DigitalResources

Electronic Survey Report 2013-010

Linguistic Comparison of Semai Dialects

Timothy C. Phillips

Linguistic Comparison of Semai Dialects

Timothy C. Phillips

SIL International®
2013

SIL Electronic Survey Report 2013-010, July 2013
© 2013 Timothy C. Phillips and SIL International®
All rights reserved

Abstract

The Semai language of peninsular Malaysia is rich in diversity: each Semai village speaks its own variety. This report[1,2] documents the various Semai dialects and explores the relationships between them, with regard to choice of words as well as the phonological changes that affect the pronunciation of each. Systematic analysis shows the dialects to be different enough from each other that it is unlikely that any one dialect can be adequately understood by all speakers. Thus, in the strictest sense of the terms, it may be argued that the Semai varieties are a cluster of closely related "languages" rather than "dialects." Following local convention, however, the term "dialect" is maintained in this paper. This report lays the foundation for establishing which dialect or dialects could be used to provide an optimal means of communication across all the Semai dialects.

[1]This research was carried out under the auspices of the Economic Planning Unit (EPU), part of the Prime Minister's Department, Malaysia. My Malaysian counterpart was Dr Wong Bee Eng, lecturer at Universiti Putra Malaysia. The original report was filed with the EPU in 2005. This version has been revised for publication in 2013.

[2]I would like to thank the people of Malaysia for welcoming me and my family to their beautiful country and for their patience as we have struggled to learn their cultures and languages. I would especially like to thank Dr Wong Bee Eng for being willing to oversee my research and for her interest in the project. I would also like to thank the JHEOA for its gracious assistance in opening up its research library and helping me to gain access to even the most remote areas. I also must thank the many colleagues who suffered through early drafts of this paper offering their suggestions, and especially Cal Rensch who spent enormous effort helping me fine-tune the final draft. Finally, I wish most of all to thank the Semai people themselves who so willingly shared their language, their time, and often even their homes and food. It is my deepest hope that the information presented in this report will be able to help in the preservation and development of the Semai language and culture.

Contents

1 Introduction

1.1 Background

The languages of the Orang Asli in Malaysia are classified into two groups: Aslian and Proto-Malay. The Aslian languages form a branch of the Mon-Khmer language family and geographically range through most of the Malay Peninsula. Aslian is divided into three groups: Northern, Central, and Southern. The following languages make up the Central Aslian group: Semai, Temiar, Lanoh, Semnam, and Sabüm.[3] The target of this research is Semai, ISO code [sea], the Aslian language that has the largest number of speakers (Lewis 2009).

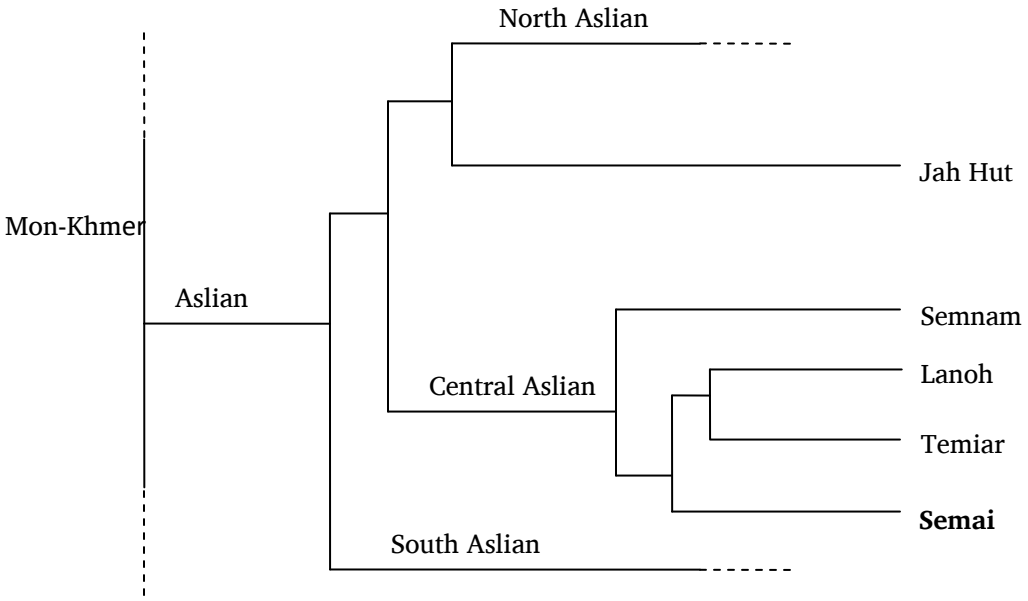


Figure 1. Aslian languages of Malaysia.[4]

The speakers of Semai are located primarily in the Malaysian states of Perak and Pahang. Traditionally, these people were swidden agriculturists living in the dense jungle, often moving their places of residence as well as moving their gardens. The Semai do hunting and gathering as well to supplement their diets. In more recent years, some Semai have resettled near towns and taken jobs as laborers, resulting in a more sedentary lifestyle.

[3]No Sabüm speakers have been located in many years; it is likely extinct. Historical documents indicate that Sabüm was quite close to Semnam (Phillips, forthcoming).
[4]This chart is derived from Phillips (forthcoming).

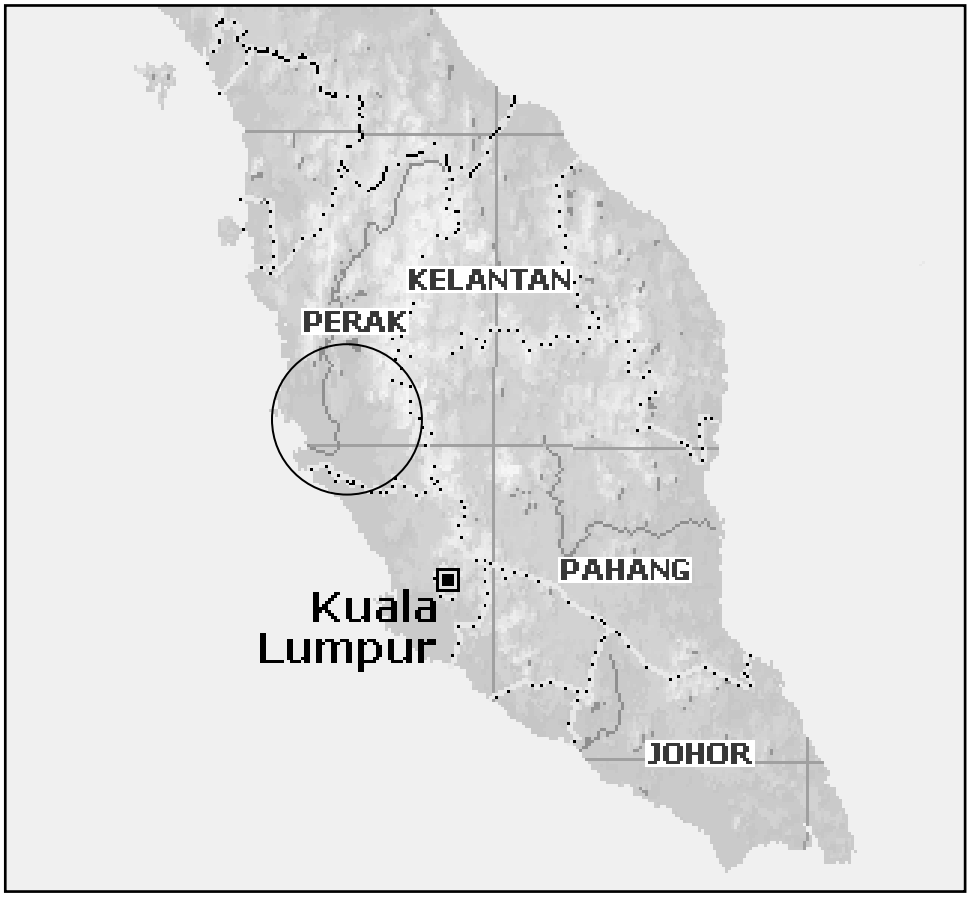


Figure 2. Approximate location of Semai on the Malay peninsula.

The exact number of speakers is difficult to ascertain because many Semai continue to move about in the deeper forest areas, and some others have moved to towns. However, the total number of Semai speakers has recently been estimated at 42,383.[5]

1.2 Preservation of the Semai language

Semai speakers are proud of their language and culture and have sought to preserve their language and culture despite interacting with other peoples for centuries. The importance of preserving indigenous languages has received much attention in recent years at the local, national, and even international level. The arguments for being involved in such preservation include the safeguarding of linguistic diversity, contributing to a knowledge base for language universals, and the belief that knowledge in and of itself is valuable. The languages of the Orang Asli in Malaysia should be considered a national treasure.

For the Semai language to be preserved, it must be studied and documented. The Semai language has been studied by only a few researchers, and while some quality work has been done,[6] the Semai language remains largely undocumented. This project collected wordlists representative of many of the dialects of Semai.[7] The wordlists were compared linguistically in order to determine how similar the

[5]According to population data for the year 2008 as provided by the JHEOA and displayed at the Orang Asli Museum in Gombak, Selangor.
[6]For example, Gérard Diffloth spent many years studying the Semai language and has written a number of papers regarding some of his findings.
[7]Diffloth (1977) has estimated there are more than forty dialects of Semai.

various dialects are. In order to preserve a written record of the Semai dialects, copies of the compiled wordlists have been turned over to the Jabatan Hal Ehwal Orang Asli as well as the Economic Planning Unit in the Prime Minister's Department.

Ultimately, for the *usage* of Semai to be preserved, some form of standardization will need to take place so that important decisions, such as orthography, can be effectively made. One of the key questions regards determining the optimal dialect or dialects that allow adequate communication with all speakers of Semai. Identification, documentation, and systematic comparison of the Semai dialects are critical first steps for standardizing Semai.

1.3 The contribution of Semai to historical linguistics

The Semai language, true to its Mon-Khmer heritage, has a rich set of vowels—nearly thirty, when counting all the nasal and length features. Furthermore, as Diffloth (1976a) has noted, Semai has preserved a number of disyllabic and polysyllabic words, features that have largely been lost in other Mon-Khmer languages in Southeast Asia. Thus the Semai people, as well as other speakers of Aslian languages in Malaysia, have much to offer humanity as we endeavor to reconstruct the history of the Mon-Khmer languages.

It is hoped that the documentation and the reconstruction of the Semai ancestor language in this report will help to further such efforts.

2 Methodology

2.1 Collection of wordlists

A wordlist of 436 words was constructed, including words from the basic 200 Swadesh wordlist, words that are typical of Southeast Asian languages, and words that are culturally and linguistically specific to the speakers of Central Aslian languages. The items in the wordlist were arranged by semantic categories and listed in Malay and English.

This wordlist was then used to elicit words from twenty-seven dialects of Semai. Dialects were selected based on a combination of information gleaned from existing literature on Semai and from asking the Semai themselves which areas spoke dialects different from their own. The following table shows the locations of the dialects selected for this research. A map showing the geographic locations of these villages is shown in Appendix A.

Table 1. Wordlist locations

Kampung	District	State
Batu 17	Batang Padang	Perak
Bidor	Batang Padang	Perak
Chinggung	Batang Padang	Perak
Cluny	Batang Padang	Perak
Rasau	Batang Padang	Perak
Sungai Bil	Batang Padang	Perak
Sungkai	Batang Padang	Perak
Tapah	Batang Padang	Perak
Gopeng	Kinta	Perak
Kampar	Kinta	Perak
Bota	Perak Tengah	Perak
Tangkai Cermin	Perak Tengah	Perak
Cenan Cerah	Cameron Highlands	Pahang
Relong	Cameron Highlands	Pahang
Renglas	Cameron Highlands	Pahang
Sungai Ruil	Cameron Highlands	Pahang
Terisu	Cameron Highlands	Pahang
Bertang	Lipis	Pahang
Betau	Lipis	Pahang
Cherong	Lipis	Pahang
Kuala Kenip	Lipis	Pahang
Serau	Lipis	Pahang
Lanai	Lipis	Pahang
Pagar	Lipis	Pahang
Simoi	Lipis	Pahang
Pos Buntu	Raub	Pahang

The wordlists were generally elicited using direct questioning in Bahasa Malaysia (Malay). Once the complete list was elicited, the data were rearranged according to the similar phonetic segments encountered. The list was then rechecked. By grouping the elicited words together according to similar sounds (for instance, all the words containing front vowels were put together), it was easier to hear the often-subtle differences between similar sounds.

In some cases a recording was also made of the same Semai speaker pronouncing the words that had just been elicited. These recordings were quite helpful in clearing up remaining inconsistencies later discovered in the elicited words, and thus often avoided the need to return to the same village for further checking.

The elicited wordlists were then used to determine the degree of linguistic similarity between dialects. The comparison of wordlists was used to determine the number of phonetically similar lexical items, to discover word families, to identify phonological changes in order to establish the linguistic relationship between the speech communities, and finally, to propose a reconstruction of several hundred lexical items for proto-Semai.

2.2 Language assistant questionnaires

Questionnaires were administered to many of the language assistants who gave the wordlists. The questions were mainly developed in order to help establish the reliability of the data. Information gathered from questionnaires was also used to determine which dialects still needed to be sampled, dialects that had not initially been selected based on the literature search.

2.3 Linguistic comparison and analysis

A variety of comparisons and analyses were carried out on the data. These included the following:

- Establishing the phonemes for each dialect
- Comparing the number of lexically similar items in each wordlist
- Determining phonological changes in each wordlist
- Correlating phonological changes and postulating a family tree for the Semai dialects
- Reconstructing lexical forms of proto-Semai
- Estimating the percentage of borrowed words for each dialect

3 Results

This section presents the results of the analytical methods used to understand the relationships that exist among the varieties of Semai represented by the wordlists collected during the current research. Despite the simplicity of the data collection method, there is a great wealth of information that can be drawn from the data.

First, although the Semai language has many dialects and variations, it is useful to consider the basic, overall phonology of the language. The various dialects are then compared to this basic phonology (see section 3.1 Basic Phonology).

Second, the wordlists are lexically compared to show the degrees of lexical similarity that exist among the different varieties of Semai (see section 3.2 Shared Word Families).

Third, an attempt is made to discern dialect boundaries by comparing shared word families; that is, looking for dialects that could be linked together by unique sets of related words (see section 3.3 Basic Lexical Comparison).

Fourth, consistent phonological changes from the norm are compared as another reflection of dialect boundaries (see section 3.4 Shared Phonological Changes).

Fifth, the historical comparative method is used to postulate a family tree for the Semai dialects and to reconstruct several hundred Proto-Semai words (see section 3.5 Comparative Reconstruction of Proto-Semai).

Lastly, some observations are made on topics that were observed but not rigorously investigated as part of this research (see section 3.7 Some Other Observations).

3.1 Basic phonology

This section presents a generalized description of words, syllables, and inventory of sounds in the Semai language.[8]

3.1.1 Words

Semai words, in good Mon-Khmer tradition, fit the following syllable template:

[8]For a more complete summary of the phonology of one dialect, see Appendix F.

$(C_3\ \ V_2\ \ (C_4)\ )$	$C_1\ \ V_1\ \ C_2$
Minor	Major

The final syllable is regarded as the major syllable; the penultimate syllable, if present, is regarded as the minor syllable. Semai words always have ultimate stress; that is, on the major syllable. While many Semai words have only one syllable, the majority of Semai words have two syllables. The minor vowel V_2 is usually very short, nonphonemic, epenthetic [ə], and its enunciation in any given word is often optional if the two consonants are easily pronounced without the epenthetic vowel. For this reason Semai roots are sometimes called "sesquisyllabic" since the minor syllable does not carry the same weight, phonetically or phonemically, as the major syllable. The following forms, examples from the Tapah dialect, are illustrative of Semai word shapes.

/liip/[9]	'to swallow'	/ŋ.kuuʔ/	'thunder'
/mɑt/	'eye'	/sə.lec/	'smooth'
/mɑ.niiʔ/	'rain'	/mər.gɑs/	'tiger'

There are a number of words that have minor syllables with minor vowel segments (V_2) other than [ə]; namely [ɑ], [i], and [u]. Diffloth (1968, 1976a) has shown that these segments may be phonologically conditioned in some cases and morphemes in other cases.[10] The following examples are from the Bota dialect.

/mɑ.muh/	'to bathe'	/sɑ.miiw/	'bear'
/pi.nuuy/	'wind (n)'	/ti.ɟiiʔ/	'snake'
/ku.reʔ/	'to dig'	/ku.rool/	'knee'

It is noteworthy that when the minor vowel segment (V_2) is [ɑ], [i], or [u], it is pronounced with greater length than when V_2 is [ə], roughly equal in length with V_1 in the major syllable when V_1 is not a long vowel. Stress remains on the ultimate syllable.

Also worth noting is that a reduced set of consonants are found in the C_4 position; namely, /r/, /l/, /m/, /n/, /ɲ/ and /ŋ/. However, the voiceless stops (/p/, /t/, /c/ and /k/) and fricatives (/s/ and /h/) do occur due to infixation, reduplication, and compound words.

While most roots are apparently either mono-, sesqui-, or disyllabic, there are examples of words with three syllables. The antepenultimate syllable acts as a minor syllable. The following examples, and for the rest of this section, are from the Terisu dialect.

/tə.mɑ.gii ʔ/	'forehead'
/bə.rɑ.poo ʔ/	'to dream'
/bə.lɑ.ʔɨɨr/	'green' (an expressive[11])
/hi.bət.bɛɛt/	'is sleeping'

[9]In this paper short vowels are represented by a single letter, and long vowels by a double letter. The latter is a departure from standard IPA. Another departure is that the palatal central approximant is represented by the symbol 'y' rather than the IPA standard 'j', which could easily be confused with the palatal voiced plosive 'ɟ' and with local orthographies, especially Bahasa Malaysia, which use 'j' to represent a voiced alveopalatal affricate. Lastly, 'ɑ' is used for the unrounded open central vowel, and 'ɒ' for the rounded open back vowel.

[10]Diffloth (1976a) claims, for example, that /-a-/ in certain minor syllables is a morpheme.

[11]Diffloth (1976b) discusses a word class called *expressives*, also known as *ideophones*. Diffloth demonstrates that expressives in Semai have a phonology that is different from other word classes, exhibiting sequences of sounds not found in the rest of Semai.

3.1.2 Syllables

In Semai every syllable has an obligatory nucleus and onset, and an optional coda. The nucleus is usually a vowel; however, there are some nasals that are syllabic as well in the minor syllable.

/m̩.pɔɔc/	'salt'	/n̩.tooy/	'big'
/ɲ̩.cees/	'shallow'	/ŋ̩.kuuʔ/	'thunder'

The onset and coda are consonants. The two basic syllable types are CV and CVC. The syllable type CV is found only in the minor syllable.[12] For example,

CV	/bɑ.hɑ.yɑʔ/	'crocodile'
	/tɑ.wɑɑk/	'butterfly'
	/kə.rɨɨk/	'to shiver'

All major syllables, and some minor syllables, have the syllable type CVC. For example,

CVC	/pɔɔc/	'to wait'
	/bə.lɨk/	'blunt'
	/kəl.ʔũũp/	'brain'

3.1.3 Basic inventory: consonants

When viewed as a whole, the Semai language can be said to have twenty-three consonants.

Table 2. Consonant phonemes in Semai

	Bilabial	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Plosive, voiceless	p	t	c	k	ʔ
Plosive, voiced	b	d	ɟ	g	
Preploded nasal[a]	^{b}m	^{d}n	^{ɟ}ɲ	^{g}ŋ	
Nasal	m	n	ɲ	ŋ	
Trill or flap		r			
Lateral approximant		l			
Fricative, voiceless			s		h
Central approximant	w		y		

[a]The preploded nasals are found only at the end of major syllables. There is good evidence to support the notion that preploded nasals occur only after oral (nonnasal) vowels. Assuming this proves to be the case, then preploded nasals would represent allophones of simple nasals at the end of major syllables, rather than contrastive phonemes. However, phonemic preploded nasals are, in fact, found in some related languages—for example, Kensiw (Bishop 1996) and Temiar (Benjamin 1976a).

[12]The Gopeng dialect, however, has lost glottal stops after long vowels in the major syllable, resulting in CV syllable types in those words. This loss of the glottal stop is further noted in section 3.4.

3.1.4 Basic inventory: vowels

When viewed as a whole, the Semai language can be said to have thirty vowels. Semai has both short and long vowels. The long vowels are not dramatically elongated. Indeed, it may be more accurate to portray the long vowels as the more "normal" and the short vowels as extra short. Overall, there are roughly twice as many words with long vowels as opposed to short in the major syllable.

Table 3. Oral vowels

Oral, long	**Front**	**Central**	**Back**
Close	ii	ɨɨ	uu
Close-mid	ee		oo
Open-mid	ɛɛ		ɔɔ
Open		ɑɑ	
Oral, short	**Front**	**Central**	**Back**
Close	i	ɨ	u
Close-mid	e		o
Open-mid	ɛ	ə	ɔ
Open		ɑ	

Table 4. Nasal vowels

Nasal, long	**Front**	**Central**	**Back**
Close	ĩĩ	ɨ̃ɨ̃	ũũ
Mid	ɛ̃ɛ̃		ɔ̃ɔ̃
Open		ɑ̃ɑ̃	
Nasal, short	**Front**	**Central**	**Back**
Close	ĩ	ɨ̃	ũ
Mid	ɛ̃	ə̃	ɔ̃
Open		ɑ̃	

3.1.5 Deviations from basic inventory: the preploded nasals

The southeastern dialects of Semai[13] generally have the full set of consonants. That is to say, they retain the preploded nasals. These dialects include Betau, Cherong, Pos Buntu, and Bertang. In one dialect, Kuala Kenip, the plosive is voiceless, but still has a voiced nasal release. For the rest of the dialects,[14] all of the preploded nasals have become simple voiceless plosives. Hence, for these dialects there are only nineteen consonants. The following examples show the form of these endings in three representative dialects.

[13]This region is the Raub district and the southeastern half of the Lipis district in Pahang.
[14]That is, the Cameron Highlands district of Pahang, the northwest part of the Lipis district in Pahang, and all dialects in Perak.

English	Malay	Voiced preploded nasals (e.g. Betau)	Voiceless preploded nasals (Kuala Kenip)	Voiceless plosive (e.g. Simoi)
'blood'	'darah'	[bə.hii^{b}m]	[bə.hii^{p}m]	[bə.hiip]
'to die'	'mati'	[dɑ^{d}n]	[dɑ^{t}n]	[dɑt]
'termite'	'anai-anai'	[gə.rɨɨ^{j}ɲ]	[gə.rɨɨ^{c}ɲ]	[gə.rɨɨc]
'to fly'	'terbang'	[hɛɛ^{g}ŋ]	[hɛɛ^{k}ŋ]	[hɛɛk]

Note that the reduction of the preploded nasal to a voiceless plosive has produced a number of homonyms in these dialects, in those cases where there already existed a phonologically similar word with a simple voiceless plosive. The following examples are given.

English	Malay	Dialects with preploded nasals (e.g. Betau)	Dialects where nasal is now a voiceless plosive (e.g. Simoi)
'foot'	'kaki'	[ɟu^{g}ŋ]	[ɟuk]
'to return'	'pulang'	[ɟuk]	[ɟuk]
'skinny'	'kurus'	[soo^{g}ŋ]	[sook]
'navel'	'pusat'	[sook]	[sook]

3.1.6 Deviations from basic inventory: shifted vowels

Two dialects surveyed have developed shifted vowels. What most dialects pronounce as a long, open central vowel [ɑɑ] is now pronounced in Gopeng and Kampar dialects as a long, open back vowel [ɒɒ]. The entry of this vowel into the back region has caused the open-mid /ɔɔ/ to be phonetically raised to [o̝o̝], and the close-mid /oo/ to be phonetically raised to [o̞o̞]. The following examples demonstrate this phenomenon.

English	Malay	Most dialects	Gopeng, Kampar	
'shoulders'	'bahu'	/gəl.pɑɑl/	/gəl.pɒɒl/	
'bone'	'tulang'	/ɟə.ʔɑɑk/	/ɟə.ʔɒɒk/	
'fire'	'api'	/ʔɔɔs/	[ʔo̝o̝s]	/ʔɔɔs/
'shadow'	'bayang'	/wɔɔk/	[wo̝o̝k]	/wɔɔk/
'woman'	'perempuan'	/kər.door/	[kər.do̞o̞r]	/kər.door/
'roof'	'atap'	/pə.look/	[pə.lo̞o̞k]	/pə.look/

3.1.7 Deviations from basic inventory: possible merged vowel

In at least one dialect near Tapah (Perak), the contrast between /oo/ and /ɔɔ/ appears to have been lost and these reflexes have merged into /oo/. So for this dialect (Batu Dua), there is evidently one less phonemic vowel than the rest. The neighboring village (Batu Tiga) still had this contrast, so it would appear at least from this data that the phenomenon is not widespread. This phenomenon was noticed during a chance encounter, however, and full wordlists from these villages were not collected.

English	Malay	Most dialects	Batu 2, Jalan Pahang (Perak)
‘hair’	‘rambut’	/sɔɔk/	/sook/
‘navel’	‘pusat’	/sook/	/sook/

3.1.8 Deviations from basic inventory: diphthongs

In general, diphthongs rarely occur in Semai. However, in the northern region of Pahang, the Telom River dialects in Lipis District and up to the eastern edge of Cameron Highlands District in Pahang, the Semai dialects (for example, Renglas, Lanai, and Serau) have a diphthong /əu/. This occurs in words where other dialects generally have words containing /oo/ before a glottal stop /ʔ/ or glottal fricative /h/.

English	Malay	Most dialects	Telom area, Lipis (Pahang)
‘dog’	‘anjing’	/cooʔ/	/cəuʔ/
‘to defecate’	‘berak’	/cɑ[h].cooh/	/cɑh.cəuh/

These same Telom River dialects (Renglas, Lanai, and Serau) also have the diphthong /ei/ before a glottal stop /ʔ/ in a few words that in other dialects generally is /ii/ or /ee/ before /ʔ/.

English	Malay	Most dialects	Telom area, Lipis (Pahang)
‘soil, earth’	‘tanah’	/tiiʔ/ or /teeʔ/	/teiʔ/
‘short’	‘pendek’	/ku.tiiʔ/ or /ku.teeʔ/	/ku.teiʔ/

3.1.9 Deviations from basic inventory: resyllabification

Diffloth (1977) noted that in the southern reaches of the Semai territory, the Semai dialects have changed /oo/ to /uwɑɑ/ before final alveolars and palatals (-t, -n, -r, -l, -s, -c, -ɲ, -y), and /ɛɛ/ to /iyɑɑ/ before final labials and alveolars (-p, -m, -w, -t, -n, -r, -l, -s). These changes, plus the fact that the southern dialects use quite a number of words unique to the region, give these dialects an especially peculiar sound. These changes were also attested in the data collected for this report, although there were not examples of the sound change before every segment listed above.

Resyllabification and sometimes simplification of the minor syllables have occurred to produce words that fit better into Semai’s sesquisyllabic constraints. Thus the consonants (‘y’ and ‘w’) in these sequences now occupy the initial consonant position in the major syllable.

English	Malay	Most dialects	Tanjung Malim area (Perak)
‘dry’	‘kering’	/soot/	/su.wɑɑt/
‘woman’	‘perempuan’	/kər.door/	/kə.du.wɑɑr/
‘to sleep’	‘tidur’	/bɛɛt/	/bi.yɑɑt/
‘bird’	‘burung’	/cɛɛp/	/ci.yɑɑp/

The approximants ‘y’ and ‘w’ are also found in the onset of major syllables in other words, so these phonological changes and resyllabification do not introduce any new phonemes.

3.1.10 Deviations from basic inventory: other vowel differences

The Telom River dialects previously mentioned (Renglas, Lanai, and Serau) have markedly different vowel qualities for certain vowels, as compared with the rest of the Semai dialects. Where most dialects exhibit the long close central vowel /ɨɨ/, in a good number of these words, the vowel is rounded to /ʉʉ/ in Lanai and Serau. In Renglas this vowel is rounded and backed to /oo/ in these same words. Furthermore, the segment that is commonly /oo/ in most other Semai dialects occurs unrounded and more central in these three dialects, as /əə/. The following examples illustrate these shifts.

English	Malay	Most dialects	Lanai, Serau	Renglas
‘house’	‘rumah’	/dɨɨk/	/dʉʉk/	/dook/
‘betel nut’	‘pinang’	/bɨ.lɨɨk/	/bɨ.lʉʉk/	/bɨ.look/
‘rat’	‘tikus’	/pə.rook/	/pə.rəək/	/pə.rəək/
‘wind’ (n)	‘angin’	/pooy/	/pəəy/	/pəəy/

Finally, it should be noted that many of the dialects showed a reduction of vowel length for long vowels before final laryngeals /h/ and /ʔ/. However, the data is not totally consistent in showing this, and there was often difficulty in hearing the distinction in vowel length. Furthermore, some dialects seem to show this phonological change more consistently than others. These changes were also identified by Diffloth (1977).

3.2 Shared word families

This section examines those dialects which are linked through shared word families. A word family in the current study is a word form of apparently common origin that is shared by a set of dialects. In the example below, the form /bɑyɑk/ is shared by the dialects represented by the designations H, V, and W, and represents a word family. The semantic item ‘shadow’ elicited two word families, the word family /bɑyɑk/ and the word family /wɔɔk/. For the current analysis, the greater number of word families shared uniquely by a set of dialects, the more those dialects are deemed related.

Consider the following three items from the wordlist, demonstrating word families that indicate distinct groupings of the various dialects.[15]

‘shadow’	/wɔɔk/	(B,C,E,F,G,J,K,L,M,N,O,Q,R,S,T,U,X,Y,Z,AA,BB,CC,DD,EE)
	/bɑyɑk/	(H,V,W)
‘black’	/bəlʔɪk/	(B)
	/bəlʔɑk/	(C,E,L,M,N,Q,R,S,T,X,BB,CC,DD,EE)
	/bɨlʔɛk/	(J,K)

[15]The capital letter codes are defined as following: B– Gopeng; C– Rasau; E– Bertang; F– Kuala Kenip; G– Tangkai Cermin; H– Cluny; J– Tapah; L– Batu 17; M– Kampar; N– Bidor; O– Bota; Q– Sungkai; R– Pos Buntu; S– Betau; T– Simoi Baru; U– Cherong; V– Chinggung; W– Sungai Bil; X– Sungai Ruil; Y– Serau; Z– Lanai; AA– Renglas; BB– Cenan Cerah; CC– Relong; DD– Terisu; EE– Pagar.

	/ɾɨŋɑh/	(F,T,Y,Z,AA)
	/rɨŋyɛh/	(G)
	/rɨŋɛh/	(O)
	/cəɾɛɛt/	(U)
	/hitɑp/	(H,V,W)
‘some’	/gɛʔ/	(B,G,J,K,L,M,O,Q,T,X,AA,BB,CC,DD,EE)
	/gɛɛʔ/	(C,E,N,U)
	/guʔgɛʔ/	(R,S,Y,Z)
	/gɛn/	(H,V,W)

In these cases, the southern dialects (H, V, and W) share a lexical innovation that is distinct from the other dialects. Indeed, there are a total of thirteen items in the wordlist for which the dialects (H, V, and W) have a shared word that is distinct from the rest of the dialects. Another nine items have a common word for the dialects (H, V, and W), but it so happens that one or two other dialects also have this word. Furthermore, the dialects (V and W) share a lexical innovation unique to only these two dialects for an additional twenty-three wordlist items. This evidence argues strongly for dialects V and W to be grouped together, with dialect H being closely related as well.

Another pair of dialects that share a lexical innovation is from the northwest area: G (Tangkai Cermin) and O (Bota). This pair shares sixteen words that are distinct from the other dialects, plus another nine words that are found in this pair and at most two other dialects.

Beyond the two dialect clusters just discussed, the picture is less clear, mostly because lexical innovations tend not to be unique for the other dialects. For example, while examination of the various word families reveals that two of the eastern dialects (F and S) very often share the same word, that word is almost always also found in a variety of other dialects, but never consistently the same set of dialects.

3.3 Basic lexical comparison

This section analyzes the wordlists by utilizing basic lexical comparison. Two aspects are considered: simple lexical similarity between each pair of wordlists and average lexical similarity for each dialect.[16]

3.3.1 Simple lexical similarity

The corresponding lexical items from each of the wordlists were compared to determine which are lexically similar and which are dissimilar. A limited number of phonological changes (see section 3.4) account for the vast majority of the differences between related words in the Semai dialects; hence it is not generally difficult to judge whether corresponding words between wordlists are related.

The percentage of words that are lexically similar and apparently cognate was calculated for each pair of dialects. The number of words compared between any two dialects averaged 410; however, the exact number varied slightly for each pair compared because each list has a few missing items. The resulting similarity percentage between each pair of dialects is presented in Table 5.

[16]In this section there are a few places where pairs of dialects that are geographically close and lexically closely related are represented by just one of the dialects. This was done to conserve space and reduce clutter in tables and figures. In these instances label Y represents both Y– Serau and Z– Lanai, and label BB represents both BB– Cenan Cerah and CC– Relong.

Table 5. Percentage of lexical similarity between Semai dialects

H– Cluny																						
59	**B– Gopeng**																					
63	76	**J– Tapah**																				
59	67	73	**L– Batu 17**																			
61	68	70	69	**M– Kampar**																		
64	67	73	68	67	**N– Bidor**																	
61	68	70	67	68	67	**G– Tangkai Cermin**																
60	68	69	65	67	68	83	**O– Bota**															
61	72	65	59	57	64	60	59	**C– Rasau**														
67	66	70	62	62	70	64	65	73	**Q– Sungkai**													
69	63	67	58	60	67	61	62	64	62	**V– Chinggung**												
69	60	64	58	59	65	60	59	61	60	82	**W– Sungai Bil**											
57	62	63	56	54	62	56	58	68	68	56	55	**U– Cherong**										
61	61	63	59	56	63	56	58	68	69	60	58	69	**R– Pos Buntu**									
59	63	64	59	56	62	55	57	70	67	60	59	69	75	**E– Bertang**								
57	64	68	60	57	62	59	60	69	69	58	56	70	71	73	**S– Betau**							
57	63	67	61	57	64	57	59	66	67	57	57	70	68	67	75	**T– Simoi Baru**						
57	61	63	59	55	62	56	59	66	67	57	57	70	70	74	77	76	**F– Kuala Kenip**					
57	65	68	62	58	62	59	59	65	70	55	55	70	67	71	75	75	77	**EE– Pagar**				
59	62	66	61	60	63	60	63	62	67	60	61	66	67	67	71	75	74	69	**Y– Serau**			
54	62	65	60	56	61	58	61	60	63	55	56	65	62	64	67	70	70	70	71	**AA– Renglas**		
56	64	67	61	57	61	57	59	60	66	55	54	66	66	67	69	73	70	74	71	75	**BB– Cenan Cerah**	
57	67	70	64	61	65	63	62	60	67	59	58	64	63	65	68	71	68	70	70	75	75	**DD– Terisu**
60	66	66	66	63	67	64	65	61	64	61	61	60	61	61	61	64	62	62	65	65	65	69 **X– Sungai**

The following dialect pairs have similarity percentages above 80 percent: G-O, V-W. Not surprisingly, both of these pairs are geographically close. The following sets all have similarity percentages between 75 percent and 80 percent: B-J, E-R, F-S-T-EE, T-Y, and AA-BB-DD. While these pairs of dialects are separated by a greater geographical distance, the pairs are always still neighbors. Figure 3 depicts pairs of lexically similar dialects with a minimum of 71 percent similarity.

