.

[50] This was elicited from someone who had moved from Sekayu to Palembang as a small child to capture how the LWC was spoken by immigrants to Palembang. His variety preserves the diphthongs.

Table 5.15 diphthongs > monophthongs

			Village	Dialect	Code	Diphth(D) Mono(M)	%Mono
MUSI	PALMBG'-LOWLAND	LOWLAND	Tanjung Batu	Penesak	PB-TB	D	--
			Pedamaran	Penesak	PB-PDR	D	25
			Burai	Penesak Burai	PB-BR	D	--
			Tanah Abang S.	Lematang Ilir	PB-TA	D	--
			Danau Rata	Lematang Ilir	PB-SR	D	--
			Modong	Belide	PB-MD	D	--
			Cambai	Belide	PB-CB	D	--
			Talang Leban	Belide	PB-TL	D	--
			Tebing Abang	Belide	BEL-TA	D	--
			Teloko	Palembang Lama	PL-TEL	MIXED	67
			Paku	Palembang Lama	PL-PAK	M	10.[51]
			Pal. Lama	Palembang Lama	PL-PL	MIXED	67
			Sarang Lang	Palembang Lama	PB-SL	MIXED	67
			Pemulutan	Palembang Lama	PB-PM2	MIXED	67
		PALEMB'G	Pel. Dalam	Palembang Lama	PB-PM	MIXED	80
			Palembang City 1	Palembang Pasar	PB-SH1	M	100
			Palembang City 2	Palembang Pasar	PB-SH2	MIXED	62
			Tangan Buntung	Palembang Pasar	PB-SH3	D	--
			Gasing Laut	Palembang Pasar	PB-GL	M	100
			Duren Gadis	Palembang Pasar	PB-DG	MIXED	50
			Air Itam	Palembang Pasar	PB-AH	D	--
			Siju	Palembang Pasar	PB-SJU	MIXED	54
			Lebung Gajah	Pesisir	CST-LG	MIXED	43
			Pangkalan Balai	Pesisir	CST-DB	MIXED	50
			Supat	Pesisir	CST-SP	D	--
			Simpang Bayat	Pesisir	CST-SB	D	--
			Sungai Menang	Pesisir	CST-SM	MIXED	33
			Both sites	Pegagan		D	
		UPPER MUSI	All nine sites	Musi Proper		D	
			All three sites	Rawas		D	
			All five sites	Col		D	
BARISAN CLUSTER		HIGHLAND	All 30 sites	Lematang Ulu, Kaur, Pekal, Bengkulu, Besemah, Semenda, Lintang, Serawai		D	
		OGANIC	All twelve sites	Ogan, Enim, Rambang		D	
MINANGKABAU			All five sites	Muko-Muko		D	
HAJI			Sukarami	HAJI	HAJ	D	
Kubu			Sungai Kijang	Kubu	KUBU	D	

5.8 Epenthesis of low vowel after high vowel in closed ultimate syllable

A limited number of southern Sumatra Malay varieties have the phenomenon of epenthesis of low(er) vowel after high vowel in a closed ultimate syllable, or what might be called 'diphthongization' of the vowel (e.g., *taxiᵊʔ* 'pull' < PM **tarik*). In defining closed syllable, the data was analyzed looking at the canonical word shape to see if there was a word final consonant, though in some cases the

[51] Only one example in this short word list for Paku.

realized word does not have the final consonant (usually **h* which frequently elides finally) or has become an approximant (typically **r* > *w*).

The research sites can be divided roughly into three categories using this as a measurement: those in which epenthesis of low vowel never occurs, those in which it rarely occurs, and varieties where it often occurs. No varieties were found where epenthesis of low vowel always occurs in word final closed with a high vowel.

The phenomenon appears to occur in a narrow belt going from Bengkulu to the areas to the west of Palembang. It is particularly centered along the Enim/Lematang River basins and in its tributaries in the Bukit Barisan mountain range, as well as in the neighboring Serawai language area. It occurs in a limited set of phonological environments, most typically when there is an /i/ or /u/ before a glottal stop or fricative (*ʔ*/*h*). Other final consonants that appear to trigger epenthesis are the nasal velar *ŋ*, and the coronals s/t. The vowel that is inserted is usually after the high vowel, and is less open, ranging from *ɪ* to *a*.

The data is presented below in charts (Table 5.16) showing where epenthesis of low vowel was noted, the number of times it occurred in the wordlist, and the epenthetic vowel and consonant that followed it. For simplification, the high vowels were simplified from their phonetic form to their phonological form (*i*, *ɪ*, *e* > *i* ; *u*, *ʊ*, *o* > *u*).

Table 5.16 occurrence of low vowel epenthesis

PALEMBANG-LOWLAND cluster, MUSI language		
Modong	**Cambai**	**Air Itam**
rare	rare	rare
1 i aʔ	1 i aʔ	1 u ɛt
i + glottal	i + glottal	u + coronal

OGANIC cluster, BARISAN language			
Muara Emil (EN-TAS)	**Indramayu (EN-TAN)**	**Damar Pura (OG-DP)**	**Belandang (OG-UO)**
frequent	frequent	rare	rare
7 i əʔ	6 i aʔ	2 i əŋ	2 i əŋ
5 i əh	13 i aŋ	1 i əʔ	2 u əʔ
12 i əŋ	5 i ah		1 u əŋ
1 u oʔ	1 i ɪs		
1 u əʔ			
1 u əh			
2 u əŋ			
i (u) + velar nasal or glottal	i + velar nasal or glottal	i + velar nasal or glottal	i/u + velar nasal or glottal

HIGHLAND cluster, BARISAN language						
Pulau Baru (PKL-PL)	**Penyandingan (SEM-PNY)**	**Sukaraja Kisam (BES-MDK)**	**Sukaraja Kikim (BES-KT)**	**Napal Melintang (SRW-NM)**	**Arahan (LT-MR)**	**Napal Putih (PKL-NP)**
rare	rare	frequent	rare	frequent	rare	rare
1 ə iŋ[52]	1 i ət	2 i e(w)	2 i aʔ	4 i aʔ	1 i ew	1 i aʔ
		1 ɪ eʔ	2 i ah	8 i a(h)		
		2 ɪ eh		1 i ə(h)		

[52] Different from other examples in that this word has the epenthetic low vowel inserted before the primary high vowel, rather than after it, in the word **dagiŋ* 'meat' realized as *dagᵊiŋ*.

		2 i eŋ		10 u aʔ		
		2 i et		5 u a(w)(h)		
				1 u iɣ[53]		
i + velar nasal	i + coronal	i + velar nasal, glottal, or coronal	i + glottal	i/u + glottal or velar	i + r	i + glottal

The appearance of the epenthetic vowel also occurs occasionally before the **r*, seen in the one example from Arahan (*iew*), two examples from Napal Melintang (*uiɣ*), and two examples from Sukaraja Kisam (*iew*). This is similar to the cases treated by Anderbeck (2008:46) in upstream Jambi Malay, where with the high vowels *u* and *i*, a low central vowel is epenthesized and pronounced concurrently with the local reflex of **r*.

Low vowel epenthesis only occurs frequently in the related dialects of BARISAN that are found between the upstream headwaters of the Ogan and Enim River systems and the highlands of the Bukit Barisan Mountains. For grouping purposes it shows a shared innovation between nearby varieties of the BARISAN language.

*5.9 PM *a > e*

Another innovation seen in some of parts of SSML is the shift of **a* > *e* in closed ultimate and in penultimate syllables, e.g., *jelan* 'path, road' < PM **jalan*, *ubɛt* 'medicine' < **ubat*. Unlike many of the other innovations examined previously, this one does not appear to be concentrated in one language variety nor in a geographic area. The innovation is seen in the MUSI varieties around the downstream Lematang River watershed (Purun, Modong, Cambai, Pangkalan Balai), the MUSI varieties of Col spoken in South Sumatra (Jukung, Taba Dendang, and Guru Agung) and to a lesser extent in the MUSI variety of Pegagan (Muara Penimbung). The innovation occurs in BARISAN in the upstream Ogan (Belandang) and Enim (Indramayu) areas, but is not found in neighboring varieties such as Semenda or Besemah. It is also found in the Kubu language in Sungai Kijang as well as the Jambi Kubu varieties to the north, Pematang Kolim, Pematang Kabau, and Dusun Tuo (Maryono, Setyonegoro & Kusmana 1997).

The innovation of **a* > *e* is also found in the Muko-Muko variety of MINANGKABAU and the nearby Pekal variety of BARISAN. However, it is not found in both the penultimate and ultimate syllables, but rather in the ultimate only, and apparently it occurs by a different process.

For nearby varieties of Malay, the innovation of **a* > *e* that occurs in ultimate syllables is discussed by Anderbeck for the Jambi Ulu dialect of Lubuk Kepayang (2008:36).

The following tendencies are noted for the **a* > *e* innovation for Enim:

- in ultimate syllables, never occurs after a voiceless consonant
- in penultimate syllables:
  - only happens in closed syllables
  - occurs after voiced consonants, but not in all cases (no conditioning environment found)
  - rarely occurs after voiceless consonants.

Taking the data from the sites visited, the **a* > *e* innovation was checked in other areas against these early tendencies to confirm or disconfirm the pattern found in Enim. The following was found:

- in ultimate syllables, CAN occur after a voiceless consonant
- in penultimate syllables:
  - usually only happens in closed syllables
  - usually occurs after voiced stops and affricates
  - occurs less frequently following liquids, nasals, glides and voiceless stops.

[53] We see only one example of epenthesis of high vowel in **təluɣ* 'egg': Serawai's *tɔlu^{i}ɣ*.

For a more detailed look at how this can be seen in the different varieties, see Table 5.17 to see the frequency of the **a > e* innovation in select representative sites.

Table 5.17 frequency of *a > e *in selected varieties*

Environment		After voiced stop/ affricate	% a > e	After nasal	% a > e	After l, *r	% a > e
Muara Penimbung (Pegagan-Musi)	PU	37	51%	14	21%	14	21
	U	27	67%	19	16%	19	42%
Cambai (Belide-MUSI)	PU	38	100%	16	0%	12	25%
	U	17[54]	94%	16	25%	30	47%
Jukung (Col-MUSI)	PU	31	87%	19	32%	18	33%
	U	--	--	--	--	--	--
Belandang (Ogan-BARISAN)	PU	34	82%	18	23%	16	25%
	U	16	56%	10	10%	19	79%
Pulau Baru (Pekal-BARISAN)	PU	--	--	--	--	--	--
	U	28	18%	16	6%	25	16%

Looking through the environments where the innovation occurs for different varieties of southern Sumatra Malay, it becomes apparent that while the innovation does not occur in a uniform fashion in each variety, the general tendencies outlined above hold.

For the MUSI varieties of Col and Belide, the tendencies mentioned for Enim were tested. For penultimate syllables, all three tendencies frequently held. The syllables were not always closed, cf. beu < *bahu. The voiced stops and affricates had a very high rate preceding the innovation, while it rarely happens after voiceless stops and affricates (there are a few exceptions such as Belandang reflecting **kasaw* 'rafter' as *kɔsaw*). The other consonants that seem to trigger the innovation are the reflexes of **r*, the liquids, glides, and nasals. There may be a relation between the sonority of the consonant and the triggering of the innovation, as these more sonorous consonants are followed by the innovation a much lower percentage of the time than the voiced stops and affricates. The **a > e* innovation also seems to occur in a larger variety of environments in the ultimate syllable than the penultimate, as it also occurs after glides and vowels.

For the Belide varieties of MUSI, (Cambai, Modong, and Talang Leban) the innovation in the penultimate syllable always occurs after a voiced stop or affricate, never occurs after a nasal, and rarely occurs after a liquid. It never occurs after a voiceless stop or affricate. In the ultimate syllable, there appears to be a greater chance of the innovation occurring, with an increase in the innovation following nasals and liquids. In the ultimate syllable of Cambai, it is also important to note that with one exception, all the recorded instances were actually of an epenthetic high vowel inserted before *a*, such as in *bər^{i}as* 'rice' < **bəras*. This vowel cluster is then postulated to simplify ultimately to *e*, as seen in other varieties. Most sites closely follow the tendencies noted in Cambai, with few occurrences of the innovation in the penultimate after a voiceless stop. In Cambai and other MUSI varieties, the exceptions are the lexical items *tanam* 'plant' and *tampi* 'winnow' which are realized as: *tænæm* or *nənɔm* and *tæmpi* or *nəmp^{e}i*. The latter lexical item is particularly interesting, as it shows up in many varieties of the MUSI language that have no other witness to the innovation. See Table 5.19.

The Pegagan variety of MUSI spoken at Muara Penimbung has an interesting correspondence in the innovation between the penultimate and ultimate, as the innovation is frequently seen in the ultimate in words where it is the penultimate. This may suggest some principle of vowel harmony.

[54] The one only affricate in this data set, *j* occuring after the nasal in *təlaɲjaŋ* 'naked' is questionable—if this is thrown out the innovation occurs 100% of the time.

Similar to Cambai, the vowel raising appears to come from a system of insertion of a high vowel before the unrounded front open vowel *a*, with a resulting *ᵉæ*.

The sites where a different mechanism appears to be triggering the innovation are those of the MINANGKABAU influenced varieties under investigation: Bengkulu, Pekal, and Muko-Muko. The Pekal sites of Pulau Baru and Napal Putih, Bengkulu (BARISAN) and the Muko-Muko sites of Suka Pindah and Pondok Lunang all have the **a* > *e* innovation in the ultimate syllable (not in the penultimate syllable), but this is apparently triggered by a process of glottalization, where the final voiceless alveolar stop or fricative in the closed ultimate syllable moves to a glottal point of articulation. The preceding *a* is consequently raised to *e*. Possible motivation for this raising would be the increased closure of the oral cavity to accommodate this transformation. However, there are a few words in which **a* > *e* where the final consonant is not present e.g., *səbəlas* 'eleven' > *səbəle* (item 207, Bengkulu) or without going to the glottal point of articulation e.g., **urat* 'vein' > *uhet* (item 227, Pulau Baru).

The innovation in Pekal is particularly interesting as there appears to be a relation between the alternation in the vowel selection and the reconstructed PM word ending. Pekal has words that end in *oʔ* (e.g., item 154, *pandoʔ* < PM **pandak*) and in *eʔ* (e.g., item 311 *buleʔ* < PM **bulət*) while both share the same vowel in the PM word. At other times, the vowel does not undergo any change (eg item 117 *buwaʔ* < PM **buah*). The apparent motivation for whether the vowel changes and the realized resulting vowel corresponds to the final consonant. The correspondence between the final consonant the resulting vowel permutation of **a* is shown in the chart below (Table 5.18 for Pekal, Table 5.19 overall).

Table 5.18 final consonant vowel endings in Pekal

PM *VC	Pekal VC	PM word	Pekal word	WL #
aʔ	aʔ	*lawaʔ 'spider'	lalawaʔ	100
ah	aʔ	*tanah 'land, soil'	tanaʔ	119
at	eʔ	*bərat 'heavy'	behеʔ	142
ak	oʔ	*anak 'child'	anoʔ	56
ək	oʔ	*u(n)tək 'brain'	otoʔ	211

Table 5.19 occurrence of *a > e *in all varieties*

Village	Variety	Language	PU a > e	a > realized	U a > e	a > realized
Modong	Belide		X	e	X	e̯a
Cambai	Belide		X	e	X	e̯a
Talang Leban	Belide		X	e	--	--
Pangkalan Balai	Pesisir		X	e	X	e
Purun	Musi Penukal		X	ə	X	ə
Muara Penimbung	Pegagan		X	æ	X	e̯æ
Jukung	Col		X	e	--	--
Taba Dendang	Col		X	e	--	--
Guru Agung	Col	MUSI	X	e	--	--
Indramayu	Enim Ulu		X	e	X	e̯a
Damar Pura	Ogan Ulu		X	ɘ	X	--
Belandang	Ogan Ulu		X	ɘ	X	ɘ
Bengkulu	Bengkulu		--	--	X	o/e
Napal Putih	Pekal		--	--	X	o/e
Pulau Baru	Pekal	BARISAN	--	--	X	o/e
Sungai Kijang	S. Sum. Kubu		X	e	X	ɔ
Pematang Kolim	Jambi Kubu		X	e	X	o
Pematang Kabau	Jambi Kubu	Kubu	X	e	X	o

Dusun Tuo	Jambi Kubu		X	e	X	o
Pondok Lunang	Muko-Muko	Minang-	X	e	X	e
Suka Pindah	Muko-Muko	kabau	X	e	X	e

See also Mckinnon's later (2012) conference paper on this topic.

5.10 Ultimate closed 'Java' schwa in Palembang Lama

An exclusive feature of the six Palembang Lama wordlists is the presence of schwa in ultimate closed syllables where all other data points reflect *a*. So, for example, we see *bənəʁ* 'true' and *maləm* 'night' where other sites have something more like *bənar* and *malam*.

Perhaps the most noteworthy fact about this aspect of Palembang Lama is that the lexemes with schwa track nearly word-for-word with those of Jakarta Malay (Betawi).[55]

This is not an unconditioned innovation, as we also see lexemes in Palembang Lama like *jahat* 'bad' and *libar* 'wide'. Nor is it phonation-driven like the innovation discussed in §5.9 above; compare *atəp* 'roof' and *isəp* 'suck'. Two opposite interpretations can be attached to this feature: that it is a retention of Proto-Malayic (and earlier from Proto-Malayo-Polynesian), or that it is a contact effect, with the source language(s) likely including Javanese. Javanese did not merge Proto-Malayo-Polynesian **ə* and **a* as did nearly all varieties of Malay.

Adelaar (1992:32–39) argues that the presence of ultimate closed schwa in many lexemes is a retention from Proto-Malayo-Polynesian, and accordingly reconstructs schwa in many Malayic proto-forms. This is perhaps the most controversial aspect of Adelaar's generally very solid reconstruction, as its justification comes nearly entirely from Jakarta Malay, a notorious contact language.[56] This is the point briefly argued by Grijns (1991; 1996); see also Adelaar's (1994) rejoinder, as well as Nothofer's support for the continuation position (1995a; 1995b).

The fact that this feature occurs exclusively in the SSML lect most heavily influenced by Javanese seems quite suggestive to us, but we will not attempt to argue here that this feature derives from Javanese (whether directly or via Jakarta Malay). Suffice it to say that this connection is noted between Palembang Lama, Jakarta Malay and Javanese. Regarding its origin, as goes Jakarta Malay, so goes Palembang Lama.[57]

One interesting wrench to throw in the works is the presence of the suffix *–kən* in one of the Palembang Lama sites (instead of the expected *–kan*). This can hardly be unrelated to the occurrence of schwa in words like *bənəʁ* and *maləm*. The problem is that neither of the possible donor languages, Jakarta Malay or Javanese, have an identical suffix (Javanese is the closest with *–akən*). Could this be evidence for the continuation view? We note that Nothofer (1995a:89–90) reconstructs **-kən* 'benefactive, causative' for Bangka Malay (although because of rather complicated morphophonemics this analysis is not completely straightforward); if this analysis is correct, we would at minimum see a significant connection between Palembang Lama and Bangka.

We now move from vowels to a look at morphophonemics and affixes.

5.11 Initial voiced stop/affricate nasal assimilation

It is well known that, in SM and other Malayic lects, voiceless stem-initial stops are completely assimilated to the homorganic nasal when prefixed by the active prefix *meN-*. For example, *kuɲah* 'chew' becomes *məŋuɲah* and *tikam* 'stab' becomes *mənikam*.

---

[55] As documented in Blust (1988).

[56] Adelaar also cites evidence from Tioman Malay (Collins 1985a) but, as mentioned in Anderbeck (2013:285), the ultimate closed schwa in that lect is due to phonation-driven vowel shift.

[57] Combining this discussion with that of the lexical (Javanese) connection between Palembang Lama and Banjar in §4.1.7, we see an intriguing 'love triangle', where Palembang Lama shares a clear phonological feature with Jakarta Malay (but not Banjar) and a clear lexical connection with Banjar (and not with Jakarta Malay).

Some varieties of southern Sumatra Malay show nasal assimilation of not just the voiceless stops and affricates, but also of the voiced stops and affricates. This seems to be another strategy for consonant cluster reduction, like the pattern seen in the innovation of nasal deletion before a voiceless stop. In some cases, the stop/affricate does not disappear entirely, but is noticeably reduced in length and prominence. See Table 5.2 and Figure 5.8.

The data for this comparison was gathered from wordlists with both the root form of the vowel and the inflected form of the verb (resulting from the addition of the transitive (*mə*)*N*+ affix). As some of the wordlists came from sources outside of WIST, such as the Holle lists or PBh, these sometimes only presented the root form or the inflected form of an unknown root. These examples were not used in the comparison. The research sites were divided into three categories on the basis of the extent of nasal assimilation: YES (above 75% nasal assimilation), NO (below 25% nasal assimilation), and mixed (25% -75% assimilation). For the varieties falling into the mixed category, the percentages are noted. If there was reduction in prominence of the voiced stop/affricate, it was counted as assimilation for the purpose of the comparison.

Another consideration was the environments in which nasal assimilation did or did not occur, particularly in the varieties that were mixed. The three voiced stops *b*, *d*, *g* and the affricate *j*[58] are the initial consonants in the root word of the 22 verbs examined. For each instance of a voiced stop/affricate, the initial consonant noted and total occurrences were summed. This total was then divided by the number of occurrences where nasal assimilation did not ever occur for that particular consonant in a particular variety. A ratio above 1 indicates a low rate of nasal assimilation and a ratio above 1 indicates a high rate of nasal assimilation. The results are displayed in Table 5.20.

Table 5.20 Occurrence of nasal assimilation before voiced stops

phone	total occurrences	occurrences when other phones are assimilated, but phone X not assimilated in lects	Ratio
b	8	11	1.4
d	2	2	1
g	8	8	1
j	4	3	.8

From the results in the table above, given the limitations of a 358 item wordlist with 22 commonly occurring verbs that meet the parameters, it was seen that *b* assimilated less often with the nasal and *j* assimilated more often. See Tables 5.21 and 5.22. A lengthier corpus might show different results.

[58] Using Indonesian orthography, rather than the International Phonetic Alphabet (IPA) *ǰ*.

Table 5.21 Nasal assimilation before stem-initial voiced stops/affricates (MUSI)

		Village	language	Code	Nasal assimilation	% assimilation if mixed
PALEMBANG-LOWLAND	LOWLAND	PENESAK	Penesak	P. Bahasa	YES	
		Pedamaran	Penesak	PB-PDR	mixed	40%
		Tanjung Batu	Penesak	PB-TB	YES	
		Burai	Penesak	PB-BR	YES	
		Tanah Abang S.	Lem Ilir	PB-TA	mixed	40%
		Danau Rata	Lem Ilir	PB-SR	NO	
		Modong	Belide	PB-MD	NO	
		Cambai	Belide	PB-CB	mixed	65%
		Talang Leban	Belide	PB-TL	YES	
		Tebing Abang	Belide	BEL-TA	YES	
	PALEMBANG	Teloko	Pal Lama	PL-TEL	YES	
		Paku	Pal Lama	PL-PAK	YES	
		Palembang Lama	Pal Lama	PL-PL	mixed	65%
		Pemulutan	Pal Lama	PB-PM2	mixed	35%
		Pelabuhan Dalam	Pal Lama	PB-PM	mixed	55%
		Palembang City 1	Pal Pasar	PB-SH1	YES	
		Palembang City 2	Pal Pasar	PB-SH2	mixed	40%
		Tangan Buntung	Pal Pasar	PB-SH3	YES	
		Gasing Laut	Pal Pasar	PB-GL	YES	
		Duren Gadis	Pal Pasar	PB-DG	YES	
		Air Itam	Pal Pasar	PB-AH	YES	
		Siju	Pal Pasar	PB-SJU	YES	
		Lebung Gajah	Pesisir	CST-LG	YES	
		Pangkalan Balai	Pesisir	CST-DB	YES	
		Supat	Pesisir	CST-SP	YES	
		Simpang Bayat	Pesisir	CST-SB	YES	
		Sungai Menang	Pesisir	CST-SM	YES	
MUSI	Rawas	Muara Rupit	Raw. Rupit	RAW-RU	YES	
		Psr Surulangun	Rawas	RAW-PS	YES	
		Pangkalan	Rawas Ulu	RAW-PN	YES	
UPPER MUSI	Musi Proper	Petunang	Kelingi	MU-KL	YES	
		Sadu	Musi Proper	MU-SA	YES	
		Prabumulih Satu	Musi Proper	MU-P1	YES	
		Purun	M. Penukal	MU-PEN	YES	
		Embacang	Musi Proper	MU-EM	YES	
		Pauh	Musi Proper	MU-PH	YES	
		Bingin Teluk	Musi Proper	MU-BT	YES	
		Sekayu	Musi Sekayu	MU-KY	YES	
		Balai Agung	Musi Sekayu	MU-BA	YES	
	Pegagan	MrPenimbung	Pegagan	MU-PG	YES	
		Pegagan	Pegagan	MU-PG2	mixed	50%
	Col	Jukung	Col	COL-LL	YES	
		Taba Dendang	Col	COL-TT	YES	
		Guru Agung	Col	COL-PUT	YES	

			COL-	YES
	Terawas	Col	BKL	
	Pelajau	Col Lembak	COL-PLJ	YES
	Pondok Kubang	Col Lembak	COL-L8	YES

Table 5.22 Nasal assimilation before stem-initial voiced stops/affricates (non-MUSI)

Lang.	Cluster	Dialect	Village	Dialect	Code	Nasal assimilation	% assimilation if mixed
BARISAN	OGANIC	Rambang	Penyandingan	Rambang	RAM-PNY	NO	
			Tambangan Ram.	Rambang	RAM-TR	YES	
			Karangan Bindu	Rambang	RAM-RK	mixed	30%
			Jemenang	Rambang	RAM-RD	NO	
		Enim	Tanjung Raja	Enim Ilir	EN-ME	NO	
			Muara Emil	Enim Tengah	EN-TAS	NO	
			Indramayu	Enim Ulu	EN-TAN	NO	
		Ogan	Rantau Alai	Ogan Ilir	OG-RA	YES	
			Pengaringan	Ogan Tengah	OG-BR	NO	
			Damar Pura	Ogan Ulu	OG-DP	NO	
			Belandang	Ogan Ulu	OG-UO	NO	
	HIGHLAND	Pekal	PBDSBB 24	Pekal	PBDSBB	YES	
			Pulau Baru	Pekal	PKL-PL	YES	
			Napal Putih	Pekal	PKL-NP	YES	
		Bengkulu	Bengkulu City	Bngkl Ind.	BNGKL	NO	
		Lintang	Lintang PBh	Lintang	-	YES	
			Batu Galang	Lintang	BES-MP	YES	
			Lbk Puding Baru	Lintang	BES-UM	YES	
			Terusan Baru	Lintang	BES-TT	YES	
		Besemah	Sukaraja Kisam	Besemah	BES-MDK	NO	
			Sukaraja	Kikim	BES-KT	NO	
			Pematang Bango	Besemah	BES-PA	NO	
			Muara Sindang	Besemah	BES-AK	mixed	35%
			Lawang Agung	Besemah	BES-BK	NO	
		Lematang Ulu	Tinggi Hari	Lem Ulu	LT-PP	mixed	35%
			Arahan	Lem Ulu	LT-MR	mixed	40%
		Semenda	Muara Sindang Tg	Semenda	SEM-PB	NO	
			Muara Dua	Semenda	SEM-MD	NO	
			Bandar Agung	Semenda	SEM-BA	NO	
			Tanggamus	Semenda	SEM-LP	mixed	35%
		Benakat	Padang Bindu	Benakat	PB-BN	YES	
		Serawai	Napal Melintang	Serawai (S.)	SRW-NM	NO	
		Kaur	Jembatan Dua	Kaur	KAU-J2	NO	
		Muko-Muko	Pondok Lunang	Muko-Muko	MUK-PL	YES	
			Suka Pindah	Muko-Muko	MUK-SP	YES	
		HAJI	Sukarami	HAJI	HAJ	NO	
		Kubu	Sungai Kijang	Kubu	KUBU	mixed	65%

The spread of this innovation is much wider than the nasal deletion before voiceless stop innovation. Like the areal feature of nasal deletion before voiceless stops, the loss of the voiced stop after a nasal is found in the language varieties of northwest South Sumatra and northern Bengkulu, but also is found throughout the Musi River basin and the coastal plain of South Sumatra.

Once again, it appears that the MUSI language is the innovator, with most dialects of MUSI showing the innovation. The innovation is also found in the Pekal variety of BARISAN, in five out of six varieties of Rejang,[59] and in the Muko-Muko variety of MINANGKABAU. Additionally, some varieties of BARISAN in close contact with MUSI varieties show the nasal assimilation innovation. Examples are Padang Bindu and the three Lintang sites of the HIGHLAND cluster, which border the MUSI language area. In the OGANIC cluster, the most downstream Ogan site, Rantau Alai, and the nearby Rambang site, Tambangan Rambang, both have the nasal assimilation innovation. The Kubu site near the Rawas speaking area also has been considerably influenced by this innovation, with 65% of the items showing nasal assimilation.

For the purpose of grouping the dialects of southern Sumatran Malay, nasal assimilation encompasses almost all of the MUSI varieties (with the exception of two LOWLAND subcluster sites). More variation is found in BARISAN, with nasal assimilation marking the Lintang and Benakat varieties as different from most of the other varieties in the HIGHLAND cluster.

5.12 Suffixes: –ka(h), -kan, -an and -a

The transitive or causitive suffix *-kan* has four different realizations in south Sumatran Malay: *-ka*, *–kan*, *-an* and *–a*. The *-kan* affix is realized as *-kan* and *-kɨn*. The *–ka* is realized as *–ka*, *-ko*, *-kah*, *-ke*, *-kɨ*, *–kə*, while the *–a* suffix is either *-o* or *-a*. These realizations of the affix are not randomly distributed, but line up according to dialect for the most part, and with clusters of the languages to a lesser extent. The primary sources for the information presented in this section are the wordlists obtained in the rapid appraisal surveys of these language varieties. These are supplemented and checked against wordlists and language descriptions from PBh publications. It is important to note that the lack of the affix in a wordlist does not necessarily mean that this affix is not present in a language, as these affixes were not specifically sought out but rather usually emerged in eliciting causative or benefactive verbs such 'bury' or 'throw'. There were eleven words from the list of 358 that typically would show this construction. Table 5.26 and Figure 5.12 should be referred to in order to compare the distribution of the realizations of the *–ka*/*-kan* suffix.

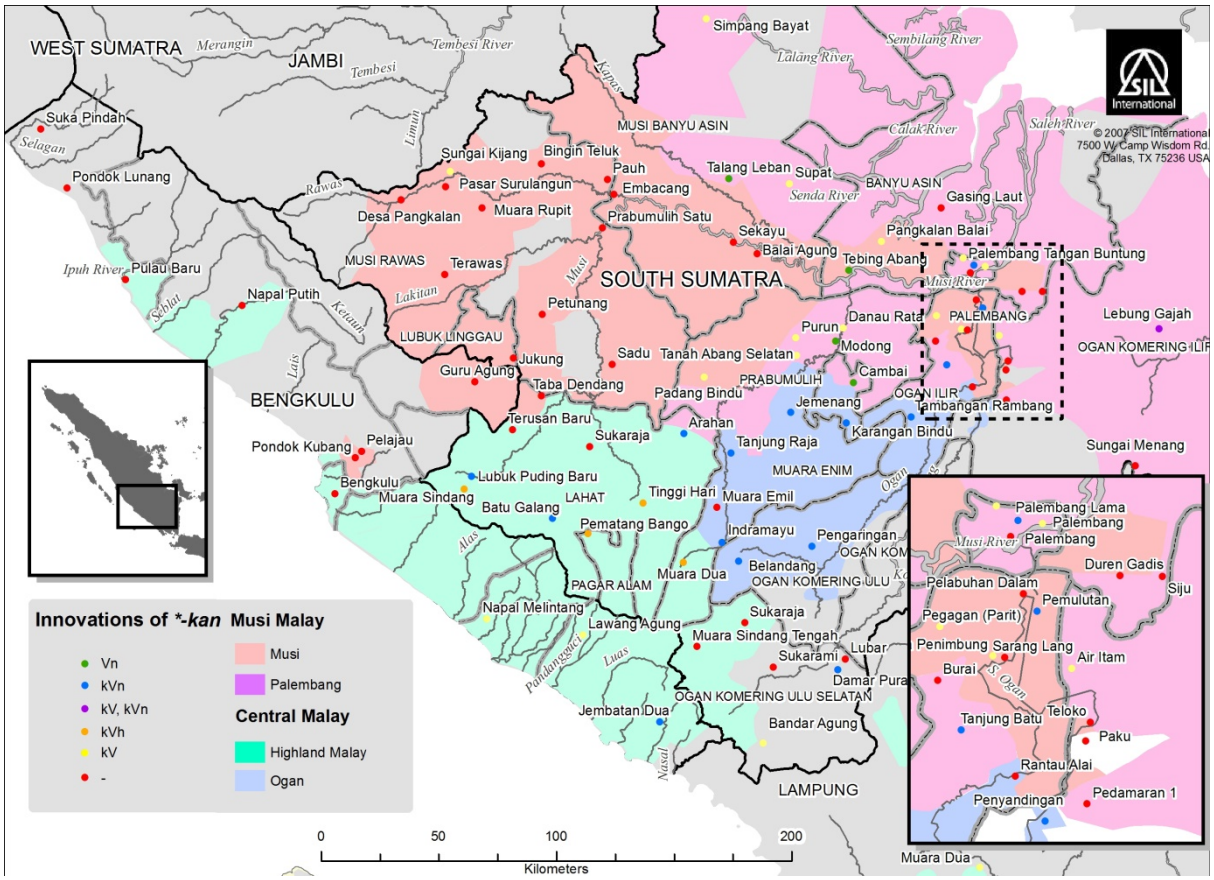
5.12.1 Background

Adelaar (1984; 1992:147–149) discusses the background of research into the transitive *–kan* marker. He reconstructs the preposition **akAn* (because PM was reconstructed to have retained schwa in final syllables, and evidence from languages like Sundanese and Javanese might suggest a PMP reconstruction of **aken*). Adelaar also leaves the door open for the preposition **kə* or a suffix **–ka*. This is on the basis of the evidence from one of the varieties of Malay used in his study, SWY (Serawai), which was also a variety included in the WIST rapid appraisal survey of Bengkulu. In his survey of Malayic lects, he finds four transitive suffixes, besides *–kan* and *-ka*, also *–an* and *–a*. In fact, he cites Medan's (1980) dialectological study as evidence that different Minangkabau dialects have all four. As it turns out, all four suffixes are also found in SSML (at least if we include Jambi Kubu as an honorary SSML member). Additionally, one of the suffixes, *-ka*, has been reanalyzed two different ways in different places.

