1 Semelai

Semelai is an Aslian language belonging to the Mon-Khmer division of Austroasiatic. It is spoken by approximately 4,103 people in Peninsular Malaysia around the shores of Tasek Bera, along the banks of the Bera, Teriang and Serting rivers in south-west Pahang and north-west Negeri Sembilan states, and the Muar River in north-west Johore (see Maps).[1]

As with all Aslian languages, there is no written tradition, or recorded history, other than oral based.

1.1 Linguistic type

Semelai exhibits many of the typological characteristics of a mainland Southeast Asian language, as well as incorporating aspects of Austronesian, notably from sustained contact with Malay. Whilst there has been considerable impact from Malay, particularly at the lexical level, Semelai remains a distinctly mainland Austroasiatic language, but at the same time raises questions with respect to our understanding of the typology of this region.

A. Phonology Semelai has a rich phonemic inventory. There are thirty-two consonant phonemes, including a series of voiceless nasals which have not been recorded previously in an Aslian language; a series of glottalised sonorants, and twenty vowel phonemes: ten oral vowels and ten phonemically nasal counterparts (§2.1).

The maximal canonical syllable is $[CV(C)]_{\sigma}$. Phonemic contrasts of both consonants and vowels are richest in the final syllable; for consonants these contrasts are maximised in the onset of the final syllable.

Words have the structure: $\omega \rightarrow (\sigma_R)^n\ (\sigma)\,'\sigma$, where: $^n \leq 2$. Reading from the right, $'\sigma$ is the final syllable – word stress is always on the final syllable, and there is no secondary stress, σ is the penultimate syllable and σ_R is the prepenultimate syllable. The $_r$ indicates significantly reduced phonotactic possibilities for both consonants and vowels, including the characteristic underspecification of vowels associated with this position. Any oral vowel phoneme has the potential to occur as the phonetic realisation of an underspecified vowel in a given environment.

The minimal word is monosyllabic, e.g. /tʰi/ 'hand', the maximal word is tetrasyllabic: /k.r.wan.ceŋ/ [kəruwanceŋ] 'coral snake sp.'

B. Morphology Semelai is an isolating language with agglutinating features. Typically for the Aslian family, but at variance with many Mon-Khmer languages, Semelai has a complex morphological system, containing a rich inventory of prefixes,

[1] The Semoq Beri do not inhabit this area (*pace* Parkin 1991: 55), but an area in north eastern Pahang.

infixes, suffixes and a circumfix. Some have their origins firmly in Austroasiatic (e.g. the nominalising affixes +n+[2] and <m> (NMZ)), others have been acquired through language contact with varieties of Malay (e.g. br- (MID) 'middle voice', and tr- (HAPP) 'happenstance'), see §3.2. Semelai has distinctive word classes (see C below and §4); the typical function of derivational morphology is to change the word class or subclass of the root, e.g. from verb to noun or vice versa, from transitive verb to intransitive verb (§5.3) or mass noun to count noun (§7.6).

There are two systems of arrangement based on the domain of attachment: a) a non-concatenative system of prefixes and infixes, which has its origins in Mon-Khmer; and b) a concatenative system of prefixes, suffixes and a circumfix acquired through contact with Austronesian, notably Malay. Both systems of attachment are prosodically driven.

Affixes may be syllabic, or they may be non-syllabic consonantal units defined in terms of syllable position, e.g. <r> 'CAUS' is a syllable coda infix $r]_{\sigma}$. Syllabic affixes are fully prespecified, e.g. $[par]_{\sigma}$ 'CAUS(ative)'.

Where affixes exhibit allomorphy, this is conditioned primarily by the syllabic structure of the root and the domain of attachment: monosyllabic roots select heavy syllable prefixes, whilst disyllabic roots select either infixes or light syllable prefixes, dependent on the individual affix. These points are illustrated in the following discussion which focuses on affixes from the non-concatenative system that take the prosodic head as the domain of attachment.

The two forms given above for the causative morpheme, tar- 'MCAUS' and <r> 'CAUS', are licit for monosyllabic and disyllabic roots respectively:

a) *Monosyllabic root*: hɒp 'to put food in (one's) mouth'
   Affix tar- 'MCAUS':
   tar-hɒp 'to put food in (s.o.'s) mouth'

b) *Disyllabic root*: jtɛk [jtɛk] 'to sleep'
   Infix <r> 'CAUS':
   j<r>tɛk [jərtɛk] 'to cause to sleep'

The most intriguing morphological process is the reduplicative process, which I term 'Copy'. This is a form of internal reduplication, whereby a phonologically underspecified morpheme template $[CC]_{\mu}$ equivalent to the maximal canonical syllable, is affixed in penultimate syllable position to the final syllable of the root. The underspecified segment, which may be a syllable onset or a coda, derives its phonemic content by copying that of the corresponding segment of the root. The underspecified vowel slot is filled by epenthesis.

