An Initial Reconstruction of Proto-Lampungic: Phonology and Basic Vocabulary

Karl Anderbeck
SIL International

Lampungic isolects such as Komering, Pubian and Lampung are spoken by approximately one and a half million people in the western Indonesian provinces of Lampung and South Sumatra. Differing subgrouping hypotheses for Lampungic have been posited, whether with Malayic, or Malayo-Javanic, or, most recently, as an isolate within Western Malayo-Polynesian (WMP).

Drawing on a recent dialect study as well as earlier sources, this study is an initial attempt to provide parameters for understanding the place of Lampungic within WMP. Specifically, it is a reconstruction of the phonology and basic vocabulary of what is called Proto-Lampungic (PLP). A principled distinction is made between pre- and post-PLP innovations. The result is a bundle of features which together can be used to define a Lampungic subgroup and to distinguish it from its neighbors. The most important of these diagnostic features involve reflexes of PMP *R, *d and *j.

Although Lampungic is one of the most phonologically conservative WMP languages, its lexicon has absorbed myriad loanwords, particularly from Indonesian and Sumatran Malay.

1. Introduction

Lampung is a set of related Austronesian language varieties spoken by approximately one and a half million people on the southern part of the island of Sumatra. There are many unanswered questions about the history of the Lampung people and their language: How long has this group been where it is? Does their language contain any evidence of past migrations, and if so, from where? What relationship does Lampung have with other Austronesian languages including its neighbors?

This comparative study does not answer all the questions above, but is aimed at providing initial parameters for understanding the history of the Lampung (LP) language or languages. Namely it is a reconstruction of parts of the phonology and lexicon of Proto-Lampungic (PLP), including how it developed from Proto-Malayo-Polynesian (PMP), and how it changed into the various forms we find currently.

* Many people have contributed in one way or another to the development of this paper. I wish to thank the Indonesian Institute of Sciences for providing visas and permissions for this research, SIL International for providing the funding, my SIL colleagues who collected much of the data used herein, Chad White who wrote much of the first draft of this paper, and the Lampungic-speaking people who were so generous with their time and help, even protecting us from being robbed of our survey equipment! I would like to single out my friend Mas Ali from Menggala, Lampung, as someone who went above and beyond to help us dig deeper into the secrets of his language.
1.1. External classification

Dyen’s *A Lexicostatistic Classification of the Austronesian Languages* (1965) was one of the first to tackle the classification of Lampungic (LP). Under his SUNDIC HESION he put the Javo-Sumatra Hesion, Sasak, Balinese, the Malayic Hesion, and the Dayak Subfamily among others. He originally grouped LP under the MALAYIC HESION along with Madurese, Achehnese, Minangkabau, and Kerinci (1965:26). Since Dyen’s research the term MALAYIC has been constricted to apply only to Malay dialects and their close kin like Iban (Adelaar 1992). Also, being that Dyen’s classifications were based on lexicostatistics and given the large amount of Malay borrowings in LP, it has since become clear to scholars that LP is not as closely related to Malayic as would be implied by lexicostatistics.

Malcolm Ross (1995:78) gives twenty-four groups for the Western-Malayo-Polynesian (WMP) languages. He writes, ‘Group 18 contains only Lampung, of extreme south-east Sumatra. Although it has been suggested in the past that it belongs to the Malayic group, current opinion regards it as not yet classified (Blust, pers. comm., Nothofer 1988).’ Adelaar (2005a) brings forward Ross’s classification of LP without further comment. The question therefore remains relatively unexplored whether LP can be subgrouped with Malay or with any other language below the (Western) Malayo-Polynesian level. I attempt to demonstrate in this paper that the defining characteristics of Lampungic are:

1) loss of PMP *h in all positions with some irregular retention word-initially;
2) PMP *q > PLP *h;
3) retention of PMP *w;
4) very limited medial nasal excrescence;
5) limited consonant cluster reduction;
6) merger of PMP *R and *r in word-initial position;
7) syncope of PMP *R in CaRaC environments;
8) non-initial PMP *(e)R > PLP *(y);
9) conditioned merger of PMP *f and *d with PLP *r;
10) retention of PMP *z as PLP *j;
11) retention of the PMP four-vowel system, and of diphthongs *-ay, *-aw and *-uy;
12) shift in some instances of PMP *-ay and *-aw to i and u respectively (irregular areal feature);
13) PMP *-iw > PLP *(y)w;
14) epenthetic semivowel w or y inserted between low-high vowel combinations;
15) Nothofer’s (1985:294) SYSTEM 3 PAN numeral system.

The question is also raised below whether any of Adelaar’s (2005a) twenty-two other WMP subgroups shares enough of these features to make a convincing case for merger with another subgroup or for positing a shared intermediate node under WMP.

1.2. History and ecology of Lampung language

Comparative linguistics, archaeology and other disciplines have given us an understanding of the origins and general migration patterns of early Austronesian-speakers and approximately when they began to move into the regions we now call island Southeast Asia. It is thought that these speakers brought agricultural technology with
them, which allowed them (or at least their languages) in many cases to overwhelm the earlier inhabitants of the areas they entered. Bellwood (1999) believes that the lowland areas of Sumatra were probably quite thinly populated due to their unsuitability for foraging prior to the advent of agriculture, so it may not have been too difficult for groups like Lampung-speakers to establish themselves in this new territory without enduring the type of language interference we see in, say, eastern Indonesia.

However, history gives us a picture of substantial later contact between Lampung and other language groups including Javanese (JAV), Sanskrit (SKT), and, of course, Malay (MAL). In fact, for most of the past two millennia Lampung must have been under at least some level of domination by either the Malay ‘port authorities’ to its north like Srivijaya, or the agrarian Javanese kingdoms to its south and east. I argue in this paper that it is Malay influence which is of the highest degree in Lampung vis-à-vis the other potential donor languages. Sumatran Malay–speaking groups border and partially surround Lampung. Additionally, in modern times we have Indonesian, the national language. In his 1976 grammar of Way Lima, Lampung, Walker writes ‘The influence of the Indonesian/Malay language on Lampung is pervasive. Contacts with Malay go back hundreds of years. In the past decades the influence of the national language is even stronger, affecting the phonology, the grammar, and the lexicon of Lampung.’ Sections 4 (PLP lexicon) and 5 (Changes from PMP to PLP) give substantial attention to the thorny issue of teasing out Malay borrowings from what is truly Lampungic.

1.3. Internal classification

On the basis of compared sound systems, lexicon, sociolinguistic attitudes, and reported and measured intelligibility, Hanawalt et al. (In progress) conclude that LP can be divided into three major dialect clusters:

1) Lampung Api;
2) Komering;
3) Lampung Nyo.

LAMPUNG API and NYO are named after their respective words for ‘what’, while KOMERING is the name of the river which forms the homeland of the northernmost dialect cluster. Lampung Api (henceforth Api) is also often referred to as PESISIR, meaning ‘coastal’, while Nyo is often referred to as ABUNG, which is an important ethnonym within the Nyo grouping. See Map 1 for a visual illustration of their locations.

There are not many strong linguistic differences between Api and Komering; their relationship is more of a language chain than two completely separate clusters. The starkest differences are between Nyo and the rest of LP. Dialect differences are discussed in further detail in §3.4.

1.4. Previous research

Hanawalt et al. (In progress) lay out in detail the various linguistic and sociolinguistic studies that have been done on Lampungic isolects, so that is not repeated here. I just mention studies that have been of particular benefit to this historical-comparative look. Pre-1950s research on Lampung was detailed in Voorhoeve’s (1955) Critical Survey of
Map 1. Major dialect groupings within LP
Studies on the Languages of Sumatra, beginning with van der Tuuk in the 1860s. Reportedly there is a grammar of Komering and a few dictionary manuscripts, but I have not had access to these works. From the same time period but more accessible are the so-called Holle Lists (Stokhof 1987), a compilation of word lists gathered mostly around the turn of the twentieth century. Among these lists are four Lampungic word lists, with a fifth ‘pretender’ identified as ‘Nasal’ discussed below.