Figure 5.12 shows the distribution of *-kan* etc. by subcluster.

[59] Of the six Rejang varieties, two are from Blust (1984), two are from McGinn (1997) and two from McGinn (2003). Rejang Musi, Rejang Pasisir, Rejang Kebanagung, and Rejong Lebong all have loss of voiced stop after nasal. Rejang Rejang is mixed (66%), while Rejang Rawas shows little evidence for loss of voiced stop before nasal (12%).

Figure 5.12 -kan *by subcluster*



5.12.2 South Barisan Malay

In the BARISAN cluster, the OGANIC cluster is consistently *–kan*, with the HIGHLAND cluster split between *–kah* and *–kan*.

5.12.2.1 OGANIC Cluster

Of the four Ogan wordlists, three of the four evince *–kan*, while the other one (Rantau Alai) was not long enough to provide evidence one way or the other. This agrees with the PBh Publications for Ogan, as *Kamus Ogan* (Aliana, Arif, et al. 1985), *Morfologi dan Sintaksis Bahasa Ogan* (Arifin et al. 1984), *and Ragam dan Dialek Bahasa Ogan* (Wahab et al. 1990), all of which give *–kan* as the suffix. The latter book touches on Rambang, using Rambang Lubai as one of its data points (point 4). This variety of Rambang also exhibits *–kan* (Wahab et al. 1990:70–72, see words 216, 223, 244, 245). Of the three Enim wordlists, two (Tanjung Raja and Indramayu) exhibit *–kan* while one (Muara Emil) does not have the affix. Checking this with the PBh publication *Geografi Linguistik Bahasa Enim* (Naning et al. 1998:204–205), there is strong confirmation of the suffix *–kan*. Of the 20 villages surveyed from the upstream to the downstream area, every single one showed the *–kan* suffix in map 147 and 148. The PBh publication *Sistem Reduplikasi Bahasa Enim* also is replete with examples of the *–kan* suffix throughout the example sentences (Arifin et al. 1988:sentences 280–298).

5.12.2.2 Highland Cluster

From the HIGHLAND cluster, Besemah, one Lematang Ulu site, Serawai, and Semenda evince *–kah*/*-ka*. Bengkulu, Lintang, Kaur and one Lematang Ulu site show *–kan*. Pekal had no evidence for either suffix in either wordlist, in this case being similar to MUSI and Rejang.

Of the lects in HIGHLAND cluster that show *–ka(h)*, the one with the largest number of speakers and the largest reach appears to be Besemah. Of the five Besemah wordlists, four show *–kah* or *–ka* as the suffix, with only Sukaraja Kisam (BES-MDK) not providing evidence. The PBh publication *Kamus Bahasa Indonesia Besemah (A-K)* (KBI-B) also supports this, showing *–ka* for the suffix in words such as *ŋadeka* for *məŋadakan* 'arrange s.t.' (Kasmansyah et al. 1999:3). The evidence from the two Lematang Ulu wordlists is mixed, with Tinggi Hari (LT-PP) using *–kah* and Arahan using *–kan*. With Arahan near the Rambang and Enim sites having *–kan*, it is understandable how it differs from the other Lematang Ulu site, which more closely follows Besemah. No PBh information was available specifically for Lematang Ulu. However, it was included in the geographic reach of Besemah in the PBh publication KBI-B which lists the spread of Besemah to the areas speaking Lematang Ulu, saying Besemah goes as far as the districts of Merapi (LT-MR) and Pulau Pinang (LT-PP) (Kasmansyah et al. 1999:vii). This agrees with *Fonologi dan Morfologi Bahasa Lematang*, which says that Lematang Ulu is the same as Besemah (Ihsan et al. 1996).

The Serawai list WIST obtained at Napal Melintang showed *–ka* while the Serawai lists from PBDSBB (lists 36-40) did not, though these lists tended to show the root form. The PBh publication *Kamus Bahasa Serawai-Indonesia (A-M)* (Aliana & al. 1985:xii) has *–ka* as the affix in the description of the language and throughout the entries. Adelaar (1992:149) lists data for Serawai with both *–kan*/*–kə*. As the southern Serawai dialect has the diphthong *-aw* as its reflex for final *–a*, it would appear that there was at some point an *h* at the end of the suffix to prevent the shift of *a* to the diphthong.

Four of the Semenda wordlists collected were sufficiently complete to check for presence of the *–kan* suffix. Three of these showed *–ka(h)* (Bandar Agung, Muara Dua, and Tanggamus), while Muara Sindang Tengah did not show evidence for the affix. Checking this against the PBh publication *Bahasa Semende* (Saleh et al. 1979:110, 119), one finds the suffix *–kah* given. A more updated PBh work by the same lead researcher, *Morfologi dan Sintaksis Bahasa Semende* (Saleh et al. 1985:227–238) also gives *–kah* in the elicited sentences. The Muara Dua wordlist evidencing *–kah* is in the Semenda research area used in the PBh studies (Saleh et al. 1979:197), where as the two sites showing *–ka* are well to the south of the PBh research area. The difference may be a dialectal one between Semenda Darat and Semenda Lembak, the Semenda dialect spoken in Ogan Komering Ulu. We conclude that the affix for Semenda is *–kah*.

In addition to these Semenda sites, the linguistically related Benakat site, Padang Bindu, realized the reflex of *–ka* as *–ko*. Remembering that the reflex for final **–a* in Benakat is *–o* and that Benakat generally preserves final **h* (cf. *ludah*, *lidah*, *mutah*) it appears that the affix was already *–ka* when it entered the grammar of Benakat. It may therefore have entered from one of the nearby MUSI dialects such as Penukal (Purun), where elision of final **h* is a strong tendency. The PBh literature *Struktur Bahasa Benakat* (Arifin et al. 2001:97–98) agrees with the evidence for a suffix of *-ko*.

Pekal also lacked evidence of *–ka* or *–kan*. Similar to Muko-Muko, the Pekal wordlists taken from Pulau Baru (PKL-PL) and Napal Putih (PKL-NP) gave no evidence for the suffix *–kan*, nor did PBDSBB (lists 24 and 26). Checking this against the PBh publication *Struktur Bahasa Pekal* (Nikelas et al. 1986), one finds that the authors of this study could not find evidence of any suffixes in Pekal. Those that were found were believed to be borrowings from Indonesian or MINANGKABAU, specifically mentioning the suffix *–kan* (p. 68-69) The sentences elicited in SBP have constructions with the *–kan* suffix in the Indonesian form given but with no suffix in Pekal form (p. 91-92). This absence of a realization of *–kan* makes Pekal similar to the nearby Muko-Muko dialect of MINANGKABAU, Rawas, Col dialects of MUSI, and Rejang. Additionally, no evidence was found of Rejang (a non-Malayic language) having the *–kan* suffix in the five Rejang wordlists examined or in *Kamus Rejang-Indonesia* (Badrul et al. 1985). This lack of a *–kan* suffix seems to run in a horizontal band from the Rawas River to the Bengkulu coast.

The HIGHLAND dialects that have *–kan* are primarily located in a narrow band going from Bengkulu City in the west across to the one Lematang Ulu site Arahan, in the east. The exception to this is Kaur to the far south. The two complete Lintang wordlists, Batu Galang and Lubuk Puding Baru, evince the *-kan* suffix, while Terusan Baru—a short 100 item wordlist—does not. The PBh publication, *Fonologi dan Morfologi Bahasa Lintang* (Nursato et al. 1989:130–131) gives *–kan* as the benefactive/causative suffix. This agrees with the PBh publication *Sistem Reduplikasi Bahasa*

Lintang which also gives *–kan* (Ihsan, Diem & Majelis 2002:28–29, 31–32). The Bengkulu wordlist by WIST shows no evidence for the suffix, but the PBh literature *Morfosintaksis Bahasa Melayu Bengkulu* (Imranuddin D et al. 1994:22–23) gives *–kan* as the benefactive/causative affix. It's worth noting that the four informants used in the PBh study had completed SMP (middle school) or higher education. This was also true of the WIST wordlist informant, who was a university student. (One might expect those with higher education in Indonesian to show grammatical interference (Gil 2005).) Kaur also has *-kan* but is geographically separated from the Bengkulu, Lintang and Lematang Ulu dialects by the Besemah, Serawai, and Semenda dialects.

5.12.3 Musi

The MUSI language shows some variation between its two constituents, the UPPER MUSI cluster and the PALEMBANG-LOWLAND cluster.

5.12.3.1 Upper Musi cluster

The WIST wordlists for the UPPER MUSI cluster show no evidence for the suffix *–kan* other than in the dialects Pegagan and Penukal. In Pegagan the *-ke* shows up both in the wordlist from WIST (MU-PG) and by Tadmor (MU-PG2). As Pegagan is lumped with Ogan and other nearby varieties by PBh, their evidence for Pegagan is found in *Ragam dan Dialek Bahasa Ogan* in data point 7, Tanjung Raja, which shows *-ke*. (Wahab et al. 1990:70–72; see words 216, 223, 244, 245). The *–ke* suffix is also found in the Penukal (Purun) subdialect of Musi Proper. Other than this, the other eight wordlists from the Musi Proper dialect of the MUSI language do not evince this suffix. However, the PBh publication *Struktur Bahasa Musi* (Gani, Ahmad & Ihsan 1981:95, 119–120) gives *–ke* as the causative/benefactive suffix with extensive examples. The authors of the SBM state that the speakers used in their study come from Musi Proper speaking areas around Sekayu. Another PBh publication *Morfologi dan Sintaksis Bahasa Musi* (Arif et al. 1985:26, 128) also has the suffix *–ke* listed as well as used in some elicited sentences. As the *-a* goes to *–e* in Musi Proper, it appears that this is an underlying *–ka* going to *–ke*. Another source, *Buku Serasan Sekate dan Penduduknya,* gives the affix of *-ke* in writing out several proverbs from Musi Proper (Haris 2006:328–333, 343). To clear up this confusion, an informant from Sekayu was contacted who reassured the researchers that there is indeed a *–ke* affix in Musi Proper. Based on the three published treatments of Musi Proper, the testimony of one of our wordlist informants, and given that the wordlists were not designed to specifically elicit the *–kan* affix, the decision is made to consider the UPPER MUSI cluster as employing the *–ke* (< *-ka*) affix.

Looking through WIST wordlists and sentence gathered for the three sites for Rawas, no clear cases of the affix for *–ka(n)* were found. However, the PBh publication, *Morfologi dan Sintaksis Bahasa Rawas* has the suffix as *–ke* without specifying if this is *–ke* or *–kə* (Aliana, Nursato, Siti Slamah Arifin, Efendi, et al. 1985:39–40, 86–87). This also seems somewhat suspect on the grounds that the reflex of final **a* for Rawas is normally *–o*. An earlier PBh publication, *Struktur Bahasa Rawas* (Saleh, Kamsari & Madjid 1984:115), says that the researchers were unable to find a local reflex of the *–kan* suffix, other than in subjects who have been educated extensively or have lived outside of the Rawas area. These use the suffix *–kan*. Given that the suffix *-kan* is unattested in the wordlists and other language varieties in the area (Col, Pekal, and Kerinci) and the conflicting witness from the PBh publications, the decision is made to keep Rawas as not having a reflex of *–kan*.

The Col variety of MUSI showed no evidence for either affix in the six wordlists taken by WIST and the data from TP 32 in *PBDSBB*. The PBh publications *Fonologi dan Morfologi Bahasa Sindang* (Arifin et al. 1996) and *Sintaksis Bahasa Sindang* (Arifin et al. 1997) also do not include the *–kan* suffix or its equivalent in the description of affixes.

5.12.3.2 Palembang-Lowland cluster

In contrast to the UPPER MUSI cluster where the suffix *–kan* is either not present or is consistently a reflex of *–ka* (*-ke*), in the PALEMBANG-LOWLAND cluster more variation is found, with both *–kan* and *–ka* evinced. PALEMBANG and the Pesisir varieties almost entirely evince *–ki*, with one

Palembang Lama dialect and Pemulutan evincing *–kan*. The PBh testimony for Palembang Pasar found in *Struktur Bahasa Melayu Palembang* described the underlying suffix *–ka* was realized as *–ke* (Dunggio et al. 1983:47–48). As the PBh publication does not differentiate in font and typeface used between *e* and *ə*, the reader is advised to take this as *–kə*. The health publication *Kampung Muarobanyu Tehindar Dari Penyakit Flu Burung* (translated from an Indonesian publication) uses the *–ki* (Matthews & Indarto 2007). Furthermore, the PBh publication *Morfologi dan Sintaksis Bahasa Melayu Palembang* uses the *–ke* affix in the example sentences printed, though as in the previous PBh publication, this should be interpreted as *–kə*/*–ki* (Aliana et al. 1987:227). Overall, these sources give solid backing to the wordlist evidence which shows the presence of the affix *–kə*/*–ki*.

The LOWLAND subcluster varieties are mixed. The four wordlists for Belide evince the *–an* ending, showing agreement with the PBh publication *Morfologi dan Sintaksis Bahasa Bilide Dialek Lembak* which also gives *–an* (Aliana, Nursato, Siti Salamah Arifin, Efendi, et al. 1985:25–26, 34–35). Both Lematang Ilir sites have the underlying form *–ka*, with Danau Rata (PB-SR) exhibiting *–ke*, and Tanah Abang Selatan (PB-TA) showing *–ki*. Danau Rata is located near the eastern border of the Pegagan language area, which also has the *–ke* suffix, while Tanah Abang Selatan is closer to the suffix used by the Pesisir areas such as Supat and Pangkalan Balai to the north. The PBh publication *Fonologi dan Morfologi Bahasa Lematang* reports that Lematang Ilir has *–kə*, apparently using the Gunung Megang dialect (Ihsan et al. 1996:83). The predicted form for the Tanjung dialect would be –ku, based on the underlying form *–ka*. It would seem that this might cause some confusion as it is also used as an abbreviated form of first person pronoun aku. The Penesak sites were mixed as well, with the Burai and Pedamaran sites not having any evidence of the suffix, while the Tanjung Batu site evinces *–kan*. The Tanjung Batu dialect agrees with the PBh publication *Morfologi dan Sintaksis Bahasa Panesak* which gives *–kan* as the causative and benefactive suffix (Purnomo et al. 2000:21, 52–53, 174). An earlier PBh publication *Struktur Bahasa Panesak* also gives *–kan* as the suffix (Gaffar et al. 1985:41). Checking these against *Ragam dan Dialek Bahasa Ogan*, which lumps the Penesak variety in with Ogan, one sees that Tanjung Batu (point 5) also has *–kan* (Wahab et al. 1990:70–72; see words 216, 223, 244, 24). All three of these publications use the Tanjung Batu dialect of Penesak as their source (which has the *–o* ending), and do not treat the Burai dialect (which has the *–e* ending). So while they provide confirmation for the data elicited for Tanjung Batu, they do not contradict the data obtained for the Burai dialect of Penesak.

5.12.4 Other lects

For the other language varieties of Bengkulu and South Sumatra that are not part of the MUSI or BARISAN languages, there was not much support for the existence of an affix, with the exception of Kubu.

5.12.4.3 Kubu (Anak Dalam)

The wordlist from the Kubu sector of Sungai Kijang (KUBU) showed *–ko* as a reflex of the suffix *–ka*. This also agreed with the PBh publication *Struktur Bahasa Kubu* (Dunggio et al. 1985:77–78) which gives *-ko*. It is worth noting that this research was conducted in the Kubu communities living around the Rawas River, which includes the area of Sungai Kijang, the WIST research site for Kubu. Checking this against the PBh publication *Dialektologi Bahasa Kubu* (Maryono, Setyonegoro & Kusmana 1997) one finds the underlying form *–ka* in most of the five villages used as samples in Jambi. Data point three (Pematang Kolim) is the closest to the Sungai Kijang site across the border in Musi Rawas, South Sumatra. The information from this book shows the affixes as follows (Table 5.23):

As can be seen from Table 5.23, the examples are not entirely consistent (cf. *memandikan* or *membelikan*). This may point to the affix as not being a regular feature of the language but something that has more recently come into use. The three data points that seem consistent reflect an underlying *–ka* suffix, with Tanjung Lebar's being realized as *–ki* (probably),[60] similar to nearby Pesisir lects in

[60] This PB publication does not differentiate between [ɨ], [e], [ə] or [ɛ].

Supat, Air Itam, and in Palembang. Pematang Kolim is identical to the nearby Kubu settlement in Sungai Kijang. We see two distinct affixes in Pematang Kabau and Dusun Tuo including the *a*/*o* suffix which seemingly does not occur in SSML proper but does in MINANGKABAU (Dusun Tuo is near Jambi's northern border).

Table 5.23 Jambi Kubu reflexes of –kan *and* –ka

Indonesian	Bukit Tembesu	Tanjung Lebar	Pematang Kolim	Pematang Kabau	Dusun Tuo
melahirkan	----	ke	ko	----	----
menguburkan	----	ke	ko	----	ko
memandikan	ko	ke	ko	kan	ko
picingkan	ke	ke	----	kan	----
menidurkan	ko	ke	----	_o	----
membelikan	kan	ke	ko	_a	_a/_o
memandikan	kan	ke	ko	kan	----/_a
dibubarkan	----	ke	----	----	----
Summary	unclear	ke	ko	unclear	ko

5.12.4.2 Muko-Muko

The Muko-Muko wordlists (TP 21, 22, 23) from PBDSBB show no evidence of *–ka* or *–kan*, agreeing with the PBh publication *Fonologi dan Morfologi Bahasa Muko-Muko* which states there is no suffix *-kan* in Muko-Muko (Aliana et al. 1993:105). The PBh book *Struktur Bahasa Muko-Muko* also has the same conclusion (Umar et al. 1986:80). In this regard Muko-Muko has more in common with nearby language varieties of MUSI and Rejang rather than other varieties of MINANGKABAU. *Geografi dialek bahasa MINANGKABAU: suatu deskripsi dan pemetaan di daerah kabupaten Pasaman* (Medan et al. 1986:41–42) shows the most common variants of the *–kan* suffix as being *–kə* and *–an*. Moussay (1998:245–249) gives *–kan* as the MINANGKABAU suffix *-kan*, agreeing with *Kamus Lengkap Bahasa Minang* (Saydam 2004). *Kamus MINANGKABAU – Indonesia* (Rusmali et al. 1985) differs, giving *–an* as the realization of *–kan*.

5.12.4.3 Haji

The wordlist for the language HAJI did not show conclusive evidence, though it had the words for 'lying' bahoŋan, budian in which *–an* might be the suffix. The PBh publication *Struktur Sastra Lisan Aji* (Ratnawaty et al. 2002) has some examples embedded in the elicited texts, but these lack consistency. Examples below:

bukakkan lawang 'open the door' (p. 21)
kunjukkan 'give' (p. 23)
kukurungkan 'I cage it' (p. 24)
usahkan 'try' (p. 39)
dighesghaskan 'to be pulled' (p. 39)
melihighken 'give birth to' (p. 45)
tapirikah 'be hurt' (p. 29)
Tuhan nurunkon 'God determines' (p. 25)
ngingokon 'think about' (p. 31)
ngaturko 'put in order' (p. 33)
campakkon 'throw out' (p. 45)
ngantakkon 'to lead someone' (p.45)
ngelunggukkon 'pile up s.t.' (p. 46)
dimasukkonnya 'place into s.t.' (p. 46)
gunakon 'use s.t.' (p. 47)
nyeghopkon 'straighten out s.t.' (p. 47)

Given the large amount of Lampungic borrowings in HAJI (Anderbeck 2007f), and given that Lampungic lects variously have both *–kən*/*-kon* and *–kə*/*-ko* suffixes (personal field data), it would seem that the native HAJI reflex is *–kan*, in spite of its lack of preponderance.

5.12.5 Clustering with -ka/-kan

Based upon the evidence from the wordlists and PBh publications, the information is summarized for the lects, clusters, and languages talked about in this section. (See Figure 5.12 and Tables 5.24, 5.25, and 5.26.)

The column to the far right in Table 5.24 shows the affix as given in the PBh literature for that dialect, which in most cases agreed with the affixes elicited in the wordlist. Please note that the table simplifies from the varieties' reflexes and shows either *–ka* or *–kan*. The tables show how while most varieties have a clear tendency (though with some exceptions), at the level of the cluster and language a clear pattern can be harder to determine. This may be in part due to the use of two different prepositions which have been cliticized as Adelaar points out.

The affix *–ka*/*-kan* appears to be useful in grouping the individual lects, of some use in making the clusters, and of less use in making conclusions regarding the overall composition of language clusters. For example, the Belide dialect has a very consistent use of the affix *–an* in the research sites visited, setting it apart from other varieties found in the MUSI language. Another example, the members of the OGANIC cluster are also uniform in their use of *–kan*. However, at the level of the cluster there does not appear to be any unifying pattern.

Trudgill and Chambers state that grammatical isoglosses, including morphological isoglosses, stratify speech communities more sharply than lexical or phonological isoglosses (1998:98–99). In the cases of the lects examined in southern Sumatra, there often seemed to be little problem in comprehension between varieties with a different suffix, as seen in the PALEMBANG varieties or in the HIGHLAND cluster varieties. Conversely, the varieties not evincing the *–ka*/*-kan* affix were seen as hard to understand by surrounding lects. The dialect with the *–an* affix, Belide, was also described as difficult to understand by other varieties which have regular contact with it.

It is particularly interesting to see this at work in the same site, such as Lebung Gajah (CSTL-LG) with both the *-kɨ* and *-kɨn* suffixes, which tends to suggest the use of both **akAn* and *kə*. So in *ɲipat*/*kɨ* 'to throw towards' one sees the locative function, while in *ɲulaʔ*/*kɨn* (to push something) one sees the change of state or to do something for someone else's benefit. In closing this section, Adelaar's statement, "There are sometimes formally different correspondences between closely related (sub)dialects…" (1992:149) certainly rings true.

Table 5.24 –ka /-kan *suffixes in MUSI language*

		Village	suffix	PBh	Name of book
PALEMBANG-LOWLAND	LOWLAND	Tanjung Batu	-kan	-kan	Morfologi dan Sintaksis BP
		Pedamaran	none		
		Burai	none	-kan	Struktur BP
		Tanah Abang S.	-ki	-kə	Fonologi dan Morfologi BL
		Danau Rata	-ke	-ku	*Tanjung dialect
		Modong	-an	-an	Morfologi & Sintaksis BBDL
		Cambai	-an		
		Talang Leban	-an		
		Tebing Abang	-an		
	PALEMB'G	Teloko	none		
		Paku	none		
		Palembang Lama	-kən		
		Sarang Lang	none		
		Pemulutan	-kan		
		Pelabuhan Dalam	none	-kə/ki	Morfologi dan Sintaksis BMP
		Palembang City 1	-ki	-kə	Struktur BMP
		Gasing Laut	none	-ki	Bird flu booklet
		Duren Gadis	none		
		Air Itam	-ke		
		Siju	none		
		Tangan Buntung	-ki		
		Palembang City 2	none		
		Lebung Gajah	-ki,-kin		
		Pangkalan Balai	-kə		
		Supat	-kə		
		Simpang Bayat	-ka		
MUSI		Sungai Menang	none		
UPPER MUSI	Rawas	Muara Rupit	none	-ke	Morfologi dan Sintaksis BR
		Pasar Surulangun	none	none	Struktur BR
		Pangkalan	none		
	Musi Proper	Purun	-ke		
		Petunang	none		
		Sadu	none		
		Prabumulih Satu	none		
		Embacang	none		
		Pauh	none		
		Bingin Teluk	none	-ke	Morfologi dan Sintaksis BM
		Sekayu	none	-ke	Struktur BM
		Balai Agung	none		
	Pegagan	Muara Penimbung	-ke	-ke	Ragam dan Dialek BO
		Pegagan	-ke		
	Col	Jukung	none		
		Taba Dendang	none	none	Sintaksis BSindang
		Terawas	none	none	Fonologi & Morfologi B.Sind.
		Guru Agung	none		
		Pelajau, P Kubang	none		

Table 5.25 –ka / -kan *suffixes in SOUTH BARISAN and other languages*

		Village	suffix	PBh	Name of book
OGANIC	Rambang	Penyandingan	none		
		Tamb. Rambang	-kan	-kan	Ragam dan Dialek BO
		Karangan Bindu	-kan		
		Jemenang	-kan		
		Tanjung Raja	-kan		
	Enim	Muara Emil	none	-kan	Sistem Reduplikasi BE
		Indramayu	-kan	-kan	Geografi Linguistik BE
	Ogan	Rantau Alai	none		
		Pengaringan	-kan	-kə	Morfologi BO
		Damar Pura	-kan	-ka	Kamus BO
		Belandang	-kan		
HIGHLAND	Pekal	Napal Putih	none		
		Pulau Baru	none	none	Struktur BP
	Bengkulu	Bengkulu City	none	-kan	Morfologi BBengkulu
	Lintang	Batu Galang	-kan	-kan	Fonologi & Morfologi BL
		Lubuk Puding Baru	-kan	-kan	Sistem Reduplikasi BL
		Terusan Baru	none		
	Besemah	Sukaraja Kisam	none		
		Sukaraja Kikim	-kah	-ka	Kamus I-BBesemah (A-K)
		Pematang Bango	-kah	-ka	Struktur BBesemah
		Muara Sindang	-kah		
		Lawang Agung	-ka		
	Lematang Ulu	Tinggi Hari	-kah		
		Arahan	-kan		
	Semenda	Muara Sindang			
		Tengah	none	-kah	Morfologi dan sintaksis BS
		Muara Dua	-kah	-kah	Struktur BS
		Bandar Agung	-ka		
		Tanggamus	-ka		
	Benakat	Padang Bindu	-ko	-ko	Struktur Benakat
	Serawai	Napal Melintang	-ka	-ka	Kamus BS-I (A-M)
	Kaur	Jembatan Dua	-kan		
	Muko-Muko	Pondok Lunang	none	none	PBSBB 21,22,23
					Fonologi & Morfologi BMM
		Suka Pindah	none	none	Struktur BMM
	HAJI	Sukarami	none	-kon	Stuktur Sastra Lisan BA
	Kubu	Sungai Kijang	-ko	-ko	Struktur BK
				-ko, -ke, -kan	Dialektologi BK

Table 5.26 Summary of –ka/-kan

Language	Decision	Cluster	Decision	Variety	Decision
MUSI	-ka	PALEMBANG	(mixed)	Penesak	-kan
				Lematang Ilir	-ka
				Belide	-an
				Palembang	-ka
		UPPER MUSI	(none)	Musi Proper	(none)
				Rawas	(none)
				Pegagan	-ka
BARISAN	(mixed)	OGANIC	-kan	Rambang	-kan
				Enim	-kan
				Ogan	-kan
		HIGHLAND	-kah	Besemah	-kah
				Semenda	-kah
				Serawai	-kah
				Lintang	-kan
				Lematang Ulu	(mixed)
				Bengkulu	-kan
Kubu	-ka				
Col	(none)				
Pekal	(none)				
Kaur	-kan				
HAJI	-kan				

5.12.6 Reconstructing the shape of pre-SSML suffixes

This section contains a brief exploration of what the phonological shape of the four suffixes may have been in the past.

The *–kan* suffix is nearly universally reflected in SSML as *–kan*; the only exception is one Palembang Lama site with *–kən* and one word in Lebung Gajah (Pesisir) with the similar *–kɨn* (see the discussion in §5.10). The reconstruction of the shape of the pre-SSML suffix depends on whether the Palembang Lama reflex is viewed as retention or innovation.

One additional issue with *–kan* is Adelaar's (1984:409) suggestion that *–kan* in SSML may have been borrowed from the standard language. Only fourteen sampled SSML sites out of 81 (about one-sixth) reflect *–kan*, and over half of these are in the OGANIC cluster. There may be something to his suggestion.

The *–an* suffix in the four Lowland sites is consistently reflected as *–an*.

The various *–a* and *–o* suffixes in Jambi Kubu mirror the lects' word-final **a* endings and should thus be reconstructed as **-a*.

The most complicated suffix to reconstruct is *–ka*/*-kah*/*-ko*/*-ke*/*-kə* etc. The likeliest conclusion is that the base form is **-ka*, but that it has been reanalyzed in the HIGHLAND area with an excrescent *h* as *-kah*, and as *–kə* in between three and nine sampled sites in the PALEMBANG subcluster. The *–kah* form should not be seen as basic, given its geographical limitation of the innovation to the HIGHLAND area, as that would require all other SSML and non-SSML Malayic lects to have lost the *–h*. The *–kə* form is more of a possibility than *-kah*, especially given Adelaar's suggestion that the suffix derives from the *kə* preposition. However, we lean away from this analysis for reasons of numerical dominance of and the geographical spread of *–ka*. Table 5.27 shows the twenty-three sampled sites with reflexes of *–ka*/*-kə*, along with the sites' reflex of **-a*. The theory is that, if a site typically reflects **-a* as *e*, and their suffix is *–ke*, we can safely assume that the underlying form is *-ka*. Of these twenty-three, fourteen or two-thirds (in white) clearly can be shown to have an underlying form of *–ka*. These fourteen come from a diverse area including the LOWLAND and PALEMBANG subclusters, UPPER MUSI, Benakat (HIGHLAND) and Kubu. Three sites' suffixes (in red), all from the

PALEMBANG subcluster, do not match with their *-*a* reflex and seem to reflect *–kə*, while six sites (in yellow) are ambiguous because their reflex of *-*a* is *ɔ*, *ɘ* or *ɨ*.

Table 5.27 Correspondences between *-a *and* -ka/-kə

Village	Subcluster /Dialect	*-a	-ka/-kə
Tanah Abang S.	LOWLAND	ɨ	-kɨ
Danau Rata	LOWLAND	e	-ke
Pelabuhan Dalam	LOWLAND	ɨ	-kə/kɨ
Palembang City 1	PALEMBANG	o	-kɨ
Air Itam	PALEMBANG	e	-ke
Tangan Buntung	PALEMBANG	o	-kɨ
Lebung Gajah	PALEMBANG	ɨ	-kɨ,-kɨn
Pangkalan Balai	PALEMBANG	ɘ	-kə
Supat	PALEMBANG	o	-kɘ
Simpang Bayat	PALEMBANG	o/a	-ka
Muara Rupit	Rawas Rupit	o	-ke
Purun	Musi Proper	e	-ke
Bingin Teluk	Musi Proper	e	-ke
Sekayu	Musi Proper	e	-ke
Muara Penimbung	Pegagan	e	-ke
Pegagan	Pegagan	e	-ke
Pengaringan	Ogan	ɘ	-kə
Padang Bindu	Benakat	o	-ko
Sungai Kijang	Kubu	o/a	-ko
Bukit Tembesu	Kubu	o	-ko, -ke, -kan
Pematang Kolim	Kubu	o/a/e	-ko
Dusun Tuo	Kubu	o/a	-ko, -a, -o

Therefore, on the basis of numerical dominance (five times greater) and geographical spread, we conclude that *–*ka*, not *–kə*, is the earlier SSML form.

5.13 Word final consonant for kinship terms

Kinship terms in southern Sumatran Malay usually end in either the glottal stop *-ʔ* (*maʔ*, *umaʔ*, *əmaʔ*, *bapaʔ*) or the velic nasal *-ŋ* (*n^duŋ*, *bapaŋ*, *maŋ*). These have been analyzed as (competing) vocative suffixes by Adelaar (1992:119). By far the most common suffix in Malayic lects is *-ʔ*; in fact, apart from Old Malayo-Javanese literature, the *-ŋ* suffix is quite rare outside SSML.

In looking through the data collected by WIST, the following kinship terms were collected through the wordlists and sentences: mother (WL59), father (WL60), grandmother (WL235), grandfather (WL236), and uncle (sentence 12). As grandmother and grandfather occur after wordlist item #200, they were not collected at some of the sites where a partial wordlist was elicited. Also it was found that the word for uncle often varied within one village, probably due to the fact that there are different words used for uncle depending upon which side of the family the uncle is from and his relative age compared to the speaker's parent. Because of this, the word 'uncle' was not used in the analysis, as one village might have both words for uncle that end with the velar nasal and with the glottal stop. For example, from *Geografi Linguistik Bahasa Enim* the mother's brother is *mintiŋ* while the father's brother is *mamaʔ* (Naning et al. 1998:map 107 & 110, village 16–Sugihwaras).

For the MUSI language, the ending for kinship terms mother, father, grandmother, and grandfather is the *ʔ*, found throughout the MUSI sites visited without exception. Due to the uniformity of the results, the MUSI language and its dialects are not shown individually in Table 5.28.

The BARISAN cluster sees a division between the OGANIC cluster and the HIGHLAND cluster. The OGANIC cluster typically evinces *ʔ*, with the exception of one Rambang site, Jemenang (RAM-RD), downstream from Lematang Ulu sites of Arahan and Tinggi Hari. This contrasts with the HIGHLAND cluster, which primarily ends in *ŋ* in the heartland of HIGHLAND and then shifts to *ʔ* as one moves out towards the coast of Bengkulu—Bengkulu, Serawai, Pekal, and Kaur have the glottal stop.

For purposes of dialect grouping, the velic nasal *ŋ* is a defining characteristic of the Semenda, Besemah, Lematang Ulu, and Lintang lects—in short the HIGHLAND varieties found in the Bukit Barisan Mountains.