With monosyllabic roots, the satisfaction of the prefixed template requires full reduplication of the root:

Monosyllabic root: sɔc 'to whistle'

a) Prefix the template $[CC]_{\mu}$:
   $[CC]_{\mu}$-sɔc

[2] Morpheme boundaries of prefixes and suffixes are represented by a hyphen '-', infixes and the circumfix by parentheses '< >', underspecified affixes are enclosed by '+' and clitic boundaries are shown by '='.

b) Onset and coda copy reduplicate the phonemic consonantal content of the root into the underspecified positions in the prefix:
sc-sɔc
c) Given that only consonants may be reduplicated, the underspecified nucleus receives phonetic content from the application of vowel realisation rules:
sc-sɔc [sicsɔc] 'to be whistling'

With disyllabic roots, the phonemic content of the penultimate syllable of the root is associated with the template. Coda copy takes place into the underspecified coda position:

Disyllabic root: catək 'to detach'
Coda copy:
a) Prefix the template $[C]_{\mu}$ to the prosodic head and associate the phonemic content of the root:
ca<$[C]_{\mu}$>tək
b) Copy the phonemic content of the root coda into the underspecified coda position C:
ca<k>tək 'to be detaching'

Coda copy may co-occur with the affixation of a partially prespecified affix like nominalising onset n- which has a morphemic template $[n\,C]_{\mu}$:

Monosyllabic root: sɔc 'to whistle'
a) Prefix the template $[n\,C]_{\mu}$ and associate the phonemic content of the root:
$[n\,C]_{\mu}$-sɔc → n C-sɔc
b) Coda copy reduplicates the phonemic consonantal content of the root into the underspecified position in the prefix:
nc-sɔc
c) The underspecified nucleus receives phonetic content from the application of vowel realisation rules:
nc-sɔc [nicsɔc] 'act of whistling'

Borrowed Malay lexemes are fully incorporated into the Semelai system. There are no restrictions on the combinatorial possibilities of indigenous and borrowed terms.

a) Coda copy
Prefixation of the template $[CC]_{\mu}$ to the prosodic head of verbal root tuɲuk 'to point out, indicate' (< Malay *tunjuk*) derives:
tu<k>ɲuk 'to be pointing'

b) Nominalisation:
Infixation of the nominalisation template $[n\,C]_{\mu}$ into the root sudu? 'spoon' (< Malay *sudu*) derives a measure noun:
s<n>u<?>du? [sənu?du?] 'spoonful'

Roots can feed multiple affixation, combining affixes from both systems of attachment. In the following example, taking the verb kɲɛp 'to blink', the causative infix, <r> 'CAUS', is a morpheme of the non-concatenative type, and the prefixed happenstance morpheme, t- 'HAPP', is from the concatenative type:

a) Infixation of <r> 'CAUS':
k<r>ɲɛp (blink<CAUS>) [kərɲɛp] 'to make blink'

b) Prefixation of t- 'HAPP':
t-k<r>ɲɛp (HAPP-blink<CAUS>) [təkərɲɛp] 'to happen to make blink'

C. Word classes Semelai has clearly distinguished word classes, both syntactically and morphologically. The open classes are nominal, verb and expressive.

The nominal superclass includes nouns, as well as the following closed classes: numerals, pronouns, demonstratives and ignoratives. Nominals function as the heads of NPs, predicates in non-verbal clauses and nominal modifiers in associative constructions.

Verbs function as predicates. The majority of verbs are clearly transitive or intransitive; there is also a small number of ambitransitive verbs. The distinction between stative and active verbs cuts across the transitive/intransitive division. Intransitive verbs also distinguish an adjective class.

Expressives[3] are iconic utterances, which function to simultaneously provide information about both the predicate and its arguments, in the form of a single lexical item. They function as clausal adjuncts, or stand alone as minor clauses.

