LUND
UNIVERSITY

Northern Aslian Linguistic Prehistory

Tracing the Effects of Contact

Joanne Yager
joanneyager@gmail.com

Supervisor:
Niclas Burenhult

Abstract

This study focuses on the genealogical and contact-based connections among the Northern Aslian varieties, which form one of three subgroups of the Aslian subbranch of the Austroasiatic stock, spoken in Peninsular Malaysia and southern Thailand. The complex patterns of contact that exist among the speakers of the closely related Northern Aslian varieties coupled with a scarcity of data and a lack of written history give rise to difficulties in the representation of the historical relationships among these varieties using traditional models of language classification. In particular, the study focuses on the variety of Menriq spoken in the resettlement area of Sungai Rual, Kelantan, Malaysia. The lexical and phonological features of this hitherto undescribed variety as well as its unexpected geographical location suggest a complex and ambiguous history. Analysis of the phonological and lexical aspects of this variety using a range of different methods suggests that its speakers may have undergone periods of both isolation from and intense contact with speakers of the other Northern Aslian varieties of the study. Furthermore, a large degree of contact among the majority of the Northern Aslian varieties is evident in the analyses, and patterns suggest the past existence of a dialect continuum stretching from central Peninsular Malaysia into southern Thailand. More recently, it would appear that this continuum has been split in two by the arrival of Jahai-speaking groups in the midst of the formerly contiguous Menriq- and Kensiw/Kintaq-speaking groups, resulting in lower levels of contact involving the varieties spoken to the north of this split. The findings suggest the importance of further study of the Menriq Rual variety and of the genealogical and contact-based patterns among the Northern Aslian varieties, as well as the urgency of documentation endeavours in this part of the world.

Table of contents

1. Introduction

1.1 Trees and waves: Traditional historical linguistic methods and the effects of contact

Historical relationships between languages are traditionally represented according to one of two models – the family tree model, representing relatively clear-cut boundaries between speaker populations, or the wave model, representing relationships resulting from expansions of speaker populations over a continuous territory (Heggarty, Maquire & McMahon, 2010). However, it is often the case that both of these processes operate within single language families, and calls have been made for a model that can account for both processes. Such a model would take into account tree-like *and* wave-like genealogical aspects of language histories as well as those aspects that have come about as a result of contact between speaker populations (Heggarty et al., 2010).

Traditional methods of language classification most often rely on the identification and exclusion of loans from analysis. In order to be considered a reliable basis for determining the genealogy of languages, the vocabulary compared must consist of actual cognates, not loans. Loans must therefore be identified as such and excluded (Campbell, 2004:126-127). However, it is well known that some degree of borrowing is evident in most languages. Situations where languages are or have been in contact with one another are widespread, if not universal (McMahon et al., 2005). Thus while the effects of contact are a vital piece of the puzzle of language histories, they also present an obstacle to traditional methods of language classification. In particular, in cases where loans are difficult to detect such as in the case of contact between closely related languages, language histories can be difficult to analyse using traditional historical linguistic methods.

The comparative method has long been the most established, and for some, the only acceptable, method of language classification. However, while classification using this technique has been successful in cases of language families with long written histories, the application of the method is much more difficult in the case of language families lacking long written histories (McMahon & McMahon, 2003). Detailed understandings of patterns of regular sound change and reliable proto-language reconstructions are required in order to best apply the comparative method. In the case of little-studied languages or extinct languages with little recorded evidence, this lack of data creates difficulties in the application of the method. Furthermore, in cases where languages have undergone situations of intense contact, especially if this contact occurs among closely related languages, loans can be almost impossible to detect, and can therefore not be excluded. In some situations all of these factors combine to make the application of the comparative method extremely problematic.

In applications of the comparative method, and even more so in lexicostatistics, classifications usually begin with the analysis of 'basic vocabulary', such as body parts and close kinship terms, since this section of the vocabulary is thought to be less prone to borrowing (Campbell, 2004). However, even basic vocabulary has been shown to be prone to borrowing to an extent: of Swadesh's (1952) 200-item list, designed to include only the most basic of vocabulary, English has twelve loans from French, and similar loan rates have been found for other Indo-European languages (McMahon & McMahon, 2003). Furthermore, in cases of borrowing among related languages, basic

vocabulary is more likely to be borrowed, and can thus not be assumed to be a reliable basis for historical linguistic analyses (McMahon & McMahon, 2003).