Table 5.28 ʔ/ŋ ending for kinship terms (BARISAN and other languages)

Language/Cluster/Variety			Village	Dialect	Code	ʔ/ŋ ending kinship term
BARISAN	OGANIC	Rambang	Penyandingan	Rambang	RAM-PNY	ʔ
			Tambangan Rambang	Rambang	RAM-TR	ʔ
			Karangan Bindu	Rambang	RAM-RK	ʔ
			Jemenang	Rambang	RAM-RD	ŋ/ʔ
		Enim	Tanjung Raja	Enim Ilir	EN-ME	ʔ
			Muara Emil	Enim Tengah	EN-TAS	ʔ
			Indramayu	Enim Ulu	EN-TAN	ʔ
		Ogan	Rantau Alai	Ogan Ilir	OG-RA	ʔ
			Pengaringan	Ogan Tengah	OG-BR	ʔ
			Damar Pura	Ogan Ulu	OG-DP	ʔ
			Belandang	Ogan Ulu	OG-UO	ʔ
	HIGHLAND	Kaur-Holle	Bintuhan	Kaur	KAUR	ʔ
		Kaur	Jembatan Dua	Kaur	KAU-J2	ʔ
		Bengkulu	Bengkulu City	Bengkulu Ind.	BNGKL	ʔ
		Lintang	Batu Galang	Lintang	BES-MP	ŋ
			Lubuk Puding Baru	Lintang	BES-UM	ŋ/ʔ
			Terusan Baru	Lintang	BES-TT	ŋ/ʔ
		Besemah	Sukaraja Kisam	Besemah	BES-MDK	ŋ
			Sukaraja	Kikim	BES-KT	ŋ/ʔ
			Pematang Bango	Besemah	BES-PA	ŋ
			Muara Sindang	Besemah	BES-AK	ŋ/ʔ
			Lawang Agung	Besemah	BES-BK	ŋ/ʔ
		Lem. Ulu	Tinggi Hari	Lematang Ulu	LT-PP	ŋ
			Arahan	Lematang Ulu	LT-MR	ŋ
		Semenda	Muara Sindang Tengah	Semenda	SEM-PB	ŋ
			Karya Nyata	Semenda	SEM-SDL	ŋ
			Penyandingan	Semenda	SEM-PNY	ŋ
			Muara Dua	Semenda	SEM-MD	ŋ
			Bandar Agung	Semenda	SEM-BA	ŋ
			Tanggamus	Semenda	SEM-LP	ŋ
		Benakat	Padang Bindu	Benakat	PB-BN	ŋ/ʔ
		Serawai	Napal Melintang	Serawai Manna	SRW-NM	ʔ
		Pekal	Napal Putih	Pekal	Pekal-NP	ʔ
			Pulau Baru	Pekal	PKL-PL	ʔ
MINANG			PBDSBB 21, 22, 23	Muko-Muko		ʔ
	Muko-Muko		Pondok Lunang	Muko-Muko	MUK-PL	ŋ/ʔ
			Suka Pindah	Muko-Muko	MUK-SP	ŋ/ʔ
		HAJI	Sukarami		HAJ	ʔ
		Kubu	Sungai Kijang		KUBU	ʔ
MUSI			All 48 sites		MUSI	ʔ

For sake of completeness, we will also mention that a different statistical measure, using phonological distance (*Levenshtein distance*; see Kessler (1995) and Nerbonne and Heeringa (1996)) was carried out on the same field data by Eldwin Truong. We take an extended quote from his unpublished conference paper (Truong 2010:3):

> For the most part, the grouping of language varieties in South Sumatra using the historical-comparative method and Levenshtein distance are substantially similar. Both analyses show two major clusters, denoted by Anderbeck and McDowell as the Musi Cluster and Central Malay Cluster. Both clusters are then subdivided into two subgroups. The Musi Cluster is divided into the Palembang subgroup and Musi subgroup, while the Central Malay Cluster is divided into the Ogan subgroup and Highland Malay subgroup. Compared with McDowell and Anderbeck's analysis, the Levenshtein distance analysis places most of the eighty-one language varieties analyzed within the same cluster, subgroup, and even smaller subgroups.

The report now shifts from presentation of linguistic to sociolinguistic data.

6 Rapid Appraisal Recorded Text Test (RA-RTT) findings

The RA-RTT is not designed to be a quantitative test. While numeric scoring was used, it is only a rough guide designed to assist the research team in gauging the comprehension of major and minor details in the story by the listeners. Various factors militate against the usage of the numerical scores as an absolute measurement of comprehension of one dialect against another.

One factor which tended to cause lower scores in one variety (the Pagar Alam dialect of BARISAN) was a poor recording volume quality when compared to the better quality of the two other varieties tested. This was alleviated by using headphones to amplify the volume for the individual listeners, a departure from the normal procedure of playing the story via cassette boom box for the assembled group. While the volume was then sufficient to be heard, the downside was only four people could listen at a time, requiring the story to be played once for the men and once for the women. In the researchers' opinion, this had little effect on the scoring.

Another variety (the Sekayu dialect of Musi Proper) evinces lexical and grammatical similarity to the primary LWC for southern Sumatra, Palembang Pasar. There was no attempt to mitigate for this factor as the RA-RTT does not claim to be a test for pure inherent comprehension, but rather a test of a community's comprehension of a given dialect, whether that be through contact or through inherent comprehension. Confirming this are the results seen in Table 6.1. The only location where the respondents had a higher percentage correct on the Musi Proper story (100%) than in the target dialect of BARISAN story (93%) was Napal Putih, located in Bengkulu Province, far from the range of the Palembang Pasar language.

As mentioned in the methodology section, the MINANGKABAU story would preferably have been done in the Bukit Tinggi dialect, the reference dialect, but the available recording was in the nearby and closely related Maninjau dialect. Native Bukit Tinggi speakers said this variety was slightly more difficult to follow, but understandable, and so the story was used. Both the Pekal and Muko-Muko respondents understood it but not perfectly. The Suka Pindah respondents scored lower than anticipated, in part due to frequent interruptions by a particular respondent and in part due to an earthquake that occurred during testing. This and their responses to the post RA-RTT questionnaire led researchers to seek out a more representative site for Muko-Muko, the village of Pondok Lunang. The researchers believe that the Pondok Lunang responses are more reliable in relation to Muko-Muko comprehension of the Maninjau dialect of MINANGKABAU.

Due to these factors the numerical results of this RA-RTT have been used only as a rough guide for relative comparison. The crux of the analysis rests on the researchers' ability to accurately gauge the group's level of difficulty in understanding the text and the respondents' own evaluations during the post RA-RTT questionnaire.

Table 6.1presents a summary of the results of the RA-RTT in terms of the percentile scores, while Table 6.2 is a qualitative evaluation of observed and reported comprehension of the texts. The blue background corresponds to low comprehension of the recorded text. Areas with high comprehension of the respective text are given an orange background. Yellow backgrounding shows the areas where the comprehension was somewhere in between.

Table 6.1 Percentage scores for major elements, RA-RTT results

	Village	**Maninjau (MINANGKABAU)**	**Sekayu (MUSI)**	**Sukaraja Kikim (Pagar Alam) (BARISAN)**
MUSI	Satu Ilir (Palembang Pasar)	-	81%	-
	Ilir Barat Satu (Palembang Pasar)	-	86%	-
	Batu Gajah (Rawas)	-	91%	-
	Pelajau (Col)	-	91%	-
BARISAN	Penanggiran (Enim/Rambang)	-	-	88%
	Jembatan Dua (Kaur)	-	-	93%
	Napal Putih (Pekal)	96%	100%	93%
MINANG-KABAU	Suka Pindah (Muko-Muko)	82%	78%	63%
	Pondok Lunang (Muko-Muko)	93%	-	-

Table 6.2 Summary of RA-RTT results

	Village	**Maninjau (MINANGKABAU)**		**Sekayu (MUSI)**		**Sukaraja Kikim (Pagar Alam) (BARISAN)**	
		Researcher's observations	Informants' self-evaluation	Researcher's observations	Informants' self-evaluation	Researcher's observations	Informants' self-evaluation
MUSI	Satu Ilir (Palembang Pasar)	-	-	Most major & 50% of minor details	Understood most, very different lang.		
	Ilir Barat Satu (Palembang Pasar)	-	-	Most major and all minor details	Understood 90%, similar lang.		
	Batu Gajah (Rawas)	-	-	Most major and all minor details	Understood all, lang. very different		
	Pelajau (Col)	-	-	Most major and minor details	Understood all, lang. almost identical		

	Village	Maninjau (MINANGKABAU)		Sekayu (MUSI)		Sukaraja Kikim (Pagar Alam) (BARISAN)	
		Researcher's observations	Informants' self-evaluation	Researcher's observations	Informants' self-evaluation	Researcher's observations	Informants' self-evaluation
BARISAN	Penanggiran (Enim/ Rambang)	-	-	Most major & 50% of minor details	NA	Almost all major and most minor details	Most of story, lang. different
	Jembatan Dua (Kaur)	-	-	-	-	All major and most minor details	Understood all, lang. different
	Napal Putih (Pekal)	Almost all major and minor details	Understood all, lang. very different	Perfect	All of story, lang. very different	All major and most minor details	Understood all, lang. very different
MINANGKABAU	Suka Pindah[61] (Muko-Muko)	Many major details, 50% of minor details	Could understand a little, very different	Many major details, all minor details	30-80% understood, lang. very different		
	Pondok Lunang (Muko-Muko)	Almost all major & 75% of minor details	Everything, but lang. very different				

[61] The researchers do not place much weight on the Suka Pindah results for the reasons provided above.

6.1.1 Survey Findings

Realizing that the RA-RTT is limited in ability, and that our research team would not have time to develop a full RTT (which is tested individually and uses quantitative scoring), the role of the RA-RTT was mainly used to confirm statements gathered during the SLQs about comprehension of other dialects in different areas and gauge attitude toward other lects.

The results displayed in Table 6.2 demonstrate a generally high comprehension rate by the tested groups of the texts presented. The most striking piece of evidence for comprehension comes from the tests which were administered within one language with a story from that same language. With the exception of Suka Pindah (an extraordinary case), all language families tested (BARISAN, MUSI, and MINANGKABAU) showed comprehension of the 'central' dialect chosen for the testing. This proved useful in testing dialects that were peripheral to the language such as Pekal and Kaur for BARISAN, Rawas and Col for MUSI, and Muko-Muko for MINANGKABAU. These locations did well in their comprehension and confirmed their statements that they could understand the dialect in question. For places like Col where conflicting answers as to comprehension were obtained, the RA-RTT provided evidence of comprehension, with the tested site believing that the Sekayu recording was from a neighboring village.

For MUSI, the RTT results showed comprehension of Sekayu in the tested sites ranging from good (PALEMBANG areas) to great (Col and Rawas areas), confirming the researchers' hypothesis that these dialects could understand the Sekayu dialect. This matches up with the language groupings made based upon the responses to the SLQs and innovations found. The Sekayu story also tested well in non-MUSI areas where it was used as a practice test before the test of the dialect in focus.

The BARISAN dialect of Besemah from Pagar Alam tested well in the three sites it was played. Penanggiran understood the story well, providing confirmation that the OGANIC cluster of BARISAN could understand the dialects from the HIGHLAND cluster. The northernmost (Pekal) and southernmost (Kaur) varieties of the HIGHLAND cluster also were able to understand the central Besemah variety of Pagar Alam. It was not tested outside of BARISAN.

The MINANGKABAU dialect of Maninjau was tested in two Muko-Muko sites and one Pekal site. It appeared to be well understood in one Muko-Muko site (Pondok Lunang), while the other site (Suka Pindah) did not score as highly. The test at the one site where the results were not as high was frequently interrupted by onlookers interjecting their comments and by an earthquake, so this is not seen as representative a score as the Pondok Lunang score. The neighboring BARISAN variety Pekal also understood the story well, which was not surprising given the influence of MINANGKABAU innovations and vocabulary in the language and their contact with MINANGKABAU speakers.

The focus of the RA-RTT in this survey was not so much to test for comprehension of different languages, but this was still done in some cases to see if there was a dramatic difference between comprehensions of languages in locations that were located at a boundary between languages. One example of this was the Pekal site Napal Putih, which was located at the periphery of the BARISAN and MINANGKABAU languages. Here the respondents scored well on both the BARISAN and MINANGKABAU dialects played, so the researchers relied on the answers to the SLQ questionnaires and phonological innovations to help in grouping Pekal. The results from the RA-RTT underscored the conclusion that there is not a sharp linguistic break between the MINANGKABAU dialect of Muko-Muko and the BARISAN dialect of Pekal, with both sharing many similar lexical items and phonological innovations.

6.1.2 RA-RTT Questionnaires

A short group questionnaire was administered after each RA-RTT (see the appendix). The questions on this questionnaire were meant to flesh out language attitudes towards the tested varieties, and to measure perceived comprehension of the texts. The assumption behind using this questionnaire in addition to the sociolinguistics questionnaire is that the RA-RTT

questionnaire more directly measures the informant's attitudes towards a particular dialect which has been directly presented to him or her. The sociolinguistic questionnaire, on the other hand, asks for the informant's attitude toward a dialect that has been presented by name only—the informant and the researcher may not always have the same perception of what a particular name is referring to.

Four questions dealt with perceived comprehension of the text (Table 6.3). The second and third questions were also presented briefly in Table 6.4 so that a side-by-side comparison could be made between the respondents' perception of comprehension and the researchers' perception. These are the results of the Musi Sekayu post-RTT questionnaire.

Table 6.3 Perceived comprehension of the Sekayu text

	According to your opinion, from where does the story-teller originate?	How much of the story did you understand?	How different is the story-teller's speech from your own?	What's the difference?
MUSI: Ilir Barat (Pbang)	Sekayu	90%	similar to Palembang	--
MUSI: Satu Ilir (Pbang)	Sekayu	Most	very different	accent
MUSI: Batu Gajah (Rawas)	Kab. Musi Banyuasin (in story)	100%	very different	because final **a* goes to *e*
MUSI: Pelajau (Lembak/Col)	Lembak (their area)	100%	almost the same	speaks slow and more like Indonesian
BARISAN: Napal Putih (Pekal)	Musi Banyuasin (in story)	100%	very different	different pronunciation
MINANGKABAU: Suka Pindah (Muko-Muko)	Lembak (Col)	30-80%	very different	everything

The only group that reported trouble understanding the Sekayu text was the MINANGKABAU group, and they did fairly well, getting most major details. One site received the RA-RTT but was not given an opportunity to complete the post RTT questionnaire (Penanggiran). It is interesting to note that many locations reported that the Sekayu text was very different based on the different pronunciation rather than different vocabulary or grammar.

Another item of interest was that both Pelajau in the Lembak/Col area and Suka Pindah (Pekal) identified the text as being from Lembak, highlighting the similarities between Sekayu and the Lembak variety of the Col dialect.

Two questions seek to gauge attitudes toward the tested variety. These answers are in response to the Sekayu text (Table 6.4):

Table 6.4 Attitudes toward the Sekayu text

	Does this storyteller speak Sekayu well? Why? / Why not?	Is the language of this story in high, mid or low register?
MUSI: Ilir Barat (Palembang Pasar)	no response	Normal
MUSI: Satu Ilir (Pbang)	Yes	Normal
MUSI: Batu Gajah (Rawas)	Yes	Normal
MUSI: Pelajau (Lembak/Col)	almost speaks Lembak well…too slow and too Indonesian	Normal
BARISAN: Napal Putih (Pekal)	Yes	Normal
MINANGKABAU: Suka Pindah (Muko-Muko)	Yes	Normal

The MINANGKABAU text was tested among the Muko-Muko and Pekal dialect test sites, but nowhere else. This is because the other dialect areas did not indicate either comprehension of MINANGKABAU or affiliation with MINANGKABAU. The tests showed that neither Pekal nor Muko-Muko felt that their lect was similar to the Maninjau dialect of MINANGKABAU. It is interesting that all responses to where the story teller originated were replied to with 'Padang' or a more general geographic term like 'West Sumatra' or the ethnic term 'Minang'. This was differed from the responses from Minang speakers originating from Bukit Tinggi and West Sumatra, who could all identify the speaker as being from the Maninjau area based upon the speaker's vocabulary, manner of speech, and certain place names.[62] This corresponds with what the researchers observed in Bengkulu: *Padang* often serves as a cover term for all things MINANGKABAU, where the finer distinctions between the different varieties of MINANGKABAU are not as important or well known as in West Sumatra. The Suka Pindah site felt they could understand only a little, though they actually understood most of the main points in the story. This points to the perceived difficulty of the Maninjau text, which even MINANGKABAU speakers from other parts of West Sumatra said was "lebih dalam" or harder to understand than the Bukit Tinggi dialect (Table 6.5).

Table 6.5 Perceived comprehension of the MINANGKABAU text

	According to your opinion, from where does the story-teller originate?	How much of the story did you understand?	How different is the story-teller's speech from your own?	What's the difference?
MINANGKABAU: Suka Pindah (Muko-Muko)	Padang	little	different	vocabulary
MINANGKABAU: Pondok Lunang (Muko-Muko)	Minang/Padang	100% but hard to understand	very different	everything
BARISAN: Napal Putih (Pekal)	Padang/West Sumatra	100%	very different	different intonation

[62] The story tellers were told not to include place names but had a hard time remembering to do this. When texts were re-elicited to obtain a text without the place names, the new text was often of a lower quality, so the initial text was used.

Unlike some lects such as Lampungic and Javanese, the MINANGKABAU influenced lects of Muko-Muko and Pekal (BARISAN) did not indicate a perception that the variety they heard was crude or refined speech, but everyday (Table 6.6). It does seem that in the MINANGKABAU heartland there is an increasing idea of some speech being kasar (crude) when they use traditional MINANGKABAU pronominals or other lexical items.[63] In this matter, Muko-Muko and Pekal follow the MUSI and BARISAN varieties which do not seem to indicate a crude or refined manner of speaking.

Table 6.6 Attitudes toward the MINANGKABAU text

	Does this storyteller speak MINANGKABAU well? Why? / Why not?	Is the language of this story in high, mid or low register?
MINANGKABAU: Suka Pindah (Muko-Muko)	yes	normal
MINANGKABAU: Pondok Lunang (Muko-Muko)	yes	normal
BARISAN: Napal Putih (Pekal)	yes, average	normal

The BARISAN variety of Besemah was tested in three BARISAN varieties and one MINANGKABAU variety. The three BARISAN sites reported good comprehension of the Pagar Alam but felt that the speech was different than their own. The MINANGKABAU site was less certain of their comprehension and felt that everything in the text was different from their lect (6.7). It was also interesting that the locations in northern Bengkulu associated the Pagar Alam text with BARISAN varieties spoken in south Bengkulu, a confirmation that those HIGHLAND varieties group together. Kaur identification of the Pagar Alam text as originating possibly from Ogan was also a confirmation that the OGANIC cluster fits in the BARISAN language.

Table 6.7 Perceived comprehension of the Pagar Alam text

	According to your opinion, from where does the story-teller originate?	How much of the story did you understand?	How different is the story-teller's speech from your own?	What's the difference?
BARISAN: Penanggiran men (Enim)	Pagar Alam	most	different	tone and pronunciation, vocabulary similar
BARISAN: Penanggiran women (Enim)	Pagar Alam	most	different	vocabulary
BARISAN: Jembatan Dua (Kaur)	Ogan or Semendo	100%	different	vocabulary and pronunciation
BARISAN: Napal Putih (Pekal)	Kaur, Besemah, Semendo, Pagar Alam	100%	very different	language, vocabulary

[63] Based on personal conversation with a native MINANGKABAU speaker and linguistics professor.

MINANGKABAU: Suka Pindah (Muko-Muko)	Kaur	0-70%	very different	everything

Two questions seek to gauge attitudes toward the Pagar Alam text (Table 6.8):

Table 6.8 Attitudes toward the Pagar Alam text

	Does this storyteller speak Pagar Alam well? Why? / Why not?	Is the language of this story in high, mid or low register?
BARISAN: Penanggiran men (Enim)	yes	crude
BARISAN: Penanggiran women (Enim)	yes	normal
BARISAN: Jembatan Dua (Kaur)	yes, speaks Ogan/Semendo well	normal
BARISAN: Napal Putih (Pekal)	speaks (southern) Bengkulu okay	normal
MINANGKABAU: Suka Pindah (Muko-Muko)	speaks Kaur well	normal

From these responses it appears that the text was received as being normal for the most part, only once being called crude. This may be due to the fact that the story was a hunting story, which might have caused the Penanggiran men to view it as less refined. A listener who overheard the story (not a respondent) also commented on the hesitancy with which the story teller spoke, saying the speaker was a poor story teller, which may have influenced the perception of the Besemah sample text. All the responses thought the text spoke language X well, but, varied in what they thought the language was. The fact that all the elicited responses are within the BARISAN grouping provides some confirmation of the researchers' grouping of these dialects.

6.1.3 RA-RTT conclusions

The observed and reported comprehension of other dialects within the proposed languages supports the proposed language groupings of BARISAN, MUSI, and MINANGKABAU. This element corresponds to the comprehension component of the Ethnologue definition of a language. This comprehension was also predicted by the positive responses to the comprehension questions found in the SLQ and the high rate of lexical similarity found between the dialects in the proposed languages. In addition to these factors, the similar phonological innovations within the languages also led the researchers' to correctly predict a fairly high comprehension rate between these varieties.

However, the RA-RTT post test responses also show that these related varieties, while having a great deal of similarity and comprehension of other varieties, still see the other varieties as 'different' to 'very different'. While nothing in either the lexicostatistics or most of the phonological innovations would point to these varieties having a large difference, the few differences that were present were enough to be seen as very salient to the speakers. The different reflexes of final **a*, the different pronouns that are sometimes used, and different pronunciation contributed to 'identity' responses where almost no varieties garnered a response as being 'similar' to the tested sites' varieties. The exception to this was the Sekayu text being listed as similar and almost the same as Satu Ilir (Palembang Pasar) and Pelajau (Col Lembak). This confirms the isolating view of language found throughout South Sumatra and Bengkulu

in the SLQ responses, where many people identified their lect by the local village or district, but rarely as part of larger languages. However, the fact that 'wrong' responses to where the text was from often elicited the correct language, though the city of origin might be incorrect offered a glimpse that the test respondents could identify some features of the tested variety and match those with features they recognized from other varieties from the same language they were more familiar with.

7 Sociolinguistic findings

This is the fourth and final chapter on findings. (Congratulations if you've made it this far!) In this section we present the results of our sociolinguistic questionnaires and interviews, divided into sections on dialectology, language use, language change, language maintenance, language attitudes, and language and media. After this chapter we attempt to bring all the strands together in our conclusions.

7.1 Dialectology

An emic (insider) understanding of dialectology was gathered through group and individual interviews. Most sites viewed their lect as being unique, and could offer a few lexical examples or phonological examples of differences between them and neighboring villages, giving the researchers a list of potential dialects as long as the list of data points. Still, the respondents could also point out areas that had similar lects to their own and with whom they communicate without resorting to a language of wider communication (LWC).

Respondents in some areas had a much better idea of how their language related to other languages, where as others had a much more fragmented view of the linguistic landscape. Factors that seemed to be influence the respondents' knowledge of other lects are the frequency of contact with outsiders and the sense of ethnic identity.

Of the different language areas visited, it seemed that the following had a strong sense of identity and could readily name differences that distinguished their language from others: the Malayic varieties of Bengkulu Province, the varieties spoken in the area of the Bukit Barisan Mountains between Bengkulu, Lampung and South Sumatra (SOUTH BARISAN MALAY), Palembang, the languages from Sekayu and upstream along the Musi River drainage system, and the languages along the Ogan river drainage system.

However, the areas that seemed to have less of an ethnic identity and also less contact with outsiders were those in the largely swampy areas stretching east from Palembang toward the coast, between Jambi and Lampung. These often had not heard of other Malayic varieties they were asked about on the questionnaire, with the exception of Bahasa Palembang. These may fit in best with a 'Pesisir' grouping, though the extent of this Pesisir grouping needs to be tested with data from the coastal regions of Riau and Jambi. Another group of lects that were harder to group included those downstream from Tanjung Enim on the Lematang River. It was discovered that these do not form a 'Lematang' speech group but are parts of other lects that extend to the Lematang River such as Belide and Penesak. They were much more fragmented in their identity, admitting that their language might be almost entirely the same as Penesak (for example) but expressing a dislike for the Penesak villages and requesting that their lect not be called *Penesak*, but rather *Meranjat*.

In these places, a frequent response to the question, "In your opinion, what is the name of the language spoken here?" was "bahasa dusun" (village language)[64] or "Bahasa Melayu". Other variants included giving the name of the subdistrict (kelurahan) or regency (kecamatan). The respondents were asked to name dialects of Malay which were exactly the same as their own, others that were slightly different but easily understood, and others that were different enough to be difficult to be understood. As the respondents were the ones who supplied the

[64] The consistent strength of village-based identity we found throughout the research area finds an echo in Barendregt's (2006:122) statement that, at least for the southern Sumatran highlands, the village is "the core of everyday political organization."

names of the dialects, not all of the possible dialects in a cluster are represented in their responses. Also, please keep in mind that these are the responses of a small group of people from each research point, and may not be representative. In these villages could be found the 'splitters' and the 'lumpers'; splitters would state that another dialect area was entirely different because of a few lexical or phonological differences, while the lumpers were highly tolerant of large divergences in lexical items and phonology, emphasizing the importance of a shared ancestry and mutual comprehension of each others lect. The responses taken from the group reflect the group consensus.

7.1.1 MUSI Language

The MUSI language is composed of the UPPER MUSI and PALEMBANG-LOWLAND clusters, with their respective lects. See Figure 8.4 and Figure 8.5.

The UPPER MUSI cluster – composed of the Pegagan, Musi Proper, Col, and Rawas lects- stretches from the headwaters of the Musi River in the lowlands at the foot of the Bukit Barisan mountains to the swamps south of Palembang. These areas showed a gradation of reported intercomprehension, based largely upon distance from the downstream area. Of these, the most upstream, Rawas and Col,[65] appeared to be the hardest to understand for others.

Col is proposed to be part of a dialect chain with Musi Proper, Rawas, and Pegagan by WIST as background research showed similar vocabulary and similar phonological innovations between Col and Musi Proper, Rawas, and Pegagan. With Col sharing innovations with Rawas and another set of innovations with Musi Proper, it was surprising to hear the frequent statement from speakers of these two varieties that they had difficulty understanding Col (particularly given the high cognacy rate between Col and these sites). However, this is apparently asymmetric, as the Col speakers did very well at understanding a Sekayu recorded text. All four of these groups: the Col, Rawas, Musi Proper, and Pegagan had distinct ethnolinguistic identities and were able to point out differences within their own lects and with other lects. The central dialect for Col is Lubuk Linggau; for Rawas, Muara Rupit; for Musi Proper, Sekayu; and for Pegagan, the area of Indralaya.

Of the Musi Proper and Rawas sites that responded concerning comprehension of Col, seven said that Col was different enough to be difficult to understand, while only three said it was only a little different and could still be understood. This raises the question of why most Rawas and Musi Proper respondents frequently stated Col was difficult to understand. One possibility was that this was asymmetric intelligibility, a pattern in which often the lects more upstream can understand the ones downstream, while the downstream can not understand the upstream. Perhaps explaining the asymmetric comprehension is the fact that none of the other Musi Proper varieties share all the innovations of the Col varieties found in the Bukit Barisan Mountains (see §5).

The questionnaires upstream Col sites provided support for this hypothesis. Of the five Col sites that completed a SLQ, three (Terawas, Jukung, and Pelajau) replied that Musi Proper is similar enough to be easily understood. The other two sites (Taba Dendang and Guru Agung) did not say that Musi Proper was hard or easy to understand, so no conclusion can be drawn from them. The interviews and background research pointed out that Col has a tendency to reduce words and phrases, which people less familiar with it have difficulty with. The comprehension between Col and Musi Proper was tested with the Musi Proper RTT in Pelajau and the high comprehension of Musi Proper by this remote Col location confirmed placing Col with Musi Proper.

The Malay variety that has the largest spread geographically, the Musi Proper dialect, is predictably the one which has more 'labels' applied to it. Musi Proper goes as far east as kecamatan Lais and stretches west along the Musi until kecamatan BTS (Cecar) and along the Rawas River and is spoken throughout kecamatan Rawas Ilir. As is common with

[65] This is the autoglotonym for these speakers, and comes from their word for 'there is not X, none' or in BI, *tidak ada*. They are also sometimes referred to as the Lembak, a term most of their population outside of Bengkulu has not heard of, or as the Sindang Kelingi (Tarp, research notes, 2006).

ethnolinguistic identity in southern Sumatra, the political unit in which a speech community is located influences its own identity. For example, the Musi Proper varieties in Pauh and Bingin Teluk refer to themselves as being 'Rawas Ilir' but also admit that their language is very close to Sekayu (Musi Proper) and more distant from Rupit (Rawas). In Petunang, an area referred to in Wurm and Hattori and the Ethnologue as 'Sindang Kelingi', the informant we spoke with called it Kelingi, saying Sindang Kelingi was further to the west and was Col. He said Kelingi was Musi Proper, with some influence from Col. This was backed up by some illustrations of Col words and innovations that appear scattered in the wordlist he gave. Other Col areas agreed, saying the language in that area was 'mixed' with Musi Proper and more like Musi Proper than Col. The lect of Penukal (Purun; MU-PEN) has also been placed with Musi Proper on the basis of their statement that their language is very similar to that spoken in Sekayu, their ease in understanding Sekayu, and their distinct identity from Lematang Ilir speakers (whom they also understand) and Belide speakers (whom they do not understand).

The LOWLAND subcluster varieties had less clear ethnolinguistic identities. These areas could describe how their dialect was similar or different to varieties nearby, nor did most see themselves as part of a larger language group. Names of lects found (for the most part) south of the Musi River include Lematang Ilir, Penesak and Belide. Some Belide villages were found in Musi Rawas (labeled as an isolated Lematang community in Wurm and Hattori) and north of the Musi River in Banyuasin. One common innovation that most of these research sites shared is the apical trill *r*. The trilled *r* seems to be one sign of the influence of the modern Palembang Pasar speech on these communities. The sociolinguistics for this group were the most difficult of all the sites surveyed, as they reported they could understand many languages that were geographically quite removed (and it was unclear if and when they had had contact with people from these groups) while at the same time reporting some difficulty understanding nearby groups from related lects. One thing that helped was looking at the varieties they believed an uninitiated child could understand—in many cases they said the other languages in the LOWLAND subcluster and Palembang Pasar were understandable for a child.

Palembang city had a clear ethnolinguistic identity. Of the varieties sampled in the city, most had the innovation of the apical trill for the reflex of **r*. The distinguishing feature of the Palembang lect to the informants we talked with was final *–o* in place of *–*a*. They made a distinction between 'village Palembang', *Palembang dusun*, stating that a number of lexical items collected by researchers from villages in Banyuasin and OKI (outside of Palembang) were not standard Palembang. They view this variety as inferior to that spoken in the city, and less refined (kasar). Those who speak the Palembang used in the villages counter saying that what they speak is closer to the original Palembang language, *Palembang asli*. Concerning *Palembang asli*, most lifelong Palembang residents the researchers interviewed felt that there were no more 'true' speakers of *Palembang Lama asli*. The uniting factor between the village and city forms of Palembang was a shared ethnolinguistic identity of speaking the Palembang language, or being within the area the Palembang language is used. The salient feature of Palembang, the *–o* reflex of * *–a*, is present in most of the research sites grouped under Palembang. Those having a different reflex but still in the grouping are there on the basis of their proximity to Palembang and their identity with Palembang. These share many common cultural traits with the Pesisir communities, as well as linguistic features.

In summary, the MUSI language has two distinct but linked clusters, the PALEMBANG-LOWLAND cluster and the UPPER MUSI cluster. The city of Palembang serves as the hinge between the two clusters geographically, with the PALEMBANG-LOWLAND cluster running north to south along the coast of South Sumatra and inland past the Musi River. The UPPER MUSI cluster follows the Musi River before the LOWLAND and PALEMBANG subclusterS separate Pegagan from the other UPPER MUSI varieties. Sociolinguistic questionnaires pointed out the links in these clusters; however, the Pesisir and Palembang dialects evinced less awareness of other lects and accordingly commented less on these, making it difficult to draw conclusions beyond their district. The predicted dialect center for the UPPER MUSI cluster is Sekayu, while for the PALEMBANG-LOWLAND cluster it is Palembang Pasar. For the overall cluster, Palembang Pasar is the predicted reference dialect due to its size, prestige, and use as

the LWC, while the more linguistically central dialect is the Sekayu dialect. Testing the Sekayu dialect among Col, Rawas, and Palembang Pasar speakers with a RTT demonstrated all three could understand it (Table 7.1).

Table 7.1 MUSI language SLQ results for comprehension

Research Site	Palembang/ Pesisir (P)	LOWLAND (L)	Musi Proper (M)	Rawas (R)	Col (C)
Palembang 1 (P)	Same		Different		
Palembang 2 (P)	Same		Different		
Supat (P)	Same	Same	Similar	Similar	Different
Tebing Abang (L)	Similar	Same	Similar	Different	Different
Tanjung Batu (L)	Similar	Same	Similar	Different	Different
Embacang (M)			Same	Similar	Different
Prabumulih 1 (M)	Similar	Different	Same	Different	Similar
Sekayu (M)	Similar	Different	Same	Different	Different
Sekayu 2 (M)	Same/ Similar	Same/ Similar	Same	Same	Similar
Petunang (M)	Similar	Similar	Same	Similar	Different
Bingin Teluk (M)			Same	Similar	Different
Muara Rupit (R)	Similar		Similar	Same	Different
Pangkalan (R)	Similar		Similar	Same/ Similar	Similar
Pasar Surulangun (R)	Similar	Different	Different	Same/ Similar	Different
Terawas (C)			Similar	Similar	Same
Taba Dendang (C)					Same/ Similar
Jukung (C)[66]			Similar	Different	Same
Guru Agung (C)				Different	Same
Pelajau (C)	Similar		Similar	Same	Same

7.1.2 South Barisan Malay Cluster

The BARISAN cluster is composed of the OGANIC cluster and the BARISAN cluster. These both had strong ethnolinguistic identities and reported good comprehension within their clusters and between the clusters. This grouping also closely corresponds to the 'Central Malay' grouping mentioned by earlier Dutch linguists (Andaya 1993:16).[67] See Figure 8.2 and Figure 8.3.

The OGANIC cluster consists of the Ogan, Enim, and Rambang lects. These are found along the Ogan and Lematang River drainage basins. The Ogan and the Enim clearly identified their own lect as Ogan and Enim respectively. Rambang had more mixed results, with some towns stating they spoke Rambang and other towns being unclear as to whether they spoke Rambang or Ogan.