Members of this class express sensate imagery – aural, visual, oral, tactile – e.g. tɔ̃ŋbɔ̃c 'short and fat (of people)'. Expressives frequently combine a cluster of properties, e.g. the lexical item ɒrʔol '(s.th.) large, dark and motionless lying submerged in the water' combines dimension, colour and position. In addition, they exhibit irregular phonology, irregular reduplication patterns and vowel alternation.

The closed classes are preposition, adverb, auxiliary, existential and ascriptive predicators, negator, connective and interjection (§4).

Lexemes must undergo derivational procedures, either morphological or syntactic, in order to function in a different word class.

D. Constituent order Semelai exhibits the characteristic head–dependent word order of the languages of mainland Southeast Asia: the attributive follows the noun (1) (§7.4); the possessor follows the possessee (2)–(3) (§7.5); the preposition precedes the noun (3)–(4) (§8), and relative clauses follow the head (5) (§11.1). At the phrase level, constituent order is fixed.

(1) wɒy ʔluh
knife be.sharp
sharp knife

(2) wɒy kmpən
knife wife
the wife's knife

(3) wɒy dɔ kmpən
knife OF wife
the wife's knife

(4) rɔm wɒy
WITH knife
with a knife

[3] 'Expressive' is the term used by scholars of Austroasiatic; the more widely used term is ideophone.

(5) wɒy mə=ʔluh
knife REL=be.sharp
the knife that is sharp

Constituent order at the clause level is fluid, with variation driven by pragmatic factors (§9). In the transitive clause the ordering is usually verb initial (6), although any argument may be pre-verbal, in which case it loses its role marking (7), see E below:

(6) ki=tikam la=knlək hn=pɔdɔŋ rɔm lmeŋ
3A=stab A=husband O=tiger WITH spear
The husband stabbed the tiger with the spear.

(7) pɔdɔŋ ki=tikam la=knlək rɔm lmeŋ
tiger 3A=stab A=husband WITH spear
The tiger, the husband stabbed with the spear.

In contrast, in the morphologically simplest, but 'statistically' marked transitive clause type – the 'universal' clause – the pattern is SVO, and there is no coding of grammatical relations:

(8) pɔdɔŋ ca sma?
tiger eat person
Tigers eat people.

In the intransitive clause, the subject may also either precede or follow the verb.

(9) kəhn swak
3S go
He went.[4]

(10) swak kəhn
go 3S
He went.

E. Morphosyntax The syntactic typology of Semelai displays an intricate mix of features based on a basic transitive/perfective–intransitive/imperfective contrast.

Semelai exhibits both head and dependent marking, e.g. the head-marking pronominal proclitics on the transitive verb and the dependent-marking proclitics on post-verbal arguments which are introduced below.

Core grammatical relations are marked by a system of clitic cross-referencing in the transitive clause. The A is cross-referenced by a pronominal proclitic on the verb (6), and proclitic la= 'A' on the optional post-verbal NP. A system of differential object marking optionally encodes the object NP with proclitic hn= 'O'. The O is not cross-referenced on the verb, see example (6).

In the intransitive clause, the subject NP does not exhibit role marking, with the exception of the third person pronominal forms which have a fused enclitic hn 'S': kəhn '3S' (< kəh=hn 3=S), dehn '3plS' (< deh=hn 3pl=S). As with O, S is not cross-referenced on the verb, see example (9) above.

[4] Third person pronouns are not distinguished for gender. Rather than using the clumsy '(s)he' for third person singular forms, they are translated throughout according to the context of the original text.

Indirect objects and obliques are coded by prepositions as in (6), with the exception of the benefactive which is not formally coded.

A semantically motivated split gives rise to two marking possibilities for intransitive verbs which express motion or emotion. The basis of the split is to mark a type of involuntary compulsion resulting from external causation as in (11). Curiously, the subject of the intransitive clause is encoded by the ergative pronominal proclitic like a transitive subject, although the clause is still monovalent, and the subject displays low agentivity (§5.5.1.2). Compare the following example with (9) above:

(11) *The following comment was made when a person walked off from the group because someone offended him.*
ki=swak
3A=go
He went off.

Semelai exhibits odd transitivity behaviour, where lower semantic transitivity is not reflected by a change in morpho-syntactic marking. The iterative suffix -iʔ 'ITER', expressing the repetition of an activity (§5.3.6), and the happenstance prefix tr- 'HAPP', expressing lack of intention or volition (§5.3.7) both signal reduced transitivity, yet the clause retains the argument coding of a prototypical transitive clause.