More recently, there have been four dictionaries made available in some form, including one Komering-to-Indonesian dictionary (Gaffar et al. n.d.) and two others focused on the Nyo dialect group (Junaiyah 1985 and Hadikusuma 1994). Noeh and Fadilah (1979) is the most comprehensive of the four and provides information on a few different Api groups as well as Nyo. The other three are not very comprehensive; Hadikusuma helpfully includes some information from fourteen dialect areas but is the thinnest of the bunch. I relied on two phonological descriptions, one of a Lampung Api area called Way Lima (Walker 1976) and the other of Komering (Abdurrahman and Yallop 1979). Arguably the most helpful resource for this study was Walker's (1975) *A Lexical Study of Lampung Dialects*, which included twelve word lists and an initial discussion of internal dialect divisions based on lexicostatistics. As part of a larger work on the people and agriculture of Sumatra, Yasuyuki Mitani (1980) wrote on the language varieties of South Sumatra Province. In his paper he commented on Walker's classification of the Lampung dialects and offered a theory of the history of the Lampung dialects. Unfortunately he did not publish most of the data upon which his conclusions were based. Aliana et al. (1986) describe thirteen speech varieties within Lampung Province using lexicostatistics to determine the most central variety. Also informative is Hadikusuma et al.’s (1996) *Adat Istiadat Daerah Lampung* (Custom and Tradition in the Lampung Area) which discusses some of the ethnic divisions and their backgrounds.

Most of the data used in this paper were gathered by SIL personnel from 2003 to 2005. Linguistic instruments included a 350-item word list, a sentence elicitation list, and targeted phonological and historical-comparative elicitation. Section 2.1 gives greater detail on the WHERE and WHY of data sources used in this paper.

1.5. Dialectology

Dialectology is concerned with defining dialect boundaries and developments within the language family. Collins (1989:237) says it this way: ‘The task of the dialectologist is to identify the splits which have yielded the contemporary network of dialects. In other words, delineating the history of a language, its diffusion, and its diversification, is the goal of dialectology.’

There are two basic models that have been followed to explain such splits: the tree model and the wave model. The tree model assumes a sharp division has occurred by a separation or migration of the language community. The wave model sees innovations like pebbles dropping into a pond of water. The ripples that are created move ever outward creating ‘a welter of isoglosses that crisscross one another’ within the area (Chambers and Trudgill 1998:91). This is sometimes referred to as diffusion.

This comparative study is firmly set within a dialectological framework. In a reconstruction of a language such as Lampung, where the entire language community lives within a single (albeit large) geographical area, one can expect to see substantial diffusion of linguistic innovations across dialect areas. The better the sampling, of both innovative and relic dialect areas, the further back in history one is able to go with a
reconstruction, and the greater one’s ability will be to accurately generalize to the whole language group.

1.6. Comparative method

This study employs the historical-comparative method to discover innovations within Lampung and retentions from PLP. In the comparative method we find regularly recurring phoneme correspondences that occur across sets of cognate forms found in the speech varieties being studied. These regular phoneme correspondence sets give an indication of what the original sound in the proto-language was. Establishing a set of these sounds we can begin to form the phonology and lexicon of the proto-language. Working from the proto-language back to the present day forms of the language we can see where changes have taken place (innovations) and where the segment has stayed the same (retentions). It is those retentions and innovations that help us determine the language classification and define dialect boundaries.

What I am seeking to define, through the use of the comparative method, is a significant intermediate stage between Proto-Malayo-Polynesian and the present where Lampung had differentiated itself from its (Western) Malayo-Polynesian kin but not yet undergone major internal dialectal differentiation. I label this stage Proto-Lampungic. That there ever was a stage where Lampung existed as an undifferentiated entity is a significant assumption, but we can see below that it holds up rather well as a working hypothesis.