The survey also found that there were differences between what the literature on Ogan considered as dialects of Ogan and what the research indicated. In the PBh book *Ragam dan Dialek Bahasa Ogan*, Rambang is listed as a dialect of Ogan (Wahab et al. 1990:17–18). However, the PBh also listed Penesak, Pegagan, and Pemulutan as dialects of Ogan, which this report disputes. Perversely, that publication's map of the Ogan dialect area does not include the

[66] One example of the subjectivity of these responses is that respondents from Jukung stated that "Musi" is hard to understand but reports the variety of speech in a Musi Proper speaking area as easily understood. The same respondents also reported using *Palembang Pasar* to speak with people from other Col areas, though they say they understand them easily.

[67] Mitani also seemed to weigh placing Ogan in his proposed Highland Malay (1980:15).

Pemulutan area and only the southern tip of the Pegagan area around Tanjung Raja, but does include the Penesak area around Tanjung Batu along with Rambang area (p.16). The PBh book *Morfologi dan Sintaksis Bahasa Ogan* also considers Pegagan to be a dialect of Ogan, and uses Pegagan to describe Ogan. The book's author defends this as Pegagan was also used in an earlier PBh study on Ogan, is closer to Palembang (where the author was based), and one of the researchers was a Pegagan speaker, as reasons why Pegagan was chosen to represent the Ogan language (Arifin et al. 1984:7–8). However, significant phonological innovations found in Pegagan are not found in Ogan, and vice versa.

The Ogan village closest to the Rambang area and the furthest downstream is the Ogan language community of Rantau Alai. The former village head helped explain the relation between Ogan and Rambang. He stated that the Rambang spoken nearby is almost the same (Penyandingan) while that spoken farthest away in Rambang Dangku is different but still comprehensible. The Enim were also regarded as being part of the Ogan ethnic group. The nearby Penesak and Pegagan, who had been mentioned as being Ogan by Mitani[68] and PBh publications, were seen as having different traditions, culture, and language. However, when asked to compare which was closer, these Ogan speakers said Pegagan seemed closer to their language than Penesak. The sites visited in the OGANIC cluster usually reported comprehension of most of the Malayic lects they were asked about, but if asked to give degrees of comprehension, rated those within the cluster as more comprehensible.

The OGANIC cluster itself serves as a lesson in the importance of collecting data along the length of a river system, as it illustrates dialect chaining well. If one data point upstream or downstream is selected as representing 'Ogan', the linguistic and sociolinguistic information obtained will not give the full picture for the dialect grouping. The Rantau Alai variety of Ogan saw itself as virtually the same as Rambang (confirmed by high cognacy rates), the midstream Pengaringan variety similar to Rambang and Enim (confirmed by high cognacy rates with nearby Enim and Rambang research sites), and the upstream Belandang variety felt Enim was similar (confirmed by the cognacy rates of the three Enim data points Enim). They all reported comprehension with Enim and Rambang. Enim also reported comprehension of Ogan and Rambang. The upstream Enim area saw Semenda as also being part of the Enim, Ogan, Rambang language grouping; not surprising given the frequent contact the upstream Enim have with Semenda.

Of the lects found in the area of LOWLAND subcluster and the OGANIC cluster, only Padang Bindu (Benakat) retained the voiced uvular fricative ʁ as the reflex. Benakat was ultimately placed in the HIGHLAND clustering, which also has the same reflex of **r*. Interestingly, some villagers from Padang Bindu town stated that their language had come from Enim. From a lexical standpoint, they are more similar with the OGANIC cluster than any other cluster. However, their village leader said that they had come from the Javanese. The origin legend for the Benakat, recorded in the PBh booklet on Benakat, mentions they came from upstream of the Lematang River and settled there (Arifin et al. 2001:8–9). This origins story fits well with the evidence for the Benakat as a HIGHLAND lect. See Table 7.2.

[68] Mitani refers to the Pegagan Ulu as possibly being part of a proposed Ogan grouping. A Rambang village visited in OKI, Penyandingan, said the name of one of the clans (*marga*) that marked their area is marga Pegagan Ulu I. For the purposes of this report, Pegagan refers to a speech community that refers to itself as Pegagan and shares the Pegagan innovations (ex. **ri>ray*).

Table 7.2 OGANIC Cluster SLQ results for comprehension

Research Site	Ogan (O)	Enim (E)	Rambang (R)	BARISAN	UPPER MUSI/ PALEMBANG
Pengaringan (O)	Same/ Similar	Similar	Similar		
Rantau Alai (O)	Same	Same	Same/ Similar	Similar	Similar
Belandang[69] (O)	Same	Similar	Similar	Similar	Similar
Muara Emil (E)	Similar	Same	Similar/ Different	Similar	Different
Indramayu (E)	Similar	Same/ Similar	Similar/ Different	Similar/ Different	Different
Tanjung Enim (E)	Similar	Same	Similar	Similar/ Different	Different
Karangan Bindu (R)			Same		
Tambangan Rambang (R)	Similar	Similar	Same	Similar	Similar/ Different
Jemenang (R)		Same	Same	Similar/ Different	Different
Penyandingan (R)	Similar	Similar	Same	Similar	
Padang Bindu (Benakat)	Different	Similar	Similar	Similar[70]	Different

The idea of inherent comprehension and ethnolinguistic self-identity were important in determining the composition of the HIGHLAND cluster, which spans the length of the Bukit Barisan range from northern Lampung to the border of Jambi and West Sumatra. The heartland of the HIGHLAND cluster is in Lahat and Pagar Alam, with the lects of Besemah, Semenda, Lintang, and Lematang Ulu all spoken here. Together they report a very high rate of intercomprehension, with little reported difficulty in speaking with one another.[71] According to the author of *Struktur Bahasa Besemah*, "Besides Besemah is another language that is the same as Besemah, but given another name by the inhabitants in that place, that is, Semende, which is spoken in Lematang Ilir Ogan Tengah (LIOT)" (Saleh, Majid & Diem 1977:8). The Malayic varieties of Bengkulu and Serawai were also included on the basis of comprehension and a shared identity with the HIGHLAND cluster. This was also the case with Kaur and Pekal, though to a lesser extent.

The research point furthest east in this cluster is PB-BEN, Padang Bindu. Prior to analysis of the wordlists and SLQ, it was originally included in the LOWLAND subcluster dialects as it is bordered by these. It also borders the Rambang language area. Their autoglottonym is Benakat, which is also the name of the subdistrict (kecamatan). They said they understood Enim, Rambang, Ogan, Lematang, and Besemah with ease but did not understand the nearby LOWLAND subcluster dialect Belide. This combined with lexicostatistic evidence and their phonology of having a voiced velar or uvular reflex for **r* (rather than the voiceless found in the Ogan lects) helped in making the decision to place Padang Bindu in HIGHLAND.

(Basilectal) Bengkulu is included in this cluster based on self-identification with HIGHLAND varieties in apposition to the MINANGKABAU varieties to the north and the intercomprehension they share with the Serawai to their south. Bengkulu appears to fit well in

[69] Balandang and Rantau Alai didn't think there were any Malayic dialects in South Sumatra that were difficult to understand.

[70] Stated that Semenda was similar enough to easily understand.

[71] See W.A. Collins (1979:Chapter One) for an excellent description of the Besemah language cluster and the other names that members of the Besemah clans use to describe themselves.

the HIGHLAND cluster based on identity and understanding. Like the everyday language spoken in Palembang Pasar, it has some unique lexical items (in this case from Padang) but adopts the phonology and much of the lexicon of the acrolect Indonesian. The speakers interviewed in Bengkulu identified more with and understood better the Malayic varieties to the south, Besemah and Serawai, than they did the more distant MINANGKABAU influenced varieties of Muko-Muko and Pekal. The mesolect of Bengkulu Indonesian is also well understood throughout the province, but this may relate more to the widespread use of Indonesian in media and education than indicating the 'centrality' of Bengkulu as a speech dialect.

Kaur was included based on the Kaur respondents reported understanding of Besemah and their performance on the Besemah RTT showing comprehension. Kaur reported that other Malayic groups in Bengkulu Province have a hard time understanding them. The interviews with Serawai, Bengkulu, and Besemah Malayic speakers confirmed this, with all saying Kaur was difficult to understand. The Serawai speakers interviewed said that the Kaur could understand Serawai fine, but that they couldn't understand Kaur. The Besemah speakers said Kaur was marginally understandable. The Kaur see themselves as part of the Malayic languages of Bengkulu, though they acknowledge the difficulty others have understanding them. While Kaur has lower levels of lexical similarity with other varieties than more central dialects of HIGHLAND, this is to be expected as the southernmost of the HIGHLAND varieties.

Similar to the case of Col, the Kaur were said to regularly shorten words and phrases, making comprehension more difficult than just looking at the root lexical form might indicate. This asymmetrical comprehension may point to a dialect chain from Bengkulu to Lampung, where those in Bengkulu and Serawai can not understand Kaur, and those who speak Besemah can understand Kaur with difficulty.

Pekal reported understanding Bengkulu and demonstrated comprehension on the Besemah RTT, so it was placed in the HIGHLAND cluster. Pekal is one of the Malayic lects of northern Bengkulu Province that has been influenced by MINANGKABAU. However, it has not been influenced to the extent of the neighboring lect, Muko-Muko. The speakers of Pekal near the coast that the researchers spoke with had a very clear ethnic identity as Pekal. They said they were different than the Muko-Muko and that their language was substantially different. A Muko-Muko speaker strongly objected to the Pekal claiming they were of different ethnicity, stating both were branches from the MINANGKABAU empire that had stretched down to Bengkulu city, but that there was a Pekal identity movement, driven by political factors. He claimed his own genealogical investigations have not found any Pekal who have family from southern Bengkulu, but instead all are from the Muko-Muko area. Somewhat negating his argument for a common ethnic identity between Pekal and Muko-Muko, he allowed that their language and traditions are quite different. These coastal Pekal reported that they follow the traditions of the Malayic speaking peoples to the south, while the Muko-Muko follow the MINANGKABAU marriage and inheritance patterns. Other Muko-Muko speakers questioned confirmed this. The Pekal stated they understood Bengkulu better and that they felt they were closer linguistically to Bahasa Bengkulu, but some also said that they can understand Muko-Muko and 'Bahasa Padang'. The Pekal community visited in the mountains had a more mixed identity, with some saying they were a mixture of Rejang and MINANGKABAU, others saying they were related to the neighboring Rejang, and others saying they were Pekal, but not related to MINANGKABAU or Rejang. These Pekal speakers demonstrated comprehension of Besemah. The Pekal are placed in the HIGHLAND cluster as the northernmost dialect, but it is important to note the extent to which they have been influenced by the nearby Rejang and MINANGKABAU languages.

One factor that may be coloring the reports from both groups has been the tension and unrest between the groups that took place in 2000 over an election. Fisherman from Muko-Muko reported that their boats were burnt while docking in Pekal ports. The Pekal questioned concerning the tensions felt things were already fine, while the Muko-Muko still seemed to harbor some animus towards the Pekal.

Muko-Muko was considered as well but is not included in the HIGHLAND cluster due to evidence that contradicts them entering the grouping. Based on phonological innovations,

relatively low lexical similarity, and a strong ethnic identity as MINANGKABAU, Muko-Muko stands outside of the BARISAN language and fits in best as a variety of MINANGKABAU. The Muko-Muko speakers interviewed by the researchers stated they understood 'Padang' better than they understood the more basilectal varieties of MINANGKABAU, likely indicating both inherent intelligibility and acquired comprehension based on exposure to the central MINANGKABAU dialect. See Table 7.3.

Table 7.3 HIGHLAND Cluster SLQ results for comprehension

Research Site	**Lematang/ Lintang/ Besemah/ Kikim (B)**	**Semenda (S)**	**Serawai**	**Bengkulu**	**Kaur**	**Pekal**
Bintuhan (Kaur)	Similar		Similar	Similar	Same	
Jembatan Dua (Kaur)	Similar				Same	
Keduran (B)	Similar/ Same	Similar	Similar	Similar	Different	Different
Napal Melintang (Serawai)	Similar	Similar	Same	Similar	Different	
Bengkulu	Similar	Similar	Similar	Same	Different	Different
Pulau Baru (Pekal)			Different	Similar		Same
Napal Putih (Pekal)				Similar		Same

Research Site	**Bengkulu**	**Pekal**	**Muko-Muko**	**MINANGKABAU**
Bengkulu	Same	Different	Different	Different
Pulau Baru (Pekal)	Similar	Same	Different	(Padang-Similar)
Napal Putih (Pekal)	Similar	Same	Similar	Similar
Suka Pindah (Muko-Muko)	Similar	Different	Same	Similar

Table 7.3 HIGHLAND Cluster SLQ results for comprehension continued

Research Site	Lematang /Lintang	Besemah/ Kikim (B)	Semenda (S)	Ogan	Enim	Rambang
(These data sites answered an SLQ focusing on dialects of South Sumatra)						
Muara Dua (S)	Same/ Similar	Same/ Similar	Same	Similar	Similar	
Karya Nyata (S)	Similar	Same/ Similar	Same	Similar		
Muara Sindang Tengah (S)	Similar	Same	Same		Similar	
Sukaraja 1 (Muara Dua Kisam) (B)	Similar	Same	Similar	Similar	Similar	
Pematang Bango (B)	Similar	Same	Similar			
Muara Sindang (B)	Similar	Same	Similar	Similar	Similar	Similar
Sukaraja Kikim (B)	Similar	Same	Similar			
Bata Galang (Lintang)	Same	Similar	Similar			
Arahan (L)	Same	Similar	Similar	Similar	Similar	

7.2 Language use

In all areas surveyed, use of the local SSML lect (L1) appeared to be stable, reflecting a diglossic situation where L1 is used in informal domains, with increasing use of L2 for more formal domains and in areas where there is contact with outsiders. The many PBh publications on the local lects of southern Sumatra rarely go in depth on language use, but do state that L1 is used for spoken communication in informal situations. Indonesian is reserved for formal occasions, with intervening events on the cline of formality show a mixture of L1 with Indonesian. For speech between people of different ethnic backgrounds in South Sumatra, Palembang Pasar is normally used for informal communication, with the use of Indonesian increasing as one goes higher on the formality cline.

The PBh publication *Kedudukan dan fungsi bahasa Palembang* is an exception to other PBh literature as it addresses language use with a thorough 110 question questionnaire administered to different segments of society in Palembang (apparently chosen, non-random sampling) to look at the use of Palembang (the study apparently did not specify which 'Palembang' variant) between ethnic groups as well as within native speakers (Arif et al. 1981). However, it is based on research from 1980, so the results, while useful as a snapshot of language use at that time, may not reflect the current reality. Presented below is one of the charts (Table 7.4) which shows the widespread use of Palembang as a LWC as well as the increasing use of Indonesian with formality.

Table 7.4 Language use in and between ethnic groups (Arif, et al. 1981: from table 23, p. 41)

Group	type of speech	Palembang	Indonesian	Other
inside ethnic group	Formal	77%	22%	1%
	informal	85%	13%	2%
between ethnic groups	Formal	25%	75%	0%
	informal	37%	59%	4%

Shifting to the current study, our research question for language use is:
People use what language(s) in what domains?
The four responses commonly encountered are given in this key:
BI=Indonesian, L1=local Malayic variety, A=Arab, PP= Palembang Pasar, C=mixture

Cell Shading Key	
	BI mentioned as dominant language
	PP mentioned as dominant language (if not L1)
	L1 mentioned as dominant language
	A(rab) mentioned as dominant language
	C Mixture of L1 and LWC/Indonesian[72]
	No dominant answer

In what domains does L1 use occur or dominate?
A series of questions was asked during the group sociolinguistics questionnaire concerning which language is usually used in each of 18 different situations, which are grouped under the domains of domestic affairs, education, public, and religion. Respondents typically listed all languages used within each situation and then were asked to specify which language was dominant. Again, these respondents do not represent a random sample but are a selected sample.

For the MUSI language, L1 is used in at least some situations within all domains in all villages. It is the dominant language for the home domain in all dialect groupings of MUSI in speaking in the home, with neighbors, in the workplace, and at the daily market. If response groups from villages were exceptions to the broader situation for its dialect grouping, this is enumerated in the parentheses for the group. In the more public portions of the domestic sphere (the weekly market and the clinic), there was less consensus, with Rawas not showing a clearly dominant language in the domain of the weekly market, and with Rawas and LOWLAND subcluster not having a dominant language in the domain of the clinic. In Col, the Palembang Pasar language was dominant in the weekly market.

For the BARISAN language, L1 is dominant in the domestic domains, but less so when it involves interactions with people from outside the community such as at the local clinic and at the market. The dialects with the strongest use of L1 were Enim, Ogan, and Lintang of the HIGHLAND cluster. For Rambang of the OGANIC cluster and for the non-Lintang dialects of the HIGHLAND cluster, L1 use was mixed with use of Indonesian or a mixture of L1 and an LWC when speaking with outsiders.

For Kubu, the results showed L1 dominant in the home but less dominant in the other domestic spheres. Only one Kubu speech community was visited in South Sumatra as the majority of this group is located in Jambi. In the community visited, only 25% of the village population were Kubu with the remainder of being Rawas speakers. This explains why Kubu was not used as much outside of working in the fields or speaking at home. However, it is likely that in any place where there would be a market or clinic the Kubu speakers will be a minority language community.

Tables 7.5, 7.6, 7.7. 7.8, 7.9, and 7.10 summarize various domains of reported language use.

[72] For Religious Domain, C= combination of Arab and Indonesian

Table 7.5 Reported language use by dialect in the domestic domain

		Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)							
Language		**MUSI**						**Kubu**	**MINANGKABAU**
SubCluster /Dialect		Musi Proper (8)	Rawas (3)	Pegagan (1)	PALEMBANG (14)	LOWLAND (9)	Col (5)	Kubu (1)	Muko-Muko (1)
Domestic	In home					(2 C)[73]			
	w/ neighbors	(1 PP)[74]			(3 C)[75]	(3 C)[76]		RAWAS	
	Workplace				(1 C)[77]	(2 C)[78]			
	Daily market	(2 C)			(2 C)	(2 C)	No daily market	RAWAS	No daily market
	Weekly market	(3 C)	1 L1/1 PP[79]		(2 C)	(1 PP/2 C)[80]	(1 C)	RAWAS	
	Clinic	3 L1/1 BI/ 4 C	1 L1/1 BI/1 C[81]		(2 BI/1 C)	3 L1/3 BI/3 C	2 L1/ 1 BI/ 1 C		

[73] Two exceptions in the LOWLAND subcluster, with Cambai (Belide) and Tanjung Batu (Penesak) saying that they use a mixture of L1 and Palembang Pasar in the home and with neighbors.

[74] Bingin Teluk was the one exception, with the respondent (in this case one person) saying he used PP to speak with his neighbors.

[75] Three exceptions to speaking L1 solely with neighbors: one Palembang City group and one group of Gandus correspondents said they used both L1 and Indonesian; one Palembang City group stated they used both Javanese and Palembang (this was a group of Javanese families).

[76] Three exceptions to speaking L1 solely with neigbors: Cambai, Purun, and Tanjung Batu said they use a combination of L1 and PP.

[77] One Palembang City group uses both Indonesian and L1 at work.

[78] Two exceptions in the LOWLAND subcluster, with Cambai (Belide) and Tanjung Batu (Penesak) saying that they use a mixture of L1 and Palembang Pasar in the work place.

[79] Rupit group did not supply an answer, Pangkalan said L1, and Pasar Surulangun said PP is used in the weekly market.

[80] Cambai and Modong report a mixture of L1 and PP at weekly market, while Tanjung Batu reports use of PP.

[81] Rupit reports use of Indonesian, Pangkalan L1, and Pasar Surulangun L1 and Indonesian.

		Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)							
Language		**SOUTH BARISAN MALAY**							
Dialect		Ogan/ Rambang (5)	Enim (3)	Lintang (3)	Besemah/ Lematang Ulu (6)	Semenda (4)	Serawai/ Bengkulu (2)	Kaur (1)	Pekal (1)
Domestic	In home								
	w/ neighbors								
	Workplace	(1 C)			3 L1/ 2 C	(1 C)			
	Daily market	2 L1/ 2 C		No daily market					
	Weekly market	2 L1/ 2 C			3 L1/ 2 BI	1 L1/2 C			
	Clinic	2 L1/2 BI/1 C			3 L1/2 BI	2 L1/ 1 BI/ 1 C	1 L1/1 BI		

In the education domain of the MUSI language speaking areas, the local MUSI dialects were dominant in use during informal times of rest at the school, with a mixture of Indonesian and the local languages being used in the instruction period in the earliest level of school by the teachers. The only dialect where this was an exception was the Pegagan community which said that Indonesian was used from the very first grade.

Kubu results for the education domain also closely follow this pattern of using L1 for the informal interaction with peers and use of a mixture of languages or of Indonesian for the more formal classroom instruction time, even at the low level of first grade. Muko-Muko differed, with L1 and Indonesian being used both in instruction and at break times. This may be from the fact that there are people from other areas such as West Sumatra in the school.

Table 7.6, Reported language use by dialect in the education domain

Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)									
Language		**MUSI**						**Kubu**	**MINANGKABAU**
SubCluster /Dialect		Musi Proper (8)	Rawas (3)	Pegagan (1)	PALEMBANG (14)	LOWLAND subcluster (9)	Col (4)	Kubu (1)	Muko-Muko (1)
Education	Used by students during school breaks	(1 C)	2 L1/1 BI		(3 BI)				
	Used by first grade teacher to explain material	2 L1/2 BI/ 4 C	1 BI/2 C		2 L1/7 BI/5 C	3 L1/4 BI/ 2 C	1 L1/1 BI/ 2 C		

Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)									
Language		**SOUTH BARISAN MALAY**							
Dialect		Ogan/ Rambang (5)	Enim (3)	Lintang (3)	Besemah/ Lematang Ulu (6)	Semenda (4)	Serawai/ Bengkulu (2)	Kaur (1)	Pekal (1)
Education	Used by students during school breaks				(2 C)	(2 BI/2 C)			
	Used by first grade teacher to explain material	(1 BI/1 C)		2 BI/1 C	(1 L1/1 BI)	(1 BI)	1 BI/ 1 C		

In BARISAN, the local lect was dominant at break times, with the exception of Semenda and Kaur. The class room instruction time shows a wider variation in what the language of instruction was, with some villages reporting use of L1, some use of both L1 and Indonesian, and some reporting use of Indonesian solely.

In the public domain of the MUSI language speaking areas, there was not a clear overall dominant language variety, but in the individual dialects there were seen preferences for certain language varieties, but with a clear increase in the use of Indonesian when compared to the domains of domestic and education.

The Kubu results for the public domain also show increased use of Indonesian. As pointed out earlier, the fact that the Kubu community is a minority in their own village means that their language is not used in ANY public context with the rest of the village, but only when it is a Kubu gathering. Muko-Muko maintained use of L1 in all public domains, though leaders' meetings also used Indonesian.

Table 7.7 Reported language use by dialect in the public domain

Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)									
Language		**MUSI**						**Kubu**	**MINANGKABAU**
SubCluster /Dialect		Musi Proper (8)	Rawas (3)	Pegagan (1)	PALEMBANG (14)	LOWLAND subcluster (9)	Col (4)	Kubu (1)	Muko-Muko (1)
Public	Traditional ceremonies	4 L1/2 BI/2 C	(1 Bi)		5 L1/5 BI/ 4 C	2 L1/3 BI/4 C	2 L1/1 BI/1 C		
	Public announcements	3 L1/2 BI/1 C	(1BI)		(3 L1/2 C)	4 L1/ 4 BI	(1 L1)	Rawas	
	Meetings of village leaders	2 L1/2 BI/4 C	(1 L1)		(1 L1/ 3 C)	(3 L1/5 BI/1 C)	(1 C)	Rawas/BI	

Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)									
Language		**SOUTH BARISAN MALAY**							
Dialect		Ogan/ Rambang (5)	Enim (3)	Lintang (3)	Besemah/ Lematang Ulu (6)	Semenda (4)	Serawai/ Bengkulu (2)	Kaur (1)	Pekal (1)
Public	Traditional ceremonies	3 L1/2 BI	(1 BI)	(1 C)	1 L1/2 BI/3 C	2 L1/1 BI/ 1 C			
	Public announcements	2 L1/3 BI	(1 BI)		(2 L1)	(1 L1)	1 BI/1 C		
	Meetings of village leaders	1 L1/2 BI/ 2 C		(1 BI)	(1 C)	1 L1/2 BI/ 1 C	1 L1/1 C		

In BARISAN, Indonesian is also in increasing use in the public domains, as it is the acrolect and carries prestige.

For MUSI, there was a decrease in use of L1 and an increase in use of Indonesian in the more formal domain of religion. Arabic is the written language used in reading the scriptures for most dialects. Note: Pegagan did not have an opportunity to respond to three of the questions in the SLQ regarding language use in the religious domain.

Table 7.8 Reported language use by dialect in the religious domain

Language most often used by village SLQ responses (number of villages) (number of exceptions to response)									
Language		**MUSI**						**Kubu**	**MINANGKABAU**
SubCluster /Dialect		Musi Proper (8)	Rawas (3)	Pegagan (1)	PALEMBANG (14)	LOWLAND subcluster (9)	Col (4)	Kubu (1)	Muko-Muko (1)
Religious	Religious programs	4 L1/ 1 A/3 BI/1 C	(1 C)		3 L1/5 BI/6 C	3 L1/3 BI/3 C	1 L1/1 BI/1 C		Muko-Muko/BI
	Praying w/ others	3 L1/2 A/ 3 BI	1 A/1 BI/1 C	NA	2 L1/1 A/ 6 BI/5 C	2 A/3 BI/4 C	1 L1/2 BI/1 C		
	Praying alone	4 L1/ 2 BI/2 C	1 L1/1 A/1 BI	NA	5 L1/2 A/ 2BI/5 C	3 L1/2 A/2 BI/2 C	1 L1/2 BI/1 C		
	Reading Scripture	(2 C)		NA	(1 BI/2 C)	5 A/1 BI/3 C	(1 C)		
	Announcements at place of worship	(1 C)	(1 L1)		(1 L1/1 BI/ 3 C)	3 L1/5 BI/1 C	(1 C)	RAWAS	
	Sermon	(2 C)	(1 BI)		(1 L1/4 C)	(2 C)	(1 C)		

In the religious domains for Kubu and Muko-Muko, increased use of Indonesian was reported for the contexts given. Interestingly for Kubu they also maintained use of L1 in their own religious programs, while MINANGKABAU had a mixture.

For the BARISAN language, the religious domain showed a decrease in use of L1 and an increase in use of Indonesian, with Arab or Arab/Indonesian translations being used for the Koran. The language of the sermon is often mixed, depending upon the origins of the speaker at the mosque.

Table 7.9 continued, Reported language use by dialect in the religious domain

Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)									
Language		**SOUTH BARISAN MALAY**							
Dialect		Ogan/ Rambang (5)	Enim (3)	Lintang (3)	Besemah/ Lematang Ulu (6)	Semenda (4)	Serawai/ Bengkulu (2)	Kaur (1)	Pekal (1)
Religious	Religious programs	(1 C)	(1 BI)		2 L1/2 BI/2 C	2 L1/1 BI/ 1 C	1 BI/1 C		
	Praying w/ others	1 A/3 BI/1 C		1 A/1 BI	2 A/1 BI/3 C	(1 A)			
	Praying alone	1 L1/2 BI/ 2 C		1 A/ 1 BI	2 L1/1 A/2 BI/1 C				
	Reading Scripture	2 A/1 BI/1 C	(1 A)		(2 C)	(1 BI)	1 A/1 BI		
	Announcements at place of worship			(1 L1)	(2 L1)	(1 C)			
	Sermon	3 BI/2 C			3 BI/3 C	(1 BI)	1 BI/1 C		

In summary, the use of L1 (basilect) decreased with the increase in formality and the use of Indonesian (acrolect) increased. The compiled results for all categories are shown below for each

BI=Indonesian, L1=local Malayic variety, A=Arab, PP=Market Palembang Pasar, C=mixture

Cell Shading Key	
	BI mentioned as dominant language
	PP mentioned as dominant language (if not
	L1 mentioned as dominant language
	A mentioned as dominant language
	C Mixture of L1 and LWC/Indonesian[82]
	No dominant answer

[82] For Religious Domain, C= combination of Arab and Indonesian

Table 7.10 Reported language use by dialect for all domains

Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)									
Language		**MUSI**						**Kubu**	**MINANGKABAU**
SubCluster /Dialect		Musi Proper (8)	Rawas (3)	Pegagan (1)	PALEMBANG (14)	LOWLAND subcluster (9)	Col (5)	Kubu (1)	Muko-Muko (1)
Domestic	In home					(2 C)			
	w/ neighbors	(1 PP)			(3 C)	(3 C)		RAWAS	
	Workplace				(1 C)	(2 C)			
	Daily market	(2 C)			(2 C)	(2 C)	No daily market	RAWAS	No daily market
	Weekly market	(3 C)	1 L1/1 PP		(2 C)	(1 PP/2 C)	(1 C, 1 L1)	RAWAS	
	Clinic	3 L1/1 BI/ 4 C	1 L1/1 BI/1 C		(2 BI/1 C)	3 L1/3 BI/3 C	2 L1/ 1 BI/ 2 C		
Education	Used by students during school breaks	(1 C)	(1 BI)		(3 BI)				
	Used by first grade teacher to explain material	2 L1/2 BI/ 4 C	1 BI/2 C		2 L1/7 BI/5 C	3 L1/4 BI/ 2 C	1 L1/2 BI/ 2 C		
Public	Traditional ceremonies	4 L1/2 BI/2 C	(1 Bi)		5 L1/5 BI/ 4 C	2 L1/3 BI/4 C	2 L1/2 BI/1 C		
	Public announcements	**3 L1/2 BI/1 C**	**(1BI)**		**(3 L1/2 C)**	**4 L1/ 4 BI**	**(1 L1, 1 C)**	Rawas	
	Meetings of village leaders	**2 L1/2 BI/4 C**	**(1 L1)**		**(1 L1/ 3 C)**	**(3 L1/5 BI/1 C)**	**(2 C)**	Rawas/BI	
Religious	Religious programs	**4 L1/ 1 A/3 BI/1 C**	**(1 C)**		**3 L1/5 BI/6 C**	**3 L1/3 BI/3 C**	**1 L1/1 BI/2 C**		Muko-Muko/BI
	Praying w/ others	**3 L1/2 A/ 3 BI**	**1 A/1 BI/1 C**	**NA**	**2 L1/1 A/ 6 BI/5 C**	**2 A/3 BI/4 C**	**1 L1/2 BI/2 C**		
	Praying alone	**4 L1/ 2 BI/2 C**	**1 L1/1 A/1 BI**	**NA**	**5 L1/2 A/ 2BI/5 C**	**3 L1/2 A/2 BI/2 C**	**1 L1/2 BI/2 C**		

	Reading Scripture	(2 C)		NA	(1 BI/2 C)	5 A/1 BI/3 C	(1 C)		
	Announcements at place of worship	(1 C)	(1 L1)		(1 L1/1 BI/ 3 C)	3 L1/5 BI/1 C	(1 C)	RAWAS	
	Sermon	(2 C)	(1 BI)		(1 L1/4 C)	(2 C)	(1 C)		

Language most often used with number of village SLQ responses (number of villages) (number of exceptions to response)									
Language		**SOUTH BARISAN MALAY**							
Dialect		Ogan/ Rambang (5)	Enim (3)	Lintang (3)	Besemah/ Lematang Ulu (6)	Semenda (4)	Serawai/ Bengkulu (2)	Kaur (1)	Pekal (1)
Domestic	In home								
	w/ neighbors								
	Workplace	(1 C)			3 L1/ 2 C	(1 C)			
	Daily market	2 L1/ 2 C		No daily market					
	Weekly market	2 L1/ 2 C			3 L1/ 2 BI	1 L1/2 C			
	Clinic	2 L1/2 BI/1 C			3 L1/2 BI	2 L1/ 1 BI/ 1 C	1 L1/1 BI		
Education	Used by students during school breaks				(2 C)	(2 BI/2 C)			
	Used by first grade teacher to explain material	(1 BI/1 C)		2 BI/1 C	(1 L1/1 BI)	(1 BI)	1 BI/ 1 C		
Public	Traditional ceremonies	3 L1/2 BI	(1 BI)	(1 C)	1 L1/2 BI/3 C	2 L1/1 BI/ 1 C			
	Public announcements	2 L1/3 BI	(1 BI)		(2 L1)	(1 L1)	1 BI/1 C		
	Meetings of village leaders	1 L1/2 BI/ 2 C		(1 BI)	(1 C)	1 L1/2 BI/ 1 C	1 L1/1 C		

Religious	Religious programs	(1 C)	(1 BI)		2 L1/2 BI/2 C	2 L1/1 BI/ 1 C	1 BI/1 C		
	Praying w/ others	1 A/3 BI/1 C		1 A/1 BI	2 A/1 BI/3 C	(1 A)			
	Praying alone	1 L1/2 BI/ 2 C		1 A/ 1 BI	2 L1/1 A/2 BI/1 C				
	Reading Scripture	2 A/1 BI/1 C	(1 A)		(2 C)	(1 BI)	1 A/1 BI		
	Announcements at place of worship			(1 L1)	(2 L1)	(1 C)			
	Sermon	3 BI/2 C			3 BI/3 C	(1 BI)	1 BI/1 C		

7.3 Language change

In looking at language vitality and language use, the surveys also looked for signs of a shift away from stable diglossia. The SLQ asked six questions regarding language use among segments of the population to the respondents representing the village.:

- What language do youths use most here?
- If children speak with their parents, what language do they use?
- Parents speak to their children in what language?
- What language do children use here when they are playing together?
- Are there words that older people use here but that are not used by younger people here? For example?
- Are there words that younger people use here but that are not used by older people here? For example?

The responses to these questions showed strong use of L1 in most villages. This is not to say these necessarily represent a good snapshot of the language situation throughout South Sumatra and Bengkulu, as for the most part these villages were chosen for dialectology purposes, as mentioned earlier. However, it does represent a good portrait of language use in rural areas with homogeneous ethnic populations.