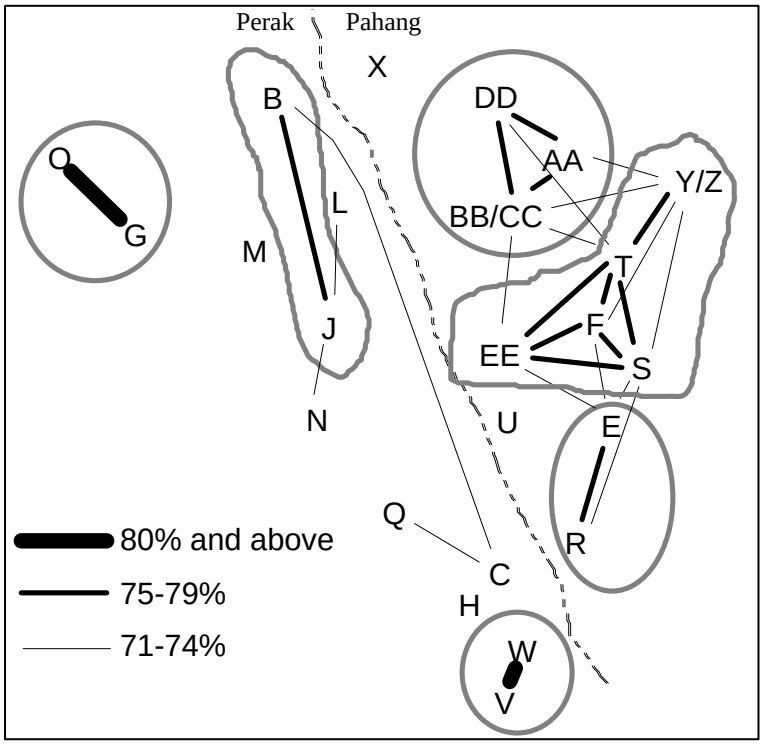


Figure 3. Map of lexical similarity between Semai dialects

While some groupings of high similarity do stand out (encircled on the map), perhaps the most salient feature of this map is the evidence of a classic dialect chain, or more accurately, a dialect network.[17] Moreover, the overall pattern shows a gap down the middle that strongly correlates with the mountain range that runs along the border between Perak and Pahang. This is unsurprising since mountains constitute natural barriers that hinder travel and thereby impede communication between dialects.

One surprise is the relatively high lexical similarity (72 percent) between dialect B (Gopeng) and C (Rasau), given that they are at opposite ends of the territory.

It should be noted that as with any analysis involving sampling, there is a range of error expected. A calculation was made based on the sample size and the degree of confidence in the reliability of the data, according to the procedure proposed by Simons (1977). Although the confidence level is estimated, it does provide some idea of what the range of error is and therefore what degree of difference is significant. For the current research, the value $N = 400$ was used (referring to the 400-plus samples in the wordlist), and a confidence level of "average" was assigned, based on the fact that all of the Semai speakers queried were adequately bilingual in Malay, and the researcher had some background in the Semai language. Using these values, the resulting calculation reveals that, when considering the lexical similarity statistics, a difference of three percentage points should be considered significant. This is an important factor when considering the previous tables and figures. In essence, what this means for this study is that one should not put too much emphasis on differences of just one or two percentage points.

[17]In a dialect network each dialect is lexically most related to its near neighbors and increasingly different with distance.

3.3.2 Average lexical similarity

Another useful indicator is the average percentage of lexical similarity with all other dialects. The average percentage for all lists is 64.1; that is, on average, 64.1 percent of the lexical items of each dialect are similar to those of other Semai dialects. The individual percentages for each dialect are shown in Table 6.

Table 6. Average lexical similarity

Percentage	Code	Location (kampung, district, state)
67.2	J	Tapah, Batang Padang, Perak
66.6	DD	Terisu, Cameron Highlands, Pahang
66.5	Y	Lanai, Lipis, Pahang
66.5	T	Simoi, Lipis, Pahang
66.4	CC	Relong, Cameron Highlands, Pahang
66.3	EE	Pagar, Lipis, Pahang
66.2	Q	Sungkai, Batang Padang, Perak
65.9	S	Betau, Lipis, Pahang
65.8	F	Kuala Kenip, Lipis, Pahang
65.8	BB	Cenan Cerah, Cameron Highlands, Pahang
64.9	AA	Renglas, Cameron Highlands, Pahang
64.8	B	Gopeng, Kinta, Perak
64.6	E	Bertang, Lipis, Pahang
64.6	N	Bidor, Batang Padang, Perak
63.7	X	Sungai Ruil, Cameron Highlands, Pahang
63.7	R	Pos Buntu, Raub, Pahang
63.4	C	Rasau, Batang Padang, Perak
63.4	U	Cherong, Lipis, Pahang
62.4	O	Bota, Perak Tengah, Perak
62.2	L	Batu 17, Batang Padang, Perak
61.7	G	Tangkai Cermin, Perak Tengah, Perak
60.6	V	Chinggung, Batang Padang, Perak
60.4	M	Kampar, Kinta, Perak
59.8	H	Cluny, Batang Padang, Perak
59.5	W	Sungai Bil, Batang Padang, Perak

These averages can be placed on the map of the Semai territory to show the geographical distribution.

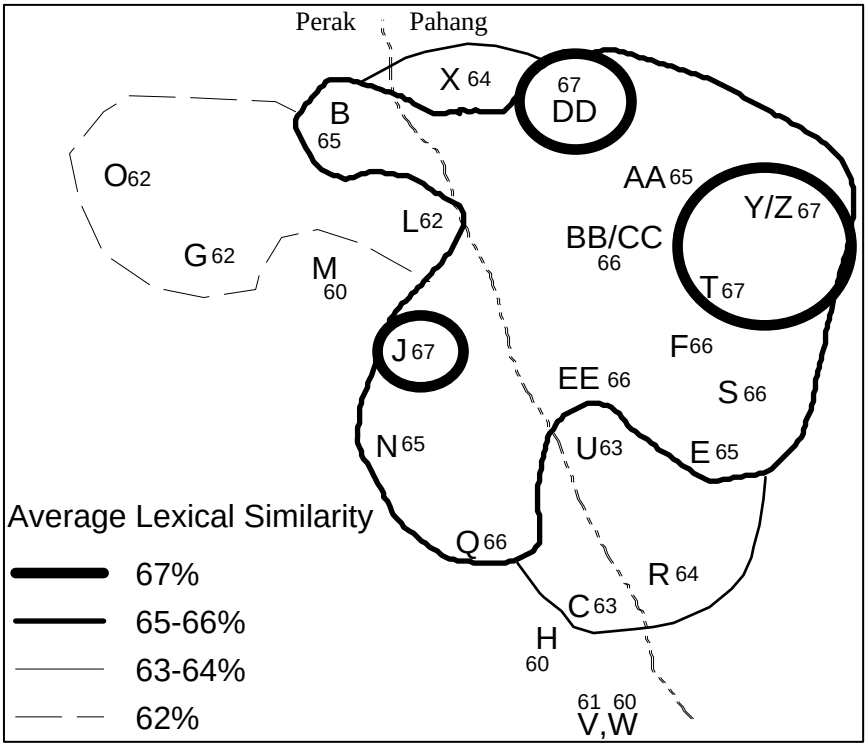


Figure 4. Distribution of average lexical similarity.

Although the average percentages do not have a particularly wide range, it is notable that Tapah (J) has the highest average lexical similarity with other dialects: 67.2 percent. Given the centrality of this dialect, plus being at the crossroads of the common travel routes, the figure is not especially surprising. However, by the same reasoning, it might be expected that Kampar (M) also might be high on the list, whereas, in fact, it is quite low. It is noteworthy that the language speaker that gave the Kampar list had been living in the Kuala Lumpur area for two years. Hence it seems reasonable to believe that he may be beginning to forget some of his language. For instance, this speaker gave the Malay word /pəlɑŋi/ for 'rainbow', whereas every other dialect gave the word /cədɑɑw/. Overall, twenty-two percent of the wordlist from the Kampar speaker were borrowed words from Malay. For comparison, the neighboring dialect of Tapah showed approximately twelve percent borrowed words. (This topic is explored further in section 3.6.)

On the other side of the ranking, Sungai Bil (W) has the lowest average lexical similarity compared with other dialects: 59.5 percent. This dialect represents the southern extreme of the range.

It is also noteworthy that the dialects that are statistically only weakly related tend to be at the extreme reaches of the overall Semai territory. This is not surprising since not only would a dialect near the outer edge of the territory have less contact with other dialects than if it were more centrally located, but such a dialect would also presumably have more contact with neighboring languages and cultures. Contact with other languages and cultures is often a major source of borrowed words.

However, once again the data also provides a few puzzling counter-examples. The Terisu (DD) dialect, at the northern extreme of the territory and known to have significant contact with its neighboring language, Temiar, has a surprisingly high position in the relatedness ranking. The Lanai (Y) dialect, at the northeast extreme, also has a surprisingly high average lexical similarity.

Overall, the averages seem low, evoking the question of why even near neighbors have at best 83 percent lexical similarity. There are a number of possible explanations, some of which will be presented here. First of all, it may indeed be the case that even near neighbors have a significant number of different words. Diffloth (1977) reported visiting 117 Semai settlements and found no two settlements to be identical in their choice of words.

Secondly, it is quite likely that a given item on the wordlist has a variety of possible correct responses due either to synonymy or to specific-generic mismatches. Synonymy can come from borrowings and from word taboos. Diffloth (1980) discusses at length Semai word taboos, especially associated with animals. Regarding specific-generic mismatches, there may be a number of words for 'spider', depending on the species or characteristics (e.g. large or small, indoor or outdoor, etc.). One village visited (Kampung Leryar, near Kampung Relong (CC)) had at least four different words for different types or sizes of spiders, while a village a few kilometers away (Renglas (AA)) had just one word for all spiders.

'spider'	/məŋ/	(smallest)	Kampung Leryar, near Relong (CC)
	/tɑ.wiik/	(small)	
	/gə.loow/	(large)	
	/məŋ.hɨ̃l/	(largest)	
	/tɑ.wiik/	(all types)	Kampung Renglas (AA)

Thirdly, some items on the wordlist sometimes present problems because they do not correspond one-to-one with Semai words or semantic categories. For instance, finding the word for 'throw' was particularly problematic. The wordlist was changed to elicit particular types of throwing, such as throwing sidearm, throwing a spear, tossing a ball, and throwing away garbage. Even then, problems emerged, which would tend to indicate the Semai do not organize their words that involve throwing the way that the Malays do. For more comments on problematic items in the wordlist, see Appendix B.

More than one Semai speaker north of the Tanjung Malim area believed that the southern Semai dialects had borrowed significant numbers of words from Temuan, the aboriginal Malay language spoken directly south of the Semai range. However, this is only hearsay evidence, and it may be hard to distinguish the true source, since Temuan and Malay are closely related.

Dialect L (Kampar) has a relatively low average lexical similarity (60 percent) with other dialects. The low average is unexpected since Kampar is centrally located along well-traveled routes. However, as previously noted, the speaker from the Kampar dialect may have been starting to forget some of his language.

The Kampar case demonstrates how a particular language speaker can have significant effect on the overall percentage due to his or her personal history and knowledge. This is an important reminder that it is not just the sample size that is important, but also that the circumstances of the individual language speakers need to be considered—factors such as how much a given speaker has moved around, which other languages he or she speaks and in what contexts, education level, where the speaker currently lives and works, and so forth.

3.3.3 Limitations of lexical similarity analysis

Lexicostatistics is but one tool for analyzing dialect similarities. While this tool is useful, it has its drawbacks. For instance, the historical linguist may readily identify the phonological changes that have taken place and thereby see the lexical similarity between two dialects for a given lexical item. However, this type of information may well be beyond the grasp of the individual speakers. For instance, in the Gopeng dialect the long /ɑɑ/ vowel has become backed and rounded to /ɒɒ/; furthermore, the glottal plosive after long vowels has often been dropped in this dialect. Hence the word /*cɑɑʔ/ 'to eat', which is still pronounced /cɑɑʔ/ in almost every dialect, is pronounced /cɒɒ/ in Gopeng. This word has sometimes been known to be mistaken for /cooh/ 'to defecate' by Semai from other dialects.

3.4 Shared phonological changes

There are a number of phonological changes that have affected various cognates in the Semai dialects. Dialects that share a phonological change are more similar, at least for that change, than dialects that

did not undergo the same phonological change. The greater number of phonological changes in one dialect but not in another, the greater the degree of distinction between those two dialects. This distinction is in addition to differences in lexical items, as discussed in section 3.3.

3.4.1 Phonological changes

The following is a list of phonological changes found in the Semai dialects studied in this survey.[18,19]

(a) Final preploded nasals became voiceless plosives (in most dialects, except S– Betau, E– Bertang, U– Cherong, R– Pos Buntu, and F– Kuala Kenip, the last of which has a voiceless plosive but is still a preploded nasal).

Example:

/*liiᵇm/[20] 'to swallow'

/liiᵇm/	(E,R,S,U)
/liiᵖm/	(F)
/liip/	(B,C,G,H,J,L,M,N,O,Q,T,V,W,X,Y,Z,AA,BB,CC,DD,EE)

(b) Final glottal stops were lost after long vowels (B– Gopeng).

Example:

/*cɛɛʔ/ 'head louse'

/cɛɛʔ/	(C,F,G,H,J,K,L,N,O,Q,R,S,T,U,W,X,Y,AA,BB,CC,DD,EE)
/cɛɛ/	(B)

(c) /*oo/ became unrounded and centralized to /əə/ (AA– Renglas) or /ɨɨ/ (Y– Serau, Z– Lanai) in all environments[21] except before final /*-h/ or /*-ʔ/, where it became the diphthong /əu/.

Example:

/cooᵍŋ/ 'rattan'

/cook/	(B,C,G,H,J,L,M,N,O,Q,T,V,W,X,BB,CC,DD,EE)
/cooᵍŋ/	(E,R,S,U)
/cooᵏŋ/	(F)
/cɨɨk/	(Y,Z)
/cəək/	(AA)

/cooʔ/ 'dog'

/cooʔ/	(C,J,K,L,M,N,O,Q,R,U,V,W,X,BB,CC,DD)
/coʔ/	(E,F,G,H,S,T,EE)

---

[18]Nearly all of the data collected was congruent with Diffloth's analysis (1977). Indeed, I have relied heavily upon Diffloth's proposed Proto-Semai forms as a basis for my hypotheses regarding proto-forms.

[19]The phonological changes are presented in order of the average number of words affected, with the change affecting the most words first. The relevance of this ordering will be addressed in the following section.

[20]The asterisk (*) is used to designate a reconstructed element of a proto language, in this case a reconstructed lexical item. A proto language is an ancestral form of a language from which modern varieties presumably have developed.

[21]In this section the phrase "in all environments" refers to the consonants on either side of the given segment.

/coo/ (B)

/cəuʔ/ (Y,Z,AA)

(d) /*ɑɑ/ became backed and rounded to /ɒɒ/ in all environments (B– Gopeng and M– Kampar).

Example:

/*kərɑɑl/ ‘man, male’

/kərɑɑl/ (C,E,F,G,H,J,L,N,O,Q,R,S,T,U,V,W,X,Y,Z,AA,BB,CC,DD,EE)

/kərɒɒl/ (B,M)

(e) /*ee/ became centralized and raised to /ɨɨ/ in all environments (B– Gopeng, M– Kampar, and J– Tapah).

Example:

/*b[ɑʔ]heeʔ/ ‘satiated’

/bɨheeʔ/ (C,F,G,N,O,T,U,V,W,AA,BB,EE)

/bɨheʔ/ (E,H,R,S,Y,Z)

/bɑʔheeʔ/ (L,Q,CC,DD)

/bɑheeʔ/ (X)

/baʔhɨɨ/ (B)

/bɑʔhɨɨʔ/ (J,M)

(f) /*ɤɤ/ became rounded to /oo/ (AA– Renglas), centralized, raised, and rounded to /~~ʉʉ~~/ (Y– Serau, Z– Lanai), but simply centralized and raised to /ɨɨ/ in all other dialects. This change happened in all environments.

Example:

/*dɤɤ^{g}ŋ/ ‘house’

/dɨɨk/ (B,C,G,H,J,L,M,N,O,Q,T,V,W,X,BB,CC,DD,EE)

/dɨɨ^{g}ŋ/ (E,R,S,U)

/dɨɨ^{k}ŋ/ (F)

/d~~ʉʉ~~k/ (Y,Z)

/dook/ (AA)

(g) /*u/ became lowered to /o/ in all environments (M– Kampar, X– Sungai Ruil, and DD– Terisu).

Example:

/*ɟu^{g}ŋ/ ‘foot’

/ɟok/ (M,X,DD)

/ɟuk/ (B,C,H,J,L,N,O,Q,T,V,W,Y,Z,AA,BB,CC,EE)

/ɟu^{g}ŋ/ (E,R,S,U)

/ɟu^{k}ŋ/ (F)

(h) Final palatal consonants /*-c/ and /*-ɟɲ/ first merged as /*-c/, and then shifted to /-t/ or /-k/;[22] also, final /*-ɲ/ shifted to /y/ (V– Chinggung, W– Sungai Bil, and some lexical items at H– Cluny). This phonological change may well still be in process, since a couple of the older people interviewed seemed to still have a few palatal final consonants, whereas the younger adults appeared to have shifted them (either in place of articulation or in manner of articulation). More research is needed on this dialect to determine the demographics (e.g. age, gender, and location) of those who are yet to shift the final palatal consonants, as opposed to those who have already shifted them.

Examples:

/*pɔɔʲɲ / 'to wait'

/pɔɔt/	(V,W)
/pɔɔʲɲ/	(E,R,S,U)
/pɔɔᶜɲ/	(F)
/pɔɔc/	(B,C,G,H,J,L,M,N,O,Q,T,X,Y,Z,AA,BB,CC,DD,EE)

/*sɛc / 'flesh'

/sɛk/	(H,V,W)
/sɛc/	(B,C,E,F,G,J,L,M,N,O,Q,R,S,T,U,X,Y,Z,AA,BB,CC,DD,EE)

/*ləmũũɲ/ 'tooth'

/ləmũũy/	(V, but also G)[23]
/ləmũũɲ/	(B,C,E,F,H,J,L,M,N,O,Q,R,S,T,U,W,X,Y,Z,AA,BB,CC,DD,EE)

(i) /*iə/ became /ii/ (dialects E, F, L, J, L, M, N, R, S, T, U, X, BB, CC, DD, EE), /ee/ (dialects B, C, G, H, O, Q, V, W, Y, Z), or /ɛɛ/ (dialect AA) in all environments.

Examples:

/*ləhiəᵍŋ/ 'saliva'

/ləhiik/	(J,K,L,M,N,T,X,Y,Z,BB,CC,DD,EE)
/ləhiiᵍŋ/	(E,S,U)
/ləhiiᵏŋ/	(F)
/ləheek/	(B,C,G,H,O,Q,V,W)
/ləhɛɛk/	(AA)

/*rəʔiəs/ 'root'

/rəʔiis/	(E,F,H,M,N,R,S,U,BB,DD,EE)
/rəʔiiy/	(J,L,CC)[24]

[22]So far there is not enough data to determine which environments produce /-t/ and which produce /-k/, if indeed the shift is predictable from the environment.

[23]The W– Sungai Bil speaker was an older woman in her fifties. As mentioned in this section, some of the older speakers still maintained some final palatal consonants, as attested in this example where one would expect /ləmũũy/ if all final palatals had shifted.

[24]In some dialects final /-s/ has become /-y/ in some words. The conditions under which this happens are unclear. Some words seem to have undergone this change completely; that is, they are consistently pronounced with a final

/rəʔees/ (B,C,G,O,Q,W,Y,Z)

/rəʔɛɛs/ (AA)

(j) /*oo/ became /wɑɑ/ before alveolar (*-t, *-^{d}n, *-n, *-r, *-l, *-s) and palatal (*-c, *-^{j}ɲ, *-ɲ) final consonants (H– Cluny, V– Chinggung, and W– Sungai Bil).

Examples:

/*cəŋroos/ 'fingernail'

/cənwɑɑs/ (H,V,W)

/cənɾoos/ (C,E,F,Q,R,S,T,U,BB,CC,EE)

/cəŋɾoos/ (B,G,M,N,O,X,DD)

/cəŋrooy/ (J,L)

/cənɾɨɨs/ (Y,Z)

/cənɾəəs/ (AA)

/*ɲĩɲõõy/ 'lip'

/ɲũwɑɑy/ (H,V,W)

/ɲĩɲõ̃õ̃y/ (B,E,F,R,S,T,Q,U,Y,Z,BB,CC,EE)

/ɲĩɲũũy/ (G,O,DD)

/ɲĩɲõõy/ (C,J,N)

/ɲũũy/ (X)

/ɲĩɲə̃ə̃y/ (AA)

(k) /*Noo/[25] became /Nuu/ (B– Gopeng, G– Tangkai Cermin, O– Bota, X– Sungai Ruil, and DD– Terisu), or /Nə̃ə̃/ (AA– Renglas), or /Nɔɔ/ (all other dialects).

Example:

/*nõõŋ/ 'path, road'

/nũũŋ/ (B,G,O,X,DD)

/nɔɔŋ/ (C,E,F,H,J,L,M,N,Q,R,S,T,U,V,W,Y,Z,BB,CC,EE)

/nə̃ə̃ŋ/ (AA)

(l) /*ɛɛ/ and /*iə/ became /yɑɑ/ before final labials (*-p, *-^{b}m, *-m, *-w) and final alveolars (*-t, *-^{d}n, *-n, *-r, *-l, *-s) (H– Cluny, V– Chinggung, and W– Sungai Bil).

Examples:

/*bɛɛt/ 'to sleep'

/biyɑɑt/ (H,V,W)

/bɛɛt/ (B,C,E,F,J,L,M,N,Q,R,S,T,U,X,Y,Z,AA,BB,CC,DD,EE)

/-y/. However, other words clearly still have a final /-s/, and another set of words seem to have free variation, sometimes pronounced with a final /-s/ and sometimes with a final /-y/ by the same speaker.

[25]The 'N' symbol represents any nasal consonant (/m, n, ɲ, ŋ/). Hence /*Noo/ means "proto-vowel /oo/ after a nasal consonant."

/*siə^{b}m/ ‘to forget’

/siyɑɑp/	(H,V,W)
/siip/	(J,L,M,N,T,X,Y,Z,BB,CC,DD,EE)
/seep/	(O,Q,B,C,G)
/sɛɛp/	(AA)
/sii^{p}m/	(F)
/see^{b}m/	(E,R,S,U)

3.4.2 Other phonological changes

Diffloth (1977) proposed several other phonological changes that were also seen in the data collected for this study. However, where there were fewer than five examples of these remaining changes in this corpus, it was deemed insufficient to present in this report.

3.4.3 Summary of phonological changes

Table 7 summarizes the phonological changes for the dialects researched in this study.

Table 7. Summary of phonological changes

Change	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**
Proto-form[a]	*-ᵇm	*-VVʔ	*oo	*ɑɑ	*ee	*ɤɤ	*u	*-c,*-ɲ	*iə	*oo /-T,-C	*Nõõ	*ɛɛ,iə /-P,-T
B– Gopeng	-p	ʔ lost		ɒɒ	~~ii~~				ee		ũũ	
C– Rasau	-p								ee		ɔ̃ɔ̃	
E– Bertang	-ᵇm								ii		ɔ̃ɔ̃	
F– Kuala Kenip	-ᵖm								ii		ɔ̃ɔ̃	
G– Tangkai Cermin	-p								ee		ũũ	
H– Cluny	-p							some	ee	wɑɑ	ɔ̃ɔ̃	yɑɑ
J– Tapah	-p				~~ii~~				ii		ɔ̃ɔ̃	
L– Batu 17	-p								ii		ɔ̃ɔ̃	
M– Kampar	-p			ɒɒ	~~ii~~		o		ii		ɔ̃ɔ̃	
N– Bidor	-p								ii		ɔ̃ɔ̃	
O– Bota	-p								ee		ũũ	
Q– Sungkai	-p								ee		ɔ̃ɔ̃	
R– Pos Buntu	-ᵇm								ii		ɔ̃ɔ̃	
S– Betau	-ᵇm								ii		ɔ̃ɔ̃	
T– Simoi	-p								ii		ɔ̃ɔ̃	
U– Cherong	-ᵇm								ii		ɔ̃ɔ̃	
V– Chinggung	-p							shifte	ee	wɑɑ	ɔ̃ɔ̃	yɑɑ
W– Sungai Bil	-p							shifte	ee	wɑɑ	ɔ̃ɔ̃	yɑɑ
X– Sungai Ruil	-p						o		ii		ũũ	
Y– Lanai	-p		~~ii~~			~~ɯɯ~~			ee		ɔ̃ɔ̃	
Z– Serau	-p		~~ii~~			~~ɯɯ~~			ee		ɔ̃ɔ̃	
AA– Renglas	-p		əə			oo			ɛɛ		ə̃ə̃	
BB– Cenan Cerah	-p								ii		ɔ̃ɔ̃	
CC– Relong	-p								ii		ɔ̃ɔ̃	
DD– Terisu	-p						o		ii		ũũ	
EE– Pagar	-p								ii		ɔ̃ɔ̃	

[a]In this header row the '-P' represents any final labial consonant (/-p, -m, -w/); the '-T' represents any final alveolar consonant (/-t, -n, -r, -l, -s/); and the '-C' symbol represents any final palatal consonant (/-c, -ɲ, -y/). Note also that change 'c' refers to where the proto-form *oo changed in all environments except before final /*-h/ or /*-ʔ/.

In Table 7 the letters in the first title row refer to the phonological changes listed in the previous section. The second title row refers to the proto-Semai form. The table is filled in where the particular dialect is different from the usual form found in the majority of other dialects. For instance, column 'f' refers to the proto-Semai vowel /*ɤɤ/. In the majority of dialects, this proto-vowel became present-day /~~ii~~/, but this is not shown in the table. Rather, the table is filled in where the proto-vowel became rounded to /oo/ (Renglas) or /~~ɯɯ~~/ (Lanai and Serau), which is different from the majority of the dialects.

When no one form is especially common, the form for each dialect is listed. For example, column 'i' refers to the proto-Semai vowel /*iə/. Since no one present-day manifestation of this proto-vowel is markedly more common than the others, the entire column is filled in.

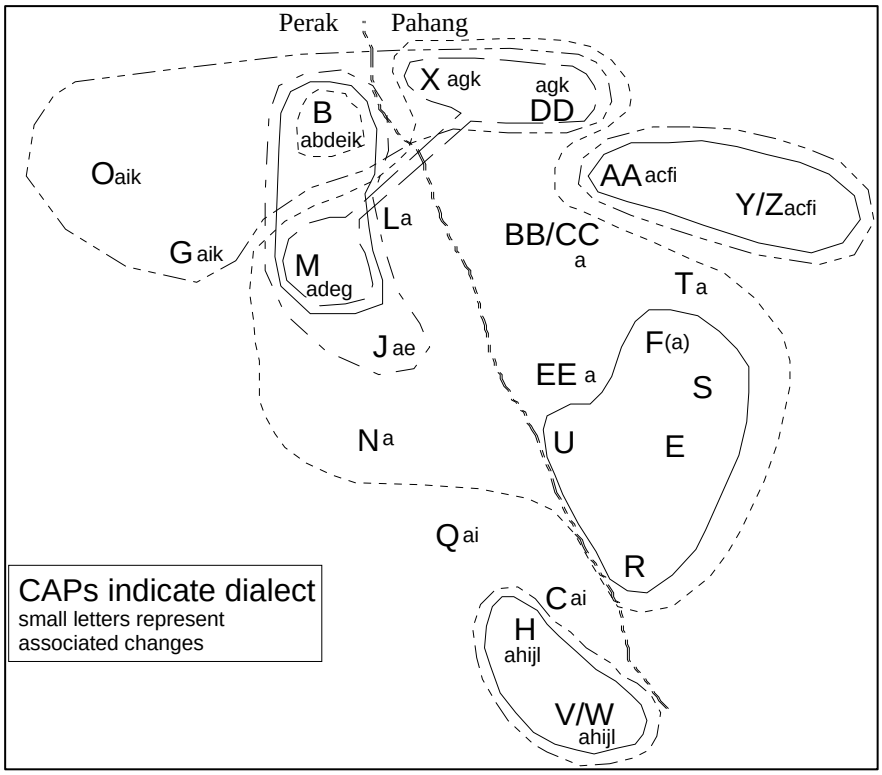


Figure 5. Geographical distribution of phonological changes.

The phonological changes can also be displayed on a map, with regions of the shared phonological changes encircled, as shown in Figure 5. Note that Figure 5 graphically demonstrates the complicated relationship of each phonological change, often resulting in overlapping circles, either partially or in full. For example, Gopeng (B) uniquely shares phonological change 'd' with its neighbor Kampar (M); that is, they both exhibit this phonological change. Hence, one of the lines encircles just these two dialects. These two dialects together also share phonological change 'e', but this change is also shared with Tapah J). Hence another circle encompasses these three dialects. For another phonological change, 'k', Gopeng shares this change with four different neighbors: Sungai Ruil, Terisu, Tangkai Cermin and Bota (X, DD, G, and O, respectively). The circle for this change only partly intersects the two circles previously mentioned.

Table 7 and Figure 5 allow us to readily see how many phonological changes exist between any one dialect and its neighbors near and far. When two dialects lie within the same set of circles—that is, there are no lines between them—then those two dialects do not have any consistent phonological change differences. The more lines between two dialects being compared, the more phonological changes that distinguish them.

If Figure 5 illustrates just one thing, it is that the dialect situation for Semai is quite complicated. There are no clusters of isogloss lines that clearly separate the Semai dialects into just two or three groupings. Indeed, there are ten distinct regions in this distribution map.

To further complicate the analysis, not all phonological changes should be given equal weight because they do not appear to have equal impact. For example, a particular phonological change may amount to a relatively minor phonetic shift (as in /*ɑɑ/ changing to /ɒɒ/), or it may instead result in a fairly radical restructuring of the affected words (as in /*oo/ changing to /*uwɑɑ/, which often resulted in other changes in these words). Furthermore, while some phonological changes impact great swaths of the lexicon (for example, the loss of the glottal stop after long vowels affects 13.3 percent of the words in this study), other changes only involved a handful of words (such as /*ɛɛ/ and /*iə/ changing to /yɑɑ/, which occurred in just 1.8 percent of the words).

Whereas it is difficult to quantify the former of these impacts (the nature of a phonological change, whether slight or radical), the latter impact (the number of words affected by a phonological change) can be calculated. Such an analysis is undertaken in the following section.

3.4.4 Aggregate phonological changes

Phonological changes cause words to take on different shapes and sounds. If two dialects do not share a given phonological change, then they will have differing forms for those words that have a common source. However, two dialects could have several phonological changes that are not shared between them and yet these changes might affect only a very small total number of words. On the other hand, two dialects could have just one phonological change that is not shared between them, yet that one change may affect a large percentage of words.

Based on the assumption that the greater total number of words impacted by phonological changes not shared between two dialects causes a greater degree of distinction between them, it is insightful to look at the relative number of words affected by given phonological changes. Note that two dialects may have both undergone several phonological changes that affected a large percentage of the words; however, if both dialects share all of the same phonological changes, then presumably they still have exactly the same form for each word they still have in common. (Obviously, this is not relevant for lexical innovations, wherein the two dialects would by definition no longer have the same word.)

One way to analyze the impact of nonshared phonological changes is to examine the average total number of words affected by phonological changes. Table 8 summarizes of the phonological changes plus the average number of occurrences per 100 words.[26]

Table 8. Percentage of words affected by each phonological change

	Phonological change	**Average per 100 words**
(a)	Preploded nasal > voiceless plosive	18.0
(b)	Glottal stop lost after long vowels	13.3
(c)	/*oo/ > /əə/ , /ɨɨ/	13.0
(d)	/*ɑɑ/ > /ɒɒ/	10.5
(e)	/*ee/ > /ɨɨ/	5.5
(f)	/*ɤɤ/ > /ʉʉ/ , /oo/	5.5
(g)	/*u/ > /o/	4.0
(h)	Final palatal consonant shifted	3.8
(i)	/*iə/ > /ii/ , /ee/ , /ɛɛ/	3.5
(j)	/*oo/ > /wɑɑ/	2.8
(k)	/*oo/ after N > /uu/ , /ɔɔ/	2.5
(l)	/*ɛɛ/, /*ie/ > /yɑɑ/	1.8

The letters associated with the phonological changes are the same as in the previous section. Note that the list of phonological changes has been arranged in order of changes that affected the most words, on average, per 100 words in this study. The phonological change that affected the most words is placed at the top of the list. These percentages can then be applied to the map of geographic distribution (Figure 5) to produce the contour map in Figure 6.

[26]This average is based on counts of words within the wordlist and not on a frequency count of any particular text.

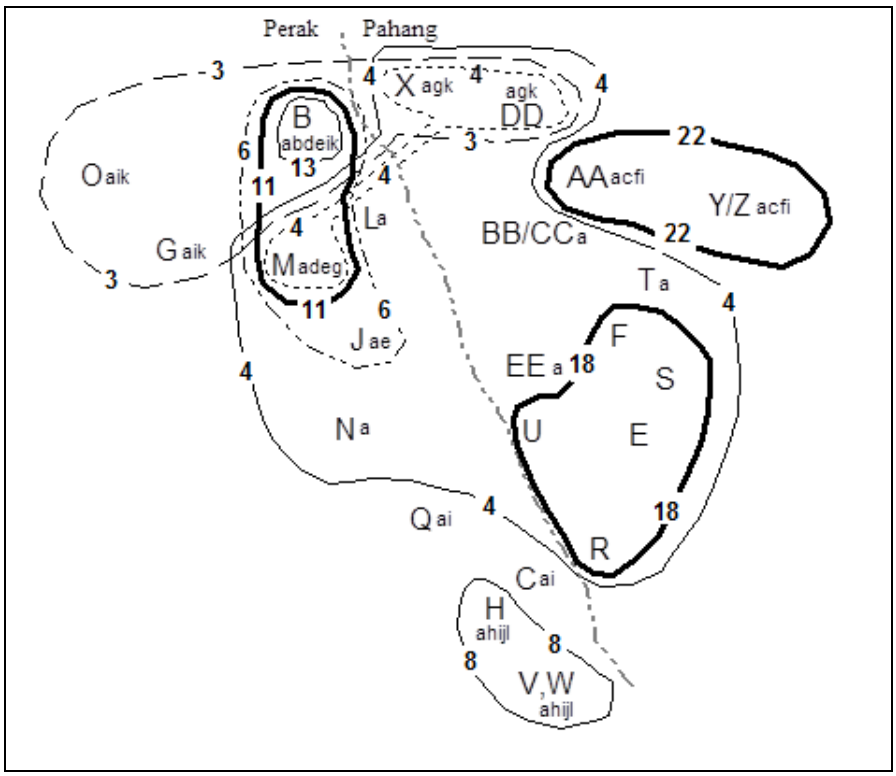


Figure 6. Contour map of average total number of phonological changes.

The ten dialect regions can now be compared to each other. (A dialect region here is defined as a set of dialects that have no phonological change boundaries between them.) If two dialects lie within the same set of circles, such as Pos Buntu and Betau (R and S, respectively), then there are no consistent phonological changes between the two dialects. If two dialect regions have just one line between them, the value associated with that contour line[27] is the average number of instances per 100 words of differences in the two regions as a result of one or more phonological changes. In the case of Sungkai (Q) and Bidor (N), for example, the line dividing them has the value of four; hence, there are approximately four instances per 100 words where there are differences in the two dialects as a result of an identified phonological change.