1.7. WordCorr

Mechanical processing of the approximately 6500 Lampungic lexemes gathered was done with a computer program called WordCorr. It allows the user to keep track of the huge amounts of data involved with historical comparative linguistics, gives a structure for the data to be entered uniformly, and allows the user to manipulate and organize data according to several different analytical viewpoints. Several different language projects can be worked on at once within WordCorr as well and passed to other linguists for review and comments. It is able to generate an exhaustive list of correspondence sets along with any conditioning environments a linguist wishes to posit, along with reconstructions of all the words used for comparison.\(^1\) I used WordCorr to keep track of correspondence sets, to help me find the regularly recurring changes from Proto-Lampungic to modern Lampungic isolects, and to reconstruct PLP phonemes and word forms.

I shall now briefly detail the steps I went through in processing the data. After the word lists had been gathered and keyboarded, they were placed together in an Excel template, then processed with a program called PalmSurv Converter (http://sourceforge.net/projects/psconverter) to get into a form which could be imported (data and metadata) into WordCorr.

The next step was to decide which lists would be included in the analysis and in which order. This is called a VIEW; decisions made related to this are discussed in §2.1. One of the decisions which had to be made was the THRESHOLD percentage at which WordCorr would include correspondence sets for analysis. For example, in a collection of ten lists, for the gloss ‘lightning’ there may be three different cognate sets, one with five representatives and the others with just a few. The default percentage in WordCorr is set at 50%, which

\(^1\) For a more complete description of WordCorr see http://wordcorr.sourceforge.net.
basically means only one correspondence set per gloss will be selected for analysis. Given the large number of word lists in this collection and a desire to be as thorough as possible, I set the Threshold at 10%. At 10%, the threshold caught for analysis any word with at least three reflexes. This seemed appropriate given that my goal was not simply to reconstruct a proto-language but also catch and analyze reflexes of higher-order reconstructions, primarily Proto-Malayo-Polynesian. This requires a larger filtered data set.

Once the cognates are set into groups, the next task was to ALIGN them so that, in a word like rua 'two', the reflexes of *r lined up with each other, *u, *a, etc. Once the segments are lined up, I assigned a protosegment to each correspondence set of each cognate group, and an environment, like word-final, intervocalic, etc. Upon completing that task for all 350 glosses, I was able to look at all the correspondence sets together and REFINe my analysis, grouping or splitting environments or protosegments and noting any irregularities. For example, when I observed that reflexes of *ə in ultimate syllables behaved a certain way sometimes but differently in others, I was able to look at all the examples together and determine that a further conditioning environment needed to be specified to explain the extant reflexes. (This particular example is discussed in §3.4.1.)

Finally, I was able to export my analyses, which appear, after a bit of massaging, in the data sections of this paper.

2. Lampung today

2.1. Data sources

The primary data sources for this study, besides those mentioned in §1.4, are twenty-three word lists, some of which are 350 items in length, others 200. See Table 1 for details, including the order in which the data in later sections is presented, and the short names by which the data points are referenced in the remainder of the paper. Map 2 shows much of the same information in visual form.

One may notice that many of the SIL word list sites are also sites of lists from Walker (1975) and wonder why we revisited those sites. This following in Walker’s footsteps, so to speak, was deliberate, because Walker’s lists (excepting Way Lima) were not elicited by him, a trained linguist, but were rather produced by native speakers from those areas but living elsewhere, written in orthographic script. So one could say that the later SIL lists from these areas are simply a refinement and check on the earlier work. The exception to this pattern is Way Lima, where Walker himself had done extensive fieldwork, and we did not sample there. The SIL teams also sampled several areas from where Walker had not published lists. I am confident that the lists used in this study provide a fairly comprehensive geographical representation of the multiplicity of LP isolects.