The Kubu population is interesting in that it is the one group that shows a language shift underway, with youths and children speaking the local MUSI variety Rawas rather than Kubu with one another, while communication between children and adults still occurs in Kubu. In the areas with high numbers of immigrants, some of the villages also replied to using Indonesian in these relationship contexts. Bengkulu city residents said that parents sometimes use Indonesian to speak to children. Karya Nyata, with a reported 50% of the population Javanese immigrants, said that parents often use Indonesian to speak to children or mix it with Semenda. One set of Palembang city respondents said that Indonesian is often mixed with the local lect in the cases of youths speaking and parents and children speaking. Some respondents from Balai Agung, near the small city of Sekayu, said that children and youths also use some Indonesian in their speech with their peers. These examples stand out against the overall trend, which shows little language change. See Table 7.11.

BI=Indonesian, L1=local Malayic variety, PP= Palembang Pasar, C=mixture

Cell Shading Key	
	BI mentioned as dominant language
	PP mentioned as dominant language (if not
	L1 mentioned as dominant language
	C Mixture of L1 and LWC/Indonesian
	No dominant answer

Table 7.11 Reported language use by dialect in the relationships domain

		Language most often by _____ when they're talking with _____.			
	(number of villages replying to SLQ)	**youth with youth**	**children with parents**	**parents with adults**	**children with children**
Dialects	Musi Proper (8)				
	Rawas (3)				
	Pegagan (1)				
	PALEMBANG (14)				
	LOWLAND subcluster (9)				
	Besemah/ Lematang Ulu (6)				
	Lintang (3)				
	Bengkulu/ Serawai (2)				
	Semenda (4)				
	Ogan/ Rambang (5)				
	Enim (3)				
	Col (5)				
	Pekal (1)				
	Kaur (1)				
	Muko-Muko (1)				
	Kubu (1)	RAWAS			RAWAS

In answer to the question of whether there were some words used by older people but not by younger people, and vice versa, there were some sites that could give examples of a few words, but the overall consensus was there was little difference in their speech. The most frequent comment was that younger people knew more Indonesian and more '*bahasa TV*' or '*bahasa Jakarta*' (TV language and Jakartan slang). They said the young people picked this up from media from Jakarta, primarily television.

7.4 Language maintenance

The section of the SLQ dealing with language maintenance asked the respondents 11 questions. The questions deal with who moved in to the community and for what reason, who left the community and for what reason, what language was used between immigrants and residents, whether immigrants learned the local lect, what languages were taught in the educational system, and about bilingualism.

According to the group responses, immigrants to their area (from western Indonesia in general, with the largest ethnic group being the Javanese) move to their areas because of transmigration (government resettlement policy), looking for work, or because of marriage to a local resident. According to respondents, these immigrants reportedly usually learn to speak the local lect. Those who do not speak the local variety use either Indonesian or Palembang Pasar to communicate with local residents. In response to the question of whether the immigrants learned the local language variety, the respondents answered yes in 88% of the cases.

Residents reportedly frequently work outside the local areas (merantau) where they use Indonesian or another language variety.

In regard to prohibitions against marriage with someone from a different faith, language, or tradition, only marriage with someone of a different faith was consistently taboo, with marriage with different ethnic groups and language groups posing little or no cultural problem.

When asked what language(s) should be taught in the schools, most residents believe that Indonesian and English should be taught in the schools, with many of these also opining that Arab be taught as the language of religion. A smaller percentage felt that the local language should be taught in the schools. In answer to the question "Is the local language taught in the schools?", only about 20% of villages responded yes. The composition of the courses in the local language is not described, but Table 7.12 shows by dialect grouping what the villages answered on their SLQ.

Table 7.12 Teaching of L1 in school

Cluster	SubCluster /Dialect	# answering yes	% of villages with local language instruction	% Cluster	% language
OGANIC	Ogan/ Rambang (5)	1	20%	13%	24%
	Enim (3)	0	0%		
HIGHLAND	Lintang (3)	1	33%	29%	
	Besemah/ Lematang Ulu (6)	3	50%		
	Semenda (4)	1	25%		
	Serawai/Bengkulu (2)	0	0%		
	Kaur (1)	0	0%		
	Pekal (1)	0	0%		
UPPER MUSI	Musi Proper (8)	0	0%	6%	18%
	Rawas (3)	0	0%		
	Pegagan (1)	0	0%		
	Col (5)	1	20%		
PALEMBANG	PALEMBANG (14)	4	29%	26%	
	LOWLAND (9)	2	22%		
Kubu (1)		0	0%		
Muko-Muko (1)		0	0%		
TOTAL		13	20%		

When asked whether there were people who could not speak Indonesian in the village, nearly 80% of villages that reported that there was a portion of the population that could not use Indonesian. This portion varied in according to the groups' estimates but the size of the population that could not speak a language other than L1 was always less than a quarter of the population.

On the question of Indonesian being used every day in the village, four of the sites surveyed replied in the affirmative, or about 6% of the villages surveyed.

From this it appears that Indonesian is REPORTEDLY understood by most people in the villages surveyed, but not used with great frequency. The draft South Sumatra and Bengkulu Bilingualism Survey report (Im & Simanjuntak 2009) goes into detail as to the actual level of Indonesian ability possessed by the speakers of Malay varieties in these regions. As mentioned earlier, the overall ability in Indonesian of the SSML speakers was measured at a level 2 on the Interagency Language Roundtable (ILR) scale, indicating while it could be used for simple daily needs, it could not be controlled well for more complex tasks.

7.5 Language attitudes

The response groups were asked four questions about language use connected to attitude. To the first question about which language they most enjoyed to use, only two mentioned a lect other than L1. One Enim group mentioned liking to use both Indonesian and Enim, while the Kubu group said they preferred to use Rawas. To the second question of what language is used when angry, all answered that L1 is used, but Kubu added Rawas as a language used as well. To the question of what language is used to when joking or saying something funny, the results were the same, with all other than Kubu using L1, and Kubu using L1 and Rawas. To the question of whether they had ever been embarrassed to use their language, 6 groups replied yes (around 9%), while the rest said no, with many in fact saying they were proud to use their language. Of those who reported being ashamed to use their language, the general reason offered was being regarded as kampungan (a country hick), not being understood, or being made fun of. This section gave the researchers the general impression that L1 remains the language of the emotions, a means of identifying themselves, and a valued means of celebrating their culture. However, the local lect also occupies a lower social status than Indonesian.

7.6 Language and media

The respondents were asked a series of questions on the SLQ about media and their language. The first question asked if people had heard their language in TV, radio, cassettes, or DVDs. The majority of village response groups replied that they had, usually in radio programs and cassette tapes of popular local songs. For Palembang, they also have some TV programs where the local language is used. The most popular response was the show *Gerebek* about police raids on criminals in the area.

When asked if people had seen a book using the local language, 13 groups responded yes (around 20%). The villages replying that they had seen a book in the local language are composed of: one UPPER MUSI site (13%), five PALEMBANG sites (36%), two LOWLAND subcluster sites (22%), one Semenda site (25%), three Besemah sites (50%), and one Serawai site (100%). These books were reported to be about songs, pantun (quatrains of verse), traditions, religion, and history. The Rawas, Pegagan, Bengkulu, Lintang, Ogan, Rambang, Enim, Col, Pekal, Muko-Muko, Kaur and Kubu reported not knowing of any books in their language varieties.

When asked what language is the easiest to understand for books, 29 groups said L1 (51%), 23 groups Indonesian (40%) and 5 groups both (9%). Of those who preferred L1, the most common reasons were that L1 is easier to understand, it is more enjoyable, and because "it's my language". Those who preferred Indonesian stated that they are already accustomed to reading in Indonesian, that L1 is difficult to read (orthography/literacy issues), and that Indonesian is the national and unifying language. Those desiring books in both languages saw diglot books as being helpful for people wanting to read Indonesian better while at the same time identifying with their own ethnolinguistic group. See Table 7.13.

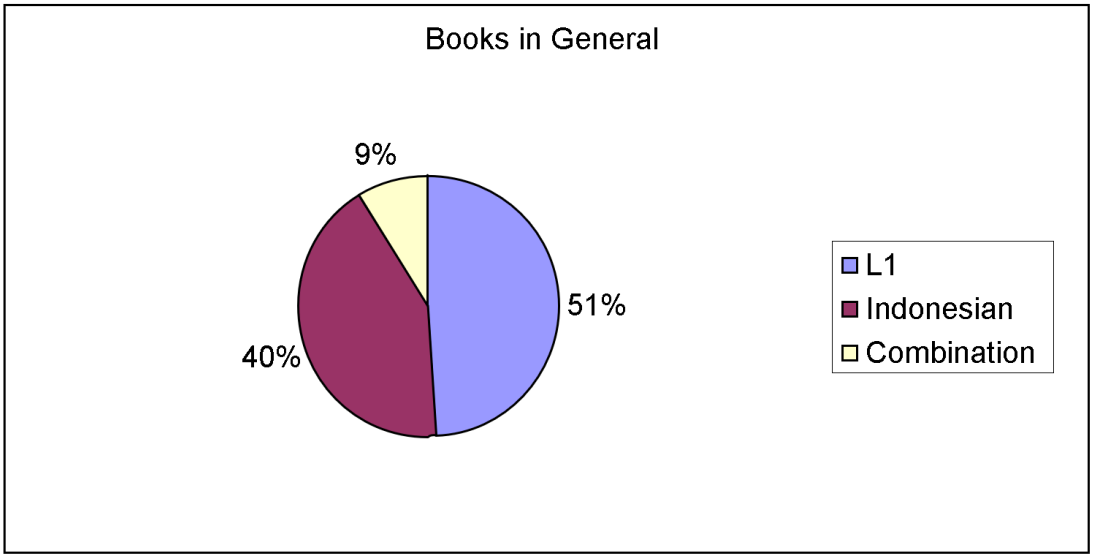
Table 7.13 Preferred language for written media

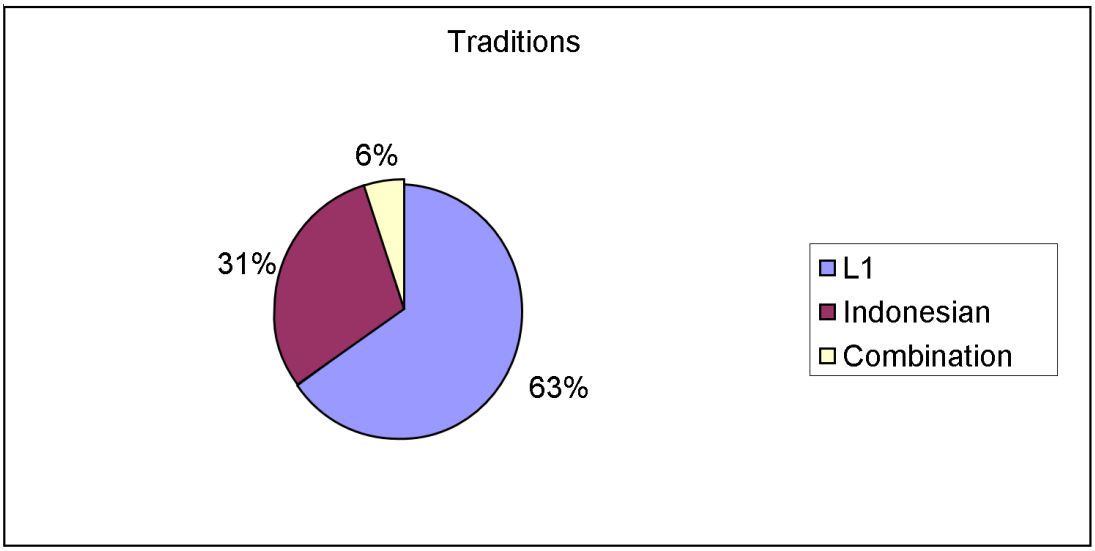
Language for Books on:	% L1	% Indonesian	% Combination	% Arab
Books in General	51	40	9	
Traditions	63	31	6	
Songs	30	36	34	
Stories	39	44	17	
Religion	3	81	13	3
Education	2	98	0	
Health/Farming	20	74	6	

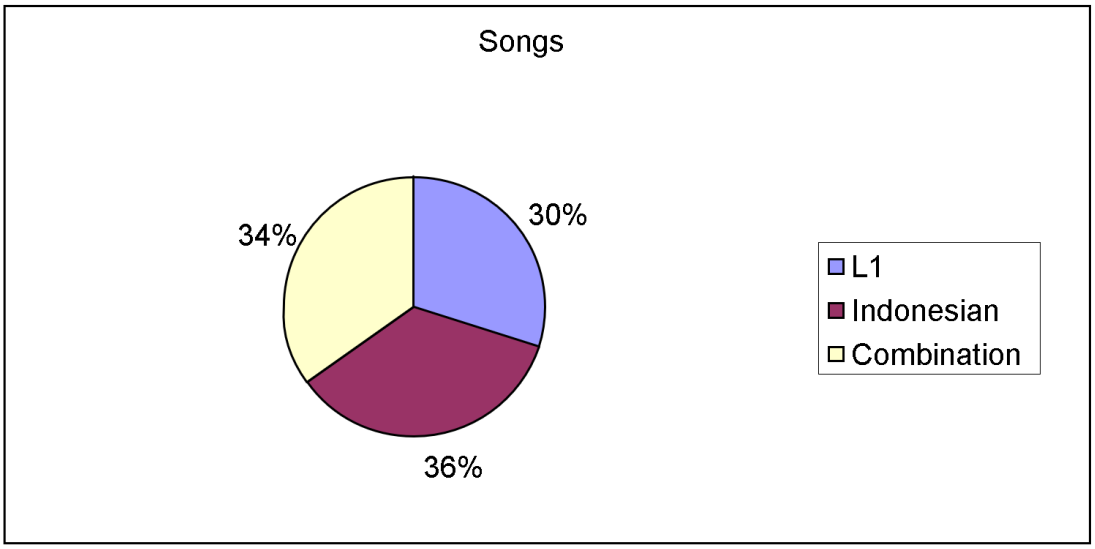
The groups were then asked a series of questions about different types of books and what language they would want the books in. For books about tradition, most preferred having their books use L1. Of the groups that answered, 40 replied they preferred L1 (63%), 19 that they preferred Indonesian (31%), and 4 that they preferred both (6%).

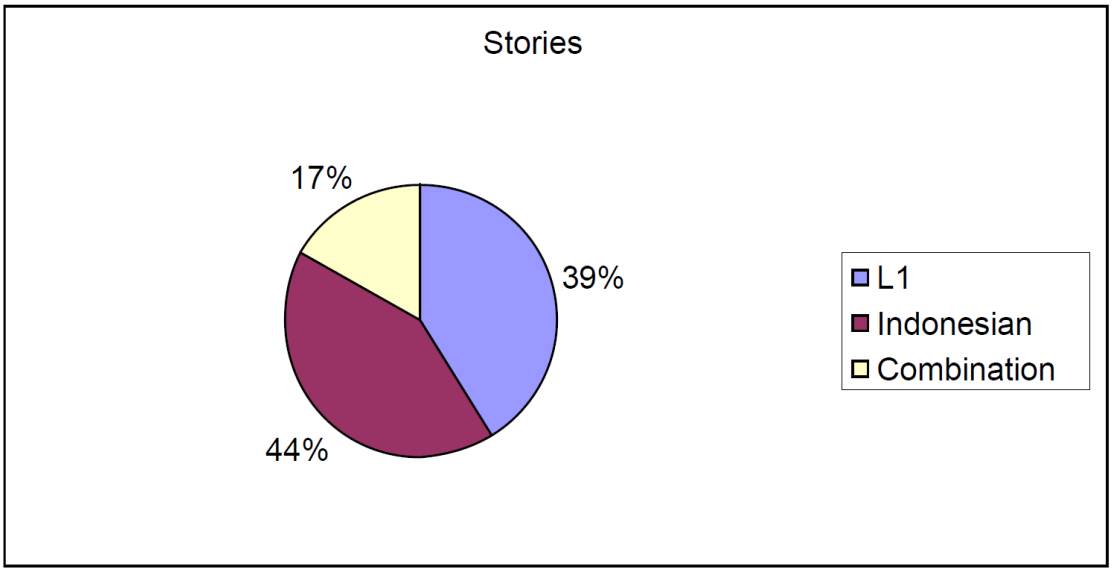
When asked about a book for songs, there was no clear winner, with 22 groups (36%) favoring use of Indonesian, 18 groups (30%) favoring use of L1, and 21 groups (34%) favoring the use of both Indonesian and L1. One frequent comment was that it depended which songs the books was about—if it was about traditional songs, L1 was good, whereas for Indonesian patriotic songs or pop tunes, Indonesian was best.

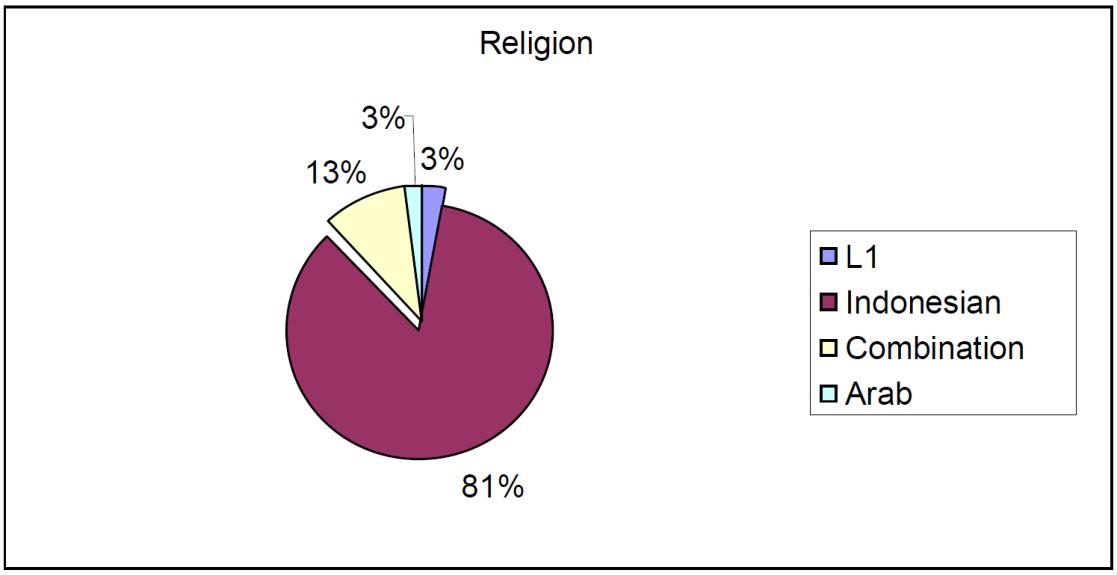
This was true with the question about a book of stories as well, with people saying it depended if the stories were local stories or national literature. Of the groups surveyed, 27 groups (44%) said that a book of stories should be in Indonesian, 24 groups (39%) said that it should be in L1, and 9 groups (17%) said that it should be in both Indonesian and L1. From this it seems that L1 still maintains some respect as means of communicating things touching on cultural identity, in spite of the handicap of not being used much in the educational system and not being used in printed materials much, if at all, in most communities.

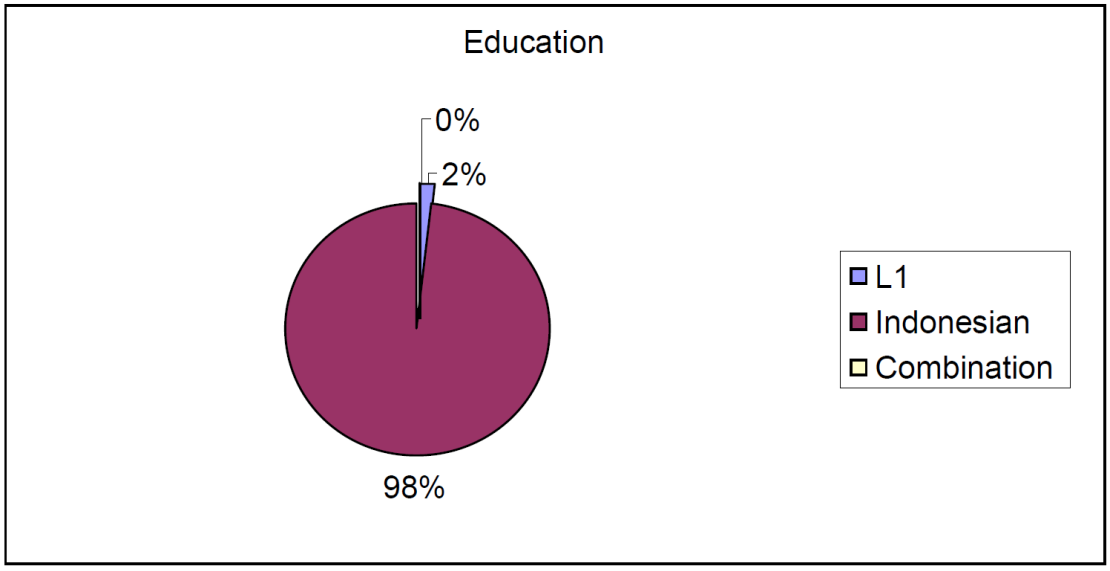

Books in General
9%
51%
40%
L1
Indonesian
Combination

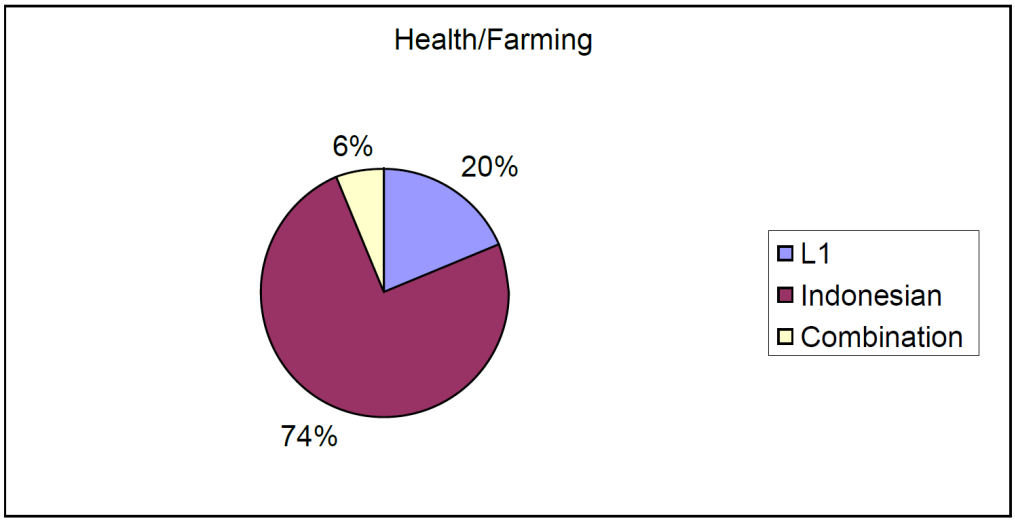

Traditions
6%
31%
63%
L1
Indonesian
Combination


Songs
30%
34%
36%
L1
Indonesian
Combination


Stories
17%
39%
44%
L1
Indonesian
Combination


Religion
3%
3%
13%
81%
L1
Indonesian
Combination
Arab


Education
0%
2%
98%
L1
Indonesian
Combination



When asked about an educational book, only one (Pegagan) out of 63 groups replied that they would want an educational book in L1. This seems to indicate the prestige that Indonesian has as the language of education.

When asked about preferred language for a religious book, the overwhelming response was for Indonesian as well. Two groups replied that they would prefer Arabic for religious materials (3%), two groups preferred L1 (3%), 49 groups preferred Indonesian (81%), and eight groups preferred books with both Arabic and Indonesian, with some of those also wanting L1 (13%). Indonesian is again seen as the fitting language for talking about religion as it fills the role of the acrolect.

When asked about books about practical matters such as farming and health, the response was again overwhelmingly in favor of Indonesian. Of 65 groups that responded, 48 (74%) preferred Indonesian, 13 groups (20%) preferred L1, and 4 (6%) preferred books using both Indonesian and L1. In connection to these topics, a book on Avian Influenza (Flu Burung) was translated into Palembang Pasar and shown in different communities in and around Palembang, as well as some of the Pesisir areas (Matthews & Indarto 2007). This book generally received a good reaction, but some thought that there was not a large benefit in having the book available in the local language rather than just using Indonesian. There were also some small disagreements about how some words should be spelled or which pronouns should be used.[83]

It should be mentioned here that, after our research, a Master's thesis was completed investigating the current uses and future prospects of written Palembang Pasar (Hall 2009). The abstract concludes with the following statement:

> This research demonstrates that people in Palembang choose to use the language of their hearts and homes in print for a variety of reasons, reasons that can best be understood through careful examination of the specific cultural environment in which texts are produced. Some of the reasons established in this study were 1) the desire to build or sustain solidarity and trust, 2) the desire to draw in readership through the shocking rarity of seeing the vernacular in print, and 3) the desire to authentically report the utterances of local people.

8 Summary and conclusions

In this report, we have provided the background and goals of this study, laid out our methodology, and presented analysis of the data gathered. In this final section, we summarize our main (dialectological) findings, compare them with other research, and reflect on some historical implications. We then present our sciolinguistic findings, particularly other, non-dialectological, topics such as language attitudes and use, and conclude with some ideas for further research.

8.1 Dialect clusters

The Malayic lects found in southern Sumatra, when they have been described, appear to have been dealt with in two manners. The first manner is that of 'lumping' where they are described as various dialects of (Sumatran) Malay. One weakness with this approach is that it does not adequately describe the diversity in the

[83] Different varieties of Palembang use different words for the third person plural pronoun.

language varieties or assert any distinct basis for the grouping other than geography. Another approach is that of 'splitting', where the individual lects are treated as distinct languages. This approach is evident in reading the different PBh publications such as "The Lembak Dialect of the Belide Language". The weakness in this approach is that the lect is treated in isolation, without reference to the other varieties around it.[84] Thus the similarities (and differences) between adjacent dialects are not mentioned, and the potential for seeing relations to other varieties is lost.

An intermediate approach, which this report attempts, examines the Malayic lects of southern Sumatra, identifies common and unique features in the varieties, and looks for connections or discontinuities between them. This approach was applied in the past in the description of 'Midden Maleisch' by Dutch ethnologists, and Mitani also applied it to dialects in South Sumatra in delineating a Highland and Lowland Malay based upon lexicostatistics.[85] This section describes the relationship between the dialects using our findings in the areas of sociolinguistics, shared phonological innovations, lexicostatistics and comprehension tests using the Rapid Assessment Recorded Text Test (RA-RTT).

Using the Ethnologue's criteria for defining a language (see §1.5.2), the SLQ was used to probe the reported comprehension between Malayic varieties and dialects within those varieties and ask questions regarding ethnolinguistic identity. The use of lexicostatistic data is compare synchronic lexical similarity as well as to predict poor comprehension on the basis of lexical dissimilarity and provide some confirmation of reports of poor comprehension between different varieties, as well as confirm to some extent reports of good comprehension. The phonological innovations help in grouping dialects that share these common innovations (and therefore often shared history), while comprehension and lexical similarity between varieties is often linked to geographical proximity in a dialect chain.

The dialect clusters proposed include most of the Malayic varieties investigated, but a few remain outside of the proposed clusters due to insufficient evidence or because of significant differences with the varieties within the clusters. Two large clusters are proposed for southern Malayic lects: SOUTH BARISAN MALAY Cluster and MUSI language. These two languages have their own two clusterings which exhibit closer relations. SOUTH BARISAN MALAY is composed of the HIGHLAND and OGANIC clusters, while MUSI is composed of the UPPER MUSI and PALEMBANG-LOWLAND clusters. See Figure 8.1, Figure 8.2, Figure 8.3, Figure 8.4, Figure 8.5, Table 8.1, and Table 8.2 for dialect groupings.

[84] Sometimes there is a map showing adjacent lects or reputed dialects within the treated variety.
[85] See Anderbeck (2008) for a discussion of the inadequacies of this approach for dialectology.