F. Pronouns The pronominal system expresses categories of first, second and third person, with an inclusive/exclusive distinction for first person, a deference distinction for first and second persons singular, and a category of unidentified agent in the third person (§6.1). Interestingly, the personal pronouns have a greater range of number possibilities in the third person, where there are independent singular and plural forms in both the pronominal proclitics and free forms, but in first and second persons this distinction is not made for the proclitic forms, and is only optionally expressed in the free forms by the enclitic =ʔen 'PL(ural)'.

G. Locative prepositions and directionals Semelai has a topologically based system of spatial deixis, manifested in locative prepositions and directionals, which conflate the locative relation with three degrees of height. These are arranged in a 'person' or 'speaker' oriented system mapped in relation to the topography of the local environment (§8).

The locative system consists of one proximate term, constituting the 'zero-point', and three distal terms (§8.5):

haʔ	'AT'	immediate location of Speaker
hɛʔ	'AT:above'	space above Speaker
tɔʔ	'AT:across'	space on the same level, but away from Speaker
cɔʔ	'AT:below'	space below Speaker

There are four terms in the directional system, two neutral terms distinguished for specificity of direction, and two deictically specified terms (§8.6):

tet	'TO:spec'
te	'TO:unspec'

leŋ ~ ləŋ	'TO:up'	upward, upstream, uphill
tə?ɛn	'TO:down'	downward, downstream, downhill

An example of a locative, cɔ? 'AT:below' and a directional leŋ ~ ləŋ 'TO:up' are shown in the following example:

(12)

ki=cʰək	hn=hayam	tɔm	cɔ?	?atɛ	leŋ	dɔl
3A=throw	O=chicken	SRC	AT:below	ground	TO:up	house

He threw the chicken from down on the ground up to the house.

Even when used to encode a core recipient relation like the indirect object, the locative preposition retains this semantic distinction, as illustrated below:

(13)

ki=cʰək	hn=hayam	cɔ?	??ɛ?
3A=throw	O=chicken	AT:below	elder.sibling

He threw the chicken down to his elder sibling.

H. Avoidance speech style Semelai has an avoidance speech style, cɒkɒp b-sener (speak MID-tease.by.allusion), a system of word substitution, utilising the normal phonology and grammar of Semelai to avoid the utterance of tabooed words, see examples (15) and (16) below. Uttering tabooed words would result in the violation of the prohibition ma=pnɔn 'prohibitions relating to the consumption of food, and the uttering of tabooed words'.

Neglecting or being unable to fulfil one's desire is also ma=pnɔn, as a speaker explains below:

(14)

məhn	da?	da?	jalu,	hayam,	da?	sot	ma=cəl,
if	NEG	EXIST	pig	chicken	NEG	permit	IRR=utter

ma=pnɔn,	yɒ?=hn	?ɒsɒl	da?,	da?	da?	hal.	kna?
IRR=taboo	but=CONN	when	EXIST	NEG	EXIST	problem	incur

skɔ?.	pɲakit,	?ɲih	dɔ	ha?	ke,	kna?	b-blyan
malevolent.spirit	illness	be.ill	OF	AT	there	incur	shamanic.ritual

If there is no pork or chicken, one shouldn't utter (the name). It is taboo. However, when there is, it doesn't matter. (When we don't have any, and yet we utter the name) (we) will be afflicted by malevolent spirits. The illness, being sick from them, is cured by a shamanic healing ritual.

When the Semelai enter the jungle to hunt, collect forest products, or prepare a swidden, it is imperative to employ this speech style. The jungle is seen as fraught with peril, so in order to avoid the danger of attack from skɔ?, 'malevolent spirits', this taboo is used. Failure to do so can result in a range of afflictions including soul-loss. Other consequences are falling victim to a tiger, crocodile, snake or centipede.

The avoidance style also exhibits variation connected to particular locations. In locations of peril, especially those connected to significant events in the creation of Tasek Bera, terms specific to the area are used. In the area of the Tembangau River, one must use cɒkɒp tʰmaŋə (speak Tembangau), where twar 'fish trap' is rhnoh, the fish species tumɒk is called jalay and ?us 'fire' is pnraŋ, not sma? gdo (person be.old).

Some speakers also extend the use of the avoidance speech style to the home, to prevent the intrusion of snakes or centipedes.