As a result, in such situations a method is needed that can separate out the effects of contact from genealogical patterns. Furthermore, since contact influences often play an important role in language histories, a model is needed that can take into account both genealogy and contact. It has been proposed that network-based models (as opposed to tree-based models) have the ability to represent both genealogical and contact-based aspects of the relationships between languages (Heggarty et al., 2010). Furthermore, the method of comparing genealogies constructed on the basis of more and less retentive subsets of vocabulary has been proposed to differentiate between the effects of genealogy and contact (McMahon & McMahon, 2003, McMahon et al., 2005). Thus rather than seeing the influence of contact as an obstacle to language classification, these influences can be used in order to give an indication of the aspects of the relationships between languages that are likely to be due to genealogy and those that are likely to have been caused by contact. This method is based on the idea that vocabulary changes and is replaced at different rates, and that certain parts of the vocabulary are more readily borrowed than others. As mentioned above, this assumption is present in the comparative method as well as in lexicostatistics, where more basic vocabulary is assumed to be a more reliable basis for analysis of the true genealogy of languages. However, the idea that different sections of the vocabulary change at different rates has mainly been used in endeavours to exclude less retentive items, rather than being used to enable representation of different aspects of language histories.

Combined, these two methods allow investigation of the patterns in lexical data as to (a) the extent of borrowing present in the data set, and (b) the connections in the data set that are grounded in genealogy and those that are more likely to be the result of contact. This kind of analysis may allow us to see traces of borrowing which are otherwise difficult to detect, giving us a more complete insight into the histories of languages.

1.2 The Northern Aslian Varieties and the Elusive Menriq Rual

In attempts to classify the Northern Aslian language varieties, the genealogy and history of these varieties has proven difficult to decipher. Speakers of these closely related varieties are involved in complex patterns of contact with one another, making the effects of this contact difficult to exclude from analysis. Research into the Aslian languages is only relatively recent, and a reconstruction of proto-Aslian is not yet available. This coupled with a lack of written history as well as a relative lack of data for many varieties has meant that a reliable classification of the Northern Aslian varieties has yet to be achieved. In the absence of reliable classifications the more neutral term 'variety' is thus used throughout the present study when referring to the linguistic entities of Northern Aslian (rather than 'language' or 'dialect' – except where referring specifically to discussions of distinct languages) in order to avoid assumptions regarding the relationships between these varieties.

The Northern Aslian varieties form one of three subgroups of the Aslian languages which are spoken in Peninsular Malaysia and southern Thailand. The Aslian language family forms part of the southern subgroup of the Mon-Khmer branch of the Austroasiatic stock. The majority of the speakers of the Northern Aslian varieties are

semi-nomadic foragers known ethnographically as the Semang, and most groups are engaged in a high degree of intra-Aslian contact (excluding a small number of groups, see discussion in 1.4 below). This pattern has been described as a 'continuous mesh of communication' (Benjamin, 2009:10) arising out of the nomadic lifestyle of the Semang, their tendency to live in small groups, and the prevalence of intermarriage between members of different language groups. The Semang traditionally live in small groups of less than 50 people, moving from one place to another after a few weeks or months, or sometimes years. It is now also common for the Semang to live in resettlement villages set up by the government (Burenhult, Kruspe & Dunn, 2011).

The linguistic variation among the Northern Aslian varieties has been described as 'as much idiolectal as dialectal' (Benjamin, 2009:20), here, variation and change are the norm. Benjamin (2009:20-21) reports claims by Northern Aslian speakers that they have 'consciously changed their way of speaking during their lives, depending on whom they married and where they wandered to'. The speakers of the Northern Aslian varieties are accustomed to high degrees of movement, where groups repeatedly disintegrate and reform in different places. Thus the speakers of these varieties are accustomed to high degrees of linguistic non-uniformity – there are reports of villages of less than 30 people where no less than six different linguistic varieties are spoken (Bishop & Peterson, 1993).