Figure 8.1 Dialect groupings

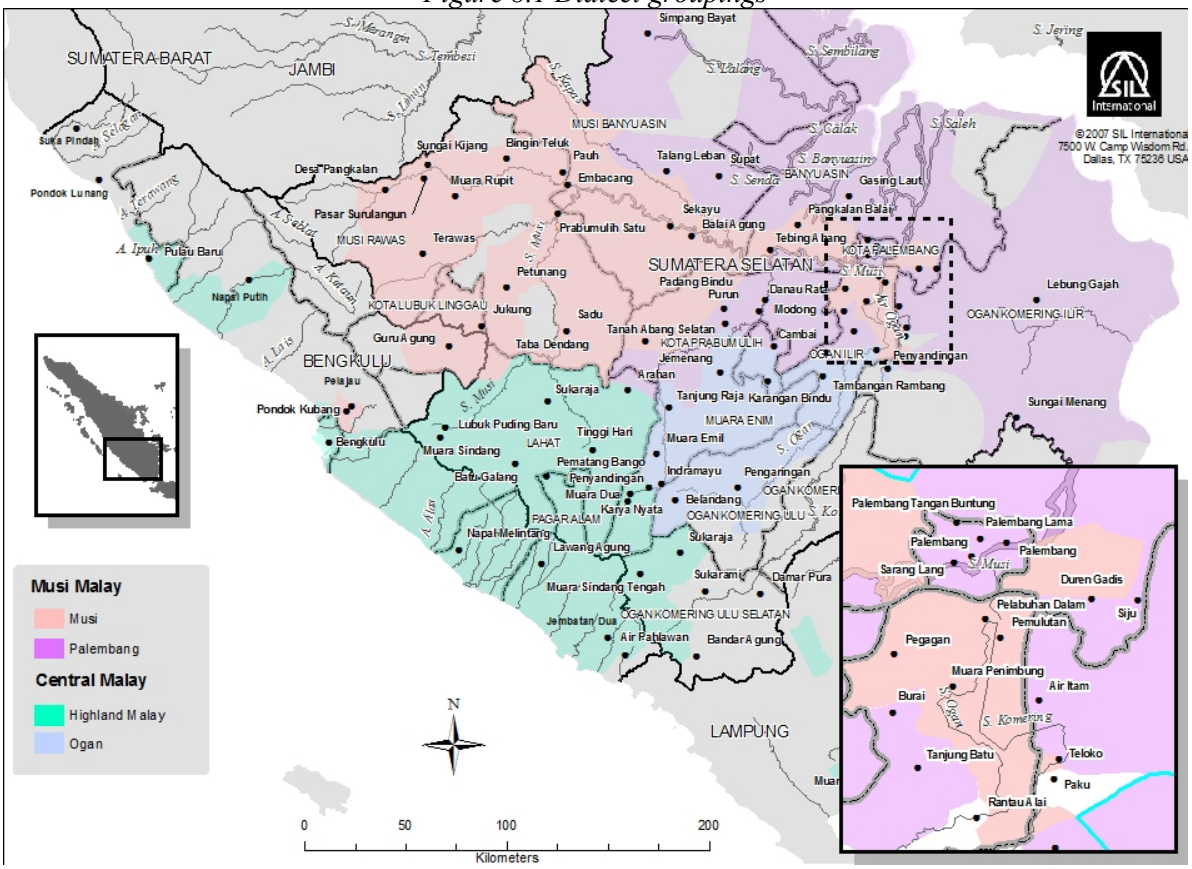


Figure 8.2 HIGHLAND Cluster dialect groupings

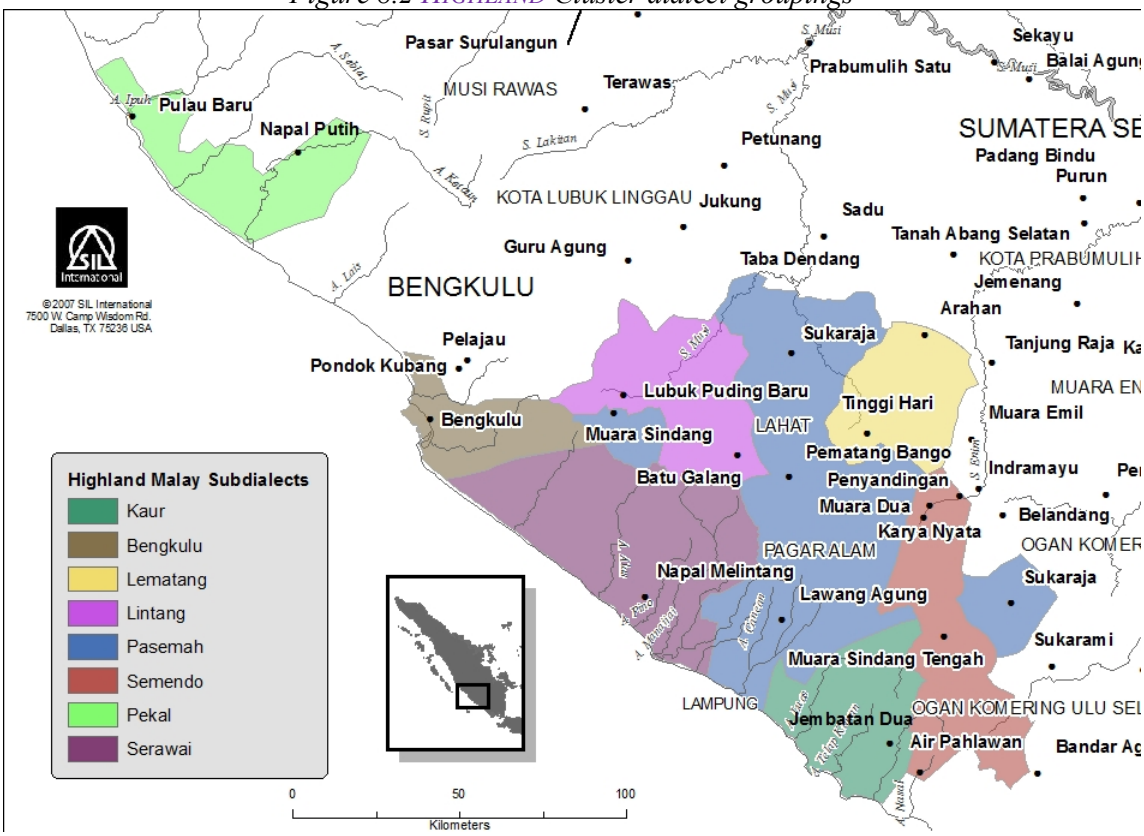


Figure 8.3 OGANIC cluster dialect groupings

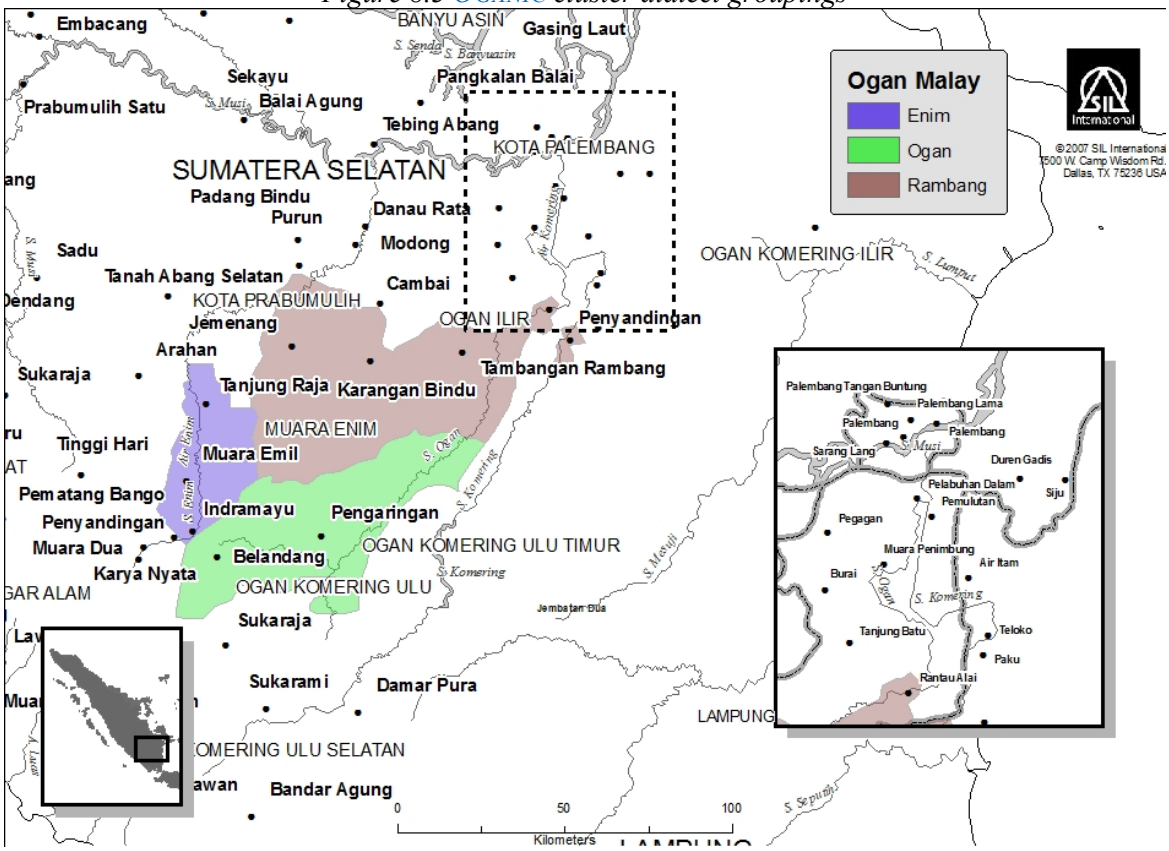


Figure 8.4 UPPER MUSI cluster dialect groupings

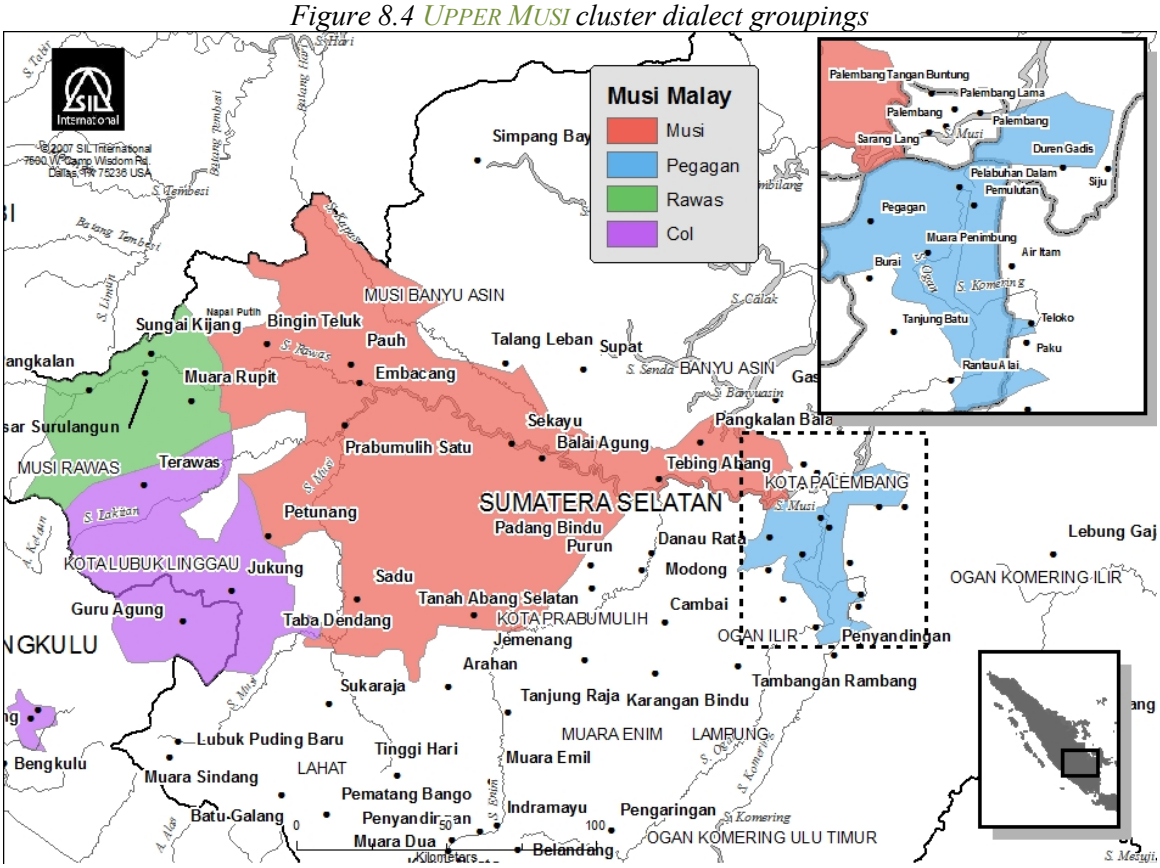


Figure 8.5 PALEMBANG-LOWLAND dialect groupings

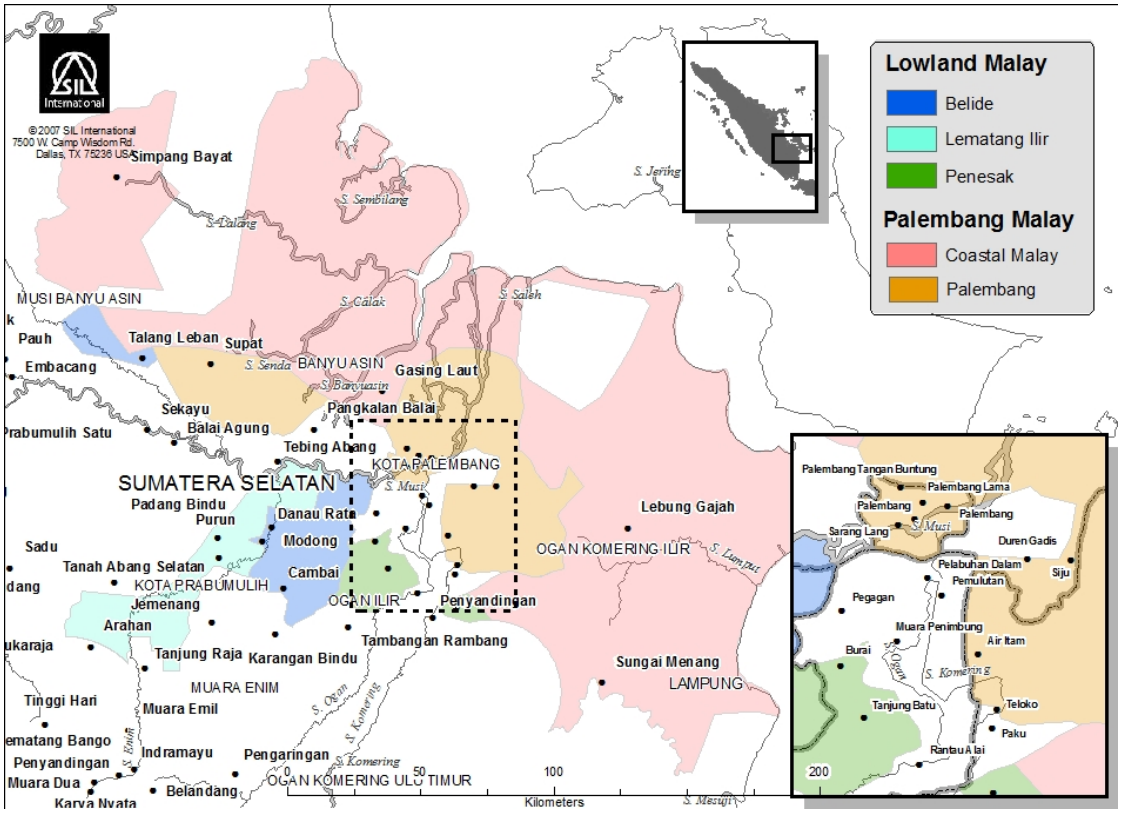


Table 8.1 Dialect groupings (MUSI)

Cluster (Subcluster)	Dialect	Village	Subdialect	Code
Palembang-Lowland (Lowland)	Penesak	Tanjung Batu		PB-TB
		Pedamaran		PB-PDR
		Burai	Burai	PB-BR
	Lem. Ilir	Tanah Abang Selatan		PB-TA
		Danau Rata		PB-SR
	Belide	Modong		PB-MD
		Cambai		PB-CB
		Talang Leban		PB-TL
		Tebing Abang		BEL-TA
(Palembang)	Pal. Lama	Teloko		PL-TEL
		Paku		PL-PAK
		Palembang Lama		PL-PL
		Sarang Lang		PB-SL
		Pemulutan		PB-PM2
		Pelabuhan Dalam		PB-PM
	Pal. Pasar	Palembang City 1		PB-SH1
		Pulau Betung		PB-BET
		Gasing Laut		PB-GL
		Duren Gadis		PB-DG
		Air Itam		PB-AH
		Palembang City 2		PB-SH2
		Tangan Buntung		PB-SH3
		Siju		PB-SJU
	Pesisir	Lebung Gajah		CST-LG
		Pangkalan Balai		CST-DB
		Supat		CST-SP
		Simpang Bayat		CST-SB
		Sungai Menang		CST-SM
Upper Musi	Rawas	Muara Rupit	Rupit	RAW-RU
		Pasar Surulangun	Tengah	RAW-PS
		Pangkalan	Ulu	RAW-PN
	Musi Proper	Petunang	Kelingi	MU-KL
		Sadu		MU-SA
		Prabumulih Satu		MU-P1
		Embacang		MU-EM
		Pauh		MU-PH
		Bingin Teluk		MU-BT
		Sekayu	Sekayu	MU-KY
		Purun	Penukal	MU-PEN
		Balai Agung	Sekayu	MU-BA
	Pegagan	Muara Penimbung		MU-PG
		Pegagan		MU-PG2
	Col	Jukung		COL-LL
		Taba Dendang		COL-TT
		Guru Agung		COL-PUT
				COL-BKL
		Terawas	Bengkulu	
		Pelajau	Lembak	COL-PLJ
		Pondok Kubang	Lembak	COL-L8

Table 8.2 Dialect groupings (SOUTH BARISAN MALAY and others)

Cluster	Dialect	Village	Subdialect	Code
Oganic	Rambang	Penyandingan		RAM-PNY
		Tambangan		
		Rambang		RAM-TR
		Karangan Bindu		RAM-RK
		Jemenang		RAM-RD
	Enim	Tanjung Raja	Ilir	EN-ME
		Muara Emil	Tengah	EN-TAS
		Indramayu	Ulu	EN-TAN
	Ogan	Rantau Alai	Ilir	OG-RA
		Pengaringan	Tengah	OG-BR
		Damar Pura	Ulu	OG-DP
		Belandang	Ulu	OG-UO
Highland	Bengkulu	Bengkulu City		BNGKL
	Pekal	Pulau Baru		PKL-PL
		Napal Putih		PKL-NP
	Lintang	Batu Galang		BES-MP
		Lubuk Puding		
		Baru		BES-UM
		Terusan Baru		BES-TT
	Besemah	Sukaraja Kisam		BES-MDK
		Sukaraja	Kikim	BES-KT
		Pematang Bango		BES-PA
		Muara Sindang		BES-AK
		Lawang Agung		BES-BK
	Lematang Ulu	Tinggi Hari		LT-PP
		Arahan		LT-MR
		Muara Sindang		
	Semenda	Tg.		SEM-PB
		Karya Nyata		SEM-SDL
		Penyandingan		SEM-PNY
		Muara Dua		SEM-MD
		Bandar Agung		SEM-BA
		Tanggamus		SEM-LP
	Benakat	Padang Bindu		PB-BN
	Serawai	Napal Melintang	Manna	SRW-NM
	Kaur	Jembatan Dua	Kaur	KAU-J2
(HAJI)	HAJI	Sukarami		HAJ
(Minang)	Muko-Muko	Pondok Lunang		MUK-PL
(Minang)		Suka Pindah		MUK-SP
(Kubu)	Kubu	Sungai Kijang		KUBU

8.1.1 Comparison with Pusat Bahasa's results

As mentioned in the introduction, large amounts of data in the form of 1000-plus-item wordlists were collected under the auspices of Pusat Bahasa's project to map the languages of Indonesia. In the provinces of South Sumatra and Bangka/Belitung alone, 117 of these lists were collected. (Unfortunately it seems that none of these data will ever be published.) The first two hundred items of the lists were used in dialectometrical calculations to identify 'subdialects', 'dialects', 'languages' and 'language stocks'.

The conclusions for the whole country (including some rather difficult-to-read maps) have been published in a recent volume (Ruskhan et al. 2008) and some additional details for South Sumatra were included in Sudarmanto (2012). Although PBh does not follow externally-recognized language classifications, the publications name six dialectometrical clusters or 'languages' in southern Sumatra which can mostly be considered Malayic, plus a seventh which is mostly Lampungic but includes the Malayic HAJI. We have included the 'dialects' listed for each language, and attempted to incorporate this report's coloring system for the lects mentioned by PBh):

1. **Bengkulu**(-Muko-Muko-Lembak-Nasal-Serawai-Pasemah-Pekal-Kaur)
2. **Basemah**(-Semende-Pegagan)
3. **Pedamaran**
4. **Lematang**(-Pegagan-Lahat-Ujan Mas Lama-Rambutan-Rambang)
5. **Malay**(-Palembang Sukabangun-Kisam-Muara Saling-Selangit-Rupit-Bentayan-Palembang 16 Ulu-Padang Bintu-Talang Ubi
6. **Ogan**(-Musi-Rawas-Col)
7. **Komering**(-Aji)

The fact that two of their 'languages' contain combinations of genetically-unrelated languages (Malayic plus Nasal, Lampungic plus HAJI) does little to inspire confidence in the remainder of the groupings. As is obvious by looking at the colors, there turns out to be no more than slight concordance between our clustering and theirs.

8.1 Miscellaneous thoughts on epicenter and migrations

What can we glean about issues like the homeland of Malayic or past migrations of speakers from the dialectological data presented herein? This could easily be the subject of a full paper (or another chapter of this report), so the following will just be a few musings on the topic. In particular, the question of homelands (whether of Malayic or Malay; see §1.2.3) is too thorny to take up here.

However, the lexicostatistical and phonological data discussed in this report can more easily be brought to bear on the 'concentric circles' hypothesis mooted in §1.2.4. What do we see as regards lexical and phonological diversity in SSML compared to other areas, particularly those between South Sumatra and the hypothesized 'canonical center' of Johor-Riau? One problem with testing this hypothesis is that we do not have at our disposal a standard wordlist from Riau or Johor. What we do have, however, is a Standard Indonesian wordlist. Inasmuch as Indonesian is a descendant of Johor-Riau Court Malay (Sneddon 2003:8), this may provide us with a rough baseline. We first look at lexicon. As noted in §4, the average percentage of shared cognates (PSC) between SOUTH BARISAN MALAY and SI is 73%, while between MUSI and SI the PSC is 78%. In the area, the most lexically aberrant lects are HAJI at 60% PSC and Kaur at 68%. When we compare these percentages to other areas (Table 8.3), we find some support for the hypothesis.

Table 8.3 Percentages of shared cognates with Indonesian, lowest to highest

Lect	PSC with SI	PSC with PM
Duano	56%	59%
HAJI	60%	63%
Kaur	68%	70%
SOUTH BARISAN MALAY	73%	77%
Jambi Ulu	73%	72%
Bangka Malay	75%	78%
MUSI	78%	79%
Jambi Ilir	79%	76%
MINANGKABAU	86%	87%
Kerinci	86%	84%
Kubu	86%	81%
Deli	87%	86%
Bintan Suku Laut	87%	n/a

With the exception of Duano in Riau, for which it has been argued that it may have a non-Malayic substratum (Anderbeck 2013), the geographical pattern is rather straightforward. The farther from the 'epicenter', the more divergent from SI.

The argument from phonology will have to be more impressionistic than precise. (While a more thorough measure like Levenshtein distance has been applied to SSML, it has not to the other lects mentioned above.) The picture that can be painted from phonology is substantially different than that of lexicostatistics. While Duano is also very phonologically aberrant from Proto-Malayic and from the 'epicenter', so is Kerinci, and MINANGKABAU is not terribly far behind. Perhaps the next most phonologically aberrant are the lects on Bangka Island (including Lom), with Pekal, Kubu, Jambi Ulu and Kaur not far behind. We can perhaps conclude from this whirlwind tour of Sumatran Malay phonologies that the 'epicenter effect' can still be seen but more weakly than in the lexicon.

Moving on to the issue of migrations, various hypotheses have been proposed for who has moved where and when. One of the later propositions is that Sumatran Malay is composed of two 'waves' of migrations (Nothofer 1995a; Nothofer 1995b), where Wave One included Palembang Lama and possibly Serawai and other unspecified areas of southern Sumatra, and Wave Two included the rest of Sumatran Malay. This hypothesis is based mostly on putative retentions (including what we here call 'Java schwa'; see §5.10) and one innovation, the loss of final **h* (see §5.1).

How does this wave theory stand up now that a substantially greater data corpus is available? Not very well, it turns out. We will discuss three issues: 'Java schwa', loss of final **h*, and the overall pattern of the SSML dialect network. Regarding the 'Java schwa', we demonstrated in §5.10 that a) this feature (whether innovation or retention) is strictly limited in the SSML corpus to Palembang Lama; and b) it is clearly connected with Jakarta. So even if an argument from retentions is not too problematic, this limits the 'wave one' area to Palembang Lama only. Given the connection to Jakarta, it could easily be argued that, even if Jakarta retains PM **ə*, the existence of this feature in Palembang Lama could be a loan from Jakarta.

Regarding loss of final **h*, Figure 5.3 shows that this innovation is also concentrated around Palembang, although not so sharply as it also appears sporadically in UPPER MUSI lects. It does *not* appear to any significant degree in South Barisan Malay (where Serawai can be found). This innovation is also found in the ancient court area of Jambi Ilir, some MINANGKABAU dialects, a minority of Bangka dialects (part of Nothofer's wave one), Belitung Malay and modern but not 1881 Sekak (Riedel 1881). Loss of final **h* is labeled a 'cosmopolitan' innovation in Anderbeck (2008:81) and we see no reason now to change that evaluation.

Third and finally, we look at the overall pattern indicative of a dialect network in SSML. Anderbeck (2008:87) employs the concept of bundled isoglosses to differentiate inmigrations of speakers from other areas (in this case, MINANGKABAU speakers into Jambi Ulu) from the normal criss-crossing of isoglosses seen in a typical dialect network. So, what sharp breaks, if any, do we see within SSML or between SSML lects and their neighbors, breaks that might indicate a history of migrations? There are a few candidates for 'bundled isoglosses': HAJI, Pekal, UPPER MUSI and Palembang Lama. We will briefly examine each in turn. HAJI turns out to be distinctive from SSML not because of its phonological innovations but because of its archaisms and extensive lexical borrowings from Lampungic (Anderbeck 2007f). Pekal is fairly phonologically distinct, but in such a way that its isoglosses criss-cross with neighboring Muko-Muko, upstream Jambi (Heinzpeter Znoj p.c.), Rejang and the HIGHLAND cluster. Nor does it share any particular configuration of innovations with faraway Malayic lects. The same thing can be said for UPPER MUSI: it has a distinctive set of innovations, but they tend to fuzz out on the edges; plus, a migration of a group that large would surely have left distinctive traces elsewhere. The final candidate with 'bundled isoglosses' is Palembang Lama. What we find there, both with final closed schwa and the numerous Javanese loanwords, is unique among SSML lects. Finally, evidence for migration! Except in this case, the newer lect in the dialectal neighborhood, Palembang Lama, is the one supposedly part of Nothofer's first, not second, wave of migration. We therefore consider the 'two wave migration' hypothesis for Sumatran Malay disproven or at least strongly disfavored.[86]

8.2 Language use

The linguistic landscape of southern Sumatra for Malayic languages exhibits stable diglossia as defined by Ferguson, with the addition of a mesolect.

[86] Adelaar (2004:18–21) explains his rejection of Nothofer's hypothesis in greater detail than is given here.

> "Diglossia is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation" (Ferguson 1959:336).

In the regions of southern Sumatra that were researched, the Malayic lects fill the role of the basilect, being spoken in the informal domains of home and neighborhood. Indonesian fills the role of the acrolect, being used in schools, most media, and religious and government domains. As discussed above, the mesolect for the province of South Sumatra is Palembang Pasar, one of several forms of what local residents call *baso P'lembang* (the Palembang language) coexisting in and around the city Palembang. Palembang Pasar is a second language used by speakers of another primary language (Komering, Javanese, Sundanese, or a different Malayic dialect). Respondents to the SLQ from outside the Palembang area frequently said they spoke Palembang Pasar to visitors from outside their village or area, even if they spoke the same variety of Malayic. For example, one Rawas speaker from the upstream area meeting another from Rupit might use Palembang Pasar to speak with his interlocutor, even though they could reportedly understand one another's lect.[87]

In the government offices visited during the 2006 and 2007 surveys, it was observed that even in the official sphere of government offices, Palembang Indonesian (an acrolectal form of Palembang Pasar) was frequently used after initial introductions were done in Indonesian.[88] In the SLQs administered in South Sumatra Province, Indonesian was indicated as the medium of interethnic communication with only one group, the Javanese, where as with all others, including the non-Malayic Komering and Rejang, Palembang is used.

The languages of wider communication depended upon the region surveyed. In Lampung Province, no mesolect was noted in the small sample of Semenda and HAJI villages. Indonesian is used as the language of wider communication (LWC) in Lampung. In Bengkulu Province, it appears that Indonesian and Bengkulu Indonesian serve as the mesolect, though Padang (a form of MINANGKABAU) is used in the sea port towns of Bengkulu in the commercial sphere as the language of wider communication in addition to the local Malayic dialects.

The extent of the use of the basilect (vernacular) depended on factors of isolation, presence of outsiders, and perceived importance of the occasion. In areas where the market was not frequented by people from different ethnic backgrounds, the basilect was used in the commercial sphere. This was also true for agricultural work. Those villages with larger numbers of inhabitants involved in industries such as energy and mining were usually exposed to people from many different ethnic backgrounds and thus would use the acrolect, Indonesian, or the LWC, Palembang Pasar in the workplace. During traditional ceremonies such as marriages, the greeting is often given in Indonesian to respect guests present, but then moves to the basilect.

The acrolect, Indonesian, is used in the domains of government, religion, and education. The *reported* ability to understand Indonesian from the respondents varies from a low of 50% to a high of 98%. As the question did not ask at what level Indonesian is spoken, no attempt to gauge 'competence' in Indonesian was made based on the SLQ, but other instruments were used. The highest reported percentages for command of Indonesian come from areas near urban centers. For sermons in the mosque and other religious events, the basilect might be used if the teacher was local, but the usual response was that Indonesian was used. The same was true for explaining subject matter in school—Indonesian reportedly is used for the most part, but *if* local teachers are available, they may explain concepts using the local lect in elementary school (SD).

8.2.1 Ability in Standard Indonesian

In order to gauge competency in Standard Indonesian (SI) of the SSML communities in question, a bilingualism survey was conducted by WIST in South Sumatra and Bengkulu in cooperation with a government department in 2008. A complete treatment of the findings is found in a draft report (Im & Simanjuntak 2009). The following is a brief summary of the findings.

---

[87] When asked why they would not use their own respective dialects, the two most common answers were to 'show respect to one's guest' by not forcing the guest to use a different dialect and 'to avoid misunderstanding', as a word might have a different meaning or connotation.

[88] For South Sumatra, Indonesian still is used at the *kabupaten* level, but at the level of the *kecamatan* and village, Palembang dominates.

The survey was designed to test bilingualism of SI in both the BARISAN and MUSI villages. With the criteria that these villages possessed a majority of residents from the language in question, the villages for BARISAN and MUSI were further divided into two categories: those closer to secondary educational facilities (less than 10 km) and those farther from said facilities (greater than 20 km). Two villages for each language were selected randomly in each category, totaling four villages per language. The testing in these villages then used systematic sampling to choose the residences that would be tested, followed by random selection of the individual in the household to be tested. The test used three instruments: the Indonesian Sentence Repetition Test (ISRT) developed by WIST[89], a guided Conversation in a Second Language (CL2), and a self evaluation questionnaire (SEQ). While the other tests were administered, the results are based of the scores of the ISRT (see Im and Simanjuntak for further discussion). The scores from the ISRT (which already has undergone a rigorous testing and modification process) correspond to the language levels standards from the International Language Roundtable, a consortium of entities interested in evaluating second language ability (Interagency Language Roundtable (ILR) n.d.).

The results of the testing in four randomly chosen MUSI sites (two near secondary schools, two far away) and the four randomly chosen BARISAN sites meeting the same criteria gave both groups an ILR level of 2 (see Table 8.4). This indicates 'Limited Working Performance', meaning:

> Unable to transfer information [in SI] reliably in most instances. May communicate some meaning when exchanges are short, involve subject matter that is routine or discourse that is repetitive or predictable, but may typically require repetition or clarification. Expression in the target language is frequently faulty (Interagency Language Roundtable (ILR) n.d.).

Table 8.4 List of randomly chosen sites for bilingualism testing

South Barisan Malay	**Musi**
Penanggiran	Satu Ilir
Lubuk Tuba	Pulau Negara
Air Dingin	Pulau Geronggang
Sukaraja Kikim	Tanjung Bali

While 'limited working performance' was the average level of competency in SI representative for the languages of BARISAN and MUSI, SI competency varied greatly between individuals. The factors that appear to have influenced performance on the test were: frequency of travel outside the area, level of formal education, age, sex, and frequency watching or listening to media in Indonesian.

All these factors are linked to the regularity with which respondents came into contact with Indonesian. The primary factor that determines individual ability in SI is education. In short, those younger, more educated, and who traveled more often or who had more contact with others outside their language scored higher on the test than the test average. It appeared that women scored lower than men due to less frequent travel outside of the area and lower educational levels. The younger generation that scored higher not coincidentally has more educational opportunities than the older generation did and less time has lapsed since they finished their schooling.

The lower functional ability in SI indicates that the primary languages will continue to be the local Malayic dialects for the immediate future, though increasing educational opportunities should result in an increasing ability in SI in the longer term.

8.3 Language maintenance/shift

No signs of language shift were noted in the ethnically homogeneous villages visited. Some signs of language shift were seen in the linguistically diverse city, where marriage to a non-Malayic speaker might result in a shift to Indonesian use in the home, but this seemed rare. In the SLQ interviews, people repeatedly stated a pride in their own language. As noted above, the diglossic situation appears stable, with little use of Indonesian

[89] For more information on the development of the Indonesian Sentence Repetition Test and its correlation with SLOPE scores, Hanawalt's manuscript book chapter "Development of the ISRT" is recommended (Hanawalt 2012).

in the home and informal domains. From all of the survey interviews, only one respondent (in a pre-interview to test his suitability to give a Palembang Pasar wordlist) reported using Indonesian in the home.[90] However, this must be weighted against the fact that the survey team selected villages that had high homogeneity in ethnic composition (with the exception of the cities of Palembang and Bengkulu and the Kubu community).

When asked about different vocabulary used by the older and younger generations, few respondents were able to give specific examples, though one answer frequently given was that younger people were using words they had seen on television, bahasa gaul (teenage slang). The students from SD (elementary school) to SMA (high school) continue to use their local lects with their friends or at home, though the children from village settings often would have to go live in the nearest urban center if they wanted to attend high school. There they become familiar with the LWC, *Palembang Pasar*. In one Pegagan village we observed children just arriving home from high school using Palembang Pasar in speaking to each other. It appears that there is not any danger of extinction of these Malayic dialects, but that there is a likelihood of increasing homogeneity as *Palembang Pasar* is used even in the furthest reaches of the Rawas River.[91]

The areas that appear to have had the greatest amount of language shift are the city of Palembang itself and the towns extending southwest of it in what has been labeled 'LOWLAND subcluster' for the purposes of grouping the lects.[92] These towns have the apical trill for **r* and greater similarity to Standard Indonesian lexicostatistically. They also have a more fragmented and a less defined cultural identity – from what could be obtained during the interviews.[93]

8.4 Further research

A number of areas seem ripe for further research. We will mention a few:

- The status of ISO 639-3 codes in Bengkulu province, namely Col [liw], Kaur [vkk], and Pekal [pel] was not examined as well as those in South Sumatra. The question of the intelligibility and strong distinct identity of these three lects vis-a-vis Musi, Barisan, and Minangkabau respectively should get another look.
- Participatory dialect mapping as well as other participatory tools (Truong & Garcez 2012) would be an excellent emic complement to the largely etic clustering and labels employed in this report.
- Similarly, much more could be done in looking at patterns of interaction including using the techniques of social network analysis (Graham 2000).
- This study has only superficially employed historical and anthropological sources to understand the roots of linguistic relationships. Certainly more could be done in this area.
- 3D dialect modeling using something like Splitstree (www.splitstree.org) could produce some interesting visualizations of the SSML dialect network, as could network and phonostatistic analysis using Gabmap.
- The grammatical aspect of our dialectology study is very shallow; much more investigation could be undertaken, particularly in the use (or lack thereof) of various affixes.
- It would be interesting to follow up on the impressions of linguistic diversity brought forth above in §8.1 with a proper statistical study of relative phonological and lexical diversity between Sumatra and Borneo.

8.5 Summary

The Malayic lects of southern Sumatra carry many labels and possess certain uniqueness lexically, phonetically, and culturally. In addition to these items and features that distinguish them, there are also many

---

[90] In his case, while he and his family grew up in Palembang, his brothers and sisters and married people from different ethnic backgrounds from outside southern Sumatra and so Indonesian was the language of the home. They were all educated at the college level.

[91] Theise researchers noted several times when taking a word list, the informant would give their '*asli*' word for an item, and then also give another commonly used word, adding that the other word had come from Palembang. These communities could still tell which ones were from their own history and which were from Palembang, but we are not sure how long that distinction will last.

[92] See Tadmor, (2001) "The Case of Palembang Malay" where de-Javanization is demonstrated for Palembang City.

[93] This was the area where people were frequently stumped by the question of what language they spoke, and would discuss it for a while or ask the village head or the visiting researcher.

innovations and parts of the lexicon these varieties share. A shared cultural identity, lexicon, comprehension and shared innovations can be found at the dialect level. The CLUSTER level has three of these four variables. An example is LOWLAND subcluster in the PALEMBANG-LOWLAND cluster, where the cultural identity was not clearly defined. The large clusters were based on these same principles, looking at shared cultural history, shared innovations, and reported comprehension, checked against lexical similarity. The clusters that seem very clearly defined by the aforementioned factors are the OGANIC, HIGHLAND, and UPPER MUSI clusters. The PALEMBANG-LOWLAND cluster, because of the different varieties of Palembang spoken, the use of Palembang Pasar as a LWC, is somewhat harder to define, as can be seen in the divergence of innovations and lexicon found in Palembang itself. The larger clusters of MUSI and SOUTH BARISAN MALAY appear defensible from a sociolinguistic perspective considering the respondents' language identity and orientation. The unique cultural background of the Kubu (Suku Anak Dalam) and the similarities in culture and language between Muko-Muko and other MINANGKABAU varieties make it unlikely that either of these varieties would be placed under MUSI or SOUTH BARISAN MALAY.

9 References

Adelaar, K. Alexander. 1984. Some Proto-Malayic affixes. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140(4). 402–421.

Adelaar, K. Alexander. 1992. *Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. (Pacific Linguistics C-119). Canberra: Australian National University.

Adelaar, K. Alexander. 1993. The internal classification of the Malayic subgroup. *Bulletin of the School of Oriental and African Studies, University of London* 56(3). 566–581.

Adelaar, K. Alexander. 1994. Mise au point sur l'histoire linguistique du malais jakartanais [History of the development of Jakarta Malay]. *Archipel* 46. 47–5.

Adelaar, K. Alexander. 1995. Minangkabau. In Darrell Tryon (ed.), *Comparative Austronesian dictionary: An introduction to Austronesian studies, part 1*, 433–442. (Trends in Linguistics Documentation 10). Berlin: Mouton de Gruyter.

Adelaar, K. Alexander. 2004. Where does Malay come from? Twenty years of discussions about homeland, migrations and classifications. *Bijdragen tot de Taal-, Land- en Volkenkunde* 16. 1–3.

Adelaar, K. Alexander. 2005. Malayo-Sumbawan. *Oceanic Linguistics* 44(2). 357–388.