One potential ‘data point’ needs to be mentioned here. One of the Holle lists (Stokhof 1987:143–157) is from the Nasal River in Bengkulu bordering the Krui area of Lampung. While the language variety represented by this list, gathered in 1895, at first blush seems to be Lampungic, I argue in §6.1 that it is most likely NOT, but does contain a very high number of Lampung loanwords including some archaisms useful for this reconstruction. As such, the NASAL list is occasionally used as a data source.

2 The SIL teams did other work in these areas including sociolinguistic questionnaires, sociolinguistic observation, recorded text testing, etc. See Hanawalt (this volume) and Hanawalt et al. (In progress) for further details.
Table 1. Word list sites

<table>
<thead>
<tr>
<th>Village</th>
<th>Short name</th>
<th>Dialect (Subdialect)</th>
<th>Source</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kayu Agung Asli</td>
<td>KAAAsli</td>
<td>Komering (Kayu Agung Asli)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Adumanis</td>
<td>Kom-Adu</td>
<td>Komering (Ulu)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Pulau Gemantung</td>
<td>KomIlir</td>
<td>Komering (Ilir)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Perjaya</td>
<td>Kom-Jaya</td>
<td>Komering (Ulu)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Damarpura</td>
<td>Kom-Dpur</td>
<td>Komering (Ulu)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Tihang</td>
<td>Daya</td>
<td>Komering (Daya)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Pilla</td>
<td>Ranau</td>
<td>Api (Ranau)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Buay Nyerupa</td>
<td>Sukau</td>
<td>Api (Sukau)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Banjar Agung</td>
<td>Krui</td>
<td>Api (Krui)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Kota Besi</td>
<td>Belalau</td>
<td>Api (Belalau)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Mesir Udik</td>
<td>WayKanan</td>
<td>Api (Way Kanan)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Kandang Besi</td>
<td>KotAgung</td>
<td>Api (Kota Agung)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Sukanegeri Jaya</td>
<td>TalaPada</td>
<td>Api (Talang Padang)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Way Lima</td>
<td>WayLima</td>
<td>Api (Way Lima)</td>
<td>Walker</td>
<td>200</td>
</tr>
<tr>
<td>Banjar Ketapang</td>
<td>Sungkai</td>
<td>Api (Sungkai)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Negeri Kepayungan</td>
<td>Pubian</td>
<td>Api (Pubian)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Tengkujuh</td>
<td>Kalianda</td>
<td>Api (Kalianda)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Nibung</td>
<td>Melintin</td>
<td>Nyo (Melinting)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Jabung</td>
<td>Jabung</td>
<td>Api (Jabung)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Paku</td>
<td>KAPend</td>
<td>Nyo (Kayu Agung Pendatang)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Nyampir</td>
<td>Sukadana</td>
<td>Nyo (Abung/Sukadana)</td>
<td>SIL</td>
<td>350</td>
</tr>
<tr>
<td>Blambangan Pagar</td>
<td>KotaBumi</td>
<td>Nyo (Abung/Kotabumi)</td>
<td>SIL</td>
<td>200</td>
</tr>
<tr>
<td>Ujung Gunung</td>
<td>Menggala</td>
<td>Nyo (Menggala/Tulang Bawang)</td>
<td>SIL</td>
<td>350</td>
</tr>
</tbody>
</table>

2.2. Phonology

This section summarizes the phonology of the Lampungic speech varieties based upon our research. Some selected differences with and between the two phonologies published by Walker (1976) and Abdurrahman and Yallop (1979) are also discussed briefly. It should be noted that Walker’s phonology was based on data from Way Lima, a village from the southern (Api) part of the region, while Abdurrahman’s is from a village in the Komering area. The phonology presented here attempts to look at the whole cluster of Lampungic speech varieties at once, thus variations between this presentation and what is actually found at any given location will differ to a small degree.

---

3 This section is taken near-verbatim from Hanawalt et al. (In progress) which functions as a sort of umbrella report for other subsidiary papers.
Map 2. LP word list sites