This analysis can be extended to dialects that are separated by more than one contour line. In principle the procedure is to add up the figures for the contour lines that are crossed. In practice this can be a little tricky since the contour map in Figure 6 is actually a simplification of what amounts to a multidimensional surface. If a line is drawn to avoid unnecessary crossing of contour lines, however, the correct average aggregate number of phonological changes can be determined. For example, when comparing Cluny (H) with Lanai (Y), a line can be drawn from H to Y that avoids going through the two southeast regions so that only two contour lines are crossed. The sum of these two lines is 30, meaning that on average there are 30 instances per 100 words where there are differences between Cluny and Serau as a result of phonological changes.

Every dialect region was compared to each of the others. The following matrix shows a summary of the average aggregate number of instances of differences per 100 words, comparing each dialect region to all the others.

---

[27]A given contour line may well represent more than one phonological change. For instance, in the northeast region dialects AA, Y, and Z are encircled by a contour line with the value 22. This actually represents the sum of three phonological changes: 'c, f, i.' These three changes have occurred in these three languages, but not outside of that area.

Table 9. Average aggregate number of instances of differences per 100 words

Q,C– Sungkai, Rasau									
3.5	**L,N,T,BB,EE– Batu 17, Bidor, Simoi Baru, Cenan Cerah, Pagar**								
2.5	6.0	**G,O– Tangkai Cermin, Bota**							
9.0	5.5	11.5	**J– Tapah**						
10.0	6.5	7.5	12.0	**X,DD– Sungai Ruil, Terisu**					
8.3	11.8	10.8	17.3	18.3	**H,V,W– Cluny, Chinggung, Sungai Bil**				
23.6	20.1	26.1	14.5	18.5	31.8	**M (Kampar**			
21.6	18.0	24.1	23.6	24.6	29.8	38.1	**S,R,U,E,F– Betau, Pos Buntu, Cherong, K. Kenip, Bertang**		
21.1	24.6	18.5	30.1	26.1	29.3	44.6	42.6	**AA,Y,Z– Renglas, Lanai, Serau**	
31.8	35.3	29.3	29.8	36.8	40.1	23.3	53.4	47.9	**B– Gopeng**

To read the matrix, simply find the intersection of any two dialect regions on the chart. For example, the intersection of Row S,R,U,E,F and Column Q,C reads 21.6, meaning that on average, there are 21.6 instances of differences per 100 shared words between dialect region E (Betau, etc.) and dialect region C (Sungkai, etc.).

A low score indicates few instances of consistent phonological change between dialects and suggests relatively good comprehension. Conversely, a high score indicates a large number of instances of consistent phonological change. Note that this number is not a percentage, since if there were enough phonological changes, there could be cases where there was an average of more than 100 instances of phonological change per 100 words, since some words have multiple phonological changes. While there are no cases of numbers greater than 100 in this chart, there are words that have multiple phonological changes between certain dialects. For example, there are two phonological changes that affect the following word, making the form from dialect Chinggung (V) different from that of dialect Betau (S).

/*siə^{b}m/ ‘to forget’ (V): /siyɑɑp/ [Preploded nasal > voiceless plosive; plus /*iə/ > /iyɑɑ/]

(S): /see^{b}m/ [/*iə/ > /ee/]

Finally, an overall average for each dialect region can be calculated from the matrix.

Table 10. Average aggregate number of instances of phonological change per 100 words

Overall average	Dialect region
36.4	B
31.6	Y,Z,AA
30.6	E,F,R,S,U
26.7	M
21.9	H,V,W
17.8	X,DD
17.0	J
15.1	G,O
14.6	L,N,T,BB,CC,EE
14.6	C,Q

Table 10 dramatically demonstrates that some dialects have many more instances of phonological change than others, when compared to all the dialect regions. For example, Gopeng (B), the dialect with the highest aggregate, has on average nearly 2.5 times as many instances of phonological change as dialects with the lowest aggregate (C, Q, L, N, T, BB, EE).

3.4.5 Summary of phonological changes

This section has looked at the aspect of consistent phonological changes from a variety of angles. These changes are important when establishing dialect boundaries as well as determining the degree of difference between dialect regions.

The previous tables, matrices, and maps provide important information that should be useful in choosing a standard dialect, if so required. Other factors being equal, dialects with a lower average aggregate number of phonological changes would be better candidates for a standard dialect, since the lower average aggregate implies there are fewer instances of nonshared phonological change, on average, with other dialects. However, it must be stressed that other factors are *not* equal, and that this is just one factor. All the factors, especially sociolinguistic factors, should be considered when making such decisions. For example, the people's attitude toward their language and dialects can turn out to be even more important than linguistic factors.[28]

3.4.6 Limitations of phonological change analysis

In addition to noting that phonological changes are just one indicator of dialect boundaries and variation, it is worthwhile to point out other limitations of this approach. First, as pointed out earlier in this section, not all phonological changes ought to be considered equal in that some changes have a much more profound impact on the forms of words than others. That is to say, some changes result in a dramatically different form of a word, compared to the same word that has not changed (or changed differently) in another dialect, while other changes amount to a slight phonetic shift, easily recognizable by speakers of other dialects. Unfortunately, it is rather difficult to quantify just how slight or radical a particular phonological change is. Sociolinguistic factors are important, since it ultimately is the perception of Semai speakers that determines whether a particular phonological change is profound or not.

[28]For instance, in countries with a caste system or with strong social classes, one would generally expect the language/dialect of the lowest caste or class to be unacceptable as the standard language/dialect, even if it were the most widely understood.

For the analysis regarding aggregate phonological changes cited earlier (Table 10), it is worth noting that the averages were determined from the roughly 436 lexical items that were used in this study. A larger corpus would theoretically produce more accurate figures. It should also be noted that each dialect region was given equal weight when determining the averages. No adjustment was attempted based on factors such as the total population in each dialect region. It may be a strategic factor to give more weight to the more populous dialect regions if a standardized form of Semai is to emerge.

Lastly, attention should be given to items of the wordlist in which dialect forms are not similar and apparently are unrelated. When the forms are dissimilar, the matter of aggregate instances of phonological changes is irrelevant. Therefore if two given dialects have a great number of unrelated lexical items, the approach of looking at aggregate instances of phonological change is much less insightful since it applies to much less data.

3.5 Comparative reconstruction of Proto-Semai

Comparative reconstruction involves using daughter languages and dialects to reconstruct the proto-language or ancestor language, a single original language from whence all of the daughter languages and dialects descended. The most important tool for reconstruction is the comparative method. The previous sections have already referred to the various consistent phonological changes found between dialects. This section will consider whether certain phonological changes that are the same or nearly the same in two dialects did in fact occur simultaneously or while those dialects were in contact.

3.5.1 Independent phonological changes

Generally, dialects that share the same set of phonological innovations are assumed to have a shared history and therefore are more closely related. There are important exceptions, however. One exception is when a phonological change is regarded as extremely common and therefore may well have occurred as an independent innovation in two or more dialects after they separated from each other. Stated another way, the phonological change did not occur before the dialects split, as might mistakenly be presumed. Losses of a final–h or final–r are examples of such changes that may occur independently.

There have been several independent phonological changes in Semai. Diffloth (1977) shows that Proto-Semai had more nasal vowels than do the current-day dialects. For instance, Proto-Semai had a long nasal vowel /ẽẽ/, which has been either raised to /ĩĩ/ or lowered to /ɛ̃ɛ̃/ in all dialects. Since only two change strategies (to eliminate the long nasal /ẽẽ/) were employed throughout the Semai dialects, it would not be surprising to find that several dialects had undergone the same phonological change independently. Indeed, Diffloth shows the raising of /ẽẽ/ in two geographically separate areas, and claims that the same phonological change occurred in both areas independently, after the dialects had split.[29]

In the current research, two of the identified phonological changes fall into this category, and as such, offer weak evidence at best of a genetic relationship. The first is phonological change (i), where /*iə/ changes to /ii/, /ee/, or in one dialect /ɛɛ/. The second is phonological change (k), where /*õõ/ becomes /ũũ/ or /ɔ̃ɔ̃/.

[29]Unfortunately, on average this vowel was elicited fewer than twice per wordlist, which was not deemed sufficient to verify Diffloth's findings. What data there was did mostly substantiate Diffloth's claim for this phonological change, however.

3.5.2 The Malay peninsula as a linguistic area

The principle that shared phonological changes indicate relatedness can also exhibit exceptions when dialects or languages come into contact and influence each other. A *linguistic area*[30] is a geographical region where languages come into contact with each other and various language features are borrowed; not just lexical items or phonological changes are borrowed, but often other features such as morphological and/or syntactic structures. Indeed, linguistic history is replete with examples of phonological changes induced by language contact. For purposes of tracing linguistic history and determining family relationships, it is critical to resolve which phonological features are inherited and which are borrowed through language contact.

There is strong evidence that Semai and many of its surrounding languages constitute a linguistic area. The Semai language has clearly borrowed many lexical items from Malay, as evidenced in the collected wordlists. In addition, Semai has borrowed the Malay suffix-*lah* (with various pronunciations), which stands out because Mon-Khmer languages are typically devoid of suffixes. Furthermore, the southern dialects near Tanjung Malim (represented by wordlists from Chinggung and Sungai Bil) are seen to be shifting the final palatal consonants to either the alveolar or velar points of articulation. This shifting is evidently due to the influence of their neighboring language, Temuan, as well as Malay. Both are Austronesian languages, and Austronesian languages in general do not have final palatal consonants. However, although it is beyond the scope of this report, it is rather interesting to note that wordlists of Temuan *do* show a few words with palatal final consonants (Baer 1999), and this feature evidently is borrowed from Semai and/or other Mon-Khmer languages in the region. Temuan is not an isolated case: Seidlitz (2005) reports that Jakun, another neighboring Austronesian language, also has some words with palatal final consonants. [31]

[30]A *linguistic area* has also been referred to in various literature by the following terms: *Sprachbund, diffusion area, adstratum relationship*, and *convergence area*.

[31]It is also possible that final palatals–c and–ɲ in Temuan and Jakun are evidence of substratum; that is, vestiges of another language (presumably an Aslian language) spoken by Temuan and Jakun speakers before they essentially switched to a variety of Malay.

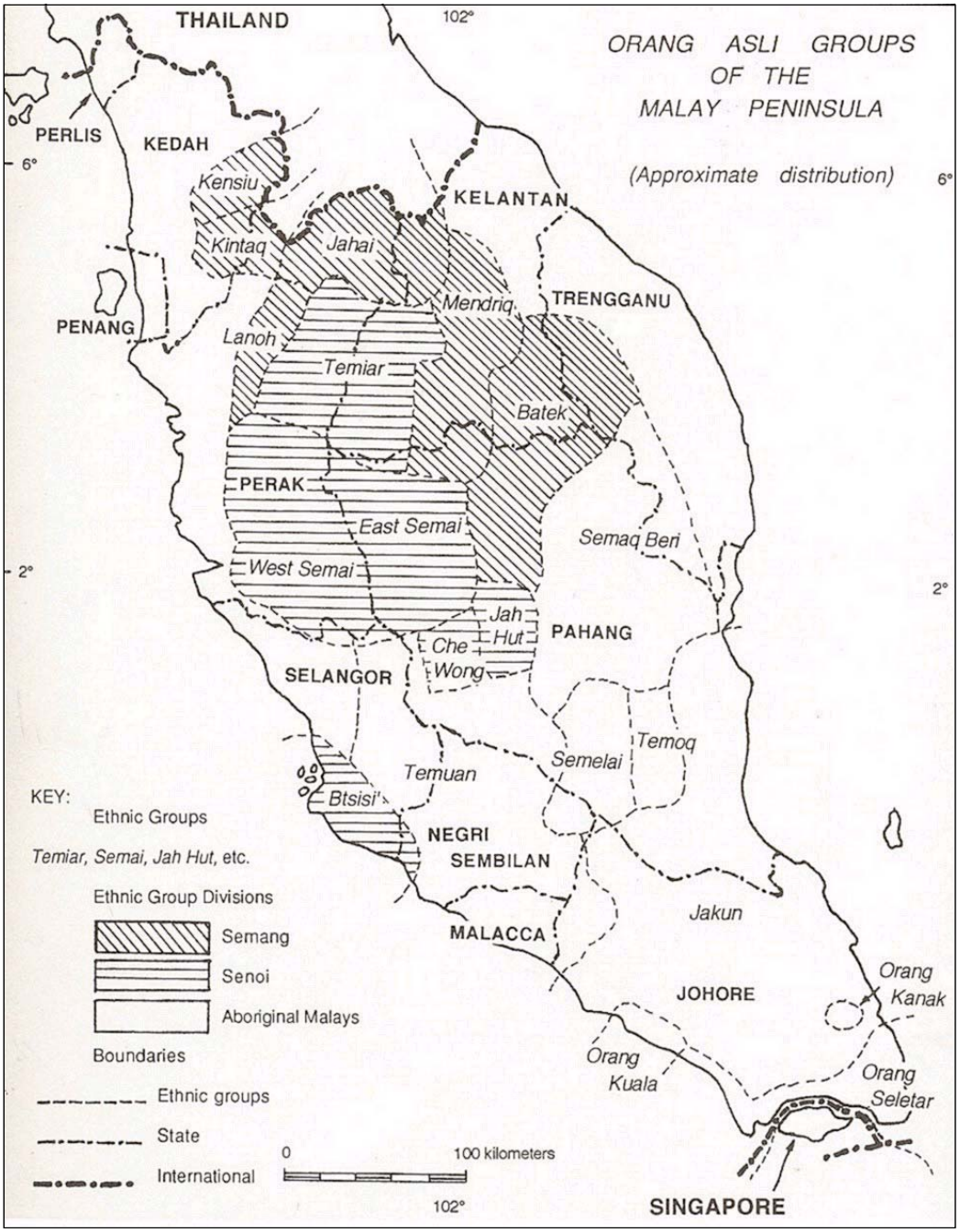


Figure 7. Geographical range of Orang Asli languages.[32]

In the linguistic area to which the Semai language belongs, probably the most interesting feature, or at least the most relevant to this discussion, is the preploded nasal. Most Aslian languages, as well as several local Malay dialects, exhibit nasal preplosion in some form or another (Phillips 2006a). Proto-Semai clearly had preploded nasals, which derived from simple nasals at some point in its history. In the present day, most dialects of Semai have reduced the preploded nasals to simple, voiceless plosives. The line dividing the region between the areas that preplode and those that have simple plosives cuts across

[32]This map is reproduced from Dentan (1997).

the Semai territory more or less from north to south along the Perak-Pahang border, but with the Cameron Highlands portion of Pahang exhibiting voiceless plosives as the Perak side does.

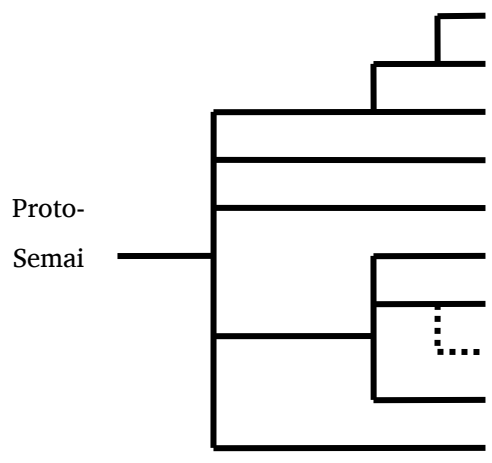
The crucial question remains of whether the preploded nasal in Semai is an inherited trait or an areal feature—a separate phenomenon sweeping across the Semai dialects. Diffloth (1977) does not address the question directly, perhaps because his paper is specific to the Semai vowels; however, his groupings imply that he views the preploded nasals as an areal feature. In particular, Diffloth treats his entire set of "NE" dialects as a single branch of Semai, even though one of the dialects (NE) in that set has preploded nasals and the others (NE_1 and NE_2) do not. Furthermore, he does not group NE with E or SE, even though the latter two also exhibit preploded nasals.

The current research contains three wordlists that seem to correspond with Diffloth's NE designations: Renglas, Lanai, and Serau. None of these wordlists contained preploded nasals. However, upon listening to the Lanai and Serau speakers talking in Semai after wordlists were elicited, it became clear that the speakers do indeed preplode. Apparently these speakers suppress the preploded nasal when giving words in isolation. Whether the suppression is intentional or not is not known. It would be good to work with these speakers again, and perhaps devise an elicitation method whereby words can be elicited in a frame of text such that the preplosion is not suppressed.

The fact that several phonological changes unique to Diffloth's NE dialects were verified in this research[33] gives valuable evidence that the NE dialects should be grouped together and not genetically split by the preplosion feature. Therefore, unless future research turns up contradicting evidence, the reduction of preploded nasals to simple plosives is best viewed as an areal feature, not a genetic one.

3.5.3 Genetic relationships of Semai dialects

Once all genetic phonological changes are identified, a "family tree" relationship of the dialects can be drawn up based on inherited phonological changes.



Dialect area	Phonological change	Areal feature
B	b,d,e	a
M	d,e,(g)	a
J	e	a
Y,Z,AA	c,f	a
X,DD	g	a
G,O	i_1	a
E,F,R,S,U		
N,L,BB,CC,EE		a
C,Q	i_2	a
H,V,W	h,j,l	a

Figure 8. Family tree for Semai dialects.

In the family tree in Figure 8, a number of observations can be made. First of all, two dialect areas are rather divergent from the rest of the dialects. The bottom group, representing the southern dialects H, V, and W, has three shared phonological changes not found in other dialects, indicating that these three dialects link back to proto-Semai and are not otherwise related to the other dialects. In the second

[33]Where data was too scant to be included in this paper, the data that did exist appeared to support Diffloth's (1977) conclusions.

dialect area, the top dialect B– Gopeng also has three phonological changes that distinguish it from most of the other dialects; however, two of the changes are also shared with M– Kampar. Gopeng and Kampar share one of their phonological changes with J– Tapah. Therefore it appears that B, M, and J could be viewed as a grouping linked by sharing change 'e'.

Secondly, two other dialect areas also exhibit phonological changes, albeit fewer, that separate them from the rest. The dialects in the northeast corner of the Semai territory (Y– Lanai, Z– Serau, and AA– Renglas) have two phonological changes not found elsewhere. The other dialect area, in the northern part of Cameron Highlands (X– Sungai Ruil and DD– Terisu), has just one phonological change setting it apart from others. This change, where short vowel /u/ goes to /o/, happens also to be found in Kampar. Since crossing genetic lines is not generally allowed when drawing family trees,[34] this change has been added only in parentheses to M– Kampar. This may be an areal feature, an independent innovation, or simply that the speaker who gave the Kampar list had been influenced by the dialect from Cameron Highlands for some reason. In any case, this phonological change should be viewed as suspect until further evidence can be brought to bear on the situation.

Next, the dialects roughly representing center and northwest regions all group together with no or few phonological changes. Two exceptions have been drawn on this tree. First, the northwest dialects G– Tangkai Cermin and O– Bota and the south-central dialects C– Rasau and Q– Bidor happen to share the same phonological change. However, it seems probable that this change represents independent innovations for two reasons: first, the geographic separation between these two regions argues against linking the two dialect areas; and second, this phonological change (where /*iə/ changes to /ii/ or /ee/) was previously mentioned as likely being a nongenetic phonological change (i.e. an independent innovation).

Finally, only one feature is shown dividing the remaining dialects, which roughly represent the geographical center of the Semai territory. This phonological change, where the preploded nasals become simple, voiceless plosives, was argued earlier to be an areal feature. So, while an areal feature is not generally included on a genetic chart, it has been included here (with a dashed line) to highlight the residual difference between these central dialects.

3.5.4 Reconstruction of Proto-Semai lexical items

Although there is some room for differing interpretations of the phonological changes that determine the genetic tree, the reconstruction of Proto-Semai is a bit more clear-cut. Indeed, even if the preploded nasals are considered to be a genetic feature rather than an areal one, the reconstruction produces the same forms. The proto-forms determined from the data in the study can be found in Appendix E. The list contains 429 lexical items. However, since many lexical items produced multiple words across the many dialects, a total of about 760 unique words were reconstructable. Of this total, 150 words are from Diffloth's data and congruent with the current findings, while 610 words are newly added.

3.6 Word borrowings

Another insightful analysis regards word borrowings. The most obvious source for Semai to borrow words is from Bahasa Malaysia (Malay), which is the national language as well as the language of wider communication. Table 11 gives the percentage of apparent borrowings from Malay found in the wordlists elicited. As noted earlier, it would be hard to distinguish between Malay and Temuan as the source of the borrowings since these two languages are so closely related.

[34]Another method of representing phonological changes, called wave-theory, has no such limitation. Circles are used to show the extent of each phonological change, and the various circles are permitted to overlap and cross each other.

Table 11. Percentage of apparent borrowings from Malay

Percent	Dialect
26.5	W– Sungai Bil
22.8	V– Chinggung
21.8	M– Kampar
21.1	H– Cluny
20.1	N– Bidor
19.2	G– Tangkai Cermin
18.2	O– Bota
16.2	L– Batu 17
16.0	Z– Serau
16.0	X– Sungai Ruil
14.7	Y– Lanai
12.5	CC– Relong
12.1	R– Pos Buntu
12.1	Q– Sungkai
12.1	J– Tapah
12.0	DD– Terisu
11.2	T– Simoi
10.7	E– Bertang
10.2	B– Gopeng
10.1	AA– Renglas
10.0	F– Kuala Kenip
9.7	C– Rasau
8.0	S– Betau
7.9	U– Cherong
7.6	BB– Cenan Cerah
7.5	EE– Pagar

A number of salient observations can be made from Table 11. In broad strokes, the dialects with the fewest borrowings tend to be in the remote interior, most often in Pahang. On the other hand, the dialects with the most borrowings tend to be in the lowland areas of Perak that have ready access to roads and the surrounding cultures.

Of particular interest are the southern dialects—Chinggung, Sungai Bil, and Cluny—which represent three out of the four highest rates of borrowings among the dialects surveyed for this study. This high rate of borrowing, coupled with the shift of final palatal consonants[35] mentioned earlier (phonological change ‘h’ in section 3.4), is a strong indicator of the significant influence that Malay or Temuan has been exerting on Semai dialects in this region.

Also exhibiting relatively high rates of borrowing are Tangkai Cermin and Bota—nineteen percent and eighteen percent, respectively. This fact, plus the fact that these dialects also share several unique

[35]Bahasa Malaysia does not have the final palatal consonants /-c/ and /-ɲ/, although these palatal consonants are common in other positions. In fact, as a rule the entire Austronesian language family has a dearth of these final palatal consonants. But a curious exception to this rule is Temuan, the Malayic language that is Semai’s immediate neighbor to the south. Temuan has several lexical items with palatal final consonants, which conceivably could have been borrowed from Semai (or another Aslian language), or be substratum.

lexical innovations, helps establish a link between them as well as a distinction of sorts from the other Semai dialects.

At the lower end of the chart, six of the seven dialects with the lowest rates of apparent borrowing from Malay are spoken in relatively remote villages in Pahang. The one surprising exception, Rasau, is in a relatively accessible area in Perak; indeed, it is rather close to the southern dialects (Sungai Bil, Chinggung, and Cluny) that have the highest rates of borrowing.

It would be premature, however, to draw strong conclusions from Table 11 until more research can be done, since many factors can influence the number of apparent borrowings. Personal factors of people providing the language data, including age, gender, education, mobility, and intermarriage, affect responses. Even fatigue can come into play—in some cases the language speaker was able to produce the Semai word quickly, going through the list for the first time; however, during subsequent checking, often several hours later, the speaker was no longer able to remember the word easily and would substitute the Malay word.

For future research it would also be useful to compare the wordlists from this research with those of other languages in the region, especially those known to be in contact with or in geographical proximity to the Semai, such as Temuan (to the south) and Temiar (to the northeast) (Benjamin 1976b:75).

3.7 Some other observations

3.7.1 Language preservation and vitality

One aim of this research project is the preservation of the national treasure that Malaysia has among its wealth of indigenous languages, such as the Semai language. Language preservation has two aspects. One aspect, to which this project has been largely dedicated, is the documentation of the language as it exists today, a snapshot to be preserved for generations to come. The second aspect of language preservation involves the continuation of the use of the language.

Although it was not the direct aim of this research project to document the vitality of Semai, several noted factors indicate Semai is indeed a living and vital language. The following four observations were made from informal discussion with the various Semai speakers who provided the wordlists for this study. First, all of the Semai speakers seemed quite proud of their language. Second, the Semai claimed that they always spoke their mother tongue when talking to other Semai, even with Semai from distant dialects, the only exception being when the dialects were just too different, and they needed to supplement their speech with Malay. Third, the Semai are diligently teaching the language to their children and expressed their desire to see their grandchildren and future generations continue to learn and use Semai. Fourth, most Semai interviewed are interested in seeing the language developed. They desire to be able to read and write their mother tongue, and in most cases are delighted to see materials such as school curricula and dictionaries in their language.

It is hoped that the results presented in this report will further the use, study, and development of the Semai language.

3.7.2 Morphology– reduplication

In the course of any research involving digging for as yet unknown treasures, there often are gems of discovery beyond the original scope of the research. Such was the case with this project. During the wordlist collection it was not uncommon to encounter words, especially verbs, in a reduplicated form. It was discovered that reduplication usually conveys the sense of continuative aspect– not an uncommon function for languages that employ reduplication. But as more and more wordlists were collected, several patterns of reduplication were noticed.

- Some dialects have more than one type of reduplication, including full reduplication of the root verb; for example, Betau has these forms.

bi.ɟɑɑ^{b}m	‘cries/cried’
bi.ɟəm.ɟɑɑ^{b}m	‘is crying’
bi.ɟə.ɟɑɑ^{b}m.ɟɑɑ^{b}m	‘cries and cries’

This last form is especially interesting in that it defies the norm for Semai word in that normally only the ultimate syllable bears stress and utilizes the full set of phonemes. Instead, in this type of reduplication, both the ultimate and penultimate syllables are exact copies, each bearing stress and neither having a reduced set of phonemes.

- Some dialects in which final preploded nasals become voiceless, have nasals reappear in partially reduplicated bases. For example, Kampar has these forms.

/bi.ɟɑɑp/	‘cries/cried’
/bi.ɟəm.ɟɑɑp/	‘is crying’

- Some dialects drastically reduce the amount of phonetic information in partially reduplicated roots. For example, in the partial reduplication in the following forms from Kampar the vowel is reduced and the oral plosive has become glottal.

/bi.ciip/	‘walks/walked’
/bi.cəʔ.ciip/	‘is walking’

It is reported that some dialects of Malay have a very similar type of reduplication (Kroeger 1989 and Zaharani 1991).

- In some dialects reduplication appears to be far less productive; a number of apparently reduplicated forms are frozen. For example, the following form from Sungai Bil contains what appears to be a frozen reduplication, but no parallel form without reduplication has been found.

/bi.tɑ.tɔh/ (but no /bi.tɔh/)	‘spits’

- Perhaps the most intriguing aspect of reduplication in Semai is apparent in sesquisyllabic roots. Rather than one of the syllables being copied, a copy of the final consonant is infixed between the initial consonant and the first consonant of the major (ultimate) syllable. Matisoff (2003) coined the term “incopyfixation” to describe this phenomenon. Here are two examples from the Betau dialect.

/bi.kə.hɔɔl/	‘to cough’
/bi.kə**l**.hɔɔl/	‘is coughing’
/bi.bə.lɨh/	‘to see’
/bi.bə**h**.lɨh/	‘is seeing’

A preliminary search of the linguistic literature on this topic reveals that this form of discontinuous reduplication is very rare, if not unique to Aslian languages.

A fuller discussion of this type of discontinuous reduplication in both Semai and in surrounding languages is presented in Phillips (2006b).

4 Conclusions

4.1 Language vitality

The Semai language is very vital– the Semai people consistently use it among themselves, and they are teaching the language to their children. The Semai people are proud of their language and see it as something that identifies them as a people. In essence, the Semai people epitomize the Malay proverb:

Bahasa jiwa bangsa
Language is the soul of the race

Bilingualism in Malay does not appear to be threatening the vitality of Semai. Although it was not the specific aim of this research, it was readily apparent that in general, the Semai people are fluent in Bahasa Malaysia.

4.2 The complex dialect situation

Another richness of the Semai language is the wide diversity of its dialects. Every village speaks its own variety, giving every region its own unique flavor. As this report has shown, most Semai locales share many words with their neighboring villages, but the differences tend to grow with distance. The result is a sort of dialect chain or dialect network spread over the Semai territory, with each village linked with its near neighbors, except where physical barriers, such as the mountain range between the states of Perak and Pahang, tend to impede travel and hence hinder communication.

As would be expected, the dialects that are the most different are at the boundaries of the Semai territory. These dialects are the farthest on average from all the dialects as a whole and also more in contact with neighboring language groups. For the Semai case, the extreme southern dialects are the most notably different, but the northwest dialects also show a number of lexical innovations, and the northeast dialects have sounds that are quite distinct from the rest.

The end result of all the differences among the Semai communities, both in unique words and in phonological changes, is that most likely, one variety of Semai could not be used effectively to serve the entire Semai-speaking population. The differences are just too great for any one Semai dialect to be understood in all regions. Furthermore, the various analyses in this report have shown that the situation is rather complex, and the Semai territory cannot be neatly divided into just two or three regions.

Ultimately, direct comprehension testing between dialects will be needed, in order to truly determine how widely understood the dialects are. The preliminary groupings outlined in this report will serve as a basis for choosing dialects for such testing.

5 Recommendations for further research

This research has laid the groundwork for much research yet to be done. First of all, if the Semai language is to be developed and its vitality maintained, some form of standardization will need to take place. This study of Semai dialects has revealed a plethora of differences from one dialect to the next. The Semai people themselves are well aware of these differences and admit some dialects are harder to understand than others. It is impractical and undesirable to develop each dialect; rather, one dialect will need to be selected for standardization. As researcher Gérard Diffloth (1968) asserts,

> In spite of many common features, the differences between the various dialects labeled Semai are so numerous… As a practical consequence, it would be extremely difficult to implement a literacy campaign in their own idiom until one of the dialects has been accepted by them as a standard.

A number of factors need to be considered in determining which dialect is to be promoted as the standard. Ideally, the standard dialect is one which is highly comprehensible[36] by all of the other dialects. While wordlists, linguistic comparative analysis, and native speaker intuition is of value in predicting comprehensibility, ultimately comprehensibility must be empirically tested.[37]

Secondly, the morphology is a good next level of research after the phonology has been investigated. The syntax and semantics of Semai also need to be studied, not only for unlocking the secrets of how the Semai language works, but also for documentation and preservation purposes. And as previously mentioned, there appears to be something intriguing about reduplication in Semai, which may turn out to have significant importance for linguists.[38]

Thirdly, a dictionary project,[39] especially one that covers the dialectical differences, would immensely aid efforts to:

- Preserve the language;
- Make the Semai language available to linguistic researchers around the world;
- Raise the status of the Semai language;
- Allow Semai people to understand one another across large dialect gaps;
- Train government officials that relate to the Semai people.

Fourthly, a project to do text collection should be initiated. Not only would this produce material for more linguistic research, but it also would encourage the preservation of the language and culture that is still known to the older generation, but is gradually being lost by the younger generations. In addition to the legacy of culture and language at risk of being forgotten, it is estimated that huge amounts of knowledge about the rainforest and its useful resources will become extinct if attempts are not made to preserve this knowledge.[40]

Fifthly, in the realm of biodiversity, a topic of ever-growing international concern, "scientists are becoming increasingly aware of the sophistication of traditional ecological knowledge among many indigenous and local communities. … Traditional healers in Southeast Asia rely on as many as 6,500 medicinal plants, and shifting cultivators throughout the tropics frequently sow more than 100 crops in their forest farms" (Posey 2001). There is an inseparable link between this indigenous knowledge of local biodiversity and linguistic diversity in that such local knowledge is critically dependent on local languages for cultural transmission between generations. In other words, preservation of the Semai language will help preserve what the Semai know about the rainforest, which in turn will aid knowledge of local biodiversity and how to maintain it.

Lastly, a sociolinguistic study should be undertaken so that the social factors, as they relate to language, can be better understood. This kind of research would seek answers to questions such as the following:

---

[36]The more common terms in linguistics literature are *intelligible* and *intelligibility*. However, this term is sometimes confused with "intelligence"; therefore, I have chosen terms such as "comprehensible" and "comprehensibility," referring to the inherent ability of people from one dialect to understand another dialect without already being familiar with that dialect.

[37]A dialect comprehension suvery was carried out in 2005–2006 and resulted in the report "Dialect comprehension survey of the Semai language" (Phillips 2006b), submitted to the Economic Planning Unit (EPU) of the Prime Minister's Department, Malaysia. The results of this report were used to choose a standard dialect for Semai, corresponding roughly to Semai dialect spoken along the lower Jalan Pahang region of Perak.

[38]The phonology of Betau Semai in Appendix F was later published (see Phillips 2007). Phillips (forthcoming) outlines the morphology of Semai as well as what is currently known of the morphology of other Aslian languages in his dissertation on Aslian. As mentioned earlier, Semai reduplication has been described in some detail (Phillips 2011).

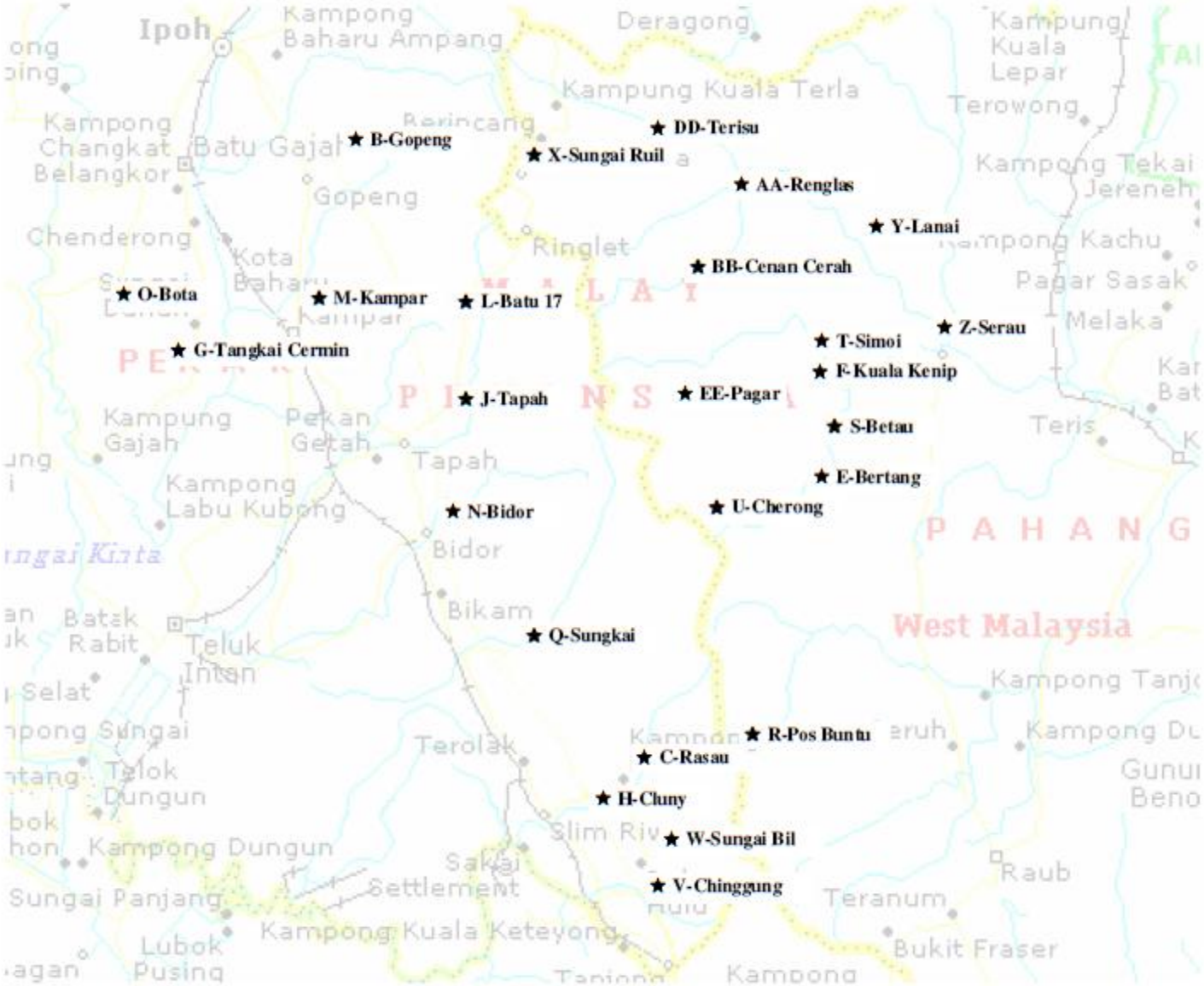
[39]A Semai dictionary project was begun in 2005 and resulted in a dictionary published in 2008 (see Basrim 2008). In 2012 a revision of this dictionary was initiated.