Adelaar, K. Alexander. 2008. On the classifiability of Malayic. In Yuri Lander & Alexander Ogloblin (eds.), *Language and text in the Austronesian world: studies in honour of Ülo Sirk*, 1–22. (LINCOM Studies in Austronesian Linguistics 06). Munich: LINCOM.

Adelaar, K. Alexander & David J. Prentice. 1996. Malay: its history, role, and spread. In Stephen Wurm, Peter Mühlhäusler & Darrell Tryon (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, vol. II Texts, 673–693. (Trends in Linguistics 13). Berlin: Mouton de Gruyter.

Aliana, Zainul Arifin & al. 1985. *Kamus bahasa Serawai-Indonesia (A-M)*. Jakarta: Pusat Bahasa.

Aliana, Zainul Arifin, Siti Salamah Arifin, Suwarni Nursato, Sungkowo Sutopo & Mardani Waif. 1987. *Morfologi dan sintaksis bahasa Melayu Palembang*. Jakarta: Pusat Bahasa.

Aliana, Zainul Arifin, R. M. Arif, Muslim Tuwi, A. Malian Erman & Ahmad Rozi Zakaria. 1985. *Kamus Ogan-Indonesia*. Jakarta: Pusat Bahasa.

Aliana, Zainul Arifin, Suwarno Nursato, Siti Salamah Arifin, Umar Efendi & Mardani Waif. 1985. *Morfologi dan sintaksis bahasa Bilide dialek Lembak*. Jakarta: Pusat Bahasa.

Aliana, Zainul Arifin, Suwarno Nursato, Siti Slamah Arifin, Umar Efendi & Mardani Waif. 1985. *Morfologi dan sintaksis bahasa Rawas*. Jakarta: Pusat Bahasa.

Aliana, Zainul Arifin, Latifah Ratnawati, Suhardi & Soedjiono Martojo. 1993. *Fonologi dan morfologi bahasa Muko-Muko*. Jakarta: Pusat Bahasa.

Andaya, Barbara Watson. 1993. *To live as brothers: southeast Sumatra in the seventeenth and eighteenth centuries*. Honolulu: University of Hawaii Press.

Anderbeck, Karl. 2007a. Initial reconstruction of Proto-Lampungic: phonology and basic vocabulary. *Studies in Philippines Languages and Cultures* 16. 41–165.

Anderbeck, Karl. 2007b. *Request for change to ISO 639-3 language code: Central Malay of Sumatra*. Dallas: ISO 639-3 Registration Authority. http://www.sil.org/iso639-3/cr_files/2007-179.pdf (2 February, 2009).

Anderbeck, Karl. 2007c. *Request for change to ISO 639-3 language code: Musi*. Dallas: ISO 639-3 Registration Authority. http://www.sil.org/iso639-3/cr_files/2007-182.pdf (2 February, 2009).

Anderbeck, Karl. 2007d. *Request for change to ISO 639-3 language code: Col*. Dallas: ISO 639-3 Registration Authority. http://www.sil.org/iso639-3/cr_files/2007-196.pdf (2 February, 2009).

Anderbeck, Karl. 2007e. *Request for change to ISO 639-3 language code: Muko-Muko*. Dallas: ISO 639-3 Registration Authority. http://www.sil.org/iso639-3/cr_files/2007-181.pdf (2 February, 2009).

Anderbeck, Karl. 2007f. Haji: one language from twelve. In Chong Shin, Karim Harun & Yabit Alas (eds.), *Reflections in Southeast Asian seas: essays in honour of professor James T. Collins*, vol. Book II, 51–91. Pontianak: STAIN Pontianak Press.

Anderbeck, Karl. 2007g. *Request for change to ISO 639-3 language code: Malay*. Dallas: ISO 639-3 Registration Authority. http://www.sil.org/iso639-3/cr_files/2007-183.pdf (2 February, 2009).

Anderbeck, Karl. 2007h. Drop, rattle, and roll: reflexes of Proto-Malayic *i, *u and *r in southern Sumatra. Paper. Paper presented at the Eleventh International Symposium of Malay/Indonesian Linguistics (ISMIL), Manokwari, Papua, Indonesia.

Anderbeck, Karl. 2008. *Malay dialects of the Batanghari River Basin (Jambi, Sumatra)*. (SIL E-Books 6). Dallas: SIL International. http://www.sil.org/silepubs/abstract.asp?id=50415 (15 August, 2008).
Anderbeck, Karl. 2013. The Malayic-speaking Orang Laut: dialects and directions for research. *Wacana*. 55.
Anderbeck, Karl & Herdian Aprilani. 2013. *The improbable language: survey report on the Nasal language of Bengkulu, Sumatra*. (SIL Electronic Survey Reports). Dallas: SIL International. http://www.sil.org/resources/publications/entry/54043 (25 February, 2014).
Anderbeck, Karl & Charlie Hanawalt. 2007. *Request for change to ISO 639-3 language code: Lampungic lects*. Dallas: ISO 639-3 Registration Authority. http://www.sil.org/iso639-3/cr_files/2007-142.pdf (2 February, 2009).
Anderbeck, Kristen. 201. *Language use and attitudes among the Jambi Malays of Sumatra*. . Vol. 21. (SIL E-Books). Dallas: SIL International.
Arifin, R. M., Sutari Harifin, Abdul Madjid, Baharuddin Nur & Gunawan. 1984. *Morfologi dan sintaksis bahasa Ogan*. Jakarta: Pusat Bahasa.
Arifin, Siti Salamah, Tarmizi Abubakar, Zahra Alwi & Ernialdi. 2001. *Struktur bahasa Benakat*. Jakarta: Pusat Bahasa.
Arifin, Siti Salamah, Tarmizi Abubakar, Kusmiarti, Hairuddin & Wowo Ario Sungkowo. 1996. *Fonologi dan morfologi bahasa Sindang*. Jakarta: Pusat Bahasa.
Arifin, Siti Salamah, Zainul Arifin Aliana, Tarmizi Mairu, Kusmiarti & Maimunah. 1988. *Sistem reduplikasi bahasa Enim*. Jakarta: Pusat Bahasa.
Arifin, Siti Salamah, Sungkowo Sutopo, Mulyadi Eko Purnomo, Sri Indrawati & Marzuan. 1997. *Sintaksis bahasa Sindang*. Jakarta: Pusat Bahasa.
Arif, R. M., Sutari Harifin, M. Yusuf Usman, Dahlia Mahabin Ayub & Latifah Ratnawati. 1981. *Kedudukan dan fungsi bahasa Palembang*. Jakarta: Pusat Bahasa.
Arif, R. M., Abdul Madjid, Tarmizi Mairu & Sudirman Sy. 1985. *Morfologi dan sintaksis bahasa Musi*. Jakarta: Pusat Bahasa.
Asmah Haji Omar. 1975. *Essays on Malaysian linguistics*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Asmah Haji Omar. 1977. *Kepelbagaian fonologi dialek-dialek Melayu*. (Penerbitan Ilmiah 1). Kuala Lumpur: Dewan Bahasa dan Pustaka.
Badan Pusat Statistik. 2004. *2000 Indonesian national census*. Jakarta: Badan Pusat Statistik Indonesia.
Badrul, Munir Hamidy, Azis Wahab, M. Yusuf Yacub, Zainal Hakim & Syafruddin. 1985. *Kamus Rejang-Indonesia*. Jakarta: Pusat Bahasa.
Barendregt, Bart. 2006. A supernatural topography of the southern Sumatran highlands. *Crossroads: An Interdisciplinary Journal of Southeast Asian Studies* 18(1). 113–125.
Bellwood, Peter. 1997. Indo-Malaysians of the last 40,000 years. In Peter Bellwood (ed.), *Prehistory of the Indo-Malaysian Archipelago*, 69–95. Honolulu: University of Hawaii Press.
Benjamin, Geoffrey. 1993. Grammar and polity: the cultural and political background to Standard Malay. In William A. Foley (ed.), *The role of theory in language description*, 341–392. (Trends in Linguistics, Studies and Monographs 69). Berlin: Mouton de Gruyter.
Bergman, Ted. 1991. Rapid appraisal of languages. *Notes on Sociolinguistics* 28. 3–11.
Bloomfield, Leonard. 1935. *Language*. London: George Allen & Unwin.
Blust, Robert. 1981. Variation in retention rate among Austronesian languages. Paper presented at the Third International Conference on Austronesian Linguistics, Bali.
Blust, Robert. 1984. On the history of the Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140(4). 422–45.
Blust, Robert. 1988. Malay historical linguistics: a progress report. In Mohd. Thani Ahmad & Zaini Mohamed Zain (eds.), *Rekonstruksi dan cabang-cabang bahasa Melayu induk*, 1–33. (Monografi Sejarah Bahasa Melayu). Kuala Lumpur: Dewan Bahasa dan Pustaka.
Blust, Robert. 199. Patterns of sound change in the Austronesian languages. In Philip Baldi (ed.), *Linguistic change and reconstruction methodology*, 231–267. (Trends in Linguistics: Studies and Monographs 45). New York: Mouton de Gruyter.
Blust, Robert. 1995. Austronesian comparative dictionary (ACD), computer files. Honolulu: University of Hawaii Department of Linguistics, ms.
Blust, Robert. 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36(1). 149–179.
Blust, Robert. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected papers from the 8th International Conference on Austronesian Linguistics*, 31–94. Taipei: Academia Sinica.

Blust, Robert. 2006. Whence the Malays? In James T. Collins & Awang Sariyan (eds.), *Borneo and the Homeland of the Malays: Four essays*, 64–87. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Boone, Douglas. 1994. Two kinds of lexicostatistical inferences. *Notes on Sociolinguistics* SI2. 73–99.
Boone, Douglas. 2007. Re: Intelligibility and comprehension. Email posting. *Language Survey SIL*.
Box, George E.P. & Norman Draper. 2007. *Response surfaces, mixtures, and ridge analyses*. 2nd ed. (Wiley Series in Probability and Statistics). Hoboken, NJ: Wiley.
Bronson, Bennet. 1977. Exchange at the upstream and downstream ends: Notes toward a functional model of the coastal state in Southeast Asia. In Karl Hutterer (ed.), *Economic exchange and social interaction in Southeast Asia: perspectives from prehistory, history, and ethnography*, 39–52. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.
Capell, A. 1982. Bezirkssprachen im gebiet des UAN. In Rainer Carle, Martina Heinschke, Peter Pink, Christel Rost & Karen Stadtlander (eds.), *Gava', studies in Austronesian languages and cultures dedicated to Hans Kähler*, 1–14. (Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg 17). Berlin: Dietrich Reimer.
Casad, Eugene. 1974. *Dialect intelligibility testing*. Norman: Summer Institute of Linguistics of the University of Oklahoma.
Chambers, J.K. & Peter Trudgill. 1998. *Dialectology*. 2nd ed. Cambridge: Cambridge University Press.
Cole, Peter, Gabriella Hermon & Yassir Tjung. 2006. Is there pasif semu in Jakarta Indonesian? *Oceanic Linguistics* 45(1). 64–9. (24 November, 2014).
Collins, James T. 1985a. The phonology of Tioman Malay and the reconstruction of Proto-Malay. In Suriya Ratanakul, David Thomas & Suwilai Premsrirat (eds.), *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt*. Bangkok: Institute of Language and Culture for Rural Development, Mahidol University. http://purl.org/sealang/collins1985phonology.pdf.
Collins, James T. 1985b. Dialek Melayu di kampung Landai, Pahang: menuju penelitian tatabahasa Melayu di kalang orang asli. *Dewan Bahasa* 29. 476–493.
Collins, James T. 1986. Kajian dialek daerah dan rekonstruksi bahasa purba. *Antologi kajian dialek Melayu*, 133–167. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Collins, James T. 1989. Malay dialect research in Malaysia: the issue of perspective. *Bijdragen tot de Taal-, Land- en Volkenkunde* 145. 235–264.
Collins, James T. 1995. *Bibliografi dialek Melayu di pulau Sumatera*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Collins, James T. & Awang Sariyan (eds.). 2006. *Borneo and the homeland of the Malays: four essays*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Collins, William A. 1979. Besemah concepts: a study of the culture of a people of South Sumatra. Berkeley: University of California, Berkeley PhD dissertation.
Dinas Pertanian Rakyat. 1975. *Monografi daerah kabupaten Bengkulu Utara propinsi Bengkulu*. Bengkulu: Dinas Pertanian Rakyat propinsi Bengkulu.
Dunggio, P. D., Suwarni N, Asnah S & Nur Indones. 1983. *Struktur bahasa Melayu Palembang*. Jakarta: Pusat Bahasa.
Dunggio, P. D., Zainan Wahab, Zainal Abidin Naning, Noor Indones & Bambang Apriay Luneto. 1985. *Struktur bahasa Kubu*. Jakarta: Pusat Bahasa.
Dyen, Isidore. 1965. *A lexicostatistical classification of the Austronesian languages*. (Indiana University Publications in Anthropology and Linguistics Memoir 19). Baltimore: Waverly Press.
Edwards McKinnon, Edmund. 1985. Early polities in southern Sumatra: some preliminary observations based on archaeological evidence. *Indonesia* 4. 1–36.
Ewing, Michael C. 2005. Colloquial Indonesian. In K. Alexander Adelaar & Nikolaus Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 227–258. (Routledge Language Family Series). London: Routledge.
Ferguson, Charles. 1959. Diglossia. *Word* 15. 325–34.
Gaffar, Zainal Abidin, Muslim Tuwi, Hasbi Yusuf, D Chairani & Makmun Rusydi. 1985. *Struktur bahasa Panesak*. Jakarta: Pusat Bahasa.
Gani, Zainal Abidin, Mustani Ahmad & Diemroh Ihsan. 1981. *Struktur bahasa Musi*. Jakarta: Pusat Bahasa.
Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17(2). 179–20.
Gil, David. 2005. Can you drink a book in Malay/Indonesian? Variation across isolect, socio-economic status and age. Paper presented at the the Ninth International Symposium on Malay/Indonesian Linguistics, Padang, West Sumatra.

Gordon, Ray Jr. (ed.). 2005. *Ethnologue: languages of the world*. 15th ed. Dallas: SIL International.
Graham, Steve. 200. Social network analysis: more toward an application to sociolinguistic research and language development assessment. In Gloria Kindell & M. Paul Lewis (eds.), *Assessing ethnolinguistic vitality: theory and practice. Selected papers from the Third International Language Assessment Conference*, vol. 3, 131–166. (Publications in Sociolinguistics). Dallas: SIL International.
Grijns, C.D. 1991. *Jakarta Malay: a multidimensional approach to spatial variation*. . Vol. 2 Vols. (Verhandelingen 149). Leiden: KITLV Press.
Grijns, C.D. 1996. Bahasa Melayu Jakarta. *Linguistik Indonesia* 14(1&2). 67–77.
Grimes, Barbara F. (ed.). 200. *Ethnologue: languages of the world*. 14th ed. Dallas: SIL International.
Hall, Jacob. 2009. Palembangese in print: an NLS look at literacies, linguacies, communicacies, and culturacies in South Sumatra. Grand Forks: University of North Dakota M.A. thesis.
Hanawalt, Charlie. 2007. Bitter or sweet?: the vital role of sociolinguistic survey in Lampungic dialectology. *Studies in Philippines Languages and Cultures* 16. 11–4.
Hanawalt, Charlie. 2012. Development of the Indonesian Sentence Repetition Test (ISRT): measuring bilingual proficiency in standard Indonesian among minority language communities. In Bahren Umar Siregar, P. Ari Subagyo & Yassir Nasanius (eds.), *Dari menapak jejak kata sampai menyigi tata bahasa: persembahan untuk Prof. Dr. Bambang Kaswanti Purwo dalam rangka ulang tahunnya ke-60*, 185–214. Jakarta: Atmajaya University.
Haris, Yusman. 2006. *Bumi serasan Sekate dan penduduknya*. Sekayu: Musi Banyuasin Regency.
Haugen, Einar. 1972. Dialect, language, nation. In Anwar Dil (ed.), *The ecology of language. Essays by Einar Haugen*, 237–254. Stanford: Stanford University Press.
Hudson, A. B. 1967. The Barito isolects of Borneo: a classification based on comparative reconstruction and lexicostatistics. Ithaca, NY: Dept. of Asian Studies, Cornell University PhD dissertation.
Hymes, Dell. 1974. *Foundations in sociolinguistics: an ethnographic approach*. Philadelphia: University of Pennsylvania Press.
Ihsan, Diemroh, Salamah Arifin, Agus Saripudin & Chuzaimah Dahlan Diem. 1996. *Fonologi dan morfologi bahasa Lematang*. Jakarta: Pusat Bahasa.
Ihsan, Diemroh, Chuzaimah Dahlan Diem & Majelis. 2002. *Sistem reduplikasi bahasa Lintang*. Jakarta: Pusat Bahasa.
Ikranegara, Kay. 198. *Melayu Betawi grammar*. (NUSA: Linguistic Studies of Indonesian and Other Languages in Indonesia 9). Jakarta: Badan Penyelenggara Seri NUSA, Universitas Katolik Indonesia Atma Jaya.
Imranuddin D, Padi Utomo, Helena, Suryam & Suryadi. 1994. *Morfosintaksis bahasa Melayu Bengkulu*. Jakarta: Pusat Bahasa.
Im, Younshim & Rindu Simanjuntak. 2009. Bilingualism survey report for language groups of South Sumatra and Bengkulu. Unpublished draft. Jakarta, ms.
Interagency Language Roundtable (ILR). n.d. About the ILR. Washington, D.C. *Interagency Language Roundtable 2006-2007*. http://www.govtilr.org/ILR_History.htm (15 July, 2008a).
Interagency Language Roundtable (ILR). n.d. ILR skill level descriptions for interpretation performance. Washington, D.C. *ILR skill level descriptions for interpretation performance*. http://www.govtilr.org/Skills/interpretationSLDsapproved.htm (15 July, 2008b).
Kamusella, Tomasz. 2012. The global regime of language recognition. *International Journal of the Sociology of Language* 2012(218). doi:1.1515/ijsl-2012-0059. http://www.degruyter.com/view/j/ijsl.2012.2012.issue-218/ijsl-2012-0059/ijsl-2012-0059.xml?format=INT (2 July, 2013).
Kasim, Yuslina, Marjusman Maksan, Syamsir Arifin & Muhamad Yamin Zailoet. 1987. *Pemetaan bahasa daerah di Sumatra Barat dan Bengkulu*. Jakarta: Pusat Bahasa.
Kasmansyah, Sungkowo Sutopo, Subadiyono, Nurbaya As'ad, Marianus Anwar & Diemroh Ihsan. 1999. *Kamus bahasa Indonesia-Besemah (A-K)*. Jakarta: Pusat Bahasa.
Kennedy, Raymond. 1962. *Bibliography of Indonesian (Sumatran) peoples and cultures*. 2nd ed. New Haven: Southeast Asia Studies, Yale University, by arrangement with Human Relations Area Files.
Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. 60–67. Dublin: EACL.
Kluge, Angela. 2007. "Sorry, could you repeat that, please?! ": Where does one language end and the next begin? Paper. Paper presented at the SIL Indonesia Eastern Indonesia Malay Seminar, Sentani, Papua.

Kozok, Uli. 2004. *The Tanjung Tanah code of law: the oldest extant Malay manuscript*. Cambridge: Cambridge University Press.
Leach, E.R. 195. *Social science research in Sarawak*. London: H.M.S.O.
Lebar, Frank. 1972. *Ethnic groups of insular Southeast Asia*. . Vol. 1: Indonesia, Andaman Islands, and Madagascar. New Haven: Human Relations Area Files Press.
Lewis, M. Paul (ed.). 2009. *Ethnologue: languages of the world*. 16th ed. Dallas: SIL International.
Lewis, M. Paul, Gary Simons & Chuck Fennig (eds.). 2013. *Ethnologue: languages of the world*. 17th ed. Dallas: SIL International.
Mahdi, Waruno. 1994. Some Austronesian maverick protoforms with culture-historical implications II. *Oceanic Linguistics* 33(2). 431–49.
Mahsun. 2005. *Metode penelitian bahasa: tahapan strategi, metode, dan tekniknya*. Jakarta: Rajawali.
Manguin, Pierre-Yves. 2009. Southeast Sumatra in protohistoric and Srivijaya times: upstream-downstream relations and the settlement of the peneplain. In Dominik Bonatz, John N Miksic, J. David Neidel & Mai Lin Tjoa-Bonatz (eds.), *Distant tales: archaeology and ethnohistory in the highlands of Sumatra*, 434–484. Cambridge: Cambridge Scholars Publishing.
Maryono, Albertus Sinaga, Agus Setyonegoro & Ade Kusmana. 1997. *Dialektologi bahasa Kubu: laporan penelitian*. Jakarta: Pusat Bahasa.
Matthews, Di & Indarto. 2007. *Kampung Muarobanyu terhindar dari flu burung*. (Trans.) Welly. Palembang: SIL International.
McDonnell, Bradley. 2014. A note on the names of Malayic isolects of southwest Sumatra. Unpublished draft. University of California, Santa Barbara, ms.
McDowell, Jonathan & Karl Anderbeck. 2020. Supplementary datafiles for McDowell and Anderbeck, The Malay Lects of Southern Sumatra. Dallas: SIL International. https://www.sil.org/resources/archives/86340 (7 November, 2020).
McGinn, Richard. 1997. Some irregular reflexes of Proto-Malayo-Polynesian vowels in the Rejang language of Sumatra. *Diachronica* 14(1). 67–107.
McGinn, Richard. 2003. Raising of PMP *a in Bukar-Sadong Land Dayak and Rejang. In John Lynch (ed.), *Issues in Austronesian historical phonology*, 37–64. (Pacific Linguistics 550). Canberra: Australian National University.
Mckinnon, Timothy. 2012. A survey of phonation driven vowel shifts in traditional Malay(ic). Presented at the Sixteenth International Symposium of Malay/Indonesian Linguistics (ISMIL 16), Kelaniya, Sri Lanka.
Medan, Tamsin. 198. *Dialek-dialek Minangkabau di daerah Minangkabau/Sumatera Barat (suatu pemerian dialektologis)*. (unpublished). Jakarta: Pusat Bahasa.
Medan, Tamsin, Marsjusman Kasan, Syamsir Arifin, Yuslina Kasim, Busri & A. Razak Sikumbang. 1986. *Geografi dialek bahasa Minangkabau: suatu deskripsi dan pemetaan di daerah kabupaten Pasaman*. Jakarta: Pusat Bahasa.
Mitani, Yasuyuki. 198. Languages of South Sumatra. In Yoshihiro Tsubouchi (ed.), *South Sumatra: man and agriculture*, 1–19. Kyoto: Center for Southeast Asian Studies, Kyoto University.
Moussay, Gerard. 1998. *Tata bahasa Minangkabau*. (Trans.) Rahayu Hidayat. Jakarta: Kepustakaan Populer Gramedia.
Mufwene, Salikoko. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.
Mühlhäusler, Peter. 1996. *Linguistic ecology: language change and linguistic imperialism in the Pacific region*. London: Routledge.
Munoz, Paul Michel. 2006. *Early kingdoms of the Indonesian archipelago and the Malay peninsula*. Singapore: Editions Didier Millet.
Naning, Zainal Abidin, Nangsari Ahmad, Surip Suwandi, Akhyar Burhan & Ansori. 1998. *Geografi linguistic bahasa Enim*. Jakarta: Pusat Bahasa.
Nerbonne, John, Wilbert Heeringa, Eric van den Hout, Peter van de Kooi, Simone Otten & Willem van de Vis. 1996. Phonetic distance between Dutch dialects. In W. Durieux, W. Daelemans & S. Gillis (eds.), *CLIN VI: Proceedings of the Sixth CLIN Meeting*, 185–202. Antwerp: Centre for Dutch Language and Speech (UIA).
Nikelas, Syahwin, Asni Ayub, Marah Rusmali Zainil & Syofyan Adam. 1986. *Struktur bahasa Pekal*. Jakarta: Pusat Bahasa.
Nothofer, Bernd. 1986. The Barrier island languages in the Austronesian language family. In Paul Geraghty, Lois Carrington & Stephen Wurm (eds.), *FOCAL II: papers from the Fourth International*

Conference on Austronesian Linguistics, 87–109. (Pacific Linguistics C-94). Canberra: Australian National University.
Nothofer, Bernd. 1988. A discussion of two Austronesian subgroups: Proto-Malay and Proto-Malayic. In Mohd. Thani Ahmad & Zaini Mohamed Zain (eds.), *Rekonstruksi dan cabang-cabang bahasa Melayu induk*, 34–58. (Monografi Sejarah Bahasa Melayu). Kuala Lumpur: Dewan Bahasa dan Pustaka.
Nothofer, Bernd. 1995a. The history of Jakarta Malay. *Oceanic Linguistics* 34(1). 86–97.
Nothofer, Bernd. 1995b. Dialek Melayu di Kalimantan dan di Bangka: misan atau mindoan? In Soenjono Dardjowidjojo (ed.), *PELLBA 8. Pertemuan Linguistik Pusat Kajian Bahasa dan Budaya Atma Jaya Kedelapan*, 53–84. Jakarta: Atmajaya University.
Nothofer, Bernd. 1997. *Dialek Melayu Bangka*. Bangi: Penerbit Universiti Kebangsaan Malaysia.
Nursato, Suwarni, Sutari Harifin, Zainan Wahab, Nangsari Ahmad & Homsen Nanung. 1989. *Fonologi dan morfologi bahasa Lintang*. Jakarta: Pusat Bahasa.
Prentice, David J. & Amir Hakim Usman. 1978. Kerinci sound-changes and phonotactics. In Stephen Wurm & Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: proceedings*, vol. Fascicle I: Western Austronesian, 121–163. (Pacific Linguistics C-61). Canberra: Australian National University.
Probst, Ulrich. 1992. Morphophonemic words in lexicostatistics: an evaluation on use and limitations of different approaches. Unpublished ms (SIL survey course materials). Horsleys Green, UK, ms.
Purnomo, Mulyadi Eko, Kusmiarti, Sri Indrawati & RHM Ali Masri. 200. *Morfologi dan sintaksis bahasa Panesak*. Jakarta: Pusat Bahasa.
Ratnawaty, Latifah, Zainul Arifin Aliana, Muslim & Ahmad Bastari Suan. 2002. *Struktur sastra lisan Aji*. Jakarta: Pusat Bahasa.
Riedel, J. G. F. 1881. Twee volksverhalen in het dialekt der Orang Lawoet of Orang Sekah van Belitoeng [Two folkstories in the dialect of the Orang Lawut or Orang Sekah of Belitung]. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 26. 264–273.
Ross, M. D. 1992. The Sound of Proto-Austronesian: An Outsider's View of the Formosan Evidence. *Oceanic Linguistics* 31. 23–64.
Ruskhan, Abdul Gaffar, Mu'jizah, Sutiman & Yeyen Maryani (eds.). 2008. *Bahasa dan peta bahasa di Indonesia*. Jakarta: Pusat Bahasa.
Rusmali, Marah, Amir Hakim Usman, Syahwin Nikelas, Nurzuir Husin, Busri, M. Yamin, Isna Sulastri, Irfani Basri & Agusli Lana. 1985. *Kamus bahasa Minangkabau-Indonesia*. Jakarta: Pusat Bahasa.
Sakai, Minako. 2006. The origin structure of *kute* among the Gumai: an analysis of an indigenous territorial institution in the highlands of South Sumatra. In Thomas Reuter (ed.), *Sharing the earth, dividing the land*, 39–64. Canberra: Australian National University.
Saleh, Yuslizal, M. Kamsari & Abdul Madjid. 1984. *Struktur bahasa Rawas*. Jakarta: Pusat Bahasa.
Saleh, Yuslizal, M. Lamsari, Abdul Madjid, Sofyan Silahiddin & Zainan Wahab. 1979. *Bahasa Semende*. Jakarta: Pusat Bahasa.
Saleh, Yuslizal, Aidy Ruslan Satun, Umar Idris & A. Kudir Ariman. 1985. *Morfologi dan sintaksis bahasa Semende*. Jakarta: Pusat Bahasa.
Saleh, Yusrizal, Lamsari Majid & Chuzaimah Dahlan Diem. 1977. *Struktur bahasa Besemah*. Palembang: Pusat Bahasa.
Sandbukt, Øyvind. 1995. The Kerinci Seblat National Park and the Kubu people: preliminary assessment and recommendations. www.warsi.or.id (5 September, 2007).
Sastra, Gusdi. 1994. Protobahasa Minangkabau – Kerinci: rekonstruksi fonologi. Jakarta: Universitas Indonesia Masters Thesis.
Saydam, Gouzali. 2004. *Kamus lengkap bahasa Minangkabau*. Padang: Pusat Pengkajian Islam dan Minangkabau.
Seidlitz, Eric. 2007. Duano: a first look at its phonology. In Chong Shin, Karim Harun & Yabit Alas (eds.), *Reflections in Southeast Asian seas: essays in honour of professor James T. Collins*, vol. Book II, 23–49. Pontianak: STAIN Pontianak Press.
Smedal, Olaf. 1987. *Lom-Indonesian-English and English-Lom wordlists accompanied by four Lom texts*. (NUSA: Linguistic Studies of Indonesian and Other Languages in Indonesia 28-29). Jakarta: Badan Penyelenggara Seri NUSA, Universitas Katolik Indonesia Atma Jaya.
Sneddon, James. 2003. *The Indonesian language: its history and role in modern society*. Sydney: University of New South Wales Press.

Stalder, Juerg. 1996. Update on rapid appraisal. *Notes on Sociolinguistics* 48. 24–28.
Stokhof, W. A. L. (ed.). 1987. *Holle Lists: vocabularies in languages of Indonesia, Vol. 10/2: Southern Sumatra*. (Pacific Linguistics D-75). Canberra: Australian National University.
Sudarmanto, Budi Agung. 2012. State of linguistic research in (South) Sumatera. Paper. Paper presented at the Current Trends of Linguistic Research of Indigenous Languages in Indonesia, Tokyo University of Foreign Studies.
Tadmor, Uri. 2001. Language contact and historical reconstruction: the case of Palembang Malay. Paper. Paper presented at the Fifth International Symposium of Malay/Indonesian Linguistics (ISMIL 5), Max Planck Institute for Evolutionary Anthropology, Leipzig.
Tadmor, Uri. 2002. Language contact and the homeland of Malay. Paper presented at the Sixth International Symposium of Malay/Indonesian Linguistics (ISMIL), Bintan Island, Riau.
Tadmor, Uri. 2003. Final /a/ mutation: a borrowed areal feature in Western Austronesia. In John Lynch (ed.), *Issues in Austronesian historical phonology*, 15–35. (Pacific Linguistics 550). Canberra: Australian National University.
Truong, Christina Lai & Lilian Garcez. 2012. Participatory methods for language documentation and conservation: building community awareness and engagement. *Language Documentation & Conservation* 6. 22–37. (31 May, 2012).
Truong, Eldwin. 201. Malay varieties of southern Sumatra: an evaluation of Levenshtein Distance as a tool for dialect classification. Paper. Paper presented at the Fourteenth International Symposium of Malay/Indonesian Linguistics (ISMIL), University of Minnesota, Minneapolis.
Umar, Mannan, Zainuddin Amir, Nasroel Malano, Anas Syafei & Agustar Surin. 1986. *Struktur bahasa Muko-Muko*. Jakarta: Pusat Bahasa.
Voorhoeve, Petrus. 1955. *Critical survey of studies on the languages of Sumatra*. (Koninklijk Instituut Voor Taal-, Land- En Volkenkunde, Bibliographical Series 1). ‘s-Gravenhage: Martinus Nijhoff.
Wahab, Zaini, Suwarni Nursato, P.D. Dunggio & Suwandi. 199. *Ragam dan dialek bahasa Ogan*. Jakarta: Pusat Bahasa.
Weintré, Johan. 2003. *A case study of the Orang Rimba in Sumatra (the nomadic Kubu society)*. Field research paper for completion of study requirements for University of New England--Australia. Yogyakarta: Universitas Gadjah Mada.
West Indonesia Survey Team. 2006. Rejang language of Sumatra: sociolinguistic description and baseline bilingualism study. Unpublished draft. Jakarta, ms.
Wetherill, G. Barrie. 1995. Research design and analysis. Ms.
Widianto, Harry. 2009. The dawn of humanity in Sumatra: arrival and dispersal from the human remains perspective. In Dominik Bonatz, John N Miksic, J. David Neidel & Mai Lin Tjoa-Bonatz (eds.), *Distant tales: archaeology and ethnohistory in the highlands of Sumatra*, 28–42. Cambridge: Cambridge Scholars Publishing.
Wouk, Fay. 2004. Transitive bare verb stems in spoken Jakarta Indonesia. In Bambang Kaswanti Purwo (ed.), *Studies in Malay dialects, part IV*, 29–5. (NUSA: Linguistic Studies of Indonesian and Other Languages in Indonesia 53). Jakarta: Badan Penyelenggara Seri NUSA, Universitas Katolik Indonesia Atma Jaya.
Wurm, Stephen & Shiro Hattori. 1983. Part II: Japan area, Taiwan (Formosa), Philippines, Mainland and Insular South-East Asia. *Linguistic atlas of the Pacific area*, vol. Sheet 41, northern part of Borneo; sheet 42, southern part of Borneo. (Pacific Linguistics C-66–7). Canberra: Australian Academy of the Humanities in collaboration with the Japan Academy.
2010. *Profil doa suku-suku yang terabaikan*. Jakarta: Indonesia Pelangi Nusantara.

Appendix 1: Rapid Appraisal Recorded Text Test (RA-RTT)

The following section has been adapted from Hanawalt et al.'s unpublished (2006) survey report on Lampungic.

The Recorded Text Test is based upon the assumption that a person's ability to retell a story heard in another speech variety corresponds to his or her ability to comprehend that speech variety. In practice, this tool can differentiate between very low levels of comprehension and moderate/high levels of comprehension. It cannot reliably distinguish between moderate and high levels of comprehension.

The rapid appraisal recorded text test (RA-RTT) was developed as a test method that is simpler and less time consuming than the classic RTT documented by Eugene Casad (1974). Its procedures, advantages and disadvantages are described in Stälder 1996. We made a few adjustments to the application of Stälder's model.