These reports reveal complex patterns of human movement as well as of linguistic contact, variation and change among the speakers of the Northern Aslian varieties. These complex patterns mean that the boundaries between idiolect, dialect and language are blurred, making this a difficult case for classification. The aim of the current study is to investigate the contact and genealogical patterns among the Northern Aslian varieties in order to add to and inform the process of the classification of these varieties. The current study focuses in particular on the Menriq Rual variety, a hitherto undescribed variety of Northern Aslian. On the basis of phonological as well as lexical analyses, and with the use of a network-based model and comparisons of more and less retentive subsets of vocabulary, the present study aims to shed light on the status and position of Menriq Rual as well as contribute to current understandings of the genealogical and contact-based connections among the Northern Aslian varieties.

1.3 The Menriq Rual Variety
The first and only known recordings of the variety of Menriq spoken in Sungai Rual were made in connection with the DoBeS-sponsored 'Tongues of the Semang' survey of the language varieties of the Semang, and consist of a single wordlist recording. The Menriq at Sungai Rual number approximately 150 and inhabit one of three permanent villages, the other two of which are primarily inhabited by Jahai, who number approximately 350. Intermarriage is common among members of the two groups, and the Menriq at Sungai Rual also have close contact with the Jahai in Perak as well as Menriq groups residing in other areas (Burenhult, n.d.).

While the speakers of this variety simply call themselves 'Menriq' (or, distinctively, as pronounced by one language consultant, 'Menrik', Burenhult, n.d.), the survey revealed several aspects of this variety that warranted closer attention. In terms of phonology the distinctive realisation of /r/ of this variety was found to diverge markedly from other recorded varieties of Menriq. In terms of lexicon the variety

appeared to diverge from other Menriq varieties, notably in its form for the first person singular (/ʔiɲ/) which is rare among the Northern Aslian varieties. Vernacular accounts of the origins of the Menriq at Rual also proved puzzling. While these Menriq themselves identify with the Menriq in other areas, to other Menriq as well as to the Jahai in other areas they are known as 'Jdek'. Asked about this term the Menriq at Rual deny that the term can be used to describe them, and suggest that the term rather applies to people in Thailand (Burenhult, n.d.). It has been suggested that this reference to a group named 'Jdek' may have some connection with the Northern Aslian variety of southern Thailand referred to in Phaiboon (2006) as Tea-De (Niclas Burenhult, personal communication). When asked about the distinctive features of their variety the Menriq at Rual suggest that they speak a conservative variety: due to an origin in places like Sungai Taku in Kuala Krai they 'speak in a way which they have forgotten elsewhere' (Burenhult, n.d.). Other reports, however, deny a connection of the Menriq at Rual with the Kuala Krai area (Burenhult, n.d.).

The geographical location of the speakers of Menriq Rual is another intriguing aspect of this variety. According to Benjamin's (1976) proposed history of the movements of the Northern Aslian speakers, the area now occupied by Jahai-speaking groups was once occupied by Menriq, who moved south to the Lah area where they are still found today following the (relatively recent) arrival of the Jahai. The existence of 'Menriq' at Sungai Rual, removed from the remaining Menriq groups and surrounded by Jahai, suggests the possibility that this group is a remnant of the Menriq who were present in the area before the arrival of the Jahai. Furthermore, Benjamin (1976) proposes the earlier existence of a dialect continuum stretching from Kensiw/Kintaq in the north to the Batek varieties in the south, including the Menriq in between. Thus the Menriq Rual variety may be the remnant of one of the varieties of this dialect continuum, closely related to, but not necessarily synonymous with, the Menriq spoken in other areas.