[40]Since this report was originally submitted in 2006, there have been several efforts to collect stories and texts in Semai. While these stories have not been officially published, some of them have been printed in limited quantities and distributed among the Semai in an effort to stimulate an interest in literacy and in language preservation. As of 2012 these efforts are ongoing.

- When people marry across dialects, which language do they speak with each other and to their children?
- What are their cultural and linguistic taboos, and how do these affect the language?
- What are the linguistic differences related to gender, age, status, and time away from the home village?

It is hoped that the research presented in this report will provide insights and pertinent information to future research about, and for the sake of, the Semai people and the country of Malaysia.

Appendix A: Map of dialects sampled



Appendix B: Wordlist employed in study

The major tool used in this research was a 429-item wordlist. The first part of this appendix presents the full wordlist that was employed, shown in Table 12. The wordlist is organized largely in semantic order i.e. words have been grouped according to similar concepts. The second part of this appendix details the various issues that arose in using the wordlist, shown in Table 13.

	Malay	**English**
1	matahari	sun
2	bulan	moon
3	bintang	star
4	langit	sky
5	air	water
6	sungai	river
7	alir, mengalir	flow, to
8a	berenang (orang)	swim, to (human)
8b	berenang (haiwan, contoh anjing)	swim, to (animal, eg. dog)
9a	apung, terapung	float, to (surface)
9b	timbul	float, to (rise to surface)
10	tenggelam	sink, to
11	mandi	bathe, to
12	cetek	shallow, to be
13	dalam (lawan cetek)	deep, to be (opposite shallow)
14	hujan	rain
15	kilat	lightning
16	guruh, guntur	thunder
17	pelangi	rainbow
18	bayang	shadow, a
19	angin	wind
22	batu	stone, a
23	gunung	mountain
24	gua batu	cave
25	tanah	soil, earth
26	lumpur	mud
27	pasir	sand
28	habuk	dust
32	garam	salt
33	jalan (kecil)	path (small)
34	api	fire, a
35	kayu api	firewood
36	abu	ashes
37	asap	smoke
38	bakar (kayu)	burn, to (wood)
39	layur; bakar (bulu)	burn off, to (feathers)

	Malay	**English**
40	hutan	forest
41	rumput	grass
42	biji, benih	seed
43	daun	leaf
44	akar	root, a
45	menggali	dig, to
46	cari, mencari	look for, to
47	bunga	flower
48	duri	thorn
49a	batang pokok	trunk, tree
49b	dahan	branch, tree
50	kulit kayu	bark, tree
51	buah	fruit
52	buah betik	papaya
53	pisang	banana
54	kelapa muda / nyiur muda	coconut (unripe)
55	kelapa tua / nyiur tua	coconut (ripe)
56	terung	eggplant
57	halia	ginger
58	cendawan	mushroom (generic)
60	kapuk / kabu (digunakan untuk membuat bantal)	kapok (cotton)
61	buluh	bamboo
62	rebung	bamboo shoot
63	rotan	rattan
64	buah pinang	betel nut
65	daun sirih	betel nut leaf
66	kapur (digunakan bersama buah pinang)	lime (used with betel nut)
67	ludah, meludah	spit, to (betel juice)
68	getah	rubber
69	toreh, potong getah	tap, to (a rubber tree)

	Malay	English
70	pokok ipoh	tree used for dart poison
71	sumpitan (alat)	blowgun
72	sumpit, menyumpit	blow, to (blowgun)
73	pangkal damak	dart head
74a	damak (yang ada racun)	dart (poison)
74b	damak (yang tiada racun)	dart (without poison)
75	tabang (bekas untuk simpan damak)	quiver
76a	tombak, lembing (dibuat daripada besi)	spear (iron)
76b	seligi (dibuat daripada buluh)	spear (bamboo)
77	buru, memburu	hunt, to
78a	bunuh, membunuh	kill, to
78b	matikan, mematikan	cause to die, to
79	tikam, menikam	stab, to
80	tembak, menembak	shoot, to
81	beruang	bear, a
82	babi	pig
83	landak	porcupine
84	monyet (ekor pendek)	monkey (short tail)
85	arnab	rabbit
86	rusa	deer (large)
87	harimau	tiger
88	ekor	tail
89	anjing	dog
90	menyalak	bark, to
91	kucing	cat
92	tikus	rat
93	gigit, menggigit	bite, to (animal)
94	burung	bird
95	terbang	fly, to
96	ayam	chicken
97	bulu (ayam)	feather (chicken)
98	telur	egg
99	sayap	wing
100	rama-rama	butterfly
101	lebah	bee
102	lalat	fly, a
104	nyamuk	mosquito

	Malay	English
105	kutu (ayam)	louse (chicken)
106	kutu (kepala)	louse (head)
107	anai-anai (sejenis semut putih)	termite
108	labah-labah	spider (small house)
109	lipas	cockroach
110	ular	snake
111	biawak	monitor lizard
112a	siput	snail
112b	siput babi	snail, garden
112c	siput air	snail, water
113	katak	frog
114a	penyu	turtle (sea)
114b	labi-labi	turtle (river)
115	ikan	fish
116	buaya	crocodile
117a	orang (asli)	person (Aslian)
117b	orang (bukan orang asli)	person (non-Aslian)
118	lelaki	man
119	perempuan	woman
120	anak	child (offspring)
121	kanak-kanak	child
122	emak/ibu	mother
123	bapa	father
124a	abang	sibling (elder brother)
124b	kakak	sibling (elder sister)
125a	adik lelaki	sibling (younger brother)
125b	adik perempuan	sibling (younger sister)
126	anak bongsu	offspring (youngest)
127	suami	husband
128	isteri	wife
129	balu (perempuan yang kematian suami)	widow
130	kawan	friend
131	nama	name
132	kepala	head
133	rambut	hair
134	sikat	comb, a
135a	botak (tiada rambut kerana rambut telah gugur)	bald, to be (natural)

	Malay	English
135b	botak (tiada rambut kerana telah dicukur)	bald, to be (shaved)
136a	menggunting rambut; potong rambut (pakai gunting)	cut hair, to (with scissors)
136b	potong rambut (pakai alat lain)	cut hair, to (with other instrument)
137	mata	eye
138	hidung	nose
139	telinga	ear
140	muka	face
141	dahi	forehead
142	kening	eyebrow
143	pipi	cheek
144	dagu	chin
145	kerongkong, tekak	throat (internal)
146a	bibir	lip
146b	bibir atas	lip, upper
147c	bibir bawah	lip, lower
148	gigi	tooth
149	lidah	tongue
150	gusi	gums
151	otak	brain
152a	tengkok	neck, (back of)
152b	leher	neck, (front of)
153	belakang (bahagian badan)	back
154	bahu	shoulders
155	ketiak	armpit
156	siku	elbow
157	tangan	hand
158	tapak tangan	hand, palm of
159	kuku	finger nail
160	jari tangan	digit, finger
161a	perut	abdomen (generic)
161b	perut (bahagian di atas pusat)	abdomen, upper
162	perut (bahagian di bawah pusat)	abdomen, lower
163	punggung	buttocks
164	pusat	navel
165	paru-paru	lungs
166	jantung	heart
167	tali perut, usus	intestines
168	hati	liver

	Malay	English
169	tulang	bone
170	tulang rusuk	rib
171	kulit (manusia)	skin (human)
174	daging (manusia)	flesh (human)
175	peha, paha	leg (upper)
176	lutut	knee
177	betis	leg (lower)
178	kaki	foot
179	tumit	heel
180	jari kaki	toes
181	kuat	strong, to be
182	letih, penat	tired, to be
183a	tidur	sleep, to
183b	tidur lena, tidur nyenyak	sleep, to (soundly)
184	berdengkur	snore, to
185	menguap	yawn, to
186	buta	blind, to be
187	nampak (contoh, nampak seekor ayam)	see a chicken, to
188	lihat	look, to
189	kenyit / kelip mata (sebelah mata)	wink, to
190	pekak, tuli	deaf, to be
191	dengar	hear, to
192	cium, menghidu (bunga)	smell, to (flower)
193a	reput	rotten/decayed, to be
193b	basi	rotten, to be (food)
194	lapar	hungry, to be
195	kenyang	full / satisfied, to be
196	makan	eat, to
197	haus, dahaga	thirsty, to be
198a	sedut, menyedut (siput air,sum-sum tulang)	suck out, to (snail, bone marrow)
198b	hisap, menghisap (gula-gula)	suck on, to (candy,sweet)
199	jilat, menjlat	lick, to
200	mabuk	drunk, to be
201	arak	liquor
202	minum	drink, to
203	telan, menelan	swallow, to
204	muntah	vomit, to

	Malay	English
205	sakit (tangan), [bukan demam]	hurt / be painful, to
206	bengkak	swollen, to be
207	gatal	itchy
208	garu, menggaru	scratch, to
209a	mencakar (seperti mencakar sampah atau daun)	rake, to
209b	kais, mengais (seperti ayam mengais untuk mencari makanan))	scratch, to (chicken)
210	sejuk	cold, to feel
211	menggigil, menggeletar	shiver, to
212	batuk	cough, to
213	bersin	sneeze, to
214	panas	hot, to be
215	air peluh	sweat
216	air liur	saliva
217	meludah (dekat)	spit close to self
218	meludah (jauh)	spit far away
219	air mata	tears (noun)
220	menangis	cry, to
221	air kencing	urine
222	darah	blood
223	tahi, najis	excrement
224	buang air besar, berak	defecate, to
225	nanah	pus
226	bisul	boil, a / abscess
227	bekas luka, parut	scar
228	ubat	medicine
229	menjampi	incant, to
230	bomoh, pawang, dukun	herbal curer, shaman
231	hidup	exist /be alive, to; live, to
232	mati	die, to
233	mengebumikan mayat, tanam mayat	bury, to (corpse)
234	tua (orang)	old, to be (animate)
235	gemuk	fat, to be
236	kurus	skinny, to be (animate)

	Malay	English
237a	tinggi (orang)	tall, to be (human)
237b	tinggi (pokok)	tall, to be (tree)
238a	pendek (orang)	short, to be (human)
239	pendek (pokok)	short, to be (tree)
240	besar	big, to be
241	kecil	small, to be
242	bernafas	breathe, to
244	tiup, meniup (api)	blow on, to (fire)
245	duduk	sit, to
246	berdiri	stand, to
247	berjalan	walk, to
248	merangkak	crawl, to
249	berlari	run, to
250	cepat	fast, to be
251	perlahan-lahan, lambat (lawan cepat)	slow, to be
252	fikir	think, to
253a	tahu, faham	know, to (thing)
253b	kenal (orang)	know, to (person)
254	lupa	forget, to
255	bermimpi	dream, to
256	pilih, memilih	choose, to
257	sayang, mengasihi, kasih	love, to
258	senyum	smile, to
259	ketawa	laugh, to
260	baik (orang)	good, to be (person)
261	jahat (orang)	bad, to be (person)
263	marah	angry, to be
264	meradang, berang, sangat marah	furious, to be
265	membohong, bercakap bohong	lie, to; fib, to
266	curi, mencuri	steal, to
267	berlawan (secara fizikal)	fight, to (physical)
268	takut	afraid, to be
269	betul	correct, to be
270	salah	wrong, to be
271	susah	difficult, to be
272	pukul	hit, to
273	bercakap	speak, to

	Malay	English
274	memanggil (untuk mendapatkan perhatian seseorang)	call, to (to get s.o. attention)
276	cerita	story
277	beritahu	tell, to
278	bersiul	whistle, to
279	jawab, menjawab	answer, to
280	nyanyi, menyanyi	sing, to
281	tari, menari	dance, to
282	jantung, buluh centong (alat muzik)	stamper, striker (musical instrument)
283	gendang	drum, a
284	bermain	play, to
285	tendang	kick, to
286	jatuh (orang)	fall, to (human)
287	gugur (buah-buahan)	fall down, to (fruit)
288	jatuh dari tempat tinggi (seperti orang jatuh dari pokok kelapa)	fall down from a height, to
289a	jinjing (seperti beg plastik)	carry in the hand, to take, to (plastic bag)
289b	mendukung (anak)	carry in arms, to (child)
290	mengambin	carry child in cloth on back, to
291	junjung	carry on head, to
292	pikul	carry on shoulder, to
293	pulang	return, to
294	datang	come, to
295	masuk	enter, to
296	tunggu	wait, to
297	bekerja	work, to
299	bayar	pay, to
300	jual	sell, to
301	beli	buy, to
302	beri, bagi	give, to
303	melempar	throw something sidearm, to
304	baling (seperti lembing)	throw (overhand), to
305	baling (seperti membaling bola dalam sepak takraw)	throw (underhand), to

	Malay	English
306	membuang (seperti sampah)	throw away, to
307	tarik	pull, to
308	tolak	push, to
309	hari	day, a
310	pagi	morning
311	tengah hari	noon
312	petang	afternoon (late)
313	malam	night
314	esok	tomorrow
315	semalam	yesterday
316	tahun	year
317	pondok sementara dalam hutan, rumah pisang sesikat	lean-to, a
322	rumah	house
323	kolong (bahagian ruang di bawah rumah)	space under house
324	atap	roof
325	tingkap	window
326	lantai	floor
327	selimut	blanket
328	tikar	mat
329	cawat	loincloth
330	kain pelikat, kain sarung lelaki	sarong (man's)
331	kain sarung perempuan	sarong (woman's)
332	ikat	tie, to
333	mengikat (mengikat dengan memusing seperti mengikat sarung lelaki)	tie by twisting, to (tie a sarong)
334	seluar	trousers
335	kotor	dirty, to be
336	cuci (tangan, pinggan)	wash, to (hands, dishes)
337	basuh (kain, pakaian)	wash, to (clothes)
338	gosok baju tanpa pakai berus (masa basuh baju)	rub, to
339	basah	wet, to be
340	kering	dry, to be
341	jemur	dry in the sun, to
342	lap,mengelap (meja)	wipe, to (table)

	Malay	English
343	sapu, menyapu	sweep, to
344	jahit, menjahit	sew, to
345	jarum	needle, a
346	memasak	cook, to
347a	masak, didih (air)	boil, to (water)
347b	rebus (ubi kayu)	boil, to (tapioca)
348	cerek	kettle
349	periuk	cooking pot
350	penuh	full, to be (container)
351a	senduk	ladle, a
351b	gayung	dipper (for bathing)
352a	lesung (dibuat daripada batu untuk menumbuk lada/rempah)	mortar (stone)
352b	lesung (untuk menumbuk padi)	mortar (for rice)
353a	antan (untuk menumbuk lada/rempah)	pestle (spices)
353b	antan (untuk menumbuk padi)	pestle (for pounding rice)
354a	tumbuk (seperti menumbuk lada/rempah)	pound in mortar, to (spices)
354b	tumbuk (seperti menumbuk padi)	pound in mortar, to (rice)
355	pisau	knife
357	tajam	sharp, to be
358	tumpul	blunt, to be
359	belah, membelah (kayu, buluh)	split, to (bamboo, wood)
360	manis	sweet, to be
361	masam	sour, to be
362	pahit	bitter, to be
363	hitam	black, to be
364	putih	white, to be
365	merah	red, to be
366	hijau	green, to be
367	kuning	yellow, to be
368	terang	bright, to be
369	gelap	dark, to be
370	baru	new, to be
371	lama (benda)	old, to be (thing)
372	bulat	round, to be
373	lurus	straight, to be
374	sempit	narrow, to be
375	tebal	thick, to be

	Malay	English
376	nipis	thin, to be (thing)
377	licin	smooth, to be
378	lebar	wide, to be
379	panjang	long, to be (thing)
380	keras	hard, to be
381	berat	heavy, to be
382	sama	same
385	lain	other
386	apa	what?
387	siapa	who?
388	bila	when?
389	berapa	how many?
391	satu	one
392	dua	two
401	tiga	three
402	empat	four
403	lima	five
404	enam	six
405	tujuh	seven
406	lapan	eight
407	sembilan	nine
408	sepuluh	ten
393	banyak (benda)	many (thing), much
394	ramai, banyak (orang)	many (human)
395	semua	all
396	sedikit	some
397	jauh	far, to be
398	dekat	near, to be
399	kanan	right (side)
400	kiri	left (side)
509	keluar	to go/come out
510	pandang dgn tepat	to stare
512	mengikuti	to follow/pursue
513	memulas (air dari kain)	to wring
514	menyembunyikan diri	to hide oneself
515	bersendawa	to burp
516	berkokok (ayam jantan)	to crow/sing (rooster)
517	kesat, kasar (tak licin)	rough (to the touch)
518	ada	there is
519	bujang (lelaki)	bachelor
520	bangun	to get up

	Malay	English
521	bau ikan segar (bukan busuk)	smell of fresh fish
528	belum masak (buah)	still unripe (fruit)
522	masak (buah)	ripe
523	menyengat	to sting

	Malay	English
524	anak yatin, piatu	orphan
525	membuat	to make
526	di belakang, selepas, terakhir	in the back, after, the last
527	senja, senjakala	to be twilight

Certain items on the wordlist caused problems. Table 13 is a compilation of the various issues that arose regarding the wordlist. These comments include items that were removed from the list, in order to help future researchers avoid some of the problems that were encountered.

Table 13. Comments on problems

No.	English	Malay	Comment
6	river	sungai	Usually the word for 'water' is given.
8b	swim (animal)	berenang (haiwan)	Usually the same word as for humans swimming.
9a 9b	float: on the surface float: rise to the surface	terapung timbul	These two words were often confusing despite explanations.
11	bathe	mandi	Care was needed to avoid eliciting the word for 'rub'.
16	thunder	guntur, guruh	The usual word given seems actually to be the name of a legendary spirit in the clouds that is responsible for making the noise.
20	east	timur	Usually BM. Removed from list.
21	west	barat	Usually BM. Removed from list.
25	earth, soil	tanah	Care was needed to stay with the BM 'tanah', and avoid the various definitions of the English 'earth'.
29	gold	emas	Usually BM. Removed from list.
30	silver	perak	Usually BM. Removed from list.
31	iron	besi	Usually BM. Removed from list.
33	path (small)	jalan (kecil)	The addition of the word 'small' was needed to avoid confusion with the alternate definition for the BM 'jalan'– that is, 'to go'.
36	ashes	abu	There did not seem to be a word for this. Rather, the word for 'dust' was usually given.
39	burn off feathers	layur, bakar bulu	Often the word 'layur' was not known. Needed to be careful with the alternate phrase 'bakar bulu' in order to be sure it was not heard as 'bakar buluh'.
49a	tree trunk	batang pokok	Often seemed to be confusing.
59	pepper (red)	lada	Usually BM. Removed from list.
61	bamboo	buluh	It appears some of the variation in the response was due to whether the speaker was referring to young or old bamboo.

No.	English	Malay	Comment
71	blow-gun	sumpitan	Care was needed to get the object and avoid getting the verb 'sumpit' (to blow).
73	dart-head	pangkal damak	Care was needed because of the confusion between the 'head' and the 'butt', as to which is the sharp end and which the dull.
74a 74b	dart (with poison) dart (without poison)	damak (yang ada racun) damak (yang tiada racun)	Some dialects did not seem to have a distinction, while others had as many as three names (e.g. with no poison, with a little poison, and with a lot of poison)
78a 78b	kill cause to die	bunuh matikan	Attempts to distinguish between these two words caused confusion.
79	stab	tikam	There may be a difference depending on the direction and force of the thrust, as well the type of instrument.
83	porcupine	landak	There appear to be two species: large and small.
84	monkey (short-tailed)	monyet (ekor pendek)	There are many names for the different types of monkeys, but apparently not for monkeys in general. Asking for the name of the short-tailed monkey helped some, but 'monyet' in BM implies a long-tailed monkey. Also, in some areas there exist more than one type of short-tailed monkey.
90	to bark	menyalak	There may be more than one type of barking, perhaps akin to 'yelp' and 'howl'.
92	rat	tikus	The response seemed to be either a particular species name, or else the general term for 'rodent'.
101	bee	lebah	It appears some dialects distinguish different species.
103	insect	serangga	Apparently there is no general term. Removed from list.
104	mosquito	nyamuk	Many dialects distinguish different species.
108	spider	labah-labah	Some dialects appeared to have a general term, while others distinguished several species.
109	cockroach	lipas	Most dialects appeared to distinguish at least large and small varieties.
111	monitor lizard	biawak	At least one dialect distinguished between the type that lives on land versus the type that lives in the water. Hence the variation in response may reflect different species.
114a 114b	turtle (sea) turtle (river)	penyu labi-labi	Since the Semai do not live near the sea, the distinction between sea and river turtles was often confusing. There may be confusion with land tortoises as well.
117b	person, non-OA	orang, bukan-OA	Apparently there is no general term; rather, the response was often 'others', 'different', or a specific people group such as 'Malay' or 'Chinese'.

No.	English	Malay	Comment
119	female	perempuan	Apparently there is no distinction between 'woman' and 'female'.
123	father	bapa	Diffloth's data implies there are different terms– perhaps one for the relationship and another for the term of address.
124a 124b	elder brother elder sister	abang kakak	There is usually just one term for 'older sibling', not distinguishing gender.
125a 125b	younger brother young sister	adik lelaki adik perempuan	There is usually just one term for 'younger sibling', not distinguishing gender.
130	friend	kawan	This word was narrowed to just 'friend', as opposed to also including 'companion'.
135a 135b	bald (natural) bald (shaved)	botak (telah gugur) botak (telah dicukur)	Most dialects did make a distinction between these two.
136a 136b	cut hair with scissors cut hair with something else	menggunting rambut potong rambut (pakai alat lain)	Attempts to make a distinction between these two usually caused confusion.
141	forehead	dahi	The response sometimes appeared to refer to the fontanel (the soft spot on a baby's head).
145 152a 152b	throat neck (back of) neck (front of)	kerongkong tengkok leher	The Semai do not appear to make all three of these distinctions.
146	mouth	mulut	There did not seem to be a general term, hence this caused confusion. Removed from the list.
147b 147c	lip, upper lip, lower	bibir (di atas) bibir (di bawah)	There did not appear to be different words for upper and lower lips.
160	finger	jari	Initially, attempts were made to distinguish between 'finger' and 'digit'; however, this usually caused confusion. Nevertheless, some of the responses implied another (or similar?) distinction that was not well understood.
161b 161c	abdomen, upper abdomen, lower	perut, di atas pusat perut, di bawah pusat	Usually there was just one general term. Attempts to elicit a distinction sometimes resulted in the words for 'chest' and 'waist', respectively.
165 166 168	lungs heart liver	paru-paru jantung hati	There was sometimes confusion regarding these internal organs. Certainly some of the confusion was due to the center of emotions, and consequently quite a number of phrases expressing emotion, being the 'heart' in English, but the 'liver' in BM.
167	intestines	tali perut, usus	One dialect made a distinction between the large and small intestines. Hence some of the variation in responses reflect this distinction.

No.	English	Malay	Comment
181	strong	kuat	The term 'kuat' seemed to be too generic, since it can have rather different meanings depending on what it refers to (e.g. people, trees, rain, etc.) Hence the term was elicited in reference to 'a strong person'.
183b	sleep soundly	tidur nyenak	There did not seem to be a special word for this. Responses tended to describe the state.
184	snore	berdengkur	The response was usually the word for 'breathe', or related to this word.
185	yawn	menguap	Needed to specify this is related to sleepiness, so that there is no confusion with being agape for other reasons, such as eating something hot or spicy.
189	wink	kelip mata	Needed to specify that this was one eye only, not blinking both eyes.
192	sniff	cium, menghidu	The example of sniffing a flower was specified so that there is no confusion with the other meaning of 'cium', i.e. 'to kiss'.
193a	decayed	reput	Needed to specify this referred to wood, for example, and not food.
193b	rotten	basi	Needed to specify this referred to cooked food that has spoiled.
195	satiated, full	kenyang	Needed to guard against confusion with the other meaning of the English word 'full', e.g. a full container.
198a	to suck out	sedut	The term 'sedut' was too general. Needed to specify the context of sucking marrow from bone, or sucking out snails
198b	to suck on	hisap	The term 'hisap' was too general. Needed to specify sucking on candy.
205	hurt	sakit	The term 'sakit' can also mean 'sickness'; hence needed to specify the example of one's hand that is hurting.
207	itchy	gatal	Needed to specify the condition of being itchy (e.g. after being bitten by a mosquito, rather than the property of causing itchiness (e.g. 'miang', the fine hair on bamboo that feels itchy).
209a	rake	cakar	This term needed the example of raking leaves or rubbish.
209b	scratch (chicken)	kais	This term needed the example of a chicken scratching for food.
210	cold	sejuk	Care was needed to separate this term from the condition of feeling cold from a fever.
234	old (animate)	tua	Sometimes the response was 'old man' and/or 'old woman', rather than the generic word for simply 'old'.
235	fat	gemuk	Care was needed to avoid confusion with the English noun, as in 'animal fat or lard'.

No.	English	Malay	Comment
237a 237b	tall (person) tall (tree)	tinggi (orang) tinggi (pokok)	Usually the response was the same for both people and trees.
238a 238b	short (person) short (tree)	pendek (orang) pendek (pokok)	Usually the response was the same for both people and trees.
243	to blow	hembus	Too generic. Removed from list.
244	to blow on	tiup	To avoid possible confusion because of the general term, the example of 'to blow on a fire' was used.
258 259	to smile to laugh	senyum ketawa	Some dialects made this distinction, but others had the same word for both.
260	good	baik	The example of 'good person' was used.
261	bad	jahat	The example of 'bad person' was used.
267	to fight	lawan, gaduh	This seemed to cause confusion, or there wasn't a ready term. Usually the response was the same as 'hit'.
274	to call	memanggil	The term 'memanggil' was too general. The example of 'to call to get someone's attention' was used.
276	story	cerita	This seemed to cause confusion, or there wasn't a ready term. Sometimes the response was the word 'legend'.
278	to whistle	bersiul	Usually the term 'bersiul' was not known. Needed to demonstrate by whistling.
286 288	to fall (animate) to fall from a height	jatuh (orang) jatuh dari tempat tinggi	Did not seem to be a distinction.
289a	to carry a bag	jinjing	Needed to specify the example of carrying a plastic bag dangling from one's hand.
289b 290	to carry in arms (child) to carry in cloth	mendukung (anak) ambin	There was sometimes unresolved confusion between these two terms despite trying to specify examples.
292	to carry on shoulder	pikul	Care was needed to avoid confusion with carrying on a pole between two people.
303 304 305 306	to throw sidearm to throw overhand to toss to throw away	lontar baling, lempar buang buang (sampah)	The Semai appear to make different distinctions; that is, what they throw and the terms for how they throw it did not seem to be organized the same as BM. Responses were almost always confused.
315	yesterday	kelmarin, semalam	The term 'semalam' may have been interpreted as meaning 'last night' only, and not the more generic 'yesterday'.
317	lean-to	pondok sementara	Needed to specify a temporary shelter when one was forced to overnight in the jungle.
333	to tie by twisting	simpul	Needed to specify the example of the twisting/tying of a man's sarong so that it does not fall down.

No.	English	Malay	Comment
336	to wash (hands)	cuci	Needed to give the example of washing hands, as opposed to clothes.
337	to wash (clothes)	basuh	Need to specify clothes.
338	to rub (clothes while washing)	gosok	The term 'gosok' was too general. The example of 'to rub while washing clothes' was used.
340 341	dry to dry	kering jemur	Needed to avoid confusion in the English between 'the state of being dry', and 'to dry something'.
346	to cook	memasak	Needed to avoid confusion in BM of the adjective 'masak', meaning 'ripe'.
347a 347b	to boil (water) to boil (tapioca)	didih rebus	Needed to distinguish between 'boiling water' and 'cooking something by boiling it in water'.
350	full	penuh	Needed to guard against confusion with the other meaning of the English word 'full', e.g. to be full (satiated).
355	knife	pisaw	Many Semai make a distinction between a knife (small) and a 'parang' (large knife, or machete). This distinction was not known for the early lists.
356	knife, hooked	pisaw	This type of knife was unknown to the Semai. Removed from list.
361	sour	masam	For many Semai, their word also could mean 'salty'.
384	also	juga	Usually BM. Removed from list.
390	count	bilang	Usually BM. Removed from list.
393	much	banyak	This distinction, as opposed to 'many', was merged into one item.
395	everyone	semua	This distinction, as opposed to 'all', was merged into one item.
511	to grope	menyentuh	Removed from list because of somewhat embarrassing (e.g., sexual) connotations.
522	ripe	masak	Needed to give the example of fruit, to avoid confusion with the other BM 'masak', meaning 'to cook'.
529	to spread	hampar	Too generic. Removed from list.

Appendix C: Language assistant questionnaires

Questionnaires were used in interviews with many of the language assistants who gave the wordlists. The questions were developed in order to establish, among other things, the reliability of the data. For example, it is reasonable to expect that a language assistant who has always lived in a certain language community and has married someone also from that community will be able to provide more reliable data than another individual who has recently moved there or has married someone from another dialect area.

A few questions also probed language attitudes and language use. Again, this helps in establishing reliability of the data. A good language assistant is one who loves his or her language and uses it as often as possible.

Lastly, a number of questions asked the language assistant about how well he or she was able to understand and/or communicate in other dialects. The answers to these questions may provide preliminary direction for future research into how well the Semai speakers are able to communicate with those from other dialects despite the many phonological and lexical differences.

The following pages provide the data that was collected. [Please note the following abbreviations: BM = Bahasa Malaysia, BS = Bahasa Semai.]