Although avoidance speech styles of this type are common amongst Aslian groups (see Diffloth 1980, and Benjamin in prep.), speech styles of this type are not exclusive to Aslian languages. The Austronesian-speaking collectors of the camphorwood found in Sumatra and Borneo and the southern part of the Malay Peninsula also used styles of this type. Skeat and Blagden (1906) dedicated a whole chapter to this particular linguistic feature.

Skeat and Blagden noted the significance of such speech styles: "The root-idea in all these languages is simple enough: it is merely the avoidance, in an indeterminate number of cases of the ordinary everyday word, and the substitution of something different and out of the common. The primary motive is ... a respectful fear of the superior powers, human, natural or supernatural, as the case may be ..." (1906: 417).

While many Aslian languages have avoidance-type speech styles (Semai (Diffloth 1980); Cheq Wong and Jah Hut (Kruspe fieldnotes)), they are usually confined to the lexical replacement of nouns as in (15). Semelai is unusual in extending this to include verbs (G. Benjamin, N. Burenhult, G. Diffloth, p.c.), see (16).

(15) The item replaced is ʔus 'fire':

hɛ=pteʔ	ʔen	[smaʔ	gdo]
1&2A=put	LOC	[person	be.old]

We put (it) in the fire.

(16) The item replaced is kʰbəs 'to die':

smaʔ	clɒy
person	die

Someone has died.

The semantic domains of nouns covered by this taboo, along with some examples, are shown in Table 1.1. Some examples of verbs, organised into semantic domains, are given in Table 1.2.

TABLE 1.1 THE SEMANTIC DOMAINS OF NOUNS IN THE AVOIDANCE SPEECH STYLE

SEMANTIC DOMAIN	REGULAR TERM	AVOIDANCE TERM
Animals	cɒŋ 'Lesser Mouse Deer'	mə=kurus jɔŋ (REL=COMP-be.thin foot) 'the one who is fine of hoof'
	tijɔ 'snake'	mə=rus kbəʔ (REL=drag body) 'the one who drags himself'
Forest produce	drɛ 'rattan'	ʔurat bri (sinew jungle) 'jungle sinews'
Consumables	ʔiɟɤ̃y 'millet'	pɒs pɔdɔŋ (tail tiger) 'tiger's tail'
	kuɲit 'turmeric'	jareʔ kaʔip (digit centipede) 'centipede's legs'
Artefacts	praho? 'dugout canoe'	dɒʔ 'large cooking vessel'
Biosphere	ʔareʔ 'rain'	mə=ramay (REL=be.many) 'the one who is many'

TABLE 1.2 THE SEMANTIC DOMAINS OF VERBS IN THE AVOIDANCE SPEECH STYLE

SEMANTIC DOMAIN	REGULAR TERM	AVOIDANCE TERM
Bodily activities or processes	cʔəc 'to shit'	wunwen
	ca 'to eat'	chɔr 'peck at'
Food preparation	cʔɔŋ 'to roast'	kwel
Cultural activities	yɒr leŋ dʔɔh (ascend TO:up swidden) 'to go to (one's) swidden'	yɒr leŋ lubuk (ascend TO:up pool) 'to go up to (one's) pool'
	cʰɔr 'fire (a swidden)'	k<l>wel (roast.(av.)<IMPERF>)
Dimension	tʰəy 'be big'	k<ʔ>bəʔ (body<HAVE>)
Colour	hitam 'be dark, black'	blhũt

One lexeme can have multiple avoidance lexemes, e.g. jalu 'pig' is dil muh (be.blunt nose) 'the nose is blunt', mə=tʰəm (REL=pound) 'the one who pounds (the earth)' and at Bukit Rok is called mə=ptidah (meaning unknown). There is also some variation in the definition of terms, e.g. grphɔp is defined by some speakers as 'hot food', others as 'to eat'.

Lexical items are drawn from the following sources:

- a lexical item which is given a new sense: dɒʔ 'cooking vessels, crockery' for prahoʔ 'dugout canoe';
- words exclusive to the register. These lexemes may have cognates in other Aslian languages: hũm 'to bathe' is mhmɛh which is possibly from Semoq Beri hmɛh 'to bathe'; kʰbəs 'to die' is clɒy, which is possibly related to Cheq Wong clɑy 'to obstruct a path'(Kruspe fieldnotes);
- a lexical item with aberrant morphology: g<r>p>hɔp (be.hot, plus infixes <r> and <p>) 'cooked food';
- Malay loans with Semelai morphology: kmreŋ (be.dry plus infix <m> 'NMZ') 'water' ← *kering* 'dry';
- a lexicalised associative phrase: kapɔh smut (egg ant) 'uncooked rice';
- a noun and an adjective: dlɔŋ raʔ-kẽt (wood COMP-be.small) '*gaharu* wood';
- a verb phrase: reŋ dlɔŋ (seek wood) 'to seek *gaharu* wood', ko=glaŋ ʔluc (3UA=swallow pass) 'to bathe'; gɒɲ daʔ br-gɒɲ (bite NEG MID-bite) 'chilli';
- a lexicalised relative clause: mə=kales plantoŋ (REL=wipe.anus chopping.board) 'the one who wiped his anus (on the) chopping board)' is kɔʔ 'Pig-tailed macaque (*Macaca nemestrina*)'.

Some terms are semantically transparent, e.g. byukuʔ 'Giant Malayan Terrapin (*Orlitia borneensis*)' is mə=ʔaɲ dɔl (REL=carry.on.back house) 'the one who carries his house on his back', whilst others require recourse to cultural knowledge. The term for the tawɔ 'White-handed gibbon (*Hylobates lar*)', whose flesh is the highest-prized of land animals, is called jlawat darat (fish sp. land) 'the jlawat of the land', the jlawat being the highest-prized of fish species.

Avoidance terms are generally based on the following:

- physical appearance: ʔurat bri (sinew jungle) for drɛ 'rattan';
- characteristic behaviour: mə=rus kbəʔ (REL=drag body) 'the one who drags himself' for snake;

- association with a traditional narrative, e.g. a lizard species, the tkeʔ, gets his name kapur kloc diŋ (lime inside bamboo), because while hiding inside bamboo he ate lime and betel nut without betel leaf and thus got his coughing-like call.

Many terms are semantically opaque to the speakers. The avoidance term for hubiʔ 'cassava' (from Malay *ubi* 'edible tuber') is ʔi=cɔm (Mr=dig up) 'Mr dig (him) up'. The Semelai lexical item for 'dig up' is bɒy, however in Mah Meri, a related Southern language, there is a cognate co̤p 'to dig', and in the Central language Jah Hut it is cwɔm.

It is possible that the presence of a system of lexeme replacement has resulted in the replacement of indigenous words by Malay loans, as with the word for cassava, above. In some instances, it would appear that avoidance terms have replaced indigenous terms in the ordinary lexicon, the lexical item for python, dageŋ jləŋ (meat be.long) has the form of an avoidance term.

1.2 The Aslian languages

Aslian languages are a sub-group of the Mon-Khmer language family. Along with Munda, Mon-Khmer is one of the two divisions that forms the Austroasiatic phylum. Austroasiatic languages are spread throughout mainland Southeast Asia, Northeast India and the Nicobar Islands.

The Austroasiatic languages have a long presence in this area, predating the Sino-Tibetan, Tai-Kadai, Miao-Yao, Austronesian, Dravidian and Indo-European families with which they coexist today. With the exception of three Mon-Khmer languages, Vietnamese, Khmer and Khasi (Assam, India), the languages are spoken by ethnic minorities.

The Mon-Khmer language family extends from Vietnam in the east, to Northeast India in the west, and from Southern China in the north, to the Aslian languages in the south, Semelai and Mah Meri (Besisi) being the southernmost members of the family. Mon-Khmer is divided into twelve branches (Diffloth and Zide 1992), although the precise relationship between the twelve is yet to be ascertained. As stated above, the Aslian branch lies firmly within the Mon-Khmer division, and is not a distinct division within Austroasiatic (Ruhlen 1986: 89-91). Aslian belongs in a Southern Branch along with the Monic languages, and possibly also those of the Nicobar islands (Diffloth p.c.).

The question of internal subgroupings of Aslian is discussed in the following section.

1.2.1 Internal subgroupings of the Aslian languages

The Aslian languages are recognised as falling into three distinct subgroups: Northern, Central and Southern Aslian (Benjamin 1976a), or Jehaic, Senoic and Semelaic (Diffloth 1976b), see Map 2. Note that both groupings are the same, the difference is only in the labels. The first set is based on predominant geographic location within the Malay Peninsula, the second on either the major language of that group (Jehai, Semelai), or 'racial' group (Senoic). The actual number of Aslian languages is yet to be determined. In Figure 1.1 nineteen languages are represented following Benjamin