The findings of the first study to compare the Menriq Rual variety with other Northern Aslian varieties (Dunn et al., 2011) suggested that the lexicon of Menriq Rual diverges substantially from other recorded varieties of Menriq. The analyses also showed a high degree of uncertainty regarding the placement of the variety among the Northern Aslian varieties. Bayesian phylogenetic inference analyses showed low probability for the subgrouping of the variety, as well as conflicting results regarding its position – Menriq Rual is shown alternatively as an outlier of the Menraq group (containing Jahai and Menriq), or as an outlier of the much wider Menraq-Batek group (containing not only Jahai and Menriq but also the numerous Batek varieties). Neighbor-Net analysis placed the variety as an outlier of the Menraq group, but showed a large degree of conflict in the connection. For this reason it is not assumed in the current study that the Menriq spoken at Rual should be treated as a variety of Menriq, and this variety is thus treated separately from the other three Menriq varieties included in the study. In the present study, as in Dunn et al. (2001), this variety is referred to as Menriq Rual after Sungai Rual, the place where it is spoken.

This variety, with its distinctive lexical and phonological features as well as its unexpected geographical location and ambiguous connections to the other Northern Aslian varieties, is believed to warrant closer attention. Study of the Menriq Rual variety has the potential to shed light on the historical movements of Northern Aslian speakers and their past and present patterns of contact. The aim of the current study is

to take a closer look at the phonological and lexical aspects of this variety, in order to investigate the genealogical and contact-based aspects of its history. While the current study is based only on the limited recordings available to date, this closer look at the data is important in order to determine whether further study of the variety is to be prioritised.

1.4 The other varieties included in the study
Wordlist data from a total of fifteen Northern Aslian linguistic varieties was included in the current study. The data for thirteen of these varieties was collected in connection with the DoBeS 'Tongues of the Semang' survey, and retrieved for use in the present study from the DoBeS archive. The data for the two additional varieties was collected by Ewelina Wnuk (for Maniq) and by Phaiboon Duangchan (for Tea-De; Phaiboon, 2006). The sources of the data included in the present study are given in Table 1 (adapted from Table 1 of Dunn et al., 2011; the two additional varieties are listed last).

Table 1. The sources of the data

Variety	Data collected by	Location
Kensiw Perak	Burenhult 2005	Sungai Lebey, Hulu Perak, Perak, Malaysia (speaker/s from: Betong, Yala, Thailand)
Kensiw Kedah	Burenhult 2005	Bukit Asu, Hulu Perak, Perak, Malaysia (speaker/s from: Lubok Legong, Baling, Kedah, Malaysia)
Kintaq	Burenhult 2005	Bukit Asu, Hulu Perak, Perak, Malaysia (speaker/s from: Lubok Legong, Baling, Kedah, Malaysia)
Jahai Banun	Burenhult 1998-2008	Sungai Banun, Hulu Perak, Perak, Malaysia (speaker/s from: Sungai Mangga, Hulu Perak, Perak)
Jahai Rual	Burenhult 2000-2006	Sungai Banun, Hulu Perak, Perak, Malaysia (speaker/s from: Sungai Rual, Jeli, Kelantan)
Menriq Rual	Burenhult 2005	Sungai Rual, Jeli, Kelantan, Malaysia
Menriq Lah	Burenhult 2006, 2008	Kuala Lah, Gua Musang, Kelantan, Malaysia
Batek Teh Taku	Burenhult 2006	Kuala Krai, Kelantan, Malaysia
Batek Teh Lebir	Burenhult 2006	Pos Lebir, Kuala Krai, Kelantan, Malaysia
Batek Deq Koh	Burenhult 2006	Kuala Koh, Gua Musang, Kelantan, Malaysia
Batek Deq Terengganu	Kruspe 2001, 2008	Sungai Berua, Hulu Terengganu, Terengganu, Malaysia (speaker/s from: Kuala Koh, Gua Musang, Kelantan)
Batek Teq	Kruspe 2008	Sungai Berua, Hulu Terengganu, Terengganu, Malaysia (speaker/s from: Kampong Sayap, Besut, Terengganu)
Ceq Wong	Kruspe 2002-2006	Kuala Gandah, Temerloh, Pahang, Malaysia
Maniq	Wnuk 2009-2011	Khao Banthad mountain range, Satun and Phattalung provinces
Tea-De	Phaiboon (Phaiboon, 2006)	Weang and Srisakorn Districts, Narathiwat Province, Thailand

These fifteen varieties can be divided into nine distinct Northern Aslian 'languages', spanning three Northern Aslian subgroups, as in 1.4.1 through 1.4.4 below. Note that the division into distinct languages followed here is based primarily on the categories recognised in Malaysian administrative practice, but that these are also thought to correspond relatively well with the linguistic situation. After Dunn et al. (2011), the

individual varieties are named in accordance with the ethnonyms used by the speakers of each variety to refer to their own group in combination with the place of residence of the speakers, while the names of the subgroups are based on the word for 'person' that is distinctive for each of the respective subgroups.