Language Code SEA-F
Gender M
Age 30+
Born where? Kg. K. Kenip, Lipis, Pahang
Grew up where? Kg. K. Kenip, Lipis, Pahang
Language used at school BM
Now lives where? Kg. K. Kenip, Lipis, Pahang
Has ego lived elsewhere? (no answer)
Where else?
Father born where? Kg. Belida, Lipis, Pahang
Father grew up where? Kg. Belida, Lipis, Pahang
Mother born where? Kg. K. Kenip, Lipis, Pahang
Mother grew up where? Kg. K. Kenip, Lipis
Spouse born where? Kg. K. Kenip Ulu, Lipis
Spouse grew up where? Kg. K. Kenip Ulu, Lipis
Ego used what lang. as a child? BS only
What language does ego use with…
Spouse BS only
Parents BS only
First Child BS only
Second Child BS only
Third Child BS only
Fourth Child BS only
Fifth Child BS only
Sixth Child BS only
What language does ego's spouse use with…
Ego BS only
First Child BS only
Second Child BS only
Third Child BS only
Fourth Child BS only
Fifth Child BS only
Sixth Child BS only
What language do the children use with…
their grandparents BS only
their friends BS only
What language does ego use to…
buy food BM only
sell to others from the [local] lang. BS only
sell to outsiders BM only
speak to teachers BM only
speak at a government office BM only
Does ego mix language with friends? No
If so, what languages?
Has ego written in the [local] language? No
If so, what?
In what language does ego…
think BS
talk about health BS
talk about finances BS
pray BS
talk about spiritual things BS
dream BS
count BM
speak when startled
speak to animals BS
Does ego want his/her grandchildren to speak the [local] language? Yes
Does ego think it is good if his/her grandchildren can read the [local] language? Yes
What advantages does ego see in being able to speak the [local] language?
identity, better relationship within community
What advantages does ego see in being able to read and write in the [local] language?
identity, language preservation
Which dialect is the hardest for this ego to understand?
Where is that dialect spoken?
What language does ego use with adults from…

The same kampung	(no answer)
Jelengkok, Betau	BS only
Simoi Baru, Lipis	BS only
Pos Buntu	BS only
Bertang	BS only
	(no answer)
Batu 17, Tapah	BS only
Bidor	BS only
Kg Chinggung, T Malim	BM only
Gopeng, Kinta	BS only
Kampar	BS only

Has ego met people from?	Did they use their own language?	How well could ego understand?
Temiar	No	0
Temuan	No	0
Ja Hut	No	0
Jakun	No	0
	No	0

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? No
Which? Radio 7
Can ego understand? Yes
Is ego interested in learning to speak Temiar, Jakun, Semelai? Yes
If so, which ones? Temiar
Other notes:

Language Code SEA-W
Gender female
Age 55
Born where? Sg. Bil
Grew up where? Sg. Bil
Language used at school tak tahu
Now lives where? Sg. Bil
Has ego lived elsewhere? No
Where else?
Father born where? Kuala Slim
Father grew up where? Kuala Slim
Mother born where? Sg. Bil
Mother grew up where? Sg. Bil
Spouse born where? Kg. Rasau
Spouse grew up where? Kg. Rasau
Ego used what lang as a child? BS only
What language does ego use with…
- **Spouse** BS only
- **Parents** BS only
- **First Child** BS only
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language does ego's spouse use with…
- **Ego** BS only
- **First Child** BS only
- **Second Child** BS only
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language do the children use with…
- **their grandparents** BS only
- **their friends** BS only

What language does ego use to…
- **buy food** BM only
- **sell to others from the [local] lang.** BS only
- **sell to outsiders** BM only
- **speak to teachers** BM only
- **speak at a government office** BM only

Does ego mix language with friends? No
If so, what languages?
Has ego written in the [local] language? No
If so, what?
In what language does ego…
- **think** BS
- **talk about health** BS
- **talk about finances** BS
- **pray** BS
- **talk about spiritual things** BS
- **dream** BS
- **count** BS
- **speak when startled** BS
- **speak to animals** BS

Does ego want his/her grandchildren to speak the [local] language? Yes
Does ego think it is good if his/her grandchildren can read the [local] language? Yes
What advantages does ego see in being able to speak the [local] language?
people understand better
What advantages does ego see in being able to read and write in the [local] language?
easy to understand
Which dialect is the hardest for this ego to understand?
Jelai (Pahang), Semai in Selangor, Terisu (Perak)
Where is that dialect spoken
Jelai (Pahang), Semai in Selangor, Terisu (Perak)
What language does ego use with adults from…

The same kampung	BS only
Kluny, Slim River	BS only
Tapah	BS only
Pos Slim, Parit	BS only
Bota, Parit	BS only
Batu 7, Kampar	BS only
Betau, Pahang	BS only
Telom, Pahang	BM only
Jelai, Pahang	BM only
Batu 17, CH	BS only
Simoi, Raub	BM only

Has ego met people from?	Did they use their own language?	How well could ego understand?
Temiar	No	0
	No	
	No	
	No	
	No	

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? Yes
- **Which?** Semai, Temiar

Can ego understand? No
Is ego interested in learning to speak Temiar, Jakun, Semelai? No
- **If so, which ones?**

Other notes:

Language Code SEA-V
Gender female
Age 24
Born where? Kg. Chinggung, Behrang, T.Malim, Perak
Grew up where? Kg. Chinggung, Behrang,
Language used at school BM
Now lives where? Kg. Chinggung, Behrang
Has ego lived elsewhere? (no answer)
Where else?
Father born where?
Father grew up where? Langkap
Mother born where? Kg. Chinggung, Behrang
Mother grew up where? Kg. Chinggung
Spouse born where? na
Spouse grew up where? na
Ego used what lang as a child? BS only
What language does ego use with…
- **Spouse** (no answer)
- **Parents** BS only
- **First Child** (no answer)
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language does ego's spouse use with…
- **Ego** (no answer)
- **First Child** (no answer)
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language do the children use with…
- **their grandparents** (no answer)
- **their friends** (no answer)

What language does ego use to…
- **buy food** BM only
- **sell to others from the [local] lang.** BS only
- **sell to outsiders** BM only
- **speak to teachers** BM only
- **speak at a government office** BM only

Does ego mix language with friends? No
If so, what languages?
Has ego written in the [local] language? Yes
If so, what? letter
In what language does ego…
- **think** BS
- **talk about health** BS (kampung), BM (with doktor)
- **talk about finances** BS(kampung), BM(outside)
- **pray** BS
- **talk about spiritual things** BS (kampung), BM (outside)
- **dream** BS
- **count** BS
- **speak when startled** BS
- **speak to animals** BS

Does ego want his/her grandchildren to speak the [local] language? Yes
Does ego think it is good if his/her grandchildren can read the [local] lang.? Yes
What advantages does ego see in being able to speak the [local] language?
easier to express self in own language
What advantages does ego see in being able to read and write in the [local] language?
good if you have 2 languages, BM & BS
Which dialect is the hardest for this ego to understand?
Where is that dialect spoken?
What language does ego use with adults from…

The same kampung	(no answer)
Pos Slim, Parit	BM only
Kampar Gopeng	BM only
Batu 7, Kampar	BM only
Kluny, Slim River	BS only
Tapah	BS only
Raub, Pahang	BM only
Jelintoh, Pahang	BM only
Cherong, Pahang	BM only
Jelai	BM only
Telom	BM only

Has ego met people from?	**Did they use their own language?**	**How well could ego understand?**
Temiar	No	0
Jakun	No	little bit, related to Temuan
	No	
	No	
	No	

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? Yes
- **Which?** Temuan, Semai;older helper: Jakun, Temuan, Semai

Can ego understand? Yes
Is ego interested in learning to speak Temiar, Jakun, Semelai? Yes
- **If so, which ones?** yang dari jauh

Other notes:

Language Code SEA-U
Gender male
Age 33
Born where? Kg. Cherong, Sinderut, Pahang
Grew up where? Kg. Cherong, Sinderut, Pahang
Language used at school BM
Now lives where? Kg. Cherong, Sinderut
Has ego lived elsewhere? No
Where else?
Father born where? Kg.Ganchar (near Cherong)
Father grew up where? Kg. Ganchar
Mother born where? Kg. Jakek, Sg. Cherong
Mother grew up where? Kg. Jakek, Sg. Cherong
Spouse born where? Kg. Cherong, Sinderut
Spouse grew up where? Kg. Cherong, Sinderut
Ego used what lang as a child? BS only
What language does ego use with…
Spouse BS only
Parents BS only
First Child BS only
Second Child (no answer)
Third Child (no answer)
Fourth Child (no answer)
Fifth Child (no answer)
Sixth Child (no answer)
What language does ego's spouse use with…
Ego BS only
First Child BS only
Second Child (no answer)
Third Child (no answer)
Fourth Child (no answer)
Fifth Child (no answer)
Sixth Child (no answer)
What language do the children use with…
their grandparents BS only
their friends BS only
What language does ego use to…
buy food BM only
sell to others from the [local] lang. BS only
sell to outsiders BM only
speak to teachers BM only
speak at a government office BM only
Does ego mix language with friends? Yes
If so, what languages? BS & BM
Has ego written in the [local] language? No
If so, what?
In what language does ego…
think BS
talk about health BS
talk about finances BS
pray BS
talk about spiritual things BS
dream BS
count BS
speak when startled BS
speak to animals BS
Does ego want his/her grandchildren to speak the [local] language? Yes
Does ego think it is good if his/her grandchildren can read the [local] language? Yes
What advantages does ego see in being able to speak the [local] language?
to not forget their own language, & knowledge (med)
What advantages does ego see in being able to read and write in the [local] language?
memberi erti terutamanya kpd yg tak tahu BM
Which dialect is the hardest for this ego to understand?
Where is that dialect spoken? Parit and Ipoh
What language does ego use with adults from…

The same kampung	BM only
Kg. Tenau	BS only
Kg. Buntu, Kelang	BS only
Kg. Buntu, Kerayong, Jelengek	(no answer)
kg. Santat, Sempoh, Labu	BS only
	(no answer)
Parit	(no answer)
Gopeng, Kampar	BM only
Slim River	mostly BS
	(no answer)
	(no answer)

Has ego met people from?	Did they use their own language?	How well could ego understand?
Temiar	No	0
Lanoh	No	0
JaHut	No	0
	No	
	No	

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? No
Which? Semai (faham), Temiar, Jakun Semelai, Temuan
Can ego understand? Yes
Is ego interested in learning to speak Temiar, Jakun, Semelai? Yes
If so, which ones? Temuan
Other notes:

Language Code SEA-S
Gender female
Age 24
Born where? Gombak
Grew up where? Gombak (till age 6) Jelangkok (RPS) Betau
Language used at school BM
Now lives where? Kg. RPS, Betau
Has ego lived elsewhere? Yes
Where else? Kg. Simoi Lama, Jelai, Kg Sinderut
Father born where? Bkt. Long, Kg. Sinderut
Father grew up where? Kg. RPS Betau, K. Lipis
Mother born where? Baru Gandang, Jelai, K. Lipis
Mother grew up where? Kg. Simoi Lama, Jelai
Spouse born where? Fiance– Kg. Simoi Lama, Jelai
Spouse grew up where? Kg. Simoi Lama, Jelai
Ego used what lang as a child? mostly BS
What language does ego use with…
- **Spouse** mostly BS
- **Parents** mostly BS
- **First Child** (no answer)
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language does ego's spouse use with…
- **Ego** mostly BS
- **First Child** (no answer)
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language do the children use with…
- **their grandparents** BS only
- **their friends** mostly BS

What language does ego use to…
- **buy food** BM only
- **sell to others from the [local] lang.** BS only
- **sell to outsiders** BM only
- **speak to teachers** BM only
- **speak at a government office** BM only

Does ego mix language with friends? Yes
If so, what languages? BS & BM
Has ego written in the [local] language? Yes
If so, what? letter, list
In what language does ego…
- **think** BS
- **talk about health** BS
- **talk about finances** BS
- **pray** BS
- **talk about spiritual things** BS
- **dream** BS & BM
- **count** BS & BM
- **speak when startled** BS
- **speak to animals** BS

Does ego want his/her grandchildren to speak the [local] language? Yes
Does ego think it is good if his/her grandchildren can read the [local] language? Yes
What advantages does ego see in being able to speak the [local] language?
Language preservation, identity
What advantages does ego see in being able to read and write in the [local] language?
Language preservation
Which dialect is the hardest for this ego to understand?
Where is that dialect spoken?
What language does ego use with adults from…

The same kampung	mostly BS
Tenau	BS only
Batu 17, CH	BS only
CH	BS only
Raub	BS only
Telom	BS only
Gopeng, Perak	(no answer)
T. Malim	BM only
Slim River	BS only
Chingugng, Kluny	BS only
	BM only

Has ego met people from?	Did they use their own language?	How well could ego understand?
	No	
	No	
	No	
	No	
	No	

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? No
- **Which?**

Can ego understand? No
Is ego interested in learning to speak Temiar, Jakun, Semelai? No
- **If so, which ones?**

Other notes:

Language Code	SEA-N
Gender	male
Age	49
Born where?	Kg. Chang Lama, Bidor, Perak
Grew up where?	Kg. Chang Lama, Bidor, Perak
Language used at school	BS & BM
Now lives where?	Kg. Chang Lama, Bidor
Has ego lived elsewhere?	No
Where else?	
Father born where?	Kg. Chang Lama, Bidor
Father grew up where?	Kg. Chang Lama, Bidor, Perak
Mother born where?	no answer
Mother grew up where?	
Spouse born where?	Kg. Chang Lama, Bidor
Spouse grew up where?	Kg. Chang Lama, Bidor
Ego used what lang as a child?	BS/BM equally
What language does ego use with…	
Spouse	BS only
Parents	BS only
First Child	BS only
Second Child	(no answer)
Third Child	(no answer)
Fourth Child	(no answer)
Fifth Child	(no answer)
Sixth Child	(no answer)
What language does ego's spouse use with…	
Ego	BS only
First Child	BS only
Second Child	(no answer)
Third Child	(no answer)
Fourth Child	(no answer)
Fifth Child	(no answer)
Sixth Child	(no answer)
What language do the children use with…	
their grandparents	BS only
their friends	BS/BM equally
What language does ego use to…	
buy food	BM only
sell to others from the [local] lang.	BS only
sell to outsiders	BM only
speak to teachers	BM only
speak at a government office	BM only
Does ego mix language with friends?	No
If so, what languages?	
Has ego written in the [local] language?	No
If so, what?	
In what language does ego…	
think	BS
talk about health	BS
talk about finances	BS
pray	BS
talk about spiritual things	BS
dream	BS
count	BS
speak when startled	BS
speak to animals	BS
Does ego want his/her grandchildren to speak the [local] language?	Yes
Does ego think it is good if his/her grandchildren can read the [local] language?	Yes
What advantages does ego see in being able to speak the [local] language?	easier to understand
What advantages does ego see in being able to read and write in the [local] language?	better
Which dialect is the hardest for this ego to understand?	Jelai, Pahang
Where is that dialect spoken?	

What language does ego use with adults from…

The same kampung	(no answer)
Tumbuh Hangat	BS only
Gopeng	BS only
Kampar	BS only
Sungkai	BS only
Kluny	BS only
Raub	BS only
Lipis	BS only
Betau	BS only
Jelai	BS only
Telom	BS only

Has ego met people from?	Did they use their own language?	How well could ego understand?
Temiar	Yes	litlle bit
Lanoh	No	0
Sabüm	No	0
Semnam	No	0
Ja Hut	No	0

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai?	Yes
Which?	some of it
Can ego understand?	No
Is ego interested in learning to speak Temiar, Jakun, Semelai?	No
If so, which ones?	
Other notes:	

Language Code SEA-T
Gender male
Age 25
Born where? Kg. Simoi Lama
Grew up where? Kg Simoi Baru
Language used at school BM
Now lives where? Kg. Simoi Baru
Has ego lived elsewhere? Yes
Where else? Kg. Jelengkek (for 2 months)
Father born where? Kg. Simoi Lama
Father grew up where? Kg Simoi Baru
Mother born where? Kg. Simoi Lama
Mother grew up where? Kg. Simoi Lama
Spouse born where?
Spouse grew up where?
Ego used what lang as a child? BS only
What language does ego use with…
- **Spouse** (no answer)
- **Parents** (no answer)
- **First Child** (no answer)
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language does ego's spouse use with…
- **Ego** (no answer)
- **First Child** (no answer)
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language do the children use with…
- **their grandparents** (no answer)
- **their friends** (no answer)

What language does ego use to…
- **buy food** (no answer)
- **sell to others from the [local] lang.** (no ans.)
- **sell to outsiders** (no answer)
- **speak to teachers** (no answer)
- **speak at a government office** (no answer)

Does ego mix language with friends? (no answer)
If so, what languages?
Has ego written in the [local] language? (no ans.)
If so, what?
In what language does ego…
- **think**
- **talk about health**
- **talk about finances**
- **pray**
- **talk about spiritual things**
- **dream**
- **count**
- **speak when startled**
- **speak to animals**

Does ego want his/her grandchildren to speak the [local] language? No
Does ego think it is good if his/her grandchildren can read the [local] language? No
What advantages does ego see in being able to speak the [local] language?
What advantages does ego see in being able to read and write in the [local] language?
Which dialect is the hardest for this ego to understand? Slim River & Batu 18, Perak
Where is that dialect spoken? Slim River & Batu 18, Perak
What language does ego use with adults from…

The same kampung	BM only
Raub, Pahang	BS only
Telom (Sg. Serau)	BS only
Kg. Sinderut	BS only
Kg Buntu	BS only
Jelai	BS only
Kg. Ulu Groh, Perak	BS only
Kluny, Slim River	BS only
Senderiang	BS only
	BS only
	BS only

Has ego met people from?	**Did they use their own language?**	**How well could ego understand?**
Temiar	No	little
Lanoh	No	0
Ja Hut	No	0
	No	
	No	

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? Yes
- **Which?** Semai, Temiar, Jakun, Semelai

Can ego understand? Yes
Is ego interested in learning to speak Temiar, Jakun, Semelai? Yes
- **If so, which ones?** all

Other notes: In question 47b (speaker only understands Semai)

Language Code SEA-Q
Gender male
Age 42
Born where? Kg. Sungai Teras, Kuala Slim, Perak
Grew up where? Kg. Sungai Teras & Trolak Timur
Language used at school BM
Now lives where? Kg. Sg. Tisung, Sungkai, Perak
Has ego lived elsewhere? Yes
Where else? Trolak Timur (10 yrs)
Father born where? Kg. Rasau, Slim
Father grew up where? Kg. Rasau, Slim
Mother born where? Trolak Timur, Slim River
Mother grew up where? Trolak Timur, Slim River
Spouse born where? Kg. Sg. Tisung, Sungkai,
Spouse grew up where? Kg. Sg. Tisung, Sungkai,
Ego used what lang as a child? BS only
What language does ego use with…
- **Spouse** BS only
- **Parents** BS only
- **First Child** BS only
- **Second Child** BS only
- **Third Child** BS only
- **Fourth Child** BS only
- **Fifth Child** BS only
- **Sixth Child** (no answer)

What language does ego's spouse use with…
- **Ego** BS only
- **First Child** BS only
- **Second Child** BS only
- **Third Child** BS only
- **Fourth Child** BS only
- **Fifth Child** BS only
- **Sixth Child** (no answer)

What language do the children use with…
- **their grandparents** BS only
- **their friends** BS only

What language does ego use to…
- **buy food** BM only
- **sell to others from the [local] lang.** BS only
- **sell to outsiders** BM only
- **speak to teachers** BM only
- **speak at a government office** BM only

Does ego mix language with friends? (no answer)
If so, what languages?
Has ego written in the [local] language? Yes
If so, what? sermon
In what language does ego…
- **think** BS & BM
- **talk about health** BS
- **talk about finances** BS
- **pray** BS
- **talk about spiritual things** BS
- **dream** BS
- **count** BS
- **speak when startled** BS
- **speak to animals** BS

Does ego want his/her grandchildren to speak the [local] language? Yes
Does ego think it is good if his/her grandchildren can read the [local] language? Yes
What advantages does ego see in being able to speak the [local] language?
Identity, secret language
What advantages does ego see in being able to read and write in the [local] language?
preservation of MT
Which dialect is the hardest for this ego to understand? na
Where is that dialect spoken?
What language does ego use with adults from…

The same kampung	BS only
Chang, Bidor	BS only
Gopeng	BS only
Kampar	BS only
Sungkai	BS only
Kluny	BS only
Raub	BS only
Lipis	BS only
Betau	BS only
Jelai	BS only
Telom	BS only

Has ego met people from?	Did they use their own language?	How well could ego understand?
Temiar	No	
Lanoh	No	
Sabüm	No	
Semnam	No	
JaHut	No	

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? No
- **Which?**

Can ego understand? No
Is ego interested in learning to speak Temiar, Jakun, Semelai? No
- **If so, which ones?**

Other notes:

Language Code SEA-R
Gender male
Age 23
Born where? Kg Pos Buntu
Grew up where? Kg Pos Buntu
Language used at school BM
Now lives where? Kg Pos Buntu
Has ego lived elsewhere? Yes
Where else? Bota (6 mos)
Father born where? Slim River
Father grew up where? Kg Pos Buntu
Mother born where? Kg Pos Buntu
Mother grew up where? Kg Pos Buntu
Spouse born where? Kg Pos Buntu
Spouse grew up where? Kg Pos Buntu
Ego used what lang as a child? BS only
What language does ego use with…
- **Spouse** BS only
- **Parents** BS only
- **First Child** BS only
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language does ego's spouse use with…
- **Ego** BS only
- **First Child** BS only
- **Second Child** (no answer)
- **Third Child** (no answer)
- **Fourth Child** (no answer)
- **Fifth Child** (no answer)
- **Sixth Child** (no answer)

What language do the children use with…
- **their grandparents** BS only
- **their friends** BS only

What language does ego use to…
- **buy food** BM only
- **sell to others from the [local] lang.** BS only
- **sell to outsiders** BM only
- **speak to teachers** BS/BM equally
- **speak at a government office** BM only

Does ego mix language with friends? Yes
If so, what languages? BM & BS
Has ego written in the [local] language? Yes
If so, what? stories, songs
In what language does ego…
- **think** BS
- **talk about health** BS
- **talk about finances** BS
- **pray** BS
- **talk about spiritual things** BS
- **dream** BS
- **count** BM
- **speak when startled** BS
- **speak to animals** BS

Does ego want his/her grandchildren to speak the [local] language? Yes
Does ego think it is good if his/her grandchildren can read the [local] language? Yes
What advantages does ego see in being able to speak the [local] language? on tape
What advantages does ego see in being able to read and write in the [local] language? on tape
Which dialect is the hardest for this ego to understand? Gopeng
Where is that dialect spoken?
What language does ego use with adults from…

The same kampung	BS only
K. Lipis	BS only
CH	BS only
Jelai	BS only
Kuala Medang	BS only
	(no answer)
Gopeng	BS only
Bota	BS only
T. Malim	BS only
Bidor	BS only
Tapah	BS only

Has ego met people from?	Did they use their own language?	How well could ego understand?
Temiar	No	0
	No	
	No	
	No	
	No	

Does ego listen to radio broadcasts in Semai, Temiar, Jakun, and Semelai? Yes
- **Which?** Semai, Temiar, Semelai

Can ego understand? Yes
Is ego interested in learning to speak Temiar, Jakun, Semelai? Yes
- **If so, which ones?** all

Other notes: Notes on 47: Semai- Gopeng: sekit, Temir (mixed with BM): sekit, Semelai:sekit (dekat dgn BM)

Appendix D: Isoglosses as determined by phonological changes

The following is a list of the phonological changes found in the Semai dialects studied in this survey.

a) Final preploded nasals became voiceless plosives (in most dialects, except S– Betau, E– Bertang, U– Cherong, R– Pos Buntu, and F– Kuala Kenip).

b) Final glottal stops were lost after long vowels (B– Gopeng).

c) /*oo/ became unrounded to /əə/ (AA– Renglas) or /ɨɨ/ (Y– Serau and Z– Lanai) in all *ɨɨ environments except before final /*-h/ and /*-ʔ/, where it became the diphthong /əu/.

d) /*ɑɑ/ became backed and rounded to /ɒɒ/ in all environments (B– Gopeng and M– Kampar).

e) /*ee/ became centralized and raised to /ɨɨ/ in all environments (B– Gopeng, M– *oo Kampar, and J– Tapah).

f) /*ɤɤ/ became rounded to /oo/ (AA– Renglas), centralized, raised, and rounded to /ʉʉ/ (Y– Serau and Z– Lanai), but simply centralized and raised to /ɨɨ/ in all other dialects. This change happened in all environments.

g) /*u/ became lowered to /o/ in all environments (M– Kampar, X– Sungai Ruil, and DD– Terisu).

h) Final palatal consonants /*-c/ and /*-ɟɲ/ first merged to /*-c/, and then shifted to /-t/ or /-k/. Also, final /*-ɲ/ shifted to /y/ (V– Chinggung and W– Sungai Bil, and some lexical items at H– Cluny). These changes occurred in all environments.

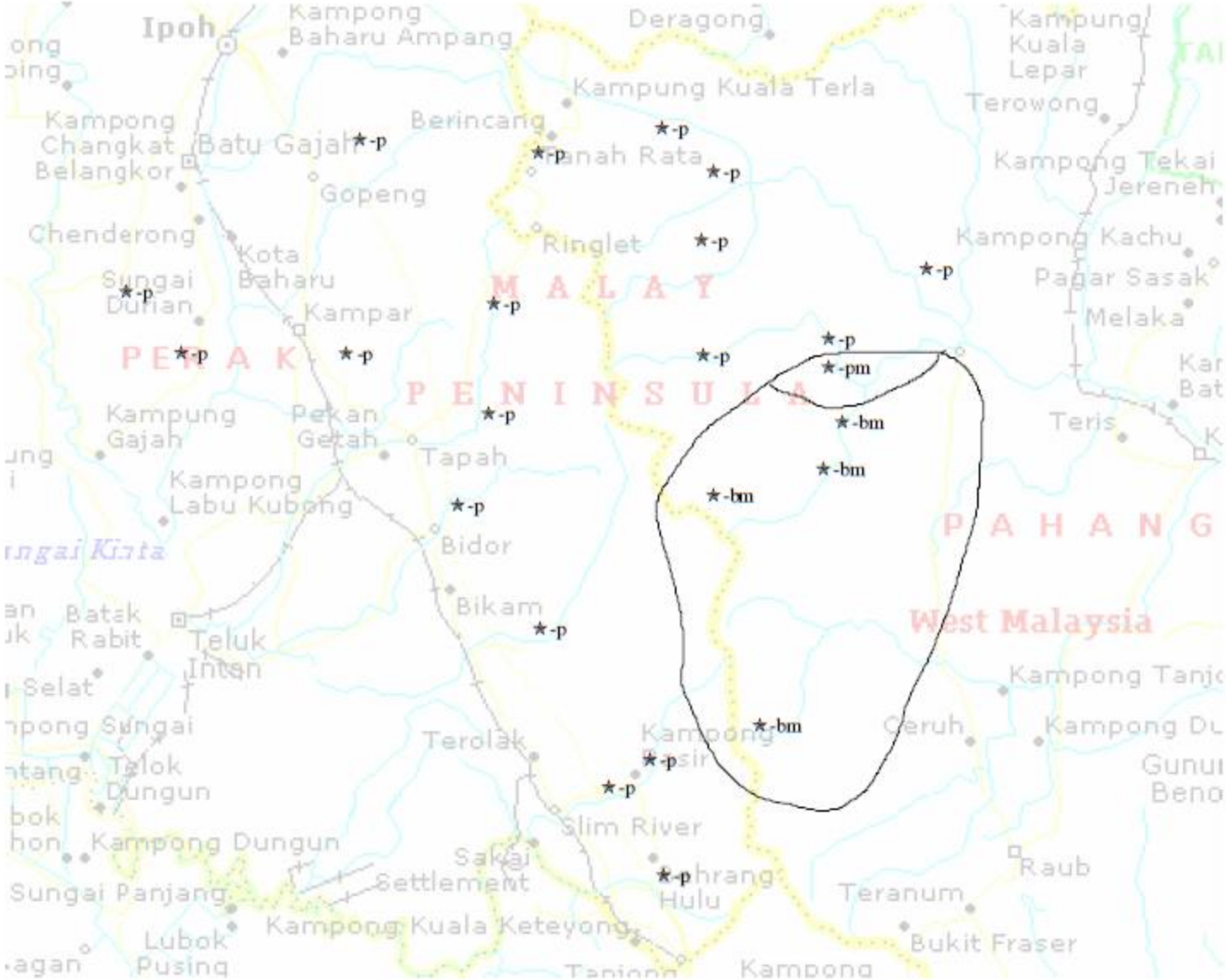
i) /*iə/ became /ii/, /ee/, or /ɛɛ/, in all environments.

j) /*oo/ became /wɑɑ/ before alveolar (*-t, *-ᵈn, *-n, *-r, *-l) and palatal (*-c, *-ʲɲ, *-ɲ, *-s) final consonants (H– Cluny, V– Chinggung, and W– Sungai Bil).

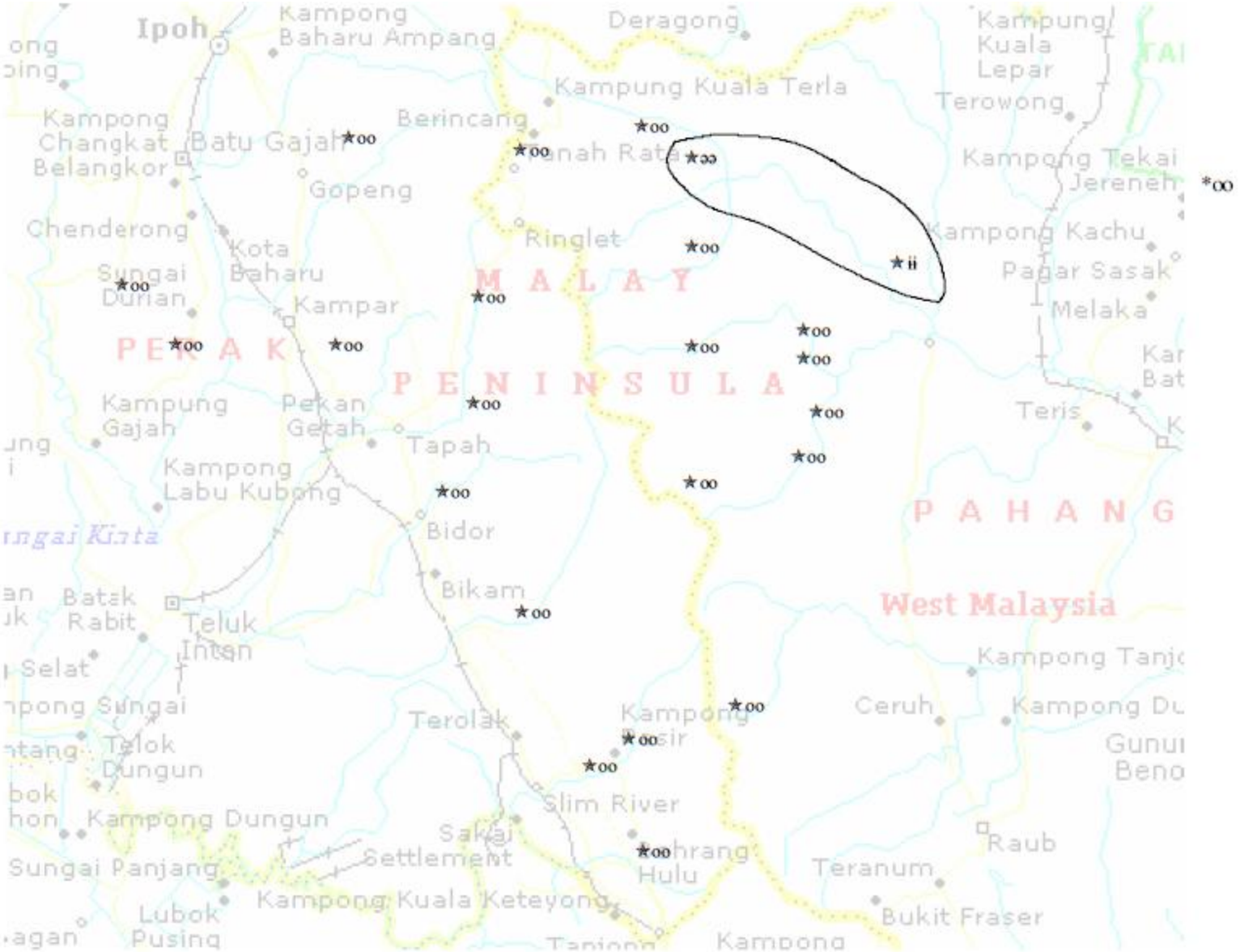
k) /*Noo/ became /Nuu/ (B– Gopeng, G– Tangkai Cermin, O– Bota, X– Sungai Ruil, DD– Terisu), or /Nə̃ə̃/ (AA– Renglas), or /Nɔɔ/ (all other dialects).

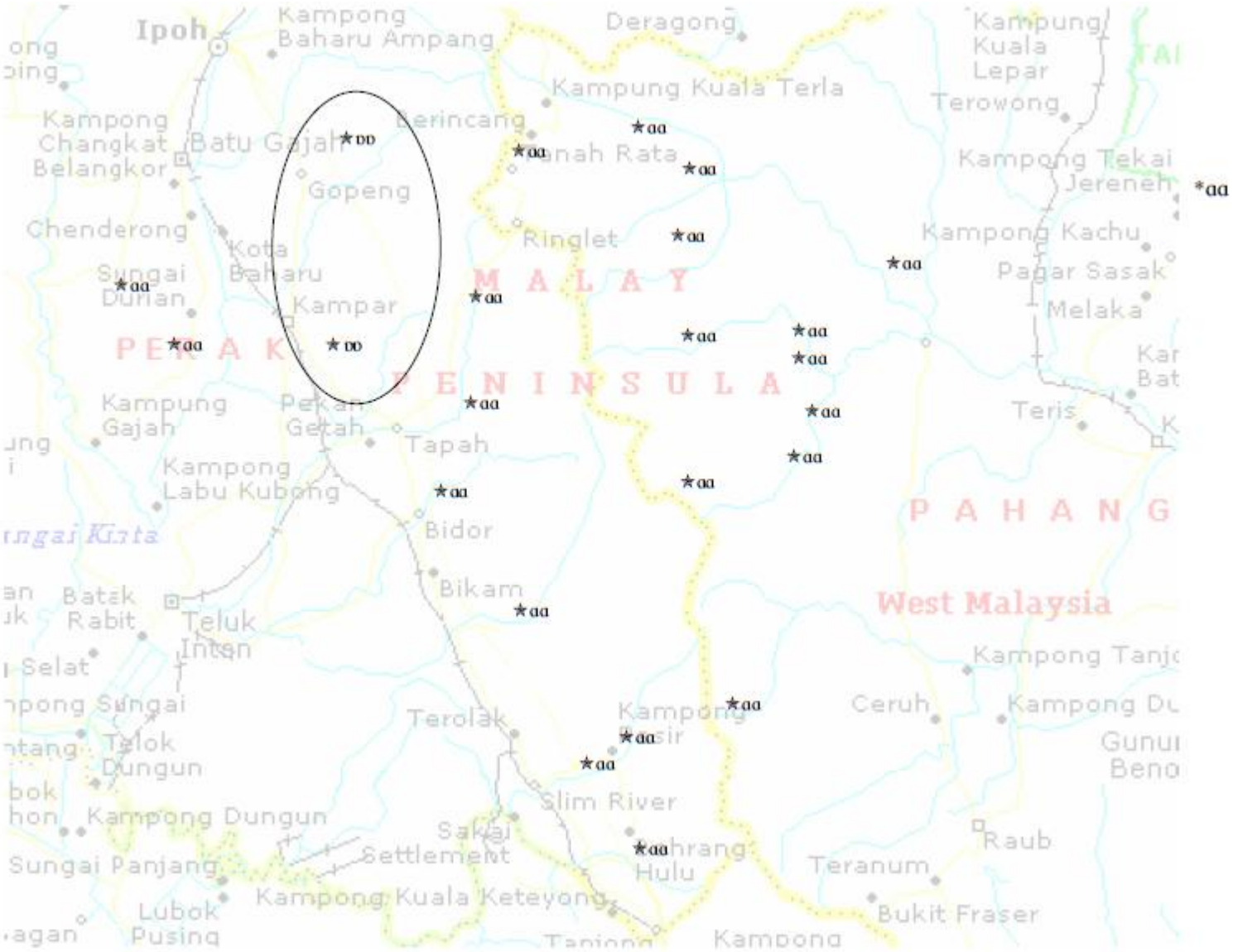
l) /*ɛɛ/ and /*iə/ became /yɑɑ/ before final labials (*-p, *-ᵇm, *-m, *-w) and final alveolars (*-t, *-ᵈn, *-n, *-r, *-l, *-s) (H– Cluny, V– Chinggung, and W– Sungai Bil).

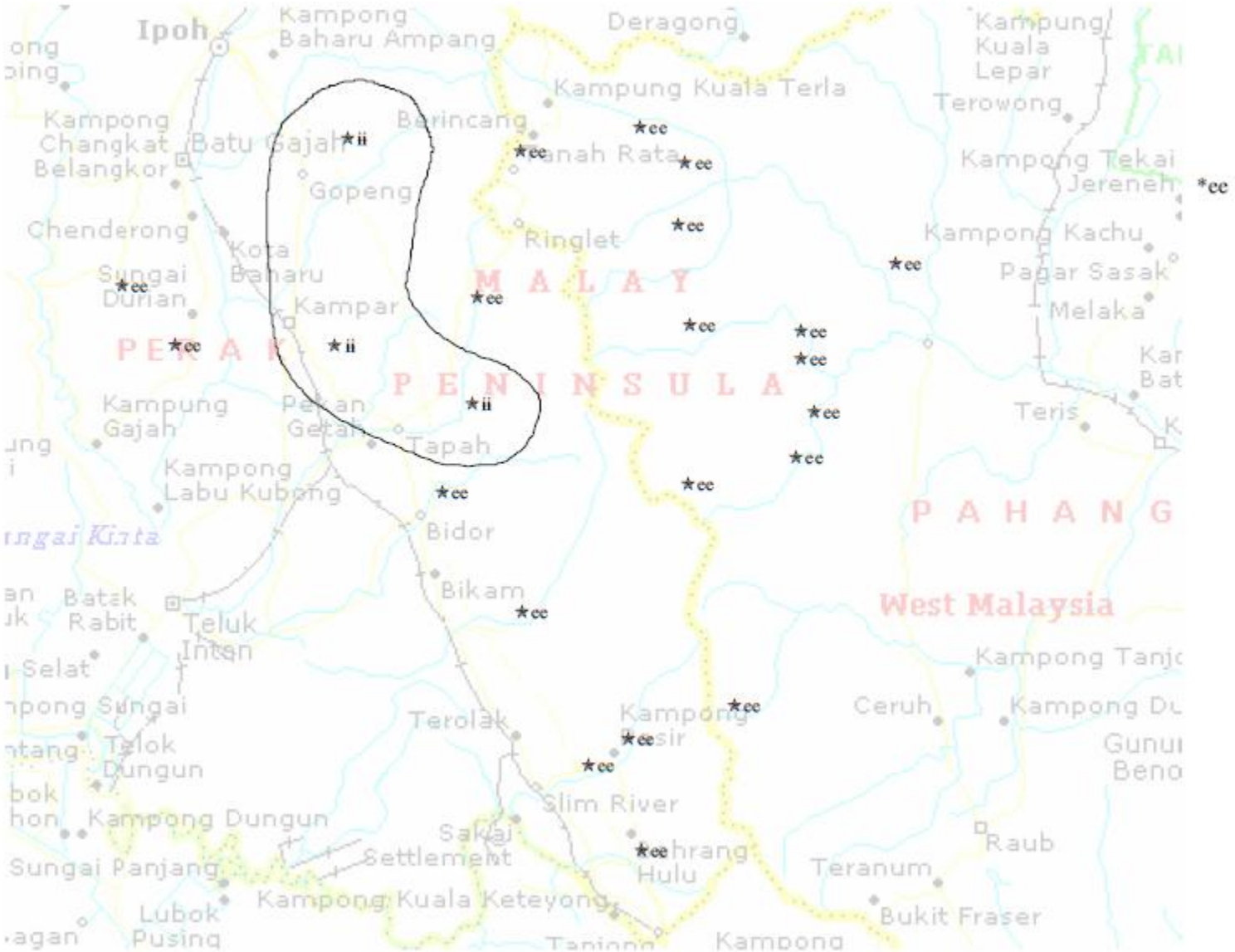
The following pages show the isogloss maps for these phonological changes.