By way of overview, the RA-RTT involves recording a short text, usually 2-3 minutes long, in the speech variety that will be tested in other areas. The text is then divided into segments short enough to retell easily. A score sheet is then made based upon those segments. This text is then tested in the same village or local area with people not involved in the recording process to determine whether any segments are unusually long or difficult for native speakers to retell. The text is then taken to other areas, and the test is administered to groups of people at each test site. These informants are scored as a group according to how well they can comprehend and then retell the story.

1 Standard Procedures

1.1 Preparation of the test

The RTT story must be a story that is rather unpredictable in its outcome, so that people from other speech varieties truly need to be able to understand the speech in order to retell the segments of the story and not simply guess at how the story unravels. Thus, famous legends and recounting of events familiar to people from other speech varieties are not good candidates as RA-RTT stories.

The story must be recorded at high quality so as not to interfere with the comprehension of those who will be tested. It may be necessary to make several recordings of the text in order to eliminate undesirable elements. Such elements would include: loanwords from the language of wider communication or the national language, place names, or other factors that could influence the listener's comprehension of the text. The process of recording the texts may take only a few hours, or multiple sessions depending on the situation.

Once the final text has been recorded, the text should be transcribed into the local speech variety, the language of wider communication and any other relevant languages for the researchers. A final check should be done to ensure that all the words are from the local speech variety and not from the language of wider communication. If necessary, the affected portion of the story can be recorded again using L1 vocabulary.

The researcher must then arrange the text so that it is divided into individual thought segments. Each segment should only be one or two sentences long. The researcher then inserts a pause between each segment in the recording. It is best to use a Mini-Disc recorder to record and edit the text digitally.

The text then needs to be tested with people from the speech variety where the text was recorded. They should not, however, have been involved in the recording process. The researchers list all the elements in the text which all home town test takers mention as they retell the story. The researcher must then write or type these core elements from each segment on a piece of paper that will be used as a score sheet. Each core element should be given a separate line on the score sheet.

The researcher should also prepare a short practice text, divided into segments, in the language of wider communication.

1.2 Administration of the test

The RA-RTT may then be tested in the desired locations. The text is tested with groups, not individuals, to ensure that difficulties with individual testing are avoided. A leader in each test location is asked to gather together a group of individuals to help in language research. The group should consist of about ten people, half men and half women, representing all age levels from fifteen years old to elderly people.

The researcher then begins the test by explaining that a short story will be played for the group. The group will be asked to retell each section of the story after they hear it. The researcher should then play the short practice text so that the informants can get used to the testing procedures.

As the informants retell each section of the test text, the researcher or his/her assistant should mark on the score sheet which elements the group as a whole retold in each segment, using a simple scoring system to account for each core element. For instance, '1' could mean that they mentioned a particular segment and '0' could mean that they did not mention it. The research team must determine ahead of time how to score uncertain responses. For instance, they could make a note of the uncertainty and then score all such occurrences consistently once all the testing is complete.

The tabulation of the score will then give the researcher a relative means for comparing comprehension ability at each location against the ability of informants at other locations.

1.3 Interpretation of scores

As the RA-RTT is a rapid appraisal tool, the results must be viewed as relative indicators of intelligibility and not as exact or quantitative measures of inherent intelligibility between two speech varieties. Similar to the classic RTT, the RA-RTT should not be viewed as a determiner of purely inherent intelligibility (see O'Leary 1994). Acquired intelligibility and the effect of relative structural similarity between other speech varieties of exposure may greatly affect the results of the RTT or RA-RTT.

2 Procedures We Followed

Because of the nature of real-life fieldwork done in a limited timeframe and with limited resources, we were unable to follow exactly every aspect of standard procedures for the RA-RTT. In the following sections is a description of how we actually administered the RA-RTT on the survey and how it diverges from standard procedures.

2.1 Text recording site selection

-Pagar Alam for Central Malay
- Sekayu for Musi Malay
- Maninjau for Minangkabau

The three sites above were chosen on linguistic, geographic, cultural, and pragmatic grounds. For RA-RTT, a story from a central dialect can be tested in the more peripheral dialect areas to test comprehension and attitude towards the central dialect.

For the three dialects chosen, Pagar Alam clearly met all four conditions mentioned above. Linguistically, Pagar Alam shows the same innovations much of the Central Malay group and was frequently mentioned when people talked about the Besemah language (*bahasa Pagar Alam*). It is the heartland of the Besemah speaking area, which is the cultural center of the Highland Malay speaking area, with other core Highland Malay groups generally acknowledging a relationship to the Besemah. Geographically, it is central to the far reaching Central Malay group, between the coastal varieties of Highland Malay to the west in Bengkulu and the varieties comprising the Ogan group to the east. Pragmatically, informants from the nearby town of Sukaraja helped the researchers find a ethnically homogeneous Besemah population to help record and home town test the story. Another pragmatic reason was that there was no clear rival to Pagar Alam from either the Ogan group or from the Highland Malay group.

Sekayu also met the same conditions. Linguistically, the most prominent innovations that can be found in Col, Pegagan, and Musi dialects are found in Sekayu. The other dialects in the Palembang subgroup of Musi Malay have few clear unifying innovations, making it difficult to choose an alternate central dialect from them. The Palembang speech varieties are very widely used as a language of wider communication, but due to the large number of Javanese and other loan words that have entered the lexicon, is less central linguistically. Culturally, the case is not as clear for Sekayu, as Palembang is by far the largest and most important city along the watershed of the Musi River, and Lubuk Linggau is a larger and more important city than Sekayu at this point. However, for the core Musi dialects, Sekayu is clearly reference, and Musi is frequently called *bahasa Sekayu* (the Sekayu language). Sekayu lies halfway between the upper reaches of the Musi river (Rawas and Col) and the lower reaches (Palembang, Pegagan, Coastal and Lowland Malay), making it appealing from a geographic standpoint. Pragmatically, the informant at Sekayu provided the researchers with an excellent text that met the requirements of the story.

For Minangkabau, the cultural center is the city of Bukit Tinggi. The story was obtained while in Bukit Tinggi and the researcher went slightly outside of the city looking for a more ethnically homogeneous area. However, the journey took the researcher unknowingly into the closely related Maninjau dialect area, where the story was obtained. Linguistically, these share the majority of sound innovations that mark Minangkabau dialects. The most widely understood Minangkabau variety is that spoken in the city of Padang, but it lacks some of the innovations and is seen as less ethnically central (somewhat the case with Palembang as well). Geographically, the Bukit Tinggi area is central to the Minangkabau language area. Pragmatically the story from Maninjau was used rather than creating a new one from Bukit Tinggi, as the Maninjau story was already prepared and the speakers from the Bukit Tinggi area could understand it and said that it was not too different.

2.2 Problems with texts

The Minangkabau RTT was recorded a year or two before the survey, while team members were in West Sumatra. They did not do hometown testing or develop a score sheet at that time, so before we left on survey we did the hometown testing with some Minangkabau speakers living in Jakarta. This was not ideal as they were out of their home area and used Indonesian more than Minangkabau in their daily lives. None of the hometown test participants was from the Maninjau dialect in which the story was recorded. They were all from other areas of West Sumatra. They all reported that the story was easy to understand, and the Maninjau dialect was quite similar to their own dialects, although not exactly the same. So the two factors of 1) hometown test participants being outside of their home area and 2) hometown test participants being from slightly different dialect areas both may have resulted in a slightly less detailed score that we developed from their answers.

The storyteller in the Besemah RTT was a young man in Pagar Alam. He had trouble thinking of a story long enough for our needs. When we recorded the story that later became the RTT, he spoke slowly and often repeated himself, trying to make the story long enough to satisfy us. This gave us a text that was not ideal because the speech was slower than normal conversational speech and the many repetitions gave listeners a chance to pick up details they may have missed in faster speech. As this was the best story we had, we developed a score sheet for it. The quality of the text may have had an effect on test results because the story seemed to be easier to understand than other RTT stories.

Due to technician error, the Besemah story was recorded at a very low volume. We increased the volume as much as possible using sound editing software, but the story was still too quiet to be heard well when played through the speakers we used for the other two RTT stories. To compensate for this, we had to use headphones so the story could be heard. As we only had four sets of headphones, we were limited to four participants per test. Depending on how much time we had in a village, we either administered the Besemah RTT once to two men and two women, or we administered it twice: first to four men, and then to four women. Using headphones had the advantage of controlling exactly who heard the text and who was able to answer. It had the disadvantage of not allowing us to note comments given by observers who listened in as with the other RTT stories .

2.3 Subject selection for testing

In each village where we administered the RTT, we asked the head of the village to gather a group of local people who would be willing to help us in our language research by listening to a story from another area. In the first village, we asked the leader to bring together some people to listen to a story from Sekayu. This created some confusion as he thought we wanted people who were known to understand Sekayu. After that, we only said that we wanted to play a story from another area and did not mention where the story was from. We did not mention where the story was from until after we had administered the RTT and then asked the participants where they thought it was from.

In each village, we asked the village leader for a group of about ten people–five men and five women–of all age groups from youths (age 15+) to elderly. Our only requirement was that all participants be locals from that village. When the group came together, we tried to ascertain through casual conversations how much they had traveled and how much exposure they had to other language varieties. We were never able to get a group of exactly five men and five women. There were usually more men than women, and anywhere from four to twelve participants in all.

2.4 Testing

During administration we played each section of the text and then asked the participants to explain it to us in Indonesian. We then played the section again and asked what other elements they had heard. If after the second time they asked for the text to be replayed, we replayed the text a maximum of one more time and made a note of this on the score sheet. If they missed a core element, we asked a general probing question, and made a note on the score sheet that we had probed.

In a few villages, the people were not very comfortable in Indonesian and so we asked them to explain the story back in the local language, having it translated by one of the more educated people there (usually the village leader). In one village, the participants felt that they were speaking standard Indonesian to us, but we could not understand them well as they were using another variety, probably Palembang Malay. There we also asked for a translator.

During administration, one or two of the more prominent people often dominated the retelling of the story. To compensate for this, we asked specific quieter individuals to take their turn to answer one or two questions. In this way we were able to elicit the participation of all group members

2.5 Post-RTT Questionnaire

After the test, we administered a brief questionnaire (see section 3.3) to find what the participants thought about the language they just heard in the story. This simple seven-question questionnaire was designed to uncover local perceptions and attitudes about the speech variety in question. We did not tell participants where the story was from until after they had answered where they thought it was from, why they thought so, and what they thought of the speech variety. Respondents were usually quite accurate in determining where texts were from.

2.6 Scoring

As participants retold each section of the RTT stories, we marked on paper score sheets each core element they mentioned. We also made notes of any unusual responses, distractions that kept them from hearing the text well, and probing questions we asked. We tabulated the scores for all the elements and presented the result as a fraction (example: 38/45 for 38 core elements mentioned out of 45 total core elements). See the datafiles in McDowell and Anderbeck (2020) for the actual scores.

3 The Data: References

3.1 Interlinear Texts of RA-RTT Stories

The black text is the recorded transcription as given, the red text is the word-for-word translation into Indonesian, and the blue text is the free translation of the text in Indonesian.

The texts are presented in the sections that were played for testing with the corresponding section number and the length of the recording noted at the heading of each section.

3.1.1 Minangkabau RA-RTT

Section 1: 14 seconds

Tahun anam puluhan maso sakolah rakyaik di nagari bayua, awak dekkarano bansaik kojuo sikola bakakurangan dalam sagalo hal. Ado satu kali pado hari jumahaik ado tantara pe-er-er-i datang,

Tahun enam puluhan masa sekolah rakyat di negeri Bayur, saya dikarenakan miskin ini juga sekolah berkekurangan dalam segala hal. Ada satu kali pada hari Jumat ada tentara pe-er-er-i datang,

Sekitar tahun enam puluhan, saya bersekolah di sekolah rakyat di daerah Bayur karena saya miskin dan sekolah itu juga berkekurangan dalam segala hal. Suatu hari, pada hari Jumat, ada seorang tentara PRRI

Section 2: 11 seconds

datang tantara pe-er-er-i masuak ka rumah sikola, lalu awak madok kainyo inyo mancaliak pulo. Kiro kiro sadang batamu pandang samo pandang ruponyo inyo mengijokan mato

datang tentara pe-er-er-i masuk ke rumah sekolah, lalu saya memandang kepadanya dia memandang pula. Kira-kira sedang bertemu pandang sama pandang rupanya dia mengedipkan mata
datang dan masuk ke gedung sekolah, lalu saya memandangnya dan dia juga memandang saya. Ketika kami saling berpandang-pandangan, rupanya dia mengedipkan mata

Section 3: 7 seconds
indak lamo antaro kami kalua sampaino di belakang menanyakan ingin kenal diak sia namo ingin kenal awak
tidak lama antara kami keluar sampainya di belakang menanyakan ingin kenal adik siapa nama ingin kenal saya
tidak lama setelah kami keluar ke belakang, dia menanyakan nama saya dan mau berkenalan dengan saya

Section 4: 13 seconds
kato awak indak ado namo do namo (?) di rumah sikola tunjuaanjolah indak baado. padahal urang tu tantara lua jadi tantara luako de karanonyo lagadang awak alun tau aa dituruikan sampai pulang
kata saya tidak ada nama loh nama (?) di rumah sekolah tunjukanlah tidak apa apa, padahal orang itu tentara luar jadi tentara luar itu karena dia sudah besar saya belum tahu apa-apa diikuti sampai pulang
saya berkata bahwa saya tidak tahu nama saya, nama saya hanya ada di sekolah. Tunjukkanlah, tidak apa-apa, katanya. Padahal orang itu tentara asing. Karena tentara itu badannya besar dan saya masih lugu, dia mengikuti saya sampai pulang

Section 5: 11 seconds
sampai bisuak dibuekno surek dek karano awak bakawan baduo urang ka kawan diagiahkano surek itu rupo sudah dibacono rupo indak kanodo dalam surek tasabuik untuak nan surang lai. bisuak diagiahkannyo kawak dikawan
sampai besok dibuatnya surat karena saya berteman berdua orang ke kawan diberikannya surat itu rupanya sudah dibacanya rupanya bukan untuk dia dalam surat tersebut untuk yang satu lagi. Besok diberikannya kepada saya oleh teman saya.
besok harinya, ia menulis surat untuk saya dan memberikannya kepada teman saya. Rupanya setelah teman saya membacanya, surat itu bukan untuk dia tapi untuk saya. Teman saya kemudian memberikan surat itu kepada saya keesokan harinya

Section 6: 10 seconds
sudah tu dibaca sudah tu dibaleh wak kecek kan dalam surek baraso awak ketek alun tau ba kirim-kirim surek. kecekno bialah ketek bana awak mantu sikola.
Sudah itu dibaca, sudah itu dibalas saya bilang kan dalam surat bahwa saya kecil belum tahu kirim kiriman surat. Katanya biarlah kecil benar saya bantu sekolah
Sesudah saya membaca surat itu, dalam balasannya saya katakan bahwa saya masih kecil dan belum tahu mengirim surat. Tentara itu berkata bahwa biarpun saya kecil, ia akan membantu menyekolahkan saya.

Section 7: 10 seconds
jadi dalam ariari karajo awak manolong urang tuo manampi menumbuak di giliangan padi atauo kincia namo atauo masin. pulang pulang sikola tolong bantu urang tuo.
jadi dalam hari-hari kerja saya menolong orang tua menampi menumbuk di penggilingan padi atau kincir namanya atau mesin. Pulang pulang sekolah menolong bantu orang tua.
Jadi saya sehari-hari bekerja membantu orang tua menampi dan menumbuk padi di penggilingan atau yang disebut kincir atau mesin. Setelah pulang sekolah, saya membantu orang tua

Section 8: 21 seconds
sahabih itu de karano urang tadi ingin jo awak juo rupo sampai bakenalan dan sampai kiro-kiro setaun lamo nagari aman nagari sudah begolak awak kawin
Sehabis itu karena orang tadi ingin juga dengan saya rupanya sampai berkenalan dan sampai kira-kira setahun lamanya negeri aman negeri sudah bergolak (perang) saya kawin
Setelah itu karena tentara tadi masih cinta dengan saya setelah satu tahun berkenalan dan karena negeri sudah aman dan tidak bergolak lagi, maka saya kemudian kawin

duo taun sudah kawin jatuah mularaik awak ko de karano diisu isuno carai, carai indak do bisa babaliak nan awak bapisaha.
dua tahun sudah kawin jatuh melarat saya ini karena digosipkan cerai, cerai tidak bisa kembali dengan saya berpisah
setelah dua tahun kawin, saya menjadi melarat karena ada gosip mau diceraikan, karena sudah tidak bisa berbaikan lagi maka saya berpisah dengannya

Section 9: 12 seconds
tingga anak surang laki-laki dari sinan jatuah batambah misikin tambah mularaik. Mancari awak bausaho untuak diri awak surang jo anak langsuang pulo sataun kawin jo nan lain.
Tinggal anak seorang laki-laki dari situ jatuh bertambah miskin tambah melarat. Mencari saya berusaha untuk diri saya seorang dan anak langsung pula setahun menikah dengan yang lain.
Dia meninggalkan seorang anak laki-laki sehingga semenjak itu saya tambah miskin dan melarat. Saya berusaha mencari laki-laki lain untuk saya dan anak saya dan setahun kemudian saya lalu menikah dengan laki-laki lain

Section 10: 11 seconds
dek karano jodoh tadi indak dikiro-kiro dapek anak batujuah tibo salapan jo anak nan tuo. sampai kini umualah anam puluh tigo masih didalam keadaan misikin atauo mularaik jo baru iduik.
Karena jodoh tadi tidak dikira-kira dapat anak tujuh. Tiba delapan sama anak yang tua. Sampai sekarang umur sudah enam puluh tiga masih dalam keadaan miskin atau melarat juga baru hidup
Karena jodoh, kami dikaruniai tujuh orang anak, delapan dengan anak yang paling tua. Walaupun sudah berumur enam puluh tiga tahun, saya masih hidup miskin dan melarat

Section 11: 11 seconds
Cumano awak inda lupo ka Tuhan badoa karano awak urang Islam ado kaimanan dan ado pulo katabahan mudah mudahan ado rasaki dan bisa awak bausaho
Cumanya saya tidak lupa kepada Tuhan berdoa karena saya orang Islam ada keimanan dan ada pula ketabahan mudah-mudahan ada rejeki dan bisa saya berusaha
Meskipun demikian, saya tidak lupa untuk berdoa kepada Tuhan karena saya orang Islam, ada iman dan ketabahan sehingga mudah-mudahan saya akan diberikan rezeki supaya bisa berusaha.

3.1.2 Sekayu RA-RTT

Section 1 (Track 1): 17 seconds
Oyong ku kak lala miniah betugas jauh di Talang Buluh tu. Itu temasuk daera tepencilnye di MUBA kak
Kakak, saya ini sudah lamah sekali bertugas jauh di Talang Buluh itu. Itu termasuk daerah terpencilnya di MUBA ini.

Nah, desa Talang Buluh tu temasuk binaan Kantor Department Sosial.
Nah, desa Talang Buluh ini termasuk pembinaan Kantor Department Sosial.

Section 2 (Track 3): 12 seconds
Amon istila itu masyarakatnye masi mayarakat teasingnye. Situ dak diterti dingin tata kerama tubuk.
Kalau istilah itu masyarakatnya itu masih masyarakat terasing. Di sana tidak mengerti peraturan kita.

Amon makai hukum die, makai hukum alam.
Kalau memakai hukum mereka, memakai hukum rimba.

Section 3 (Track 5): 15 seconds
Perna sekali pada waktu kami bengaran pergi ke situ dadang uang tenga jalan ngunde kecipek
Pernah sekali terjadi pada waktu kami pertama kali pergi ke sana dihadang orang di tengah jalan yang membawa genapan rakitan

takate panjang sehinggonye laghasnye kak almon la netir uli baek carik bekanti bayan lanan.
berlaras sangat panjang sehingga saya ini gemetaran oleh karena cara suami saya (yg bersama saya).

Die pacak ngomong akhirnye bole liwan.
Dia bisa berbicara sehingga kami boleh lewat.

Section 4 (Track 7): 23 seconds
Amon turutke dak ngolong lagi ke situ. Tapi ole kenek nekat kak tadi akhirnye sampai mika kuka masi Talang Buluh tu la.
Kalau dituruti kita tidak kembali lagi ke sana, tetapi oleh karena nekat ini tadi, akhirnya sampai sekarang saya masih di Talang Buluh itu lah.

Na, dai carik kehidupan situ masyarakatnye tu, tanah banyak, betanam segan.
Nah, dari cara kehidupan masyarakatnya itu, tanah banyak (tetapi) bercocok tanam, malas (mereka).

Misalnye, numpang bae amonade lo metik bae, tinggal metik hasil baek.
Misalnya, menumpang saja, kalau ada kesempatan, ambil saja, tinggal memetik hasil saja.

Section 5 (Track 9): 14 seconds
Amon cari kami nale mandi kayo jau misalnye tu mandinye tu ke paye dalam pulo para jauh sekitar due ratus meter.
Kalau cara kami mau mandi, ke air jauh, misalnya mau mandi ke sungai kecil dalam kebun karet sejauh sekitar 200 meter.

Na mandi kok amon dak berombongan ku takut pulek ndak ku baek
Nah, kalau tidak bersama-sama orang lain, saya tidak berani mandi sendiri, walaupun mau.

Section 6 (Track 11): 15 seconds
Pernah olang mintar mandi ujo anak muridku "Bu, jangan mandi Bu, ade rimau, laperangan aking tuenye imau tu
Pernah ketika (saya) berangkat mandi, kata murid saya, "Bu, jangan mandi Bu, ada harimau, sudah kuning-kunigan, sangat tua harimau itu.

Itu la tue niyan. La koning bulu bulunye tu. La hasil, dak mandi sehaghai-haghai takunt.
Itu sudah tua sekali. Sudah kuning-kuningan bulunya. Alhasil, tidak mandi sehari itu (karena) takut.

Section 7 (Track13): 11 seconds
Na dem tu perna pulek sekali aghai nan de ke masi kurang nia masyarakat situ tau dingan aturannye.
Pernah, sudah itu pernah juga, sekali hari menandakan masih kurang sekali masyarakat di sana tahu dengan aturannya.

Ade budak belage duma sekola.
Ada anak sekolah berkelahi di rumah sekola.

Section 8 (Track 15): 10 seconds
Na budak belage kak melapor dengan baknye. Baknye kak sek dak ditau makmane carek singgo dateng-dateng ngunde kecepeng.
Nah, anak yang berkelahi ini mengadu kepada bapaknya. Bapaknya ini, karena tidak tahu cara (untuk urus masalah anak)
sehingga datang membawa senapan rakitan.

Section 9 (Track 17): 10 seconds
Singgonye guru hak dang dalam kelas kak niter dak bae pingsan. Datang ku masuk ke kelas satu tu panggilku budaknye.

Sehingga guru yang di dalam kelas ini gemeteran, tidak saja, pingsan. Datang dan masuk saya kelas satu dan saya panggil anak itu.

Section 10 (Track 19): 8 seconds
Ujoku "Jangan datang datang baknga ndak ngunde kecepeng mitu urusan dumah sekola kak urusan dumah sekola.
Saya bilang, "Jangan datang-datang bapakmu ke sekolah (kalau dia) mau membawa senapan rakitan seperti itu. Urusan rumah sekolah ini urusan (untuk) rumah sekolah.

Panggil baknge tu kuma sekola."
Panggil bapakmu ke rumah sekola."

Section 11 (Track 21): 9 seconds
Na, kunyogku budak tu manggil baknye dak datang datang kumah sekola.
Nah, saya menyuruh anak itu memanggil bapaknya, (tetapi dia) tidak datang ke rumah sekolah.

Na, kateku "Nga ndak bauba dak? Masi dak, nga ndak nakal duma sekola?
Nah, saya bilang "Kamu mau berubah, tidak? Masih mau tidak kamu mau nakal di rumah sekolah?"

Section 12 (Track 23): 8 seconds
Uju die, "dak!" "Amon dak na nga, sekola la. Amon ke nga masih nak nakal, nga dak usa sekola la."
Kata dia, "Tidak!" "Kalau kamu tidak mau (nakal), sekolah lah! Kalau kamu masih mau nakal, kamu tidak usah sekolah."

Na, upetnye budak tu bauba.
Nah, rupanya, anak itu sudah berubah.

3.1.3 Besemah RA-RTT

Section 1
ini cerite ku dulu e masalah beburu kelas 3 SMP beburu masih kecik sand di ghumah tu mbata anjing due ni mangke ini kan ape kicek-kicek anjing
ini cerita ku dulu masalah berburu kelas 3 SMP berburu masih kecil dari di rumah itu bawa anjing dua ini, kemudian ini kan apa kecil-kecil anjing
Ini ceritaku waktu dulu berburu ketika aku kelas 3 SMP. Aku berburu waktu masih kecil dari rumah membawa dua ekor anjing yang masih kecil-kecil.

Section 2
nge tu sate mbak balau buatane sandi di pagar ini ni pagar ghumah sekolahan ni guleh ngambek pule titu tok balau luncuk tu
waktu itu satu bawa tombak buatan dari pagar ini pagar rumah sekolah ini dapat ngambil juga itu untuk tombak lancip itu
Waktu itu aku membawa satu tombak yang dibuat dari pagar ini pagar sekolah yang aku ambil juga untuk tombak lancip itu,

Section 3
terus tu ke dusun gabung nga kance-kance udem gabung nga kance anjing la cukup la siap gale nyela berangkat tujuan kami dudu ke Ayek Petay di daerah Ayek Petay
setelah itu ke dusun bergabung dengan kawan-kawan sudah bergabung dengan kawan anjing sudah cukup sesudah siap semua sudah itu berangkat tujuan kami dulu ke Air Petay di daerah Air Petay
kemudian aku pergi ke dusun bergabung kawan-kawan dan semua anjing-anjing mereka, sesudah cukup (lengkap) dan sudah siap semua, kami berangkat ke tujuan ke Air Petay di daerah Air Petay.

Section 4
nah terus tu la lame titu e la nulusughi pamah kapo la naik tebing bulan puasa la payah benagh ame kekanceghani ade nek dek puase nek puase tu gi kami gha due a
kemudian sudah lama itu menjelajahi rawa itu dan naik tebing bulan puasa capek benar kalau teman-teman ada yang tidak puasa waktu puasa itu cuma kami berdua
Kemudian setelah lama menjelajahi rawa itu dan naik/memanjat tebing, karena masih bulan puasa, kawan-kawan menjadi sangat capek yang tidak puasa waktu itu cuma kami berdua.

Section 5
mangke itu masang lapon dai kami masang lapon mangke ngibogh ghepohan tu ade la anjing ne negogh tu la bejagale anjing tu nyela la lame betatay
sudah itu memasang jebakan, kami memasang jebakan sudah nyerang semak-semak itu ada anjing yang tahu sudah itu berlarian anjing sudah lama kemudian berkelahi
Setelah itu kami memasang jebakan (perangkap) ketika menembus semak-semak itu ada anjing yang tahu (ada sesuatu!), kemudian anjing itu berlari, tidak lama kemudian berkelahi.

Section 6
payah bagong tu engkas kanye bagong tu la kincot pule laudem di geget anjing ku rase nyela ini kami ajong anjing tulah dide kami tujah kapo
capek babi itu ternyata babi itu pincang (luka) juga sudah digigit anjing kurasa sudah itu kami suruh anjing itu lah tidak kami tombak
Ternyata ada babi yang capek dan pincang (terluka!). Aku rasa (babi) itu sudah digigit anjing kami menyuruh anjing itu, (babi) itu tidak kami tombak.

Section 7
mangke ini ngape udem anjing la di tambang gale agi kami tetai bagong tu di enjokah nga anjing tadi khan diket-diket sughang khan tape nengkoni dek keruan anjing ku ni maseh kecek kecek due butek
sudah ini sesudah anjing sudah diikat seluruhnya kami potong babi itu dikasih sama anjing tadi sedikit-sedikit semua masih kecil-kecil tahu anjingku ini masih kecil-kecil dua ekor
Kemudian setelah anjing-anjing itu kami ikat semuanya, kami memotong babi itu dan memberikannya kepada anjing tadi diberikan sedikit-sedikit karena anjing masih kecil aku tahu anjingku yan gdua ekor ini masih kecil-kecil.

Section 8
kami enjok daging bagong tadi kan dek tau lamenye tu pingsan anjing ku tu sutek kan pakse dipitong
kami kasih daging babi tadi tidak taunya lama-lama itu pingsan anjingku itu satu dipaksa digending
Kami memberikan daging babi tadi, tidak taunya lama-kelamaan anjingku itu pingsan sehingga yang satu terpaksa digendong.

Section 9
anjing ku tu masih bejagal masih mitung anjing la bingung aku ini ka lok mane ka di tinggalka dek ka nade masih ku pitong anye masih ngibogh agi kami ni ngibogh agi ni terus tu kance ni la kepayahan
anjing aku itu masih berlari masih digendong anjing sudah bingung aku ini harus bagaimana mau tinggalkan tidak mungkin masih kugendong setelah itu berburu lagi kami ini berburu lagi setelah itu teman ini sudah capek
Yang satu masih bisa berlari, tapi yang satu harus aku gendong. Aku kemudian bingung harus bagaimana, mau ditinggalkan tidak mungkin; (anjing) itu masih kugendong dan kemudian (kami) berburu lagi, kemudian teman saya ini sudah capek.

Section 10
tape bulan puase kance ni dek puase ade batang jambu ngembek buah jambu dai sambel begahdu
waktu bulan puasa teman ini tidak puasa ada pohon jambu ngambil buah jambu dulu sambil istirahat
Waktu bulan puasa teman ini tidak puasa, ada sebatang pohon jambu; diambilnya buah jambu itu dulu sambil istirahat.

Section 11
la lama tu ini ase nak mecah puase anye masih ku tahan-tahan ka maseh nyampai di ghumah tu
sesudah itu ada rasa mau batal puasa tapi masih kutahan-tahan ketika masih sampai di rumah itu
Kemudian, (aku) merasa ingin batal puasa, tapi masih aku tahan-tahan sepanjang perjalanan ke rumah.

Section 12
anjing tadi la negogh agi la bejagale agi di pamah tadi siko e tu dide dapat tuapebesak masok pamah dide dapat
anjing tadi berburu lagi sudah berlari lagi di rawa tadi satu tidak dapat habis besar masuk rawa-rawa tidak dapat
Anjing tadi berburu lagi dan berlari di rawa-rawa, tapi yang satunya tidak. Karena anjing yang satu ini mengejar babi besar tapi tidak dapat karena babi masuk rawa-rawa.

Section 13
la lame tu berupok-rupok la payah aku ni tadi masih ngecai anjing nga aku la kata payahe mane mitong anjing naik tughun tebing mane nyuguh-nyuguh batang kawe
sudah lama berpikir-pikir sudah capek aku tadi masih megang anjingnya aku sudah capek mana gendong anjing naik turun tebing apalagi menembus hutan pohon kopi
Tidak lama kemudian, (aku) pikir aku sudah capek dari tadi masih megang (menggendong) anjingnya. Aku sudah capek menggendong anjing naik turun tebing, apalagi menembus (melalui) hutan pohon kopi.

Section 14
kami tu dek lame la lame jage anjing ku tadi anye maseh agak lesu anye masih pacak be jalan la nyampai di jalan aspal tu begahdu dai
kami itu tidak lama sadar anjing aku tadi tapi masih agak lesu tapi masih bisa jalan telah sampai di jalan aspal itu istirahat dulu
Tidak lama kemudian, anjing itu sadar walaupun masih agak lesu; masih bisa jalan dan ketika sampai di jalan bersapal, kami istirahat.

END OF RTT TEXT

The following is the last part of the Besemah recorded story (not tested as required number of sections already available).

jerang nginai ini anjing tadi ade dide nek kurang nge tape sebelum pegi tadi diceki gale kudai babe butek babe butek balek ni tadi pule la nyampai jalan aspal tadi di kinai anjeng nge tadi la cukup ape belom ditunggu ka kudai engkase la cukup aku ni tadi maseh puase ini tape la nyampai Dasi di jalan ni la bebuke dek biye mbatak ayek dide makanan dek biye ade daging bagong mbatakan anye dide kami makan daging bagong nek anjeng bae titu la nyampei di dusun tu la dek tahan agi kan tape dusun kami nga dusun tetangge jauh juge la belaghi aku tu belaghi sand di dusun tetangge tu li saking hak bebuke nyela udemitu dide nganju agi lasung udem malame tu la gheit gale badan aku tu badan la saket gale kan mane la cughing gale puase ni tadi ini la ka tak payahe nahan kanye udem itulah

3.3 Post-RTT Questionnaire Template

Nama Peneliti: _____ Tanggal: _____

Desa: _____ Kecamatan: _____

Kabupaten: _____ Bahasa/Dialek_____

Jumlah orang yang hadir:_____ LL Tua _____ P Tua _____ LL muda _____P muda

II. RTT OUTAKE INFORMATION

Menurut Bapak/Ibu, berasal dari mana orang yang bercerita tadi? _____

Apa yang membuat Bapak/Ibu berpikir demikian? _____

Apakah orang itu berbicara bahasa _____ dengan baik? Ya / Tidak

Jika 'tidak': Mengapa?_____

Berapa banyak yang Bapak/Ibu bisa mengerti dari cerita barusan?

Semua / sebagian besar / sedikit / sangat sedikit / tidak sama sekali

Apakah bahasa yang dipakai dalam cerita itu halus, kasar, atau sedang-sedang saja? _____

Apakah cara dia berbicara sangat sedikit berbeda atau sangat berbeda dari cara Bapak/Ibu bicara? _____

Dimana letak perbedaannya? _____

***Optional questions**:*

Apakah orang-orang dari daerah itu sering datang kemari? Ya / Tidak

Dalam keluarga Anda atau di lingkungan tetangga sekitar, apakah ada:

Wanita-wanita dari daerah sana yang menikah disini? Ya / Tidak

Keluarga dari daerah sana yang pindah kemari? Ya / Tidak

Apakah Anda sering berbicara dengan mereka dalam bahasa_____? Ya / Tidak

Notes:

Appendix 2

Other appendices (wordlists, lexicostatistical matrix, RTT scores) can be found in the datafiles in McDowell and Anderbeck (2020).