1.4.1 Varieties of the Menraq group

Menriq

Three varieties of Menriq were included in the study: Menriq Lah, Batek Teh Lebir and Batek Teh Taku. Note that while the speakers of the second and third of these varieties call themselves Batek Teh, after Benjamin (1976) and Dunn et al. (2011) the varieties they speak are considered to be varieties of Menriq. The variety termed Menriq Lah is spoken by the approximately 150 inhabitants of the long-term Menriq-only village Kuala Lah. The Menriq at Lah have close contact with the inhabitants of Sungai Rual as well as with the Batek Teh (Burenhult, n.d.).

The variety termed Batek Teh Lebir is spoken by the approximately 50 inhabitants of one of two permanent villages at Pos Lebir. The Batek Teh at Pos Lebir have close contacts with the approximately 300 Batek Deq who inhabit the other of the two villages, as well as other Batek Deq, the Menriq at Kuala Lah and the Batek Teh at Sungai Taku. The variety termed Batek Teh Taku is spoken by the dozen or so inhabitants of the small and inaccessible hamlet Sungai Taku. The village has existed for at least 50 years but its population is decreasing with movements of the inhabitants to other places such as Pos Lebir and Sungai Rual (Burenhult, n.d.). The Batek Teh at Sungai Taku have close contacts with the Menriq at Lah and the Batek Teh at Pos Lebir.

Jahai

The two varieties of Jahai included in the present study were found in Dunn et al. (2011) to share a high rate of lexical similarity, supporting speaker claims that 'all Jahai speak the same way' (Burenhult, n.d.). The Jahai Rual variety was recorded in Sungai Rual, where its speakers number approximately 350, and live in close contact with the 150 Menriq of Sungai Rual (see section 1.3 above). The other Jahai variety of the current study, Jahai Banun, was recorded in Sungai Banun, a village primarily inhabited by Jahai. The Jahai have extensive contact with the Temiar, a Central Aslian speaking group who inhabit the area to the south of Jahai territory.

1.4.2 Varieties of the Batek group

Several Batek varieties have been identified in the literature. Benjamin (1976) includes 'Bateg Deq' and 'Bateg Nong', also identifying a variety known as 'Mintil' to be a variety of Batek. The present study includes data from three Batek varieties – the varieties of Batek Deq spoken in Kuala Koh in Kelantan and in Sungai Berua, Terengganu, as well as the Batek Teq variety. The Batek are in total estimated to number roughly 1500 (Benjamin, 2009). The speakers of the Batek Deq Koh variety are the approximately 50 inhabitants of the semi-sedentary camp Kuala Koh, however consultants from this group report that there are several other villages where a variety 'identical' to their own is spoken (Burenhult, n.d.). The Batek Deq Terengganu variety was recorded at Sungai Berua in Terengganu. The speakers of this variety originate in Kelantan, and are known to have contact with speakers of Batek Igaq, Batek Teq and Batek Teh (Nicole Kruspe, personal communication). Batek Teq is a moribund variety spoken by a small number of families in northern Terengganu. The Batek Deq at Kuala Koh consider the Batek Teq variety 'different', but intelligible (Burenhult, n.d.),

however other reports suggest that these varieties are not mutually intelligible (Nicole Kruspe, personal communication). The Batek Deq consultants at Kuala Koh do not distinguish between Menriq and Jahai groups, suggesting that these Batek Deq are not involved in contact with groups of Menriq and Jahai.