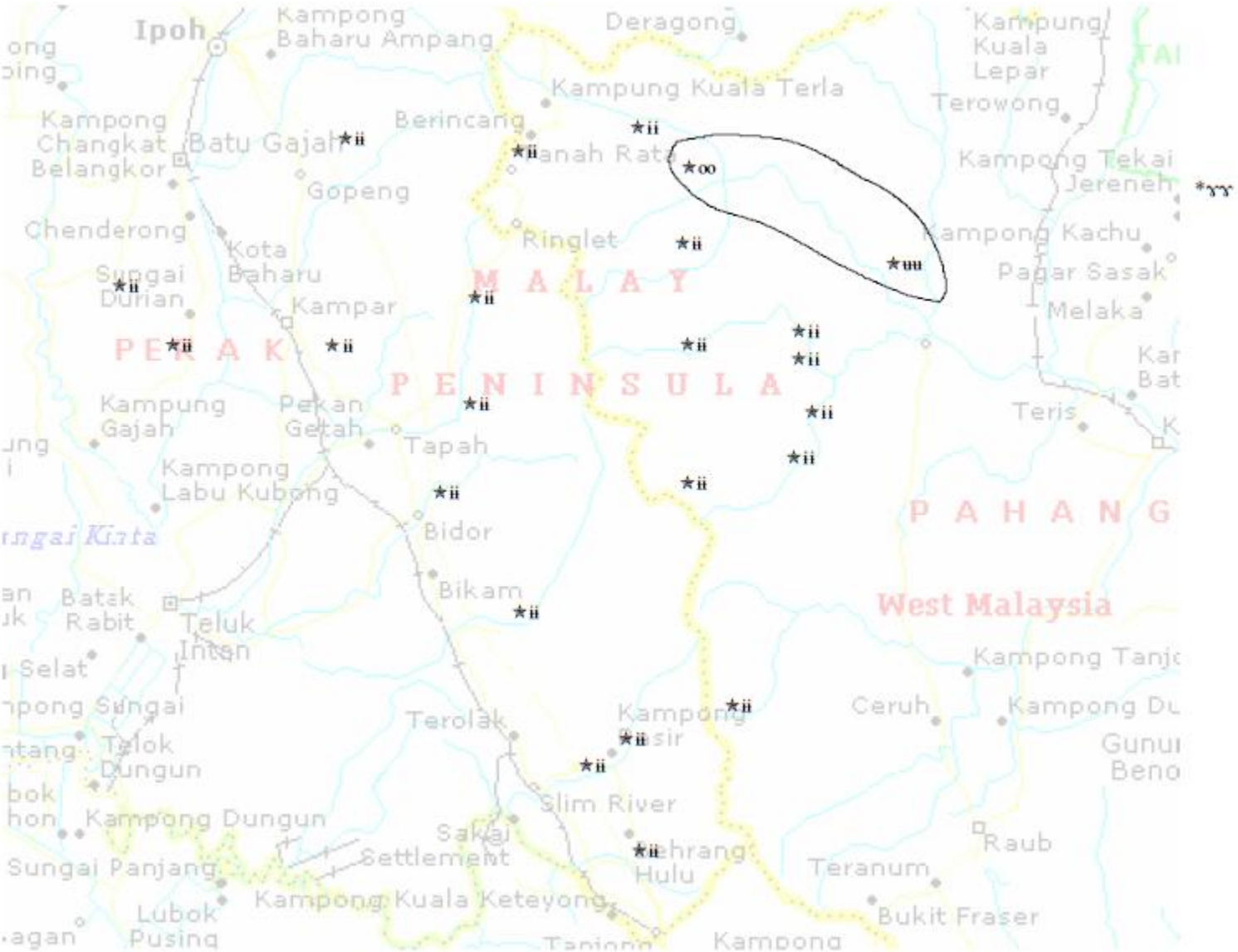

Nasal endings
Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Berincang
Kampong Changkat Belangkor
Batu Gajah
Tanah Rata
Kampong Tekai Jereneh
Gopeng
Chenderong
Ringlet
Kampong Kachu
Kota Baharu
Pagar Sasak
Sungai Durian
Kampar
MALAY
Melaka
PERAK
PENINSULA
Kampung Gajah
Pekan Getah
Tapah
Teris
Kampong Labu Kubong
PAHANG
Bidor
Bikam
West Malaysia
Batek Rabit
Teluk Intan
Selat
Kampong Tanjo
Terolak
Kampong
Ceruh
Kampong Du
Telok Dungun
Gunur
Beno
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Hulu
Teranum
Kampong Kuala Keteyong
Bukit Fraser
Lubok Pusing
Kampong
-p
-pm
-bm

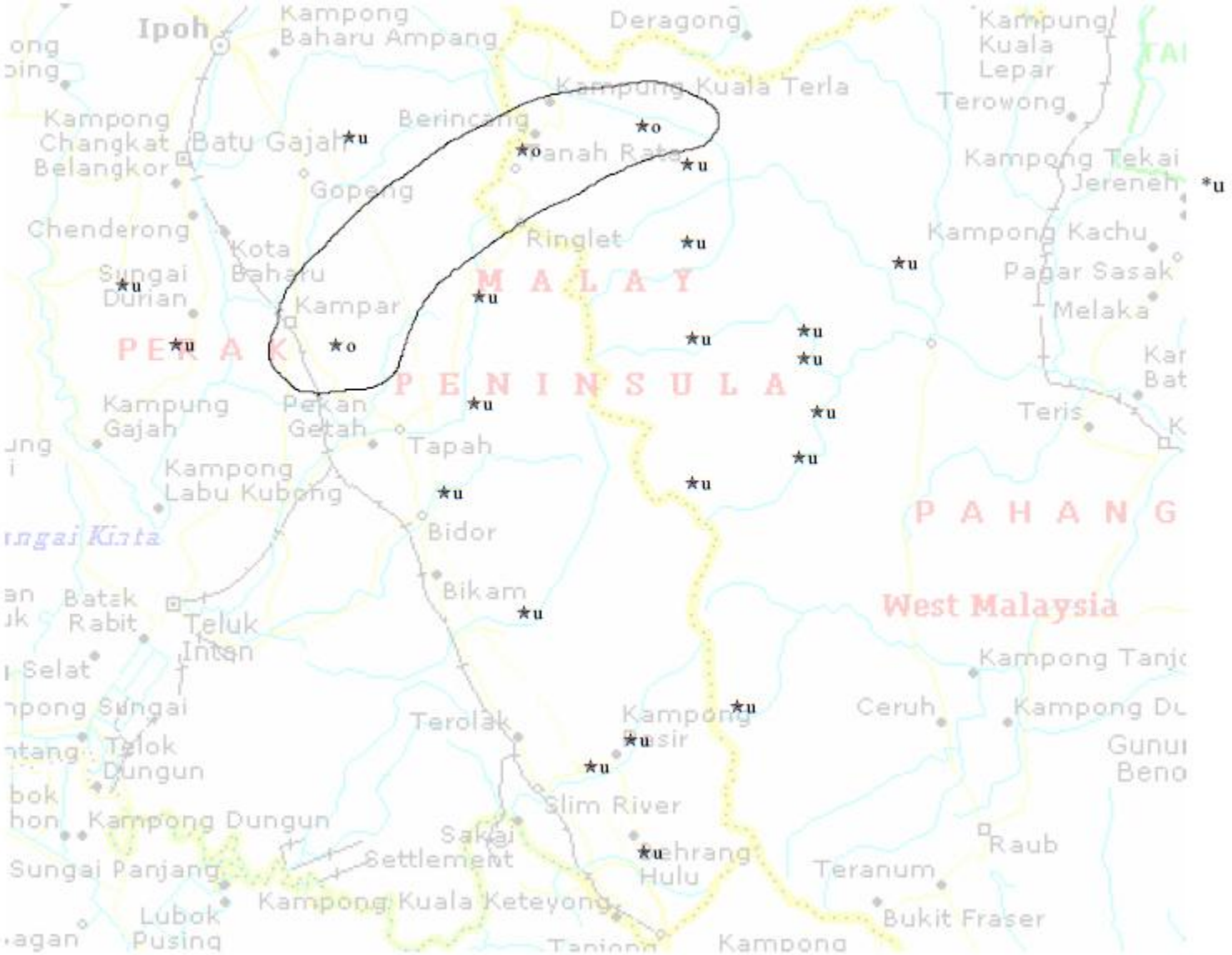

Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Kampong Changkat Belangkor
Batu Gajah
Gopeng
Berincang
Tanah Rata
Kampong Tekai Jereneh
Chenderong
Kota Baharu
Ringlet
Kampong Kachu
Pagar Sasak
Sungai Durian
Kampar
MALAY
PENINSULA
PERAK
Melaka
Kampung Gajah
Pekan Getah
Tapah
Teris
Kampong Labu Kubong
Bidor
PAHANG
Bikam
West Malaysia
Batek Rabit
Teluk Intan
Selat
Kampong Tanjo
Terolak
Kampong Besir
Ceruh
Kampong Du
Telok Dungun
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Hulu
Teranum
Kampong Kuala Keteyong
Bukit Fraser
Lubok Pusing
Kampong
none
Glottal stop endings

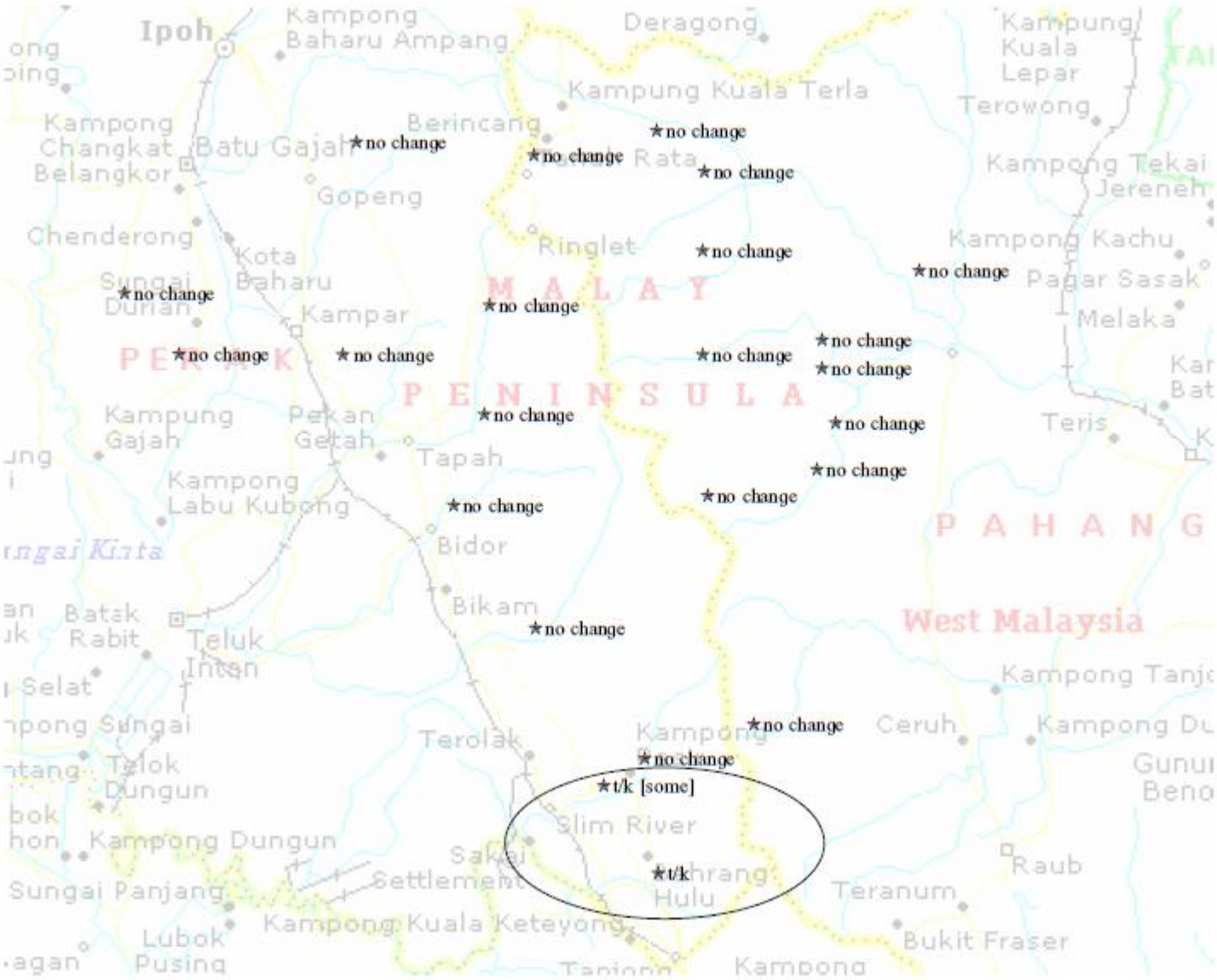

Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Kampong Changkat Batu Gajah
Berincang
Belangkor
Gopeng
Kampong Tekai
Jereneh
Chenderong
Ringlet
Kampong Kachu
Kota Baharu
Sungai Durian
Kampar
Pagar Sasak
Melaka
PERAK
MALAY PENINSULA
Kampung Gajah
Pekan Getah
Tapah
Teris
Kampong Labu Kubong
Bidor
PAHANG
Batak Rabit
Teluk Intan
Bikam
West Malaysia
Selat
Kampong Tanjo
Terolak
Ceruh
Kampong Du
Telok Dungun
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Hulu
Teranum
Kampong Kuala Keteyong
Lubok Pusing
Bukit Fraser
Kampong

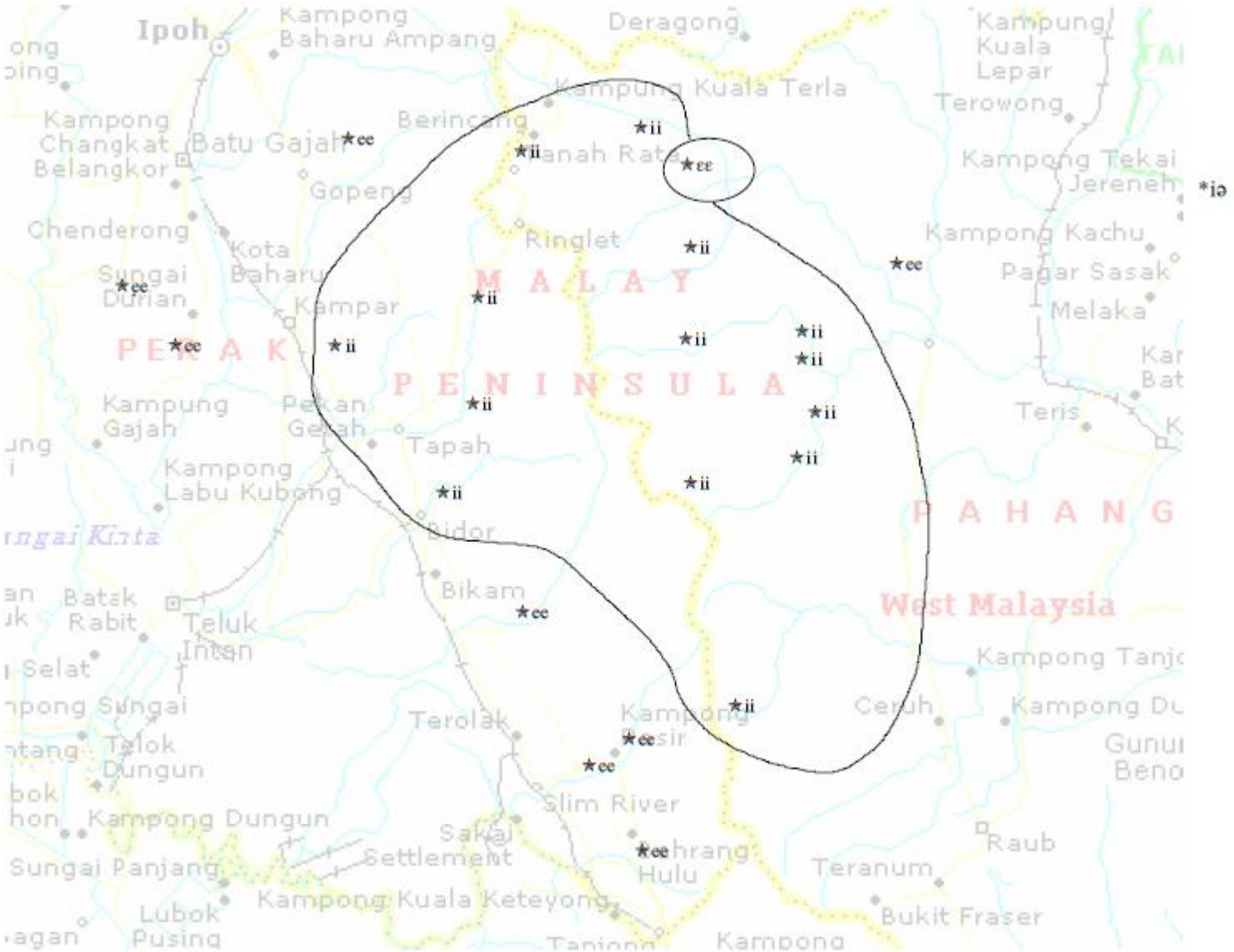

Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Kampong Changkat Belangkor
Batu Gajah
Gopeng
Kampong Tekai Jereneh
Chenderong
Kota Baharu
Ringlet
Kampong Kachu
Pagar Sasak
Sungai Durian
Kampar
M A L A Y
Melaka
P E N I N S U L A
Kampung Gajah
Tapah
Teris
Kampong Labu Kubong
Bidor
P A H A N G
Bikam
West Malaysia
Batak Rabit
Teluk Intan
Kampong Tanj
Selat
Ceruh
Kampong Du
Terolak
Kampong
Gunu
Beno
Telok Dungun
Slim River
Kampong Dungun
Raub
Sungai Panjang
Teranum
Kampong Kuala Keteyong
Bukit Fraser
Lubok Pusing

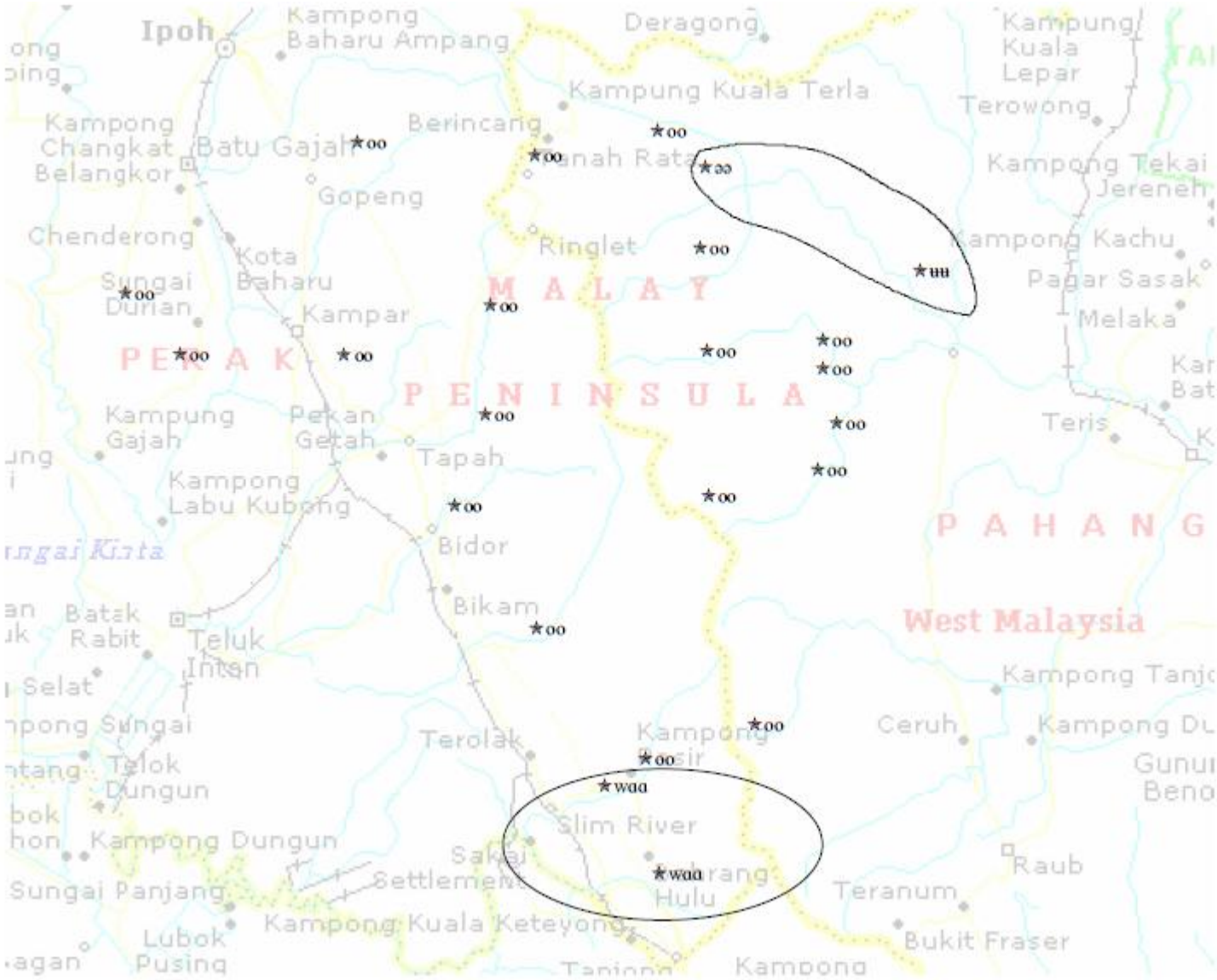

Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Terla
Kampung Kuala Lepar
Terowong
Kampong Changkat Batu Gajah
Belangkor
Gopeng
Kampong Tekai Jereneh
Chenderong
Kota Baharu
Ringlet
Kampong Kachu
Pagar Sasak
Sungai Durian
Kampar
MALAY
PENINSULA
Melaka
PERAK
Kampung Gajah
Tapah
Teris
Kampong Labu Kubong
PAHANG
Bidor
Bikam
West Malaysia
Batak Rabit
Teluk Intan
Kampong Tanjo
Selat
Kampong Sungai
Terolak
Kampong
Ceruh
Kampong Du
Telok Dungun
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Hulu
Teranum
Kampong Kuala Keteyong
Bukit Fraser
Lubok Pusing
Kampong


Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Kampong Changkat Batu Gajah
Berincang
Kampong Tekai
Belangkor
Gopeng
Jereneh
Chenderong
Ringlet
Kampong Kachu
Kota Baharu
Pagar Sasak
Sungai Durian
M A L A Y
Kampar
Melaka
P E R A K
P E N I N S U L A
Kampung Gajah
Pekan Getah
Tapah
Teris
Kampong Labu Kubong
P A H A N G
Bidor
Bikam
West Malaysia
Batak Rabit
Teluk Intan
Kampong Tanj
Selat
Ceruh
Kampong Du
Terolak
Kampong
Telok Dungun
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Hulu
Teranum
Kampong Kuala Keteyong
Bukit Fraser
Lubok Pusing
Kampong

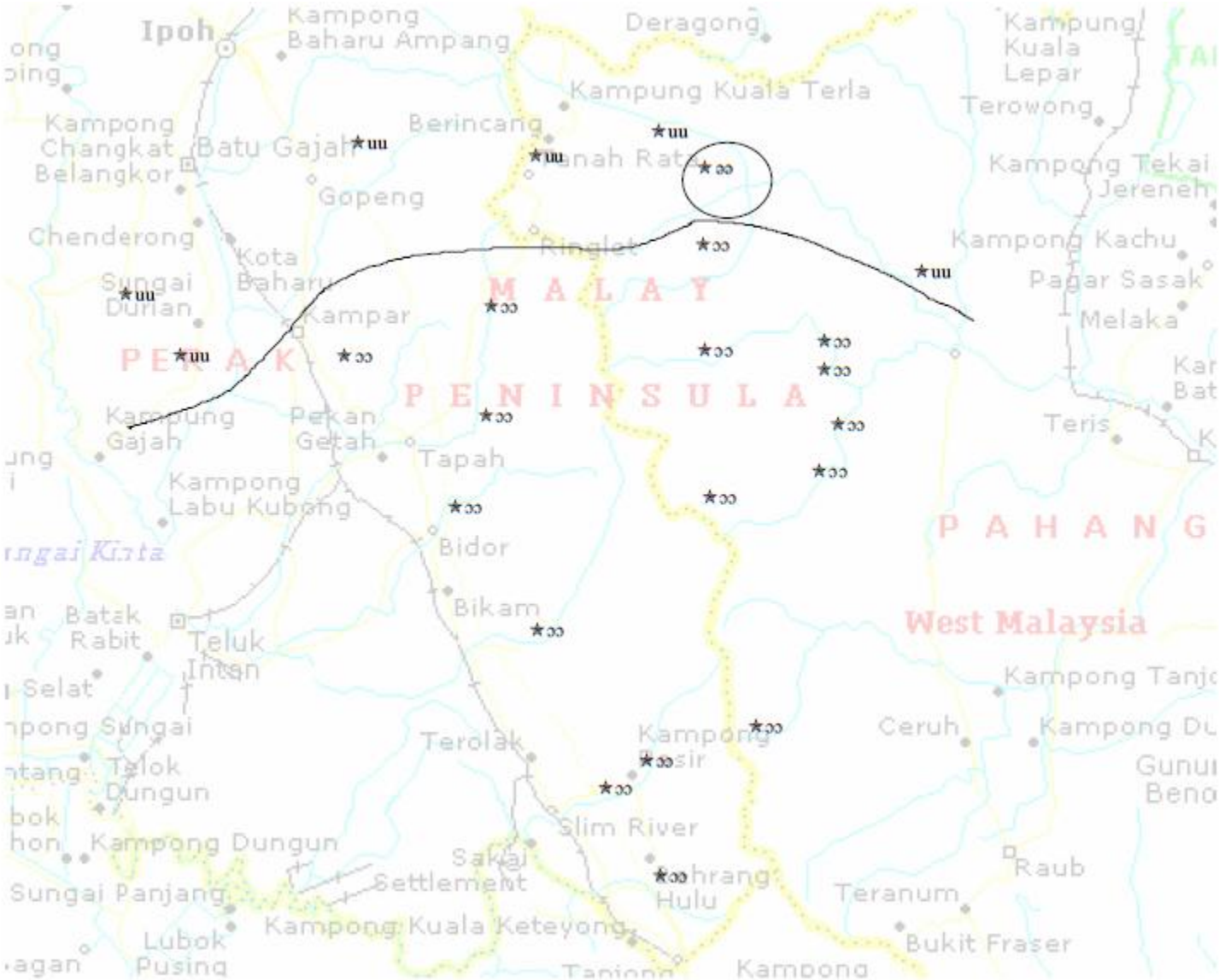

Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Kampong Changkat Belangkor
Batu Gajah
Berincang
Kampong Tekai
Jereneh
Gopeng
Chenderong
Ringlet
Kampong Kachu
Kota Baharu
Pagar Sasak
Sungai Durian
Kampar
M A L A Y
Melaka
P E R A K
P E N I N S U L A
Kampung Gajah
Pekan Getah
Tapah
Teris
Kampong Labu Kubong
Bidor
P A H A N G
Bikam
West Malaysia
Batak Rabit
Teluk Intan
Selat
Kampong Tanjo
Terolak
Kampong
Ceruh
Kampong Du
Telok Dungun
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Lubok Pusing
Kampong Kuala Keteyong
Teranum
Bukit Fraser
Kampong


Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Kampong Changkat Belangkor
Batu Gajah
Berincang
Rata
Kampong Tekai Jereneh
Gopeng
Chenderong
Ringlet
Kampong Kachu
Kota Baharu
Sungai Durian
Kampar
Pagar Sasak
Melaka
PERAK
MALAY PENINSULA
Kampung Gajah
Pekan Getah
Tapah
Teris
Kampong Labu Kubong
Bidor
PAHANG
Bikam
West Malaysia
Batak Rabit
Teluk Intan
Selat
Kampong Tanjo
Terolak
Kampong
Ceruh
Kampong Du
Telok Dungun
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Hulu
Teranum
Kampong Kuala Keteyong
Bukit Fraser
Lubok Pusing
Kampong
★no change
★t/k [some]
★t/k
*-C (final palatals)

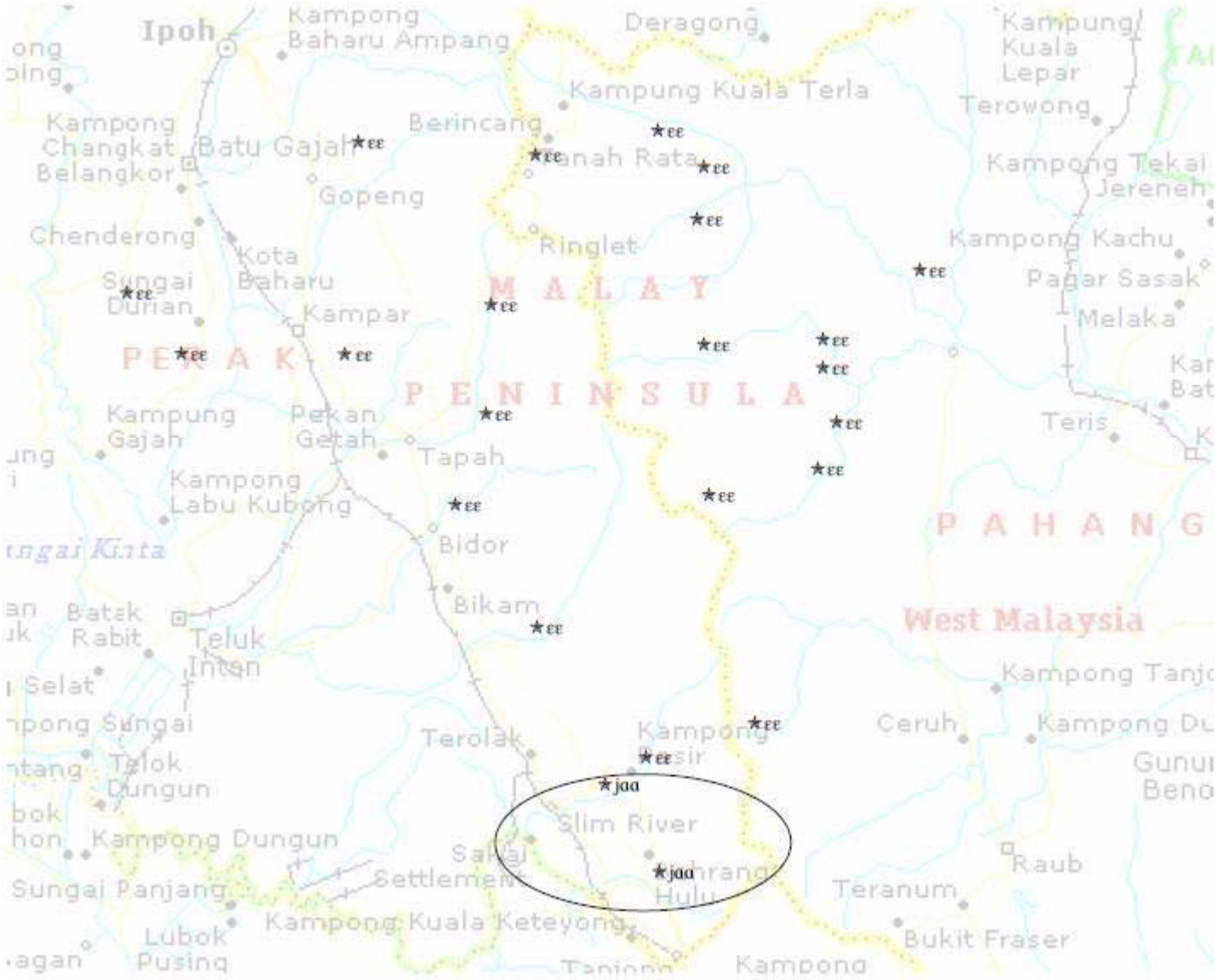

Ipoh
Kampong Baharu Ampang
Deragong
Kampung Kuala Lepar
Kampung Kuala Terla
Terowong
Kampong Changkat Batu Gajah
Berincang
Kampong Tekai
Jereneh
Belangkor
Gopeng
Chenderong
Ringlet
Kampong Kachu
Kota Baharu
Pagar Sasak
Sungai Durian
Kampar
MALAY
Melaka
PERAK
PENINSULA
Kampung Gajah
Pekan Gelah
Tapah
Teris
Kampong Labu Kubong
PAHANG
Bidor
Batak Rabit
Bikam
West Malaysia
Teluk Intan
Selat
Kampong Tanjo
Terolak
Kampong
Cenuh
Kampong Du
Telok Dungun
Slim River
Kampong Dungun
Sakai Settlement
Raub
Sungai Panjang
Hulu
Teranum
Kampong Kuala Keteyong
Bukit Fraser
Lubok Pusing
Kampong
ii
ee
iə



*oo (before alveolars and palatals)



*Noo (oo after nasal)



*ɛɛ (before alveolars and labials)

Appendix E: Proto-Semai lexical items

Table 14 lists some proto-Semai words reconstructed on the basis of material included in the Semai wordlists. A lexical item has been reconstructed only when cognate items were found in two or more areas. Naturally, no reconstruction was attempted when a lexical item was found in only one dialect. Note that the semantic range of a cognate in one language or dialect often does not completely match that of the cognate in another. Therefore, the gloss postulated for a reconstructed lexical item may differ from that of one or more of the cognates and may also differ from that of the wordlist item. When it seems clear that the gloss of the reconstructed lexical item is different from that of the wordlist item, the gloss postulated for the reconstructed item is enclosed in parentheses immediately following the item. Borrowings from other languages are included but not marked separately.

The symbols for proto-Semai proposed by Diffloth are used so as not to introduce conflicting symbols into the literature. Where Diffloth has already postulated a proto-form, it is listed here followed by a dagger (†).