1.4.3 Varieties of the Maniq group
Kensiw/Kintaq

Two recordings of the Kensiw varieties are included in the present study, as well as one recording of the Kintaq variety. The Kensiw speakers are thought to number 240 in total, and the Kintaq number 132 (Benjamin, 2009). The two Kensiw varieties are those of Hulu Perak and of Baling in Kedah, and the Kintaq variety was also recorded in Kedah. The groups appear to have extensive contact with one another. While there are reports of Jahai living in Kensiw or Kintaq villages, and vice versa (Burenhult, n.d.), the extent of contact between these two groups is not known. There is no known present-day contact between the Kensiw or Kintaq and the Menriq or Batek, neither do the Kensiw or Kintaq groups appear to be in contact with the Maniq or speakers of other closely related Northern Aslian varieties of Thailand.

Maniq

The Maniq data used in the current study collected by Ewelina Wnuk was not included in Dunn et al.'s (2011) analysis, however this variety is synonymous with Dunn et al.'s Ten'en variety, and is thus considered, along with the Kensiw and Kintaq varieties, to belong to the Maniq subgroup. The Maniq are an isolated group of Orang Asli in southern Thailand, who have no known contact with other Northern Aslian groups of Malaysia or southern Thailand (Ewelina Wnuk, personal communication; Bishop & Peterson, 1993).

1.4.4 Other varieties
Ceq Wong

The Ceq Wong inhabit the southern foothills of Gunung Benum in Central Peninsular Malaysia. They are thought to number approximately 400 (Nicole Kruspe, personal communication), have no known contact with other groups of Northern Aslian speakers, and do not know of other Aslian groups apart from the neighbouring Jah Hut and Temuan (Kruspe, 2009). The data used in the current study was recorded at Kuala Gandah in Temerloh, Pahang.

Tea-De

The Tea-De data used in the current study was that published in Phaiboon (2006). While this data did not include all of the items of the wordlist that formed the basis of the data set for the present study, items overlapping in the two lists were relatively numerous and thus a preliminary comparison could be made. Tea-De is one of the four Northern Aslian varieties identified by Phaiboon in southern Thailand, and is spoken by 'nomadic, foraging people' in the Weand and Srisakorn districts of the Narathiwat province (Phaiboon, 2006:207). The combined population of the Northern Aslian speaking groups of southern Thailand has been estimated to be roughly 200. Very little is known about the sociolinguistic situation of these groups, however it is believed that they are not involved in contact with the Northern Aslian speaking groups of Malaysia (Bishop & Peterson, 1993).

The geographical distribution of the Northern Aslian varieties is shown in Figure 1 (from Dunn et al., 2011:293). The Central and Southern Aslian languages are also

shown. The Maniq variety of the present study is represented by the area labelled
'Ten'en' on this map, and the group speaking the Menriq Rual variety is found close to
the Thai border, within the area marked Jahai.