Table 14. Reconstructed proto-Semai words

1	sun	*mɑ̃t ɟiis, *pəŋrɑᵍŋ
2	moon	*gəcɛɛʔ†
3	star	*pərlooy, *bintɑᵍŋ
4	sky	*bɑliiᵍŋ†, *rɑhuuʔ, *səwiik
5	water, river	*teew†
7	flow	*gəl
8	swim (human or animal)	*lɔɔy
9	float (on surface)	*timũl
10	sink	*kɑrɑp, *cəlɔk, *sənrɑp, *təŋlɑᵇm
11	bathe	*mɑ̃hmũh
12	shallow	*n̥cees†
13	deep (opposite shallow)	*ɟəreeʔ
14	rain	*mɑnĩĩʔ†, *pɤ̆ɤ̆k
15	lightning	*bɯɯc
16	thunder	*ɟəʔɑɑr† (thunderstorm), *ŋ̊kuuʔ
17	rainbow	*cədɑɑw
18	shadow	*wɔɔk
19	wind	*pooy†, *pəs
22	stone	*gəsuk, *bɑtuʔ
23	mountain	*looᵈn, *ɟəlmɔ̃l
24	cave	*gəpgɯɯp
25	earth	*tiəʔ†
26	mud	*pɑyɑɑʔ
27	sand	*səmbɛɛy
28	dust	*hɑbuʔ
32	salt	*n̥pɔɔc
33	path (small)	*nõõŋ

34	fire	*ʔɔɔs†
35	firewood	*cəʔɔɔs, *raŋãs
36	ashes	*habuuʔ
37	smoke	*cas†, *cuul
38	burn (wood)	*tɔɔt, *ɟɔp, *ɲõh
39	burn off (eg. feathers)	*raᵍŋ, *rəᵇm
40	forest	*ɟərəs, *sərak, *darat
41	grass	*səʔɛp, *ɲõõʔ, *lalaak
42	seed	*mãt, *kəbɯɯʔ, *bənĩh
43	leaf	*səlaaʔ
44	root	*rəʔiəs†, *ɟaŋkar, *cooᵍŋ
45	dig	*cooᵇm†
46	look for	*kɛɛʔ
47	flower	*buŋããʔ
48	thorn	*ɟərlaaʔ†
49a	trunk, tree	*keeᵍŋ†, *kənẽẽŋ†, *laloow, *cətəᵍŋ
49b	branch, tree	*kənẽẽŋ ɟəhuʔ, *caaᵍŋ
50	bark	*cəkɔɔp, *cawɔɔk, *hɔɔʔ
51	fruit	*pəlɛɛʔ†
52	papaya	*pəlɛɛʔ pətik
53	banana	*təlɛɛy†
54	coconut (unripe)	*ɟəʔɤɤr / *ɟəʔɯɯr ɲ̣ɲãm, *niyɔɔʔ ɲ̣ɲãm
55	coconut (ripe)	*ɟəʔɤɤr / *ɟəʔɯɯr ŋ̣raaʔ
56	eggplant	*təroᵍŋ
57	ginger	*kərɟaaʔ
58	mushroom	*bətees†
60	kapok	*kabuuʔ
61	bamboo	*ʔawaat, *pəʔooʔ, *leew
62	bamboo shoot	*rəbooᵍŋ
63	rattan	*cooᵍŋ
64	betel nut	*bəlɤɤᵍŋ
65	leaf (betel)	*sirih, *kaluk, *ɟərak
66	lime	*kɔp
67	spit (betel juice)	*[gə]tɔh†
68	rubber	*cəbɤɤr
69	tap (a tree)	*giiʔ†, *caᵍŋ, *kooh, *motõŋ, *waar
70	ipoh tree	*keeᵍŋ, kənẽẽŋ† (tree), *dɔɔk (poison), *tənãk
71	blow gun	*bəlaaw
72	blow	*təʔuh†, *puut
73	darthead (blunt end)	*bərɔl†, *paŋkal
74a	dart (with poison)	*rooᵍŋ, *siseey, *saseeʔ

74b	dart (without poison)	*səmõõc
75	quiver	*lək
76a	spear (iron)	*tarɔk
76b	spear (bamboo)	*mataa?
77	hunt	*lɛɛp, *?umããl
78	kill	*pərda^{d}n
79	stab	*cɔk
80	shoot	*bədil
81	bear (honey)	*bərwɔ^{g}ŋ, *samĩĩl, *bahwoow
82	pig (wild)	*gaaw, *lɤɤ?, *?uyɔ̃ŋ, *məŋhar
83	porcupine, large	*kuus†, *bah gorɛk
83a	porcupine, small	*pacɔɔr, *lokaa? dak
84	monkey (short-tailed)	*dɔ^{g}ŋ†, *raaw, *hɤɤl, *cabɛh
85	rabbit	*?arnãp
86	deer	*rusa?, *bah taba?
87	tiger	*raak, *mərgas, *ɲõõ?, *?atãɲ
88	tail	*səntaa?
89	dog	*coo?†
90	bark (to)	*ɟəl†
91	cat	*kuci^{g}ŋ
92	rat	*kənĩə?†, *pərook†
93	bite	*kap†
94	bird	*cɛɛ^{b}m†
95	fly (v)	*hɛɛ^{g}ŋ†
96	chicken	*puk
97	feather, body hair	*səntɔɔl
98	egg	*pənlɤɤ^{g}ŋ†, *kət†
99	wing	*kənyɛɛ^{g}ŋ
100	butterfly	*kərəkbaak, *tawaak, *sirooy
101	bee	*səmɤ̃ɤ̃c† (stinging insect), *luwɛɛy, *dənĩŋ
102	fly [n]	*rooy
103	insect	*cɛh, *las
104	mosquito	*kəmẽẽt†, *səbɛ^{j}ɲ†, *hɛp, *galul, *kəbo?
105	louse (chicken)	*cɛɛ?, *bəruuy, *?amãc
106	louse (head)	*cɛɛ?
107	termite	*gərɤɤ^{j}ɲ
108	spider (small, house)	*tawii^{g}ŋ†, *tapə^{g}ŋ, *məŋhãl, *manĩŋ
108a	spider (large)	*b[a]ŋhɯɯl
109	cockroach	*sɔr, *gasur, *papɔh
110	snake	*tiɟi?, *taɟu?
111	lizard (monitor)	*gəriə^{g}ŋ / gəree^{g}ŋ, *p[a]ree?, *tərakɔ̃l, *harɛk

112b	snail (garden)	*too?, *kaloo?
112c	snail (water)	*too?, *kaloo?, *kaa? kɔp, *kəmbuweey
113	frog	*tabɛɛk, *karoo?, *kaa? ɟuk, *?op, *sɛɛt, *bɛr, *kaa? səkŋũk
114a	turtle (sea)	*pə?a[a]s, *pəɲũũ?, *koh
114b	turtle (river)	*pə?a[a]s, *gəlpiil, *kaa? hɔɔ?, *kaa? paal
115	fish	*kaa?†
116	crocodile	*bahaya?
117a	person (OA)	*səŋ?ɔ̃ɔ̃y†, *mããy
117b	person (non-OA)	*mããy kilɯɯ?
118	man	*kəraal
119	woman, female	*kərdoor†
120	child (offspring)	*kənõõn
121	children	*say$\tilde{\varepsilon}\tilde{\varepsilon}$t / *saɲ$\tilde{\varepsilon}\tilde{\varepsilon}$t
122	mother	*?amẽẽ?, *?amɛ̃h
123	father	*bənẽẽ? (one's own), *?abee?, ?abʌʌh, *?aya h, *?apa?
124	sibling, elder brother or sister	*tənɛ̃?
125	sibling, younger	*bənãŋ , *ba^{g}ŋ (not one's own)
126	sibling, youngest	*?iluc, *ləc
127	husband	*gənsiir†
128	wife	*kənãh
129	widow	*balu?, *ɟandɨɨ?
130	friend	*rənãmĩ?, *kawa^{d}n
130x	companion	*roop
131	name	*mũh, *mɨ̃h
132	head	*kuuy†
133	hair	*sɔɔk†
134	comb	*sikat
135a	bald (natural)	*laac†, *pərlõõŋ
135b	bald (shaved)	*lɔɔc
136	scissors	*gunti^{g}ŋ
136a	cut hair (with scissors)	*ko[o]h†, *tas†
137	eye	*mãt
138	nose	*mɔ̃h
139	ear	*n̪ta^{g}ŋ
140	face	*rənẽẽs, *?ulah
141	forehead	*təmãgi[i]?, tahagi?, *labuu^{d}n (fontanelle)
142	eyebrow	*səntɔl mãt, *səmpooy mãt, *sətsĩĩt mãt
143	cheek	*[ka]mĩ$\tilde{\ə}$ŋ †, *kapɔ̃ɔ̃?†
144	chin	*ɟəŋkaa?, *caka?

145	throat	*gərgook, *kəruŋkuk, *taŋɨ̃n, *ruŋkuk
146	mouth	*ṃpaak
147	lip	*ɲəyɲõõy†
148	tooth	*ləmũũɲ†
149	tongue	*ləntaak†
150	gums	*ləɲsiʲɲ†
151	brain	*kəl?oo^{b}m
152a	neck, (back of)	*taŋə̃n†
152b	neck, front of	*cəŋõk
153	back	*cəloo^{d}n†, *kənõ̃k
154	shoulders	*gəlpaal†, *pɔɔk†
155	armpit	*kərandook, *sənõ?
156	elbow	*kanãŋ
157	hand	*tə^{g}ŋ
158	palm (of hand)	*paal, *tapar, *tapak
159	nail (finger)	*cəŋroos
160	finger	*cənrɛɛ^{g}ŋ, *ɟaras
161a	abdomen	*kət†, *?ɛc
161b	abdomen (upper)	*kət†, *ṇtəh (chest)
162	abdomen (lower)	*wɔ^{g}ŋ†
163	buttocks	*kiət†, *saŋkil
164	navel	*sook, *padeek
165	lungs	*suup
166	heart	*noos†, *ɟantu^{g}ŋ
167	intestines	*coo^{g}ŋ ?ɛc, *?ɛc pədɯɯl, *?ɛc wɛɛ^{d}n
168	liver	*riis, *kəlaap
169	bone	*ɟə?aa^{g}ŋ†
170	rib	*cərəs†
171	skin	*gətɤɤ?†, *kapoo^{b}m
173	dry	*soot†, *tiil†
174	flesh	*sɛc†
175	thigh	*ləmpa[a]?
176	knee	*kurool
177	calf , lower leg	*kəmõõŋ†
178	foot	*ɟu^{g}ŋ†
179	heel	*canõõŋ, *kulɛ̃ɛ̃?, *kənũũl
180	toes	*cənrɛɛ^{g}ŋ, *ɟaras
181	strong	*kuwat, *təgar
182	tired	*gəhɛɛl, *səlɛɛy
183	sleep	*bɛɛt†
184	snore	*səmãŋãr

185	yawn	*kihɔɔy
187	see	*nɛ̃ɛ̃ŋ†, *bəlɤɤh, *loow
188	look	*nɛ̃ɛ̃ŋ†, *tiyõõw
189	wink	*kəbɛr†, *kəɲẽp mãt
190	deaf	*pɛʔ gərteeᵍŋ, *tip ṇtaᵍŋ
191	hear	*gərteeᵍŋ, *cərŋɑɑy, *təŋhɛ̃ɛ̃n
192	smell	*ŋõõy†, *ɲũũy†, *ɟəhũũt, *ʔõŋ, *hõõn, *səʔɤɤt / *səʔɯɯt
193a	rotten, decayed	*sɔɔʔ†, *bɑbəh
193b	rotten (food)	*sɔɔʔ†, *bɑriʔ
194	hungry	*cuwɑɑᵍŋ, *cəmɤ̃ɤ̃s / *cəmɯ̃ɯ̃s
195	satiated	*b[ɑʔ]heeʔ†
196	eat	*cɑɑʔ
197	thirsty	*səlɔɔt
198a	suck out (snail, bone marrow)	*huc†, *hõõt, *səbɔɔᵍŋ, *səpõõt
198b	suck on (candy, sweet)	*səbɔɔᵍŋ†, *bɤɤʔ, *kɑmõm, *lõõt
199	lick	*bɤɤc†
200	drunk	*bɤɤl†
201	liquor	*teew bɤɤl
202	drink	*ŋõõt†
203	swallow	*liiᵇm
204	vomit	*kɤɤʔ
205	hurt	*pəhoot, *ɲãʔɲĩĩt
206	swollen	*ʔɑs
207	itchy	*bəhɛʲɲ, *gəsɛc
208	scratch (v)	*gees, *geeh, *gih
209a	rake	*pɤɤs†, *kɑpããc, *kɑtããc
209b	scratch (chicken)	*pɤɤs, *kəwããc
210	cold, to feel	*dəkɑt† (cold, feverish), *səŋẽẽc
211	shiver	*kərɤɤᵍŋ, *kərɔᵍŋ
212	cough	*kəhɔɔl
213	sneeze	*rɑmɔ̃h, *gəcɑs, *bərsiᵈn
214	hot	*bəkɛɛt, *pəliit
215	sweat	*səŋʔəᵇm, *məŋkɛɛt
216	saliva	*ləhiəᵍŋ†
217	spit	*[gə]tɔh†
219	tears	*teew mãt
220	cry	*ɟɑɑᵇm†
221	urine	*nɔ̃ɔ̃m†
222	blood	*bəhiiᵇm†
223	excrement	*ʔɛc

224	defecate	*cahcoh
225	pus	*bəmbɤɤᵍŋ
226	boil, abscess	*dəkɔɔʔ
227	scar	*təmããʔ, *hɔl, *waak, *bəkat, *dɛl
228	medicine	*pəŋlaay
229	incant	*cagɔh, *təhɔr, *bicaaw
230	shaman, herbal curer	*halaaʔ, *pawãŋ
231	live, to	*rees† (to be alive, erect), *suuy
232	die	*daᵈn†
233	bury	*cooᵇm†
234	old (animate)	*liiw†, *ŋ̣raʔ, *manãh, *tataaʔ, *ɟaɟaaʔ
235	fat	*hiiᵈn, *bəcɔʔ, *baʔcooʔ, *hiit
236	skinny	*soor, *sooᵍŋ
237	tall (animate, inanimate)	*cərəᵍŋ, *lasiik
238	short (animate, inanimate)	*patiəʔ, *kutiəʔ, *latẽʔ, *lẽẽŋ
240	big	*ṇtooy
241	small	*macɔ̃ɔ̃t, *ṇcɛ̃n, *kaɟək, *lalaʔ
242	breathe	*ləhə̃m
244	blow on	*təhool
245	sit	*gɤɤy
246	stand	*ɟəᵍŋ
247	walk	*ciip†
248	crawl	*wɯɯᵈn†
249	run	*ɟar†
250	fast	*[ga]gɛc, *dəras, *culas, *təʔɤɤc / *təʔɯɯc
251	slow	*liŋãh, *liiw, *pərlahaᵈn, *manɛ̃ɛ̃n, *lamãt
252	think	*səŋ̃ĩĩʔ† (awareness), *ɟooᵍŋ
253	know	*səŋ̃ĩĩʔ† (awareness), *panẽẽy
254	forget	*siəᵇm
255	dream	*bəraʔpooʔ, *ṃpooʔ, *bakeeʔ
256	choose	*pileh
257	love	*hooʔ
258	smile	*[kər]ɲĩm, *kərmĩĩʔ, *luk†
259	laugh	*luk†
260	good	*bɔɔr, *ʔilɔʔ
261	bad (person)	*dɤɤs, *nẽc
263	angry	*rɔɔʔ, *mãyɔɔʔ, *lɛs
264	furious	*bəʔal, *cəkɔh
265	fib, lie	*looᵈn, *lənõõn, *dəʔah
266	steal	*siəc†
267	fight	*kɤɤh, *lahiiʔ

268	afraid	*səŋɔ̃h
269	correct	*kəmə̃l, *kənã?
270	wrong	*sɛɛ?
271	difficult	*pɑyɑh, *susɑh
272	hit	*kɤɤh, *tɔɔ?, *pɤɤt / *pɯɯt
273	speak	*yɑɑp, *doot, *cɑɑl
274	call (get s.o.'s attention)	*cəree^{g}ŋ
276	story	*cərmɔr†, *cənãã̃l
277	tell	*?ɔɔk pɑnẽẽy
278	whistle	*hɔɔ^{g}ŋ
279	answer	*bərlaat, *bərlɑɑ^{g}ŋ, *doot, *ɟawap, *balas
280	sing	*ɟula^{g}ŋ, *laguu?, *ɲaɲĩĩ
281	dance	*gəsɑɑ^{g}ŋ, *ruŋgen, *ɟugɛt
282	stamper , striker	*cəntɤɤ^{g}ŋ
283	drum	*tagɤɤh, *tubuk, *gito?, *rəbanã?
284	play	*mãn, *too^{d}n
285	kick	*təɲã̃k, *təndɑ[^{g}]ŋ, *sipɑ?
286	fall down (animate, inanimate)	*ɟɤɤr†, *təgɔɔh, *yɔɔ^{g}ŋ
288	fall down from a height	*yɔɔ^{g}ŋ, *təgɔɔh
289	take	*kɔɔt†, *?ãã̃ɲ, *ciduu^{b}m
289a	carry in the hand, take	*taɟɑɑk, *?ãã̃ɲ
289b	carry in arms (child)	*cəduu^{b}m†, *bɔ?†
290	carry in cloth	*bɔ?†, *gɔɔl
291	carry on head	*taŋkɔ̃l, *ɟunɟu[^{g}]ŋ
292	carry on shoulder	*gulɑ^{b}m, *gɔl
292a	carry on shoulder, two people	*g[ɑ]ndɑr
293	return	*ɟuk, *lɔɔs
294	come	*hool†, *təl
295	enter	*mõõc, *pəlɤɤt
296	wait	*pɔɔʲɲ
297	work	*kərɑ?ɟɑ?
298	earn	*kə^{b}m
299	pay	*bɑyɑr
300	sell	*ɟuwɑl
301	buy	*bəli?, *kɔɔt
302	give	*?ɔɔk
303	throw (s.t. sidearm)	*pɛc, *guwɑl, *rɑwɛɛ^{g}ŋ, *pɑk
304	throw overhand	*sɑdɑk, *paŋkɑɑh
305	toss (throw underhand)	*gɑs, *tɑlɛɛh
306	throw away	*gɑs, *wees

307	pull	*ɟɛɛk
308	push	*dus, *təkuuᵍŋ, *sər
309	day	*ɟiis†, *harii?
310	morning	*pəgəlap, *huplɤɤp, *bəla ɟiis, *biyah, *[hu]pɤɤr / *[hu]pɯɯr
311	noon	*ɟəl?ɛɛᵍŋ, * təŋãh harii?, *məŋkɛɛt
312	afternoon	*duuy, *bərkes, *bərkəs
313	night	*kəlɛ̃m, *səŋə̃p, *səŋõõp, *bəŋõ̃õ̃t, *ditəp, *gəlap
314	tomorrow	*həplɤɤp, *ɲãhlɤɤp, *hupərdɔh, *cəhɔy, *hupɯɯr [dɔh], *yapɯɯr
315	yesterday	*səŋõ̃õ̃n, *duuy nɛ̃ɛ̃?, *n[ũ?]təᵇm
316	year	*tahuuᵈn
317	lean-to	*dɤɤᵍŋ rəp, *dɤɤᵍŋ danəm, *dɤɤᵍŋ sapaw, *pəndɤɤr / *pəndɯɯr
322	house	*dɤɤᵍŋ
323	space under house	*kərɔɔᵇm dɤɤᵍŋ†
324	roof	*pəlooᵍŋ†
325	window	*tiŋkap
326	floor	*ciŋkaar, *riəs
327	blanket	*tiluuᵇm, *?abaᵈn, *gəbaar
328	mat	*cəruu?
329	loincloth	*laᵈn
330	sarong, men's	*?abaᵈn, *?abaᵈn kəbɤɤ? / kəbɯɯ?
331	sarong, women's	*?abaᵈn bati?
332	tie	*bək, *kool, *sək
333	tie by twisting	*kool, *beer, *loor
334	trousers	*sərwal, *səluwar
335	dirty	*bəcɔɔt, *nɛ̃ɛ̃c, *bərit
336	wash	*suuc
337	wash (clothes)	*sah
338	rub	*gasak, *ɲĩc
339	wet	*kə?ããc, *tə?ãc, *pɤɤk
340	dry, to be	*soot†
341	dry, to	*tiil†
342	wipe	*giit, *sapuu?, *lap
343	sweep	*pɤɤs†, *sapuu?
344	sew	*cuᵍŋ, *ciəᵍŋ, *ɟayet
345	needle	*ɟaruᵇm
346	cook	*bərcɛɛᵈn†
347a	boil (water)	*gaaᵇm, *nãm
347b	boil (tapioca)	*gaaᵇm, *rəbus

348	pot	*cəreʔ
349	cooking pot	*leeʔ, *pəryuk, *təŋlɔ̃ŋ
350	full	*təbeeᵍŋ̍, *kəᵇm
351a	ladle	*səndoʔ, *mənũũh, *suduuʔ
351b	dipper (for bathing)	*c[ən]iboʔ, *səndok, *tɑkɔɔʔ
352a	mortar (stone)	*guul̥, *siŋkɑlɑᵈn, *ləsuŋ
352b	mortar (for rice)	*guul̥
353a	pestle (spices)	*kənɤ̃ɤ̃h, *rənẽẽʔ, *kənũũr
353b	pestle (for pounding rice)	*kənɤ̃ɤ̃h, *rənẽẽʔ
354a	pound (in mortar, spices)	*sɛh, *gərlit
354b	pound in mortar (rice)	*sɛh
355	knife	*yooc, *yɤɤc, *pisɑw
357	sharp	*cəbɑt
358	blunt	*bələk
359	split	*pɔk, *kɑh, *bəlɑh
360	sweet	*cəʔɛɛt
361	sour	*kəʔɯɯr̥, *kəʔɤɤᵇm / *kəʔɯɯᵇm
362	bitter	*kədɛc
363	black	*bəlʔɑk, *rəŋãh, *hitɑᵇm
364	white	*bəyɛɛk̥, *bəyɤɤk, *bəkɔ̃ɔ̃ʔ, *bəkəl, *puteh
365	red	*cəhɛɛr̥ (bright red), *cəŋõõl, *rəŋããɲ, *ɟəŋhããŋ, *pətlããc
366	green	*bəlɑʔɯɯr, *hiɟɑw
367	yellow	*rə[t]mɛ̃ɛ̃t̥ (yellow, turmeric), *kəlooy, *cəmcɑk, *pəhook
368	bright	*loow
369	dark	*gəlap, *kəlɑp, *səŋũũr, *səŋə̃p, *səŋũũp
370	new	*pɑɑy
371	old (thing)	*liiw̥, *mɑnãh, *ŋ̍rɑʔ
372	round	*kəbɤɤʔ / *kəbɯɯʔ, *bulɑt, *bəlnɔ̃l
373	straight	*təgɑʔ, *ɟəlɔɔy, *ɟəlpɔ̃ŋ, *lurus, *ɟəlʔɛɛk
374	narrow	*ŋ̍kɑp, *ʔĩĩt, *ʔĩĩp, *səmpit, *ŋ̍ẽt
375	thick	*təbɑl, *n̥seeʔ
376	thin	*n̥sɛɛy, *nĩpis
377	smooth	*səlɛc
378	wide	*ləgɑʔ, *luwɑs, *lebɑr
379	long	*cərəᵍŋ
380	hard	*cəgɛh, *təgɑr
381	heavy	*n̥n̥ɤ̃ɤ̃h
382	same	*sɑmããʔ
385	other	*kilɤɤʔ / *kilɯɯʔ, *lɑyiᵈn, *pɑsik
386	what?	*mããh̥, *ɟɑlɤɤʔ, *booʔ, *məlɤ̃ɤ̃ʔ

387	who?	*booʔ, *booʔ ʔimããy
388	when?	*lumpuuʔ, *bilɨɨ, *mãpuʔ
389	how many?	*m[ã]riəᵇm†, *mãgit
391	one	*nãnɤ̃ɤ̃ʔ, *sɑɑʔ
392	two	*nããr, *duwɑɑʔ
393	many (things), much	*ɟəʔɔɔy, *kəᵇm, *səmɤ̃ɤ̃y / *səmɯ̃ɯ̃y
394	many (human)	*ɟə[n]ʔɔɔy, *kəᵇm
395	all	*dic†, *səmũwɑɑʔ
396	some	*gɛʔ†
397	far	*n̥n̥ããʔ
398	near	*n̥n̥ɔ̃ɔ̃n, *rɤɤʔ / *rɯɯʔ
399	right (side)	*[kən]tɔᵇm
400	left (side)	*[kən]wiəl
401	three	*nĩʔ, *tigɑɑʔ
402	four	*m̥pãt
403	five	*limããʔ
404	six	*n̥nãm
405	seven	*tuɟoh
406	eight	*lɑpɑn
407	nine	*səmbilɑn
408	ten	*səpuloh
509	exit, go/come out	*hool†
510	stare	*lɯɯᵈn†
511	grope	*pɯɯᵇm†
512	follow, pursue	*bəsɯɯh†, *dɑlɑk, *ɟooy
513	wring	*riət†, *putɛ̃ɛ̃t
514	hide oneself	*bərceeᵇm†
515	burp	*gərʔʌʌp†
516	crow, sing	*tɑdʌʌr†, *cɑbɛh
517	not oily; ‘squeaky clean’	*cəkʉt†
518	there is	*mãŋ†
519	bachelor†	*litɔw†, *buɟɑᵍŋ, *ɟuleey
520	get up	*kus†, *kuuʲɲ
521	smell of fresh fish	*pəlʔĩs†, *pəlʔĩh†
522	ripe	*nɤ̃ɤ̃ᵇm†
523	sting, to	*sɤɤc†
524	orphan	*rəknẽẽk†
525	make, to	*bəh†, *ʔuuy
526	last, after, in the back	*kətnĩət†
527	twilight	*yʉp†
528	unripe	*kəlbəᵈn†

Appendix F: Phonology of Semai, Betau Dialect

F.1 Introduction

The purpose of this paper is to briefly describe the phonology of the Betau dialect of Semai. Semai is categorized with the following genetic affiliation:

Austroasiatic
Mon-Khmer
Aslian
Central Aslian
Semai

Semai is spoken by approximately 34,000 people,[41] who live mostly in the remote areas of the Malaysian peninsula, in the states of Perak and Pahang. Linguistic and anthropological research indicates that the Semai and other Aslian groups on the peninsula lived there long before the arrival of the current Malay-speaking population.[42]

F.2 Word

Semai words, in good Mon-Khmer tradition, fit the following syllable template:

$(C_3 \quad V_2 \quad (C_4))$	$C_1 \quad V_1 \quad C_2$
Minor	Major

The final syllable is regarded as the major syllable and the penultimate syllable, if present, is regarded as the minor syllable. Semai words always have ultimate stress; that is, on the major syllable. While many Semai words have only one syllable, the majority of Semai words have two syllables. The minor vowel V_2 is usually very short, nonphonemic, epenthetic [ə], and its enunciation in any given word is often optional if the two consonants are easily pronounced without the epenthetic vowel. For this reason Semai roots are sometimes called "sesquisyllabic" since the minor syllable does not carry the same weight, phonetically or phonemically, as the major syllable. The following forms are illustrative of Semai word shapes.

/liim/[43]	'to swallow'	/ŋ.kuʔ/	'thunder'
/mãt/	'eye'	/sə.lɛc/	'smooth'
/mã.nĩĩʔ/	'rain'	/gə.riiŋ/	'monitor lizard'

There are a number of words that have minor syllables with minor vowel segments (V_2) other than [ə]; namely [ɑ], [i], and [u]. It is claimed by some (Diffloth 1968) that these segments may be phonologically conditioned in some cases and morphemes in other cases.[44]

/ɟɑ.liʔ/	'what'	/tɑ.pəŋ/	'small spider'

[41]According to "Population of Orang Asli Sub-Groups, 1999" Keene State University Orang Asli Archive website.
[42]Carey (1976).
[43]In this paper the short vowels will be represented by a single letter, and the long vowels by a double letter. The latter is a departure from standard IPA. Another departure is that the palatal central approximant is represented by the symbol 'y' rather than the standard 'j', which could easily be confused with the palatal voiced plosive 'ɟ' and with local orthographies, especially Bahasa Malaysia. Further, 'r' is used to symbolize the flap. Lastly, the symbol 'ɑ' is used for the unrounded open central vowel, and 'ɒ' for the rounded open back vowel.
[44]Diffloth (1976a) claims, for example, that /-a-/ in certain minor syllables is a morpheme.

/ki.lɨʔ/	‘other’	/li.tow/	‘bachelor’
/ku.tiʔ/	‘short’	/ku.rool/	‘knee’

It is noteworthy that when the minor vowel segment (V_2) is [ɑ], [i], or [u], it is pronounced with greater length than when V_2 is [ə], roughly equal in length with V_1 in the major syllable when V_1 is not a long vowel. Stress remains on the ultimate syllable.

While most roots are apparently either mono-, sesqui-, or disyllabic, there are examples of words with three syllables.

/nə̃.rɑ.mĩʔ/	‘friend’ (possibly a nominalization of Malay *ramah* ‘friendly’)
/bə.rɑ.poʔ/	‘to dream’
/kə.rək.bɑɑk/	‘butterfly’ (possibly from proto-Austronesian **kali-baŋbaŋ*)
/bə.lɑ.ʔɨɨr/	‘green’ (an expressive[45])
/pə.nə̃.mə̃ʔ/	‘full moon’

F.3 Syllable

In Semai every syllable has an obligatory nucleus and onset, and an optional coda. The nucleus is usually a vowel; however, there are some nasals that are syllabic as well in the minor syllable.

[m̩.pɔɔc]	/m.pɔɔc/	‘salt’	[ŋ.kuʔ]	/ŋ.kuʔ/	‘thunder’
[n̩.ɲɨ̃h]	/ɲ.ɲɨ̃h/	‘heavy’	[n̩.tooy]	/n.tooy/	‘big’

The onset and coda are consonants. The two basic syllable types are CV and CVC. The syllable type CV is found only in the minor syllable. For example,

CV	/pə.nə̃.mə̃ʔ/	‘full moon’
	/pɑ.cɔɔr/	‘porcupine (small variety)’
	/kɑ.rəŋ/	‘to shiver’

All major syllables, and some minor syllables, have the syllable type CVC. For example,

CVC	/wɑɑt/	‘to split’
	/tɔh/	‘to spit’
	/ɟə.ʔɑɑŋ/	‘bone’
	/kəl.ʔoom/	‘brain’
	/kər.door/	‘woman’

F.4 Phonemes

Betau Semai in this treatment has forty-five phonemes: nineteen are consonants, fourteen are oral vowels, and twelve are nasal vowels. The consonants are /p, t, c, k, ʔ, b, d, ɟ, g, m, n, ɲ, ŋ, s, h, l, r, y, w/, the oral vowels are /i, ii, ee, ɛ, ɛɛ, ɨɨ, ə, ɑ, ɑɑ, u, uu, oo, ɔ, ɔɔ/, and the nasal vowels are /ĩ, ĩĩ, ẽ, ɛ̃ɛ̃, ɨ̃ɨ̃, ə̃, ɑ̃, ɑ̃ɑ̃, ũ, ũũ, ɔ̃, ɔ̃ɔ̃/. Tables 15–17 show the Betau Semai phonemes.

---

[45]Diffloth (1976b) discusses a word class for Mon-Khmer languages called *expressives*. He claims that expressives in Semai have a phonology that is different from other word classes, exhibiting sequences of sounds not found in the rest of Semai.

Table 15. Consonants

	Bilabial	Alveolar	Palatal	Velar	Glottal
Plosive, voiceless	p	t	c	k	ʔ
Plosive, voiced	b	d	ɟ	g	
Nasal	m	n	ɲ	ŋ	
Fricative, voiceless			s		h
Trill or flap		r			
Lateral approximant		l			
Central approximant	w		y		

Table 16. Oral vowels

Oral, long	**Front (unrounded)**	**Central (unrounded)**	**Back (rounded)**
Close	ii	ɨɨ	uu
Close-mid	ee		oo
Open-mid	ɛɛ		ɔɔ
Open		ɑɑ	
Oral, short	**Front (unrounded)**	**Central (unrounded)**	**Back (rounded)**
Close	i		u
Mid	ɛ	ə	ɔ
Open		ɑ	

Table **17. Nasal vowels**

Nasal, long	**Front (unrounded)**	**Central (unrounded)**	**Back (rounded)**
Close	ĩĩ	ɨ̃ɨ̃	ũũ
Mid	ɛ̃ɛ̃		ɔ̃ɔ̃
Open		ɑ̃ɑ̃	
Nasal, short	**Front (unrounded)**	**Central (unrounded)**	**Back (rounded)**
Close	ĩ		ũ
Mid	ɛ̃	ə̃	ɔ̃
Open		ɑ̃	

The following is a description of the articulatory features of Betau Semai phonemes and their allophones, as well as a description of their distributions.

F.4.1 Consonants

Voiceless plosives

The voiceless plosives /p, t, c, k/ all have two allophones, released and unreleased. The unreleased allophone occurs word final, and the released allophone occurs elsewhere.

/p/ [p,pˀ] is a voiceless unaspirated bilabial plosive and occurs in syllable onsets and codas.

[pɑ.yɑʔ]	/pɑ.yɑʔ/	‘mud’
[pən.lɑɑy]	/pən.lɑɑy/	‘medicine’
[m̩.pɔɔcˀ]	/m.pɔɔc/	‘salt’
[c͡çəp. c͡çiipˀ]	/cəp.ciip/	‘walking’
[c͡çər.kɨɨpˀ]	/cər.kɨɨp/	‘to stab’

/t/ [t, tˀ] is a voiceless unaspirated alveolar plosive and occurs in syllable onsets and codas.

[təᵍŋ]	/təŋ/	‘hand’
[tə.leey]	/tə.leey/	‘banana’
[sən.tɔɔl]	/sən.tɔɔl/	‘feather’
[sootˀ]	/soot/	‘to be dry’
[giitˀ]	/giit/	‘to wipe’

/c/ [c̆, cˀ] is a voiceless unaspirated palatal plosive with two allophones. It occurs in syllable onsets slightly affricated as [c͡ç], and in syllable codas as the unreleased palatal plosive [cˀ].

[c͡ço?]	/coo?/	‘dog’
[c͡çɛ.?ɛɛtˀ]	/cə.?ɛɛt/	‘sweet’
[mã. c͡çɔ̃ɔ̃tˀ]	/mã.cɔ̃ɔ̃t/	‘small’
[tã.?ãɑ̃cˀ]	/tə.?ãɑ̃c/	‘wet’
[secˀ]	/sɛc/	‘flesh’

/k/ [k, kˀ] is a voiceless unaspirated velar plosive and occurs in syllable onsets and codas.

[kə.nɑ̃h]	/kə.nɑ̃h/	‘wife’
[kɑ?]	/kɑ?/	‘fish’
[bə.kɛɛtˀ]	/bə.kɛɛt/	‘hot’
[ŋ.ku?]	/ŋ.ku?/	‘thunder’
[lən.tɑɑkˀ]	/lən.tɑɑk/	‘tongue’
[bə.ləkˀ]	/bə.lək/	‘blunt’

/ʔ/ [ʔ] is a voiceless glottal plosive and occurs in syllable onsets and codas.

[ʔɑ.beʔ]	/ʔɑ.beeʔ/	‘father’
[ʔɔɔkˀ]	/ʔɔɔk/	‘to give’
[kəl.ʔooᵇm]	/kəl.ʔoom/	‘brain’
[səŋ.ʔɔ̃ɔ̃y]	/səŋ.ʔɔ̃ɔ̃y/	‘person’

[po.ʔoʔ]	/pə.ʔooʔ/	‘bamboo’
[sən.tɑʔ]	/sən.tɑʔ/	‘tail’

Voiced plosives

The voiced plosives /b, d, ɟ, g/ occur only in syllable onsets.

/b/ [b] is a voiced bilabial plosive and occurs in syllable onsets.

[bi.hii^{b}m]	/bə.hiim/	‘blood’
[bɛɛt˺]	/bɛɛt/	‘to sleep’
[bəm.bɨɨ^{g}ŋ]	/bəm.bɨɨŋ/	‘pus’
[c͡çə.bɑt˺]	/cə.bɑt/	‘sharp’

/d/ [d] is a voiced alveolar plosive and occurs in syllable onsets.

[dɨɨ^{g}ŋ]	/dɨɨŋ/	‘house’
[də.kɔɔʔ]	/də.kɔɔʔ/	‘boil, abscess’
[c͡çə.dɑɑw]	/cə.dɑɑw/	‘rainbow’
[kər.door]	/kər.door/	‘woman, female’

/ɟ/ [ɟ͡ʝ] is a voiced palatal plosive and occurs in syllable onsets. Akin to its voiceless counter part, its articulation is slightly affricated.

[ɟ͡ʝɑɑ^{b}m]	/ɟɑɑm/	‘to cry’
[ɟ͡ʝər.lɑʔ]	/ɟər.lɑʔ/	‘thorn’
[tɑ. ɟ͡ʝuʔ]	/tɑ.ɟuʔ/	‘snake’
[kər. ɟ͡ʝɑʔ]	/kər.ɟɑʔ/	‘ginger’

/g/ [g] is a voiced velar plosive and occurs in syllable onsets.

[guʔ.gɛʔ]	/guʔ.gɛʔ/	‘some’
[gən.siir]	/gən.siir/	‘husband’
[mã.git˺]	/mã.git/	‘how many?’
[tə.gɔh]	/tə.gɔh/	‘to fall down’

Nasals

The nasal consonants /m, n, ɲ, ŋ/ all have two allophones, ”plain” and preploded. The preploded allophone occurs word final but only after nonnasal vowels, and the plain allophone occurs elsewhere. The plosive in these preploded nasals is always at the same point of articulation as the following nasal. There are also cases where the nasal segment bears syllabicity in words where the nasal segment stands alone in the minor syllable.

/m/ [m, ^{b}m] is a bilabial nasal and occurs in syllable onsets and codas.

[mãt˺]	/mãt/	‘eye, seed’
[tə.mã.giʔ]	/tə.mã.giʔ/	‘forehead’
[səm.pooy]	/səm.pooy/	‘eyebrow’
[bəm.bɨɨ^{g}ŋ]	/bəm.bɨɨŋ/	‘pus’

[nɔ̃ɔ̃m]	/nɔ̃ɔ̃m/	‘urine’
[kəl.ʔoo^{b}m]	/kəl.ʔoom/	‘brain’
[sən.ʔə^{b}m]	/sən.ʔəm/	‘sweat’
[m̩.pɔɔc$^{˺}$]	/m.pɔɔc/	‘salt’

/n/ [n, ^{d}n] is an alveolar nasal and occurs in syllable onsets and codas.

[nãɑ̃r]	/nãɑ̃r/	‘two’
[lən.tɑɑk$^{˺}$]	/lən.tɑɑk/	‘tongue’
[mɑ̃.nĩĩʔ]	/mɑ̃.nĩĩʔ/	‘rain’
[dɑ.nɑ̃n]	/dɑ.nɑ̃n/	‘corpse’
[hɔ̃ɔ̃n]	/hɔ̃ɔ̃n/	‘to smell something’
[loo^{d}n]	/loon/	‘mountain’
[lɑ.pɑ^{d}n]	/lɑ.pɑn/	‘eight’ (Malay: *lapan*)
[n̩.tooy]	/n.tooy/	‘big’
[n̩.seey]	/n.seey/	‘thin’

/ɲ/ [ɲ, ^{j}ɲ] is a palatal nasal and occurs in syllable onsets and codas.

[ɲãʔ]	/ɲãʔ/	‘far’
[ɲĩ.ɲɔ̃ɔ̃y]	/ɲĩ.ɲɔ̃ɔ̃y/	‘lips’
[sə.ɲũũp$^{˺}$]	/sə.ɲũũp/	‘dark’
[lə.mũũɲ]	/lə.mũũɲ/	‘tooth’
[gə.rɨɨ^{j}ɲ]	/gə.rɨɨɲ/	‘termite’
[bə.he^{j}ɲ]	/bə.hɛɲ/	‘itchy’
[n̩.ɲɨh]	/ɲ.ɲɨh/	‘heavy’

/ŋ/ [ŋ, ^{g}ŋ] is a velar nasal and occurs in syllable onsets and codas.

[ŋɔ̃ɔ̃t$^{˺}$]	/ŋɔ̃ɔ̃t/	‘to drink’
[sər.ŋĩĩʔ]	/sər.ŋĩĩʔ/	‘to think’
[nɔ̃ɔ̃ŋ]	/nɔ̃ɔ̃ŋ/	‘path’
[səŋ.ʔɔ̃ɔ̃y]	/səŋ.ʔɔ̃ɔ̃y/	‘person’
[tɑ.pə^{g}ŋ]	/tɑ.pəŋ/	‘spider’
[bəm.bɨɨ^{g}ŋ]	/bəm.bɨɨŋ/	‘pus’
[ŋ̩.kuʔ]	/ŋ.kuʔ/	‘thunder’

Fricatives

/s/ [s] is a voiceless palatal fricative and occurs in syllable onsets and codas.

[suuc$^{˺}$]	/suuc/	‘to wash’
[sər.ŋĩʔ]	/sər.ŋĩʔ/	‘to think’
[gən.siir]	/gən.siir/	‘husband’

[gə.sɑɑᵍŋ]	/gə.sɑɑŋ/	‘to dance’
[nɔ̃ɔ̃s]	/nɔ̃ɔ̃s/	‘heart’
[rə.ʔiis]	/rə.ʔiis/	‘root’

/h/ [h] is a voiceless glottal fricative and occurs in syllable onsets and codas.

[hɔ̃ɔ̃n]	/hɔ̃ɔ̃n/	‘to smell something’
[hən.lɨɨpˀ]	/hən.lɨɨp/	‘tomorrow’
[gə.hɛɛl]	/gə.hɛɛl/	‘tired’
[lə.hiiᵍŋ]	/lə.hiiŋ/	‘saliva’
[rɑ.mɔ̃h]	/rɑ.mɔ̃h/	‘to sneeze’
[sɑh]	/sɑh/	‘to wash clothes’

Trill / Flap

/r/ [r] is an alveolar flap and occurs in syllable onsets and codas. This segment is sometimes articulated as a trill, especially during slow or emphatic speech, but never in contrast to a flap.

[rɑɑkˀ]	/rɑɑk/	‘tiger’
[rət.riitˀ]	/rət.riit/	‘wringing’
[gər.ʔɨɨpˀ]	/gər.ʔɨɨp/	‘to burp’
[ʄ͡ʝ͡ʝər.lɑʔ]	/ʄər.lɑʔ/	‘thorn’
[ʄ͡ʝ͡ʝər. ʄ͡ʝ͡ʝar]	/ʄər.ʄar/	‘running’
[c͡çə.koor]	/cə.koor/	‘to rake’

Approximants

/l/ [l] is a lateral approximant and occurs in syllable onsets and codas.

[lukˀ]	/luk/	‘to laugh’
[ləɲ.siʲɲ]	/ləɲ.siɲ/	‘gums’
[kəl.ʔooᵇm]	/kəl.ʔoom/	‘brain’
[gəl.pɑɑl]	/gəl.pɑɑl/	‘shoulder’
[sən.tɔɔl]	/sən.tɔɔl/	‘feather’

/y/ [y] is a palatal central approximant and occurs in syllable onsets and codas.

[yɨɨcˀ]	/yɨɨc/	‘knife’
[sɑ.yɛ̃ɛ̃tˀ]	/sɑ.yɛ̃ɛ̃t/	‘child’
[bɑ.hɑ.yɑʔ]	/bɑ.hɑ.yɑʔ/	‘crocodile’
[sən.mɨ̃ɨ̃y]	/sən.mɨ̃ɨ̃y/	‘many people’
[suuy]	/suuy/	‘to live’

/w/ [w] is a labiovelar central approximant and occurs in syllable onsets and codas.

[wɔɔkˀ]	/wɔɔk/	‘shadow’

[wɑɑtˀ]	/wɑɑt/	‘to split’
[səl.wɑl]	/səl.wɑl/	‘trousers’
[kən.wiil]	/kən.wiil/	‘left side’
[teew]	/teew/	‘water’
[bə.lɑɑw]	/bə.lɑɑw/	‘blowgun’

F.4.2 Oral vowels

As with most languages of Mon-Khmer heritage, Semai is endowed with many vowels. However, while most Mon-Khmer languages have multiplied their vowel inventories with registers, tones, and diphthongs, Semai’s multitude of vowels is due to its having not only long and short vowels, but also both oral and nasal vowels.

The long vowels are not dramatically elongated. Indeed, it may be more accurate to portray the long vowels as the more “normal,” and the short vowels as extra short. Overall, there are roughly twice as many words with long vowels as opposed to short in the major syllable.

The epenthetic vowel in the minor syllable often undergoes vowel harmony with the major vowel when the intervening consonant is glottal (‘ʔ’ or ‘h’). This is discussed further in section 6 of this appendix.

Long oral vowels

/ii/ [ii] is a close front unrounded long vowel. It occurs only in closed major syllables.

[tiil]	/tiil/	‘to dry’
[gəl.piil]	/gəl.piil/	‘river turtle’
[liiᵇm]	/liim/	‘to swallow’

/ee/ [ee, e] is a close-mid front unrounded long vowel. It occurs only in closed major syllables. Usually this segment is articulated as a long vowel, but before glottal final consonants ‘h’ and ‘ʔ’, it is shortened to the allophone [e].

[teew]	/teew/	‘water’
[bə.tees]	/bə.tees/	‘mushroom’
[tə.beeᵍŋ]	/tə.beeŋ/	‘full’
[tɑ.leh]	/tɑ.leeh/	‘only, just’
[ɟ͡ɟə.reʔ]	/ɟə.reeʔ/	‘deep’

/ɛɛ/ [ɛɛ] is an open-mid front unrounded long vowel. It occurs only in closed major syllables.

[c͡çɛɛᵇm]	/cɛɛm/	‘bird’
[bə.kɛɛtˀ]	/bə.kɛɛt/	‘hot’
[lɛɛpˀ]	/lɛɛp/	‘to hunt’

/ɨɨ/ [ɨɨ, ɨ] is a close central unrounded long vowel. It occurs only in closed major syllables. Usually this segment is articulated as a long vowel, but before glottal final consonants ‘h’ and ‘ʔ’, it is shortened to the allophone [ɨ].

[sɨɨcˀ]	/sɨɨc/	‘to sting’
[pə.lɨɨtˀ]	/pə.lɨɨt/	‘to enter’

[ʔi.pɨɨr]	/ʔi.pɨɨr/	‘morning’
[kɨh]	/kɨɨh/	‘to hit’
[lɨʔ]	/lɨɨʔ/	‘pig’

/ɑɑ/ [ɑɑ] is an open central unrounded long vowel. It occurs only in closed major syllables.

[pɑɑy]	/pɑɑy/	‘new’
[ʄ͡ʝ̞ɑ.ʔɑɑᵍŋ]	/ʄə.ʔɑɑŋ/	‘bone’
[kə.rɑɑl]	/kə.rɑɑl/	‘man, male’

/uu/ [uu] is a close back rounded long vowel. It occurs only in closed major syllables.

[kuuy]	/kuuy/	‘head’
[ʄ͡ʝ̞ə.huuʔ]	/ʄə.huuʔ/	‘wood’
[nũm.puuʔ]	/nũm.puuʔ/	‘when’

/oo/ [oo, o] is a close-mid back rounded long vowel. It occurs only in closed major syllables. Usually this segment is articulated as a long vowel, but before glottal final consonants ‘h’ and ‘ʔ’, it is shortened to the allophone [o].

[looᵈn]	/loon/	‘mountain’
[n̤.tooy]	/n.tooy/	‘big’
[gər.gookˀ]	/gər.gook/	‘throat’
[koh]	/kooh/	‘to cut’
[bə.rɑ.poʔ]	/bə.rɑ.pooʔ/	‘to dream’

/ɔɔ/ [ɔɔ] is an open-mid back rounded long vowel. It occurs only in closed major syllables.

[sɔɔkˀ]	/sɔɔk/	‘hair’
[pɔɔʲɲ]	/pɔɔɲ/	‘to wait’
[ki.hɔɔy]	/ki.hɔɔy/	‘to yawn’

Short oral vowels

/i/ [i] is a close front unrounded vowel. It occurs in both open (minor) syllables and closed (minor and major) syllables.

[ki.lɨʔ]	/ki.lɨɨʔ/	‘other’
[bə.dil]	/bə.dil/	‘to shoot’
[ku.tiʔ]	/ku.tiʔ/	‘short’

/ɛ/ [ɛ, e] is a open-mid front unrounded vowel. It occurs only in closed major syllables. This segment is usually realized as [ɛ], but before a palatal consonant, it is raised to the allophone [e].

[c͡çɛʔ]	/cɛʔ/	‘louse’
[c͡çə.gɛh]	/cə.gɛh/	‘hard’
[guʔ.gɛʔ]	/guʔ.gɛʔ/	‘some’
[pən.lɛs]	/pən.lɛs/	‘dart head’
[sə.lecˀ]	/sə.lɛc/	‘smooth’

[bə.heʲɲ]	/bə.hɛɲ/	‘itchy’

/ə/ [ə] is a close-mid central unrounded vowel. It occurs in both open (minor) syllables and closed (minor and major) syllables. In the minor syllable, this segment appears to be in free variation with [ɨ]. In the minor syllable, [ə] is often very short and sometimes dropped in fast speech where the resulting consonant cluster is easy to pronounce without the vowel, such as when the onset of the major syllable is a lateral approximant or a flap.

[ləkˀ]	/lək/	‘quiver’
[sən.ʔəᵇm]	/sən.ʔəm/	‘sweat’
[sə.lɑʔ] ~ [slɑʔ]	/sə.lɑʔ/	‘leaf’

/ɑ/ [ɑ] is an open central unrounded vowel. It occurs in both open (minor) syllables and closed (minor and major) syllables.

[dɑᵈn]	/dɑn/	‘to die’
[bɑ.hɑ.yɑʔ]	/bɑ.hɑ.yɑʔ/	‘crocodile’
[c͡çɑ. c͡çoh]	/cɑ.cooh/	‘to defecate’
[bəl.ʔɑkˀ]	/bəl.ʔɑk/	‘black’

/u/ [u] is a close back rounded vowel. It occurs in both open (minor) syllables and closed (minor and major) syllables.

[su.dɑʔ]	/su.dɑʔ/	‘bamboo spear’
[kus]	/kus/	‘to get up’
[tɑ. ʄ͡ʝuʔ]	/tɑ.ɟuʔ/	‘snake’
[gɑ.sur]	/gɑ.sur/	‘cockroach’

/ɔ/ [ɔ] is an open-mid back rounded vowel. It occurs in closed (minor and major) syllables.

[kə.rɔm.dɛɛkˀ]	/kə.rɔm.dɛɛk/	‘armpit’[46]
[bɔʔ]	/bɔʔ/	‘to carry’
[kən.tɔᵇm]	/kən.tɔm/	‘right (side)’
[gɔl]	/gɔl/	‘to carry on shoulder’
[ʄ͡ʝə.lɔy]	/ɟə.lɔy/	‘straight’

F.4.3 Nasal vowels

The nasal vowels are of two types: those that are automatically nasalized after a nasal consonant, and others that are unpredictable and therefore contrastive. Both types are treated together here. While the evidence supports that nasality is phonemic, nasal vowels (especially the unpredictable type) are rare, and minimal pairs of words showing contrast with the oral vowels are rather hard to find. The following examples, while not minimal pairs, clearly show that nasality is phonemic.

[lɔ̃ɔ̃tˀ]	/lɔ̃ɔ̃t/	‘to suck on something’
[sə.lɔɔtˀ]	/sə.lɔɔt/	‘thirsty’
[lɔɔcˀ]	/lɔɔc/	‘shaved bald (head)’

[46]But this may be a frozen form of *kərɔm* ‘under’ plus *dɛɛk* ‘???’.

[tə̃.ʔə̃ə̃c˥]	/tə.ʔə̃ə̃c/	‘wet’
[ku.wɑɑc˥]	/ku.wɑɑc/	‘to scratch like a chicken’
[kə.rɑ.sə̃h]	/kə.rɑ.sə̃h/	‘rough to the touch’
[sɑh]	/sɑh/	‘to wash (e.g. hands)’
[tɑŋ.kɔ̃l]	/tɑŋ.kɔ̃l/	‘to carry on head’
[sən.tɔɔl]	/sən.tɔɔl/	‘feather’
[gɔl]	/gɔl/	‘to carry on shoulder’
[pəlʔĩ̈h]	/pəl.ʔĩ̈h/	‘smell of freshly cut fish’
[gi̇ih]	/gi̇ih/	‘to scratch’

Long nasal vowels

/ĩ̈ĩ/ [ĩ̈ĩ] is a close front unrounded long nasal vowel. It occurs only in closed major syllables.

[mĩ̈ĩŋ]	/mĩ̈ĩŋ/	‘cheek’
[kə.mĩ̈ĩl]	/kə.mĩ̈ĩl/	‘above’
[rək.nĩ̈ĩk˥]	/rək.nĩ̈ĩk/	‘orphan’

/ɛ̃ɛ̃/ [ɛ̃ɛ̃] is an open-mid front unrounded long nasal vowel. It occurs only in closed major syllables.

[sɑ.yɛ̃ɛ̃t˥]	/sɑ.yɛ̃ɛ̃t/	‘child’
[rə.mɛ̃ɛ̃t˥]	/rə.mɛ̃ɛ̃t/	‘yellow’
[nɛ̃ɛ̃ŋ]	/nɛ̃ɛ̃ŋ/	‘to see’

/ɨ̃ɨ̃/ [ɨ̃ɨ̃] is a close central unrounded long nasal vowel. It occurs only in closed major syllables. Usually this segment is articulated as a long vowel, but before glottal final consonants ‘h’ and ‘ʔ’, it is shortened to the allophone [ɨ̃].

[sən.mɨ̃ɨ̃y]	/sən.mɨ̃ɨ̃y/	‘many (people)’
[sə.mɨ̃ɨ̃c˥]	/sə.mɨ̃ɨ̃c/	‘stinging insect’
[cə.mɨ̃ɨ̃s]	/cə.mɨ̃ɨ̃s/	‘hungry’
[mɨ̃h]	/mɨ̃ɨ̃h/	‘name’
[mə̃.mɨ̃h]	/mə̃.mɨ̃ɨ̃h/	‘to bathe’

/ə̃ə̃/ [ə̃ə̃] is an open central unrounded long nasal vowel. It occurs only in closed major syllables.

[tə̃.ʔə̃ə̃c˥]	/tə.ʔə̃ə̃c/	‘wet’
[nə̃ə̃r]	/nə̃ə̃r/	‘two’

/ũũ/ [ũũ] is a close back rounded long nasal vowel. It occurs only in closed major syllables.

[lə.mũ̈ũ̈ŋ]	/lə.mũ̈ũ̈ŋ/	‘tooth’
[ʔi.nũ̈ũ̈y]	/ʔi.nũ̈ũ̈y/	‘something made’
[kə.nũ̈ũ̈r]	/kə.nũ̈ũ̈r/	‘type of rattan used as a spice’

/ɔ̃ɔ̃/ [ɔ̃ɔ̃] is an open-mid back rounded long nasal vowel. It occurs only in closed major syllables.

[hɔ̃ɔ̃n]	/hɔ̃ɔ̃n/	‘to smell/sniff something’

[lɔ̃ɔ̃tˈ]	/lɔ̃ɔ̃t/	‘to suck on something’
[mɑ̃. c͡çɔ̃ɔ̃tˈ]	/mɑ̃.cɔ̃ɔ̃t/	‘small’
[ɲĩ.ɲɔ̃ɔ̃y]	/ɲĩ.ɲɔ̃ɔ̃y/	‘lip’

Short nasal vowels

/ĩ/ [ĩ] is a close front unrounded nasal vowel. It occurs in both open (minor) syllables and closed (minor and major) syllables.

[ɲĩ.ɲɔ̃ɔ̃y]	/ɲĩ.ɲɔ̃ɔ̃y/	‘lip’
[pəl.ʔĩh]	/pəl.ʔĩh/	‘smell of fresh fish’
[nĩʔ]	/nĩʔ/	‘three’

/ɛ̃/ [ɛ̃] is an open-mid front unrounded nasal vowel. There were only three examples in the data, all in closed major syllables.

[tə.nɛ̃ʔ]	/tə.nɛ̃ʔ/	‘elder sibling’
[ʔɑ.mɛ̃ʔ]	/ʔɑ.mɛ̃ʔ/	‘mother’
[ɲɛ̃tˈ mɑ̃tˈ]	/ɲɛ̃t mɑ̃t/	‘to wink an eye’

/ə̃/ [ə̃] is an open-mid central unrounded nasal vowel. So far it has only been found in two examples, both in minor syllables.

[nə̃.rɑ.mĩʔ]	/nə̃.rɑ.mĩʔ/	‘friend’
[mə̃n.mɑ̃n]	/mə̃n.mɑ̃n/	‘playing’

/ɑ̃/ [ɑ̃] is an open central unrounded nasal vowel. It occurs in both open and closed syllables, and in both major and minor syllables.

[mɑ̃tˈ]	/mɑ̃t/	‘eye’
[mə̃n.mɑ̃n]	/mə̃n.mɑ̃n/	‘playing’
[kə.rɑ.sɑ̃h]	/kə.rɑ.sɑ̃h/	‘rough to the touch’
[mɑ̃.mɨ̃h]	/mɑ̃.mɨ̃̃h/	‘to bathe’

/ũ/ [ũ] is a close back rounded nasal vowel. It occurs in both open and closed syllables, and in both major and minor syllables.

[mũc]	/mũc/	‘to get up very early in the morning’
[nũk]	/nuk/	‘type of bird’
[nũm.puuʔ]	/nũm.puuʔ/	‘when’

/ɔ̃/ [ɔ̃] is an open-mid back rounded nasal vowel. It occurs only in closed major syllables.

[tɑŋ.kɔ̃l]	/tɑŋ.kɔ̃l/	‘to carry on head’
[sə.ŋɔ̃h]	/sə.ŋɔ̃h/	‘fear’
[rɑ.mɔ̃h]	/rɑ.mɔ̃h/	‘to sneeze’

F.5 Distribution charts of Betau Semai phonemes

The distribution of the phonemes is notably different for each position in the Semai word, which once again is:

$(C_3 \quad V_2 \quad (C_4))$	$C_1 \quad V_1 \quad C_2$
Minor	Major

The consonants have the following distribution:

Consonant C_1

	Bilabial	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Plosive, voiceless	p	t	c	k	ʔ
Plosive, voiced	b	d	ɟ	g	
Nasal	m	n	ɲ	ŋ	
Fricative, voiceless			s		h
Flap		r			
Lateral approximant		l			
Central approximant	w		y		

Consonant C_2

	Bilabial	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Plosive, voiceless	p	t	c	k	ʔ
Nasal	m	n	ɲ	ŋ	
Fricative, voiceless			s		h
Flap		r			
Lateral approximant		l			
Central approximant	w		y		

Consonant C_3

	Bilabial	Alveolar	Palatal	Velar	Glottal
Plosive, voiceless	p	t	c	k	ʔ
Plosive, voiced	b	d	ɟ	g	
Nasal	m	n	ɲ	[ŋ][a]	
Fricative, voiceless			s		h
Flap		r			
Lateral approximant		l			
Central approximant[b]	(w)		(y)		

[a]The segment /ŋ/ is evidently rare as it was not found in the data; however, it is expected for symmetry reasons.
[b]Note: In the C3 position, the central approximants (/w/ and /y/) were only as a result of reduplication.

Consonant C_4[a]

	Bilabial	Alveolar	Palatal	Velar	Glottal
Plosive, voiceless	(p)	(t)	(c)	(k)	ʔ
Nasal	m	n	ɲ	ŋ	
Fricative, voiceless			(s)		(h)
Flap		r			
Lateral approximant		l			

[a]In the C_4 position, all occurrences of the voiceless stops (/p/, /t/, /c/ and /k/) and fricatives (/s/ and /h/) appeared to be due to infixation, reduplication, and compound words.

The vowels have the following distribution:

Vowel V1, oral[a]

	Front (unrounded)		Central (unrounded)		Back (rounded)	
Close	i	ii		ɨɨ	u	uu
Close-mid		ee				oo
Open-mid	ɛ	ɛɛ		ə	ɔ	ɔɔ
Open			ɑ	ɑɑ		

[a]The segment [ʊ] is found in some borrowed words, and reflects the Malay pronunciation.

Vowel V1, nasal

	Front (unrounded)		Central (unrounded)		Back (rounded)	
Close	ĩ	ĩĩ		ɨ̃ɨ̃	ũ	ũũ
Mid	ɛ̃	ɛ̃ɛ̃	ə̃		ɔ̃	ɔ̃ɔ̃
Open			ɑ̃	ɑ̃ɑ̃		

Vowel V2, oral[a]

	Front (unrounded)	Central (unrounded)	Back (rounded)
Close	i		u
Mid		ə	
Open		ɑ	

[a]There was one example of [ɪ], but it was in a closed syllable of a borrowed word, reflecting the Malay pronunciation. Segments [ɛ] and [ɔ] were found, but only as vowel harmony with the major syllable across glottal segments /h/ or /ʔ/.

Vowel V2, nasal

	Front (unrounded)	Central (unrounded)	Back (rounded)
Close	ĩ	ɨ̃	ũ
Open		ɑ̃	

F.6 Prosodic features

F.6.1 Vowel harmony

The epenthetic vowel in the minor syllable often undergoes vowel harmony with the major vowel when the intervening consonant is glottal (/ʔ/ or /h/). This feature results in free variation of pronunciation for such words, with a tendency toward the epenthetic central vowel when the word is articulated slowly and carefully, but a harmonized vowel when articulated during normal or fast speech.

[ɟə.ʔɑɑ^{g}ŋ] ~ [ɟɑ.ʔɑɑ^{g}ŋ]	/ɟə.ʔɑɑŋ/	‘bone’
[pə.ʔoʔ] ~ [po.ʔoʔ]	/pə.ʔooʔ/	‘bamboo’
[tə.hɔr] ~ [tɔ.hɔr]	/tə.hɔr/	‘to incant’
[pə.hoot˺] ~ [po.hoot˺]	/pə.hoot/	‘to be hurting’

F.6.2 Preploded nasals

This dialect is interesting in that it is representative of a relatively small area of the Semai territory that has preserved the preploded nasals. Most of the Semai dialects have reduced these phonemes to simple voiceless plosives. It is clear that these phonemes used to be simple nasals in the past, both from comparison to other Mon-Khmer languages and to words borrowed from Malay. For example,

[lɑ.pɑ^{d}n]	/lɑ.pɑn/	Malay: *lapan* (‘eight’)
[ku. c͡çi^{g}ŋ]	/ku.ciŋ/	Malay: *kucing* (‘cat’)
[ɟ͡ɟɑ.ru^{b}m]	/ɟɑ.rum/	Malay: *jarum* (‘needle’)

This feature of preplosion has sometimes been labeled “pre-denasalization.” This is an apt term, since preploded nasals seem be part of a prosodic tendency to prevent nasalization from spreading leftward in the word. Indeed, those dialects that have reduced the nasal to a simple voiceless plosive have removed nasalization from the oral major syllable altogether, making the whole syllable nonnasal. In the Betau dialect preploded nasals are found only after oral vowels. Note the following examples, which have nasal vowels (or else vowels nasalized by the preceding nasal consonant) and hence do not have preplosion.

[hɔ̃ɔ̃n]	/hɔ̃ɔ̃n/	‘to smell’
[mĩĩŋ]	/mĩĩŋ]	‘cheek’
[nɛ̃ɛ̃ŋ]	/nɛ̃ɛ̃ŋ/	‘to look’

There is one example, however, that appears to be an exception to the rule; namely, a simple nasal is found after an oral vowel.

[pər.həm]	/pər.həm/	‘to breathe’

It is possible that the major vowel was slightly nasalized, or that the preceding ‘h’ comes into play somehow.

F.7 Residue

F.7.1 Glottal-final words

A great many of the long vowels in words ending with glottal segments /-h/ or /-ʔ/ appear to have been shortened and hence phonetically are now short vowels.[47] This helps explain why the long vowels /oo/, /ee/, and /ɨɨ/ have shortened vowel allophones in these positions. This conclusion would be more satisfactory, however, if all long vowels were shortened before these glottal-final segments. However, the language assistant did pronounce a few /-ʔ/-final words with other long vowels (although not always consistently). For example,

[47]This phenomenon was discussed by Diffloth (1977).

[mɑ̃.nɪ̃ɪ̃ʔ] ~ [mɑ̃.nɪ̃ʔ]	/mɑ̃.nɪ̃ɪ̃ʔ/	‘rain’
[sər.ŋɪ̃ɪ̃ʔ] ~ [sər.ŋɪ̃ʔ]	/sər.ŋɪ̃ɪ̃ʔ/	‘to think’
[də.kɔɔʔ] ~ [də.kɔʔ]	/də.kɔɔʔ/	‘boil’ (n)
[ʄ͡ʝə.huuʔ]	/ʄə.huuʔ/	‘wood, tree’

The conclusion is that either this phonological change is not complete or not consistent in this dialect, or else perhaps this speaker has picked up pronunciations from speakers of other dialects, and uses these pronunciations occasionally. It would be good to check these words with other speakers of this dialect.

For the sake of this paper, vowels that were sometimes heard as phonetically long before glottal stop have been marked as long vowels phonemically. Likewise, vowels never heard long before a final glottal stop have been marked as short vowels phonemically, except of course for the long vowels /oo/, /ee/, and /ɨɨ/, which do not have phonemic short vowel counterparts.

F.7.2 Syllabic nasals

Words with syllabic nasals at the beginning seem to be preceded by a glottal stop, which would allow the minor syllable in these words to be analyzed as /ʔəŋ-/. Further research is needed to see if this analysis is correct. A study of the morphology of the language, particularly reduplication and affixation, should shed light on these words.

F.7.3 Final consonants

In general, all Semai words have final consonants. However, one word appears to defy this rule:

[pɑ.nii]	/pɑ.nii/	‘to know’[48]

In other dialects of Semai this word is ‘pɑ.nɛ̃ɛ̃y’ or ‘pɑ.nɛ̃ɛ̃y’. Apparently a phonological change in this dialect has shifted the proto-vowel ‘*ɛ̃ɛ̃’ to ‘ɪ̃ɪ̃’ in this language, with the odd side effect of essentially swallowing the final semi-vowel ‘y’ in this word, or at least rendering the final ‘y’ indiscernible from the vowel. Not only would it be interesting to find more words of this type, but there may well be a parallel situation with the back vowels, if the proto-vowel ‘*õõ’ is shifted to ‘ʊ̃ʊ̃’ before ‘w’.

[48]This word was most likely originally borrowed from the Malay word *pandai* ‘intelligent’.

References

Baer, Adela. 1999. *A Temuan-English-Malay lexicon.* Adela S. Baer papers, Orang Asli archive. Keene, New Hampshire. http://www.keene.edu/library/orangasli/lexicon.pdf, accessed 2005.

Basrim bin Ngah Aching, ed. 2008. *Kamus Semai*. Bangi, Malaysia: Institut Alam dan Tamadun Melayu [Institute of the Malay World and Civilization].

Benjamin, Geoffrey. 1976a. An outline of Temiar grammar. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.), *Austroasiatic studies, Part I,* 129–187. Honolulu: University Press of Hawaii.

Benjamin, Geoffrey. 1976b. Austroasiatic subgroupings and prehistory in the Malay peninsula. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.), *Austroasiatic studies, Part I,* 37–128. Honolulu: University Press of Hawaii.

Bishop, Nancy. 1996. A preliminary description of Kensiw (Maniq) phonology. *Mon-Khmer studies: A Journal of Southeast Asian linguistics and languages* 25:227–253.

Carey, Iskandar. 1976. *Orang Asli: The aboriginal tribes of peninsular Malaysia.* Oxford: Oxford University Press.

Dentan, Robert Knox, Kirk Endicott, Alberto G. Gomes, and M. B. Hooker. 1997. *Malaysia and the original people: A case study of the impact of development on indigenous peoples.* Boston: Allyn and Bacon.

Diffloth, Gérard. 1968. Proto-Semai phonology. *Federation museums journal* 13:65–74. Kuala Lumpur, Malaysia: Muzium Negara.

Diffloth, Gérard. 1976a. Minor-syllable vocalism in Senoic languages. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.), *Austroasiatic studies, Part I,* 229–247. Honolulu: University Press of Hawaii.

Diffloth, Gérard. 1976b. Expressives in Semai. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.), *Austroasiatic studies, Part I,* 249–264. Honolulu: University Press of Hawaii.

Diffloth, Gérard. 1977. Towards a history of Mon-Khmer: Proto-Semai vowels. *Tonan Azia Kenkyu [Southeast Asian studies]* 14:463–495.

Diffloth, Gérard. 1980. To taboo everything at all times. *Proceedings of the Berkeley Linguistic Society* 6:157–165. Berkeley.

Kroeger, Paul R. 1989. Discontinuous reduplication in vernacular Malay. *Proceedings of the Berkeley Linguistic Society* 15. Berkeley.

Lewis, M. Paul, ed. 2009. *Ethnologue: Languages of the world*. Sixteenth edition. Dallas: SIL International.

Matisoff, James. 2003. Aslian: Mon-Khmer of the Malay peninsula. *Mon-Khmer Studies: A Journal of Southeast Asian linguistics and languages* 33:1–57.

Phillips, Timothy C. 2006a. A survey of nasal preplosion in Aslian languages. Paper presented at the International Conference on Indigenous People, Kuala Lumpur, Malaysia. 5 July 2006.

Phillips, Timothy C. 2006b. Dialect comprehension survey of the Semai language. Unpublished manuscript. Economic Planning Unit, Prime Minister's Department, Malaysia.

Phillips, Timothy C. 2007. A brief phonology of Betau Semai. In Shin Chong, Karim Harun, Yabit Alas, and James T. Collins (eds.), *Reflections in Southeast Asian seas*. Essays in Honour of Professor James T. Collins 2. Pontianak, Indonesia: STAIN Pontianak Press.

Phillips, Timothy C. 2011. Retention and Reduction in Reduplication of Semai. In Sophana Srichampa, Paul Sidwell, and Kenneth Gregerson (eds.), *Austroasiatic studies: Papers from ICAAL4,* 178–179. Mon-Khmer Studies Journal Special Issue 3. Dallas: SIL International; Bangkok: Mahidol University; Canberra: Pacific Linguistics.

Phillips, Timothy C. Forthcoming. Proto-Aslian: Towards an understanding of its historical linguistic systems, principles, and processes. Ph.D. dissertation. Universiti Kebangsaan Malaysia.

Keene State University. 2002. *Population of Orang Asli Sub-Groups, 1999*. Orang Asli archive. http://www.keene.edu/library/OrangAsli/pop2000.doc, accessed 7 February 2005.

Posey, Darrell A. 2001. Biological and cultural diversity: The inextricable, linked by language and politics. In Luisa Maffi (ed.), *On biocultural diversity: Linking language, knowledge, and the environment,* 379–396. Washington, DC: Smithsonian Institution Press.

Seidlitz, Eric. 2005. A study of Jakun Malay: Some aspects of its phonology. M.A. thesis. Universiti Kebangsaan Malaysia.
Simons, Gary F. 1977. Tables of significance for lexicostatistics. In Richard Loving and Gary F. Simons (eds.), *Language variation and survey techniques.* Work papers in Papua New Guinea Languages 21, 75–106. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
Zaharani bin Ahmad. 1991. *The phonology and morphology of the Perak dialect.* Kuala Lumpur: Dewan Bahasa dan Pustaka.