Figure 1. The Geographical Distribution of the Aslian languages
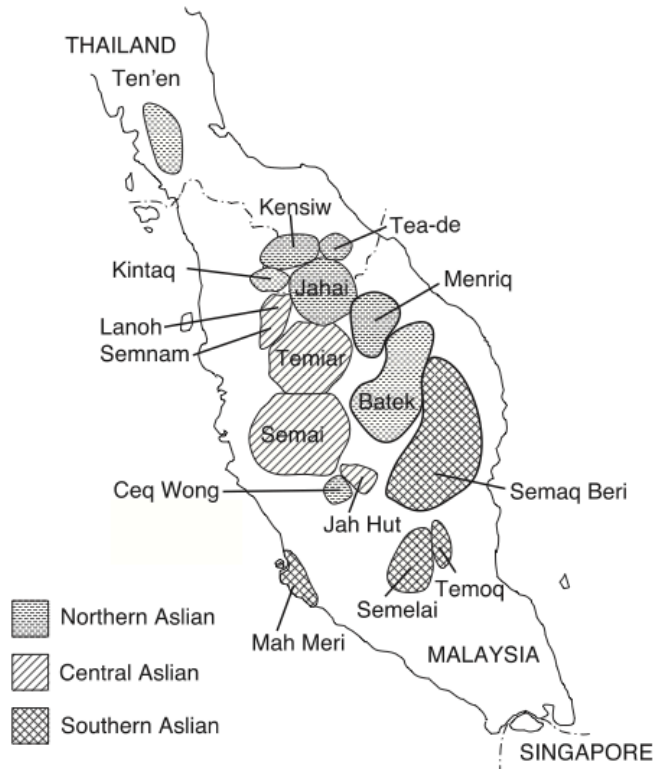
1.5 Classification of the Northern Aslian varieties
Attempts to classify the Aslian languages have been made by Benjamin (1976) using
lexicostatistical methods, by Diffloth (1975) using the comparative method, and most
recently, by Dunn et al. (2011) using distance-based phylogenetic algorithms. All three
studies point to a difficulty in the subgrouping of the Northern Aslian varieties. Ceq
Wong is consistently shown as an outlier to the Northern Aslian group, however the
subgrouping of the remaining Northern Aslian varieties differs somewhat among the
studies. While Diffloth (1975) proposes a primary split between Batek on the one hand
and Kensiw-Kintaq-Jahai-Menriq on the other, followed by a secondary split
separating Kensiw-Kintaq and Jahai-Menriq, Benjamin (1976) proposes three major
Northern Aslian subgroups: a western subgroup containing Kensiw and Kintaq (which
are classified as dialects of a single language), an eastern subgroup containing Batek
Deq, Batek Nong and Mintil (the second and third of which are not included in the
present study), and a third subgroup containing Jahai and Menriq.

While Benjamin's (1976) lexicostatistical analysis suggested that the Central and
Southern Aslian subgroups formed tree diagrams in a relatively straightforward
manner, the Northern Aslian varieties proved 'more recalcitrant to subgrouping'
(Benjamin 1976:60), the Jahai and Menriq varieties proving particularly difficult to
place. He suggests that the high levels of borrowing among the Northern Aslian

varieties mean that the connections between them are more accurately represented by a meshwork-like relationship than by a traditional tree diagram. The nomadic lifestyle of the majority of the Northern Aslian groups also caused difficulty in determining past patterns of migration and speaker group contact, as speaker movements have meant that the present-day locations of closely related varieties are not necessarily contiguous, and traces of influences from past contact may be found in varieties now spoken in areas that are distant from one another.

In contrast to Benjamin's (1976) and Diffloth's (1975) proposed subgrouping of the Northern Aslian varieties, Dunn et al.'s (2011) findings suggest a primary split between Kensiw, Kintaq, Maniq and Ten'en (a subgroup they term Maniq) on the one hand, and Batek, Jahai and Menriq (a subgroup they term Menraq-Batek) on the other. However, the analysis points to considerable uncertainty when it comes to the subgrouping within the Menraq-Batek group. The authors suggest that this uncertainty may be the result of high degrees of contact among the speakers of these varieties. This would also explain the short branch lengths that were found within this group of varieties, which it is proposed may be due to a slower rate of linguistic change resulting from this high degree of contact. The authors propose that the relationships between these varieties may be more accurately represented as a dialect continuum than as a tree, echoing the suggestion made by Benjamin (1976).

The situation of the Northern Aslian varieties is thus one where a lack of written history and a shortage of data for several varieties, coupled with a high degree of both present-day and past contact among these closely related varieties, leads to difficulties in classification with traditional methods. Moreover, it has been suggested that the relationships between the Northern Aslian varieties are difficult to represent using the tree model of linguistic phylogeny. The aim of the current study is to offer new insights into these connections, and in particular into the place of the Menriq Rual variety among them. It is believed that the case of the Northern Aslian varieties is one that is well suited for analysis using network-based models combined with the method of comparing more and less retentive subsets of vocabulary.

1.6 Previous Studies Using the Proposed Method
With the aim of testing the idea that differences in the rate of replacement of vocabulary items have an impact on the resulting phylogenetic representation of the relationships between languages, McMahon & McMahon (2003) compared trees based on subsets of more and less conservative vocabulary. The vocabulary subsets were chosen based on Lohr's (1999, as cited in McMahon & McMahon, 2003) research into the rates of replacement of different vocabulary items. Lohr tested the rates of replacement of a large number of meanings by comparing their reconstructability for the proto-languages for four different language families as well as their retentiveness, or the rates of replacement of forms for these meanings over time within Indo-European. Those meanings which could be reconstructed for all four proto-languages and which had the lowest rates of replacement for Indo-European were considered the most basic and stable. McMahon & McMahon used these findings to select two subsets of Dyen, Kruskal and Black's (1992) widely-used 200-item version of the Swadesh list. The thirty most basic and stable items were chosen to form what they termed the 'hihi' list: those items of the list scoring highest on reconstructability and retentiveness. The 23 least basic and stable items formed their 'lolo' list: the items scoring lowest on reconstructability and retentiveness.

The authors constructed phylogenetic trees for Indo-European based on their two sublists and found that the trees differed in a way which suggested that what was seen were traces of borrowing in the history of the languages. They found that the trees did not differ greatly where no borrowing had occurred in the data, but where borrowing had occurred, the position of the borrowing languages in the tree based on the less basic vocabulary gave an indication of the source of the borrowing. One example of this was the position of Rumanian within the Romance language group. On the basis of the lolo list Rumanian was shown to be marginal to the Romance group, whereas on the basis of the hihi list it was much more integrated within the group, forming a subgroup with Ladin and Sardinian. Other differences between the trees involved the position of English and that of Frisian. In all of these cases, languages which are known to have had a large degree of contact with some other language/s were found to move closer to the source of this contact in the trees drawn on the basis of the less basic vocabulary.

McMahon et al. (2005) conducted a follow-up study based on these findings, using Neighbor-Net network drawing software (Bryant & Moulton, 2004). Here the method was applied to a historical linguistic question that has proved difficult to answer with the use of other methods: the question of whether the Quechua and Aymara language families of the Andes are genetically related or connected only by contact. The authors found that there was a great difference in the distance between the root nodes of the two families in the networks drawn on the basis of the two sublists. In the network based on the less basic vocabulary this distance was just over 20% while in the network based on the more basic list it was 54.4%. Thus the lexical distance between the two language families was found to be much smaller for the less basic items, suggesting that the similarities between the two families may be traced to the effects of contact rather than to a common genealogical origin.

This contrasted with the position of languages known to be related within the subgroups: these languages were shown to be positioned *further* from the root node of their respective families on the basis of the less basic list. A similar pattern was demonstrated by the authors in the case of the relative distances between Greek and a selection of the Romance languages: the less basic list resulted in a greater distance between these subgroups, not a smaller distance as was the case of the two Andean subgroups. They conclude that in cases of common ancestry the use of less basic vocabulary will result in greater distances between languages since this vocabulary is more prone to change. Thus where languages show the opposite pattern it is likely that contact rather than common ancestry is the explanation for the lexical similarities between the languages. Furthermore, languages which have been affected by contact tend to move in the direction of the source language/s in networks based on less basic vocabulary.

1.7 The Neighbor-Net Method
In recent years a range of different computational techniques for the estimation of phylogenetic relationships has been introduced. The difference between Neighbor-Net and many other methods of phylogenetic analysis is that rather than constructing trees Neighbor-Net constructs *networks*, that is, the method does not force a treelike shape onto the data. If the data contains only relatively straightforward connections it will automatically result in a treelike network, but if the data contains conflicting

13

information about the relationships between taxa this will result in a network with more reticulations that is thus less treelike in shape. Thus rather than simply constructing the most plausible tree such networks have the ability to give a representation of the multiple alternative trees that are feasible on the basis of the data. This feature means that Neighbor-Net has a clear advantage when it comes to representing relationships between languages whose evolutionary histories are not 'treelike' (Bryant & Moulton, 2004).

The way that Neighbor-Net constructs networks differs from that of other network-drawing software. In Neighbor-Net the taxa (here, the taxa are the languages to be compared) are each represented from the beginning by one node. The program then uses a set of weights to select the three nearest nodes and collapse these three into two linked nodes, continuing the process until only two or three nodes remain. This process is illustrated in Figure 2 (adapted from Figure 2 in Bryant & Moulton, 2004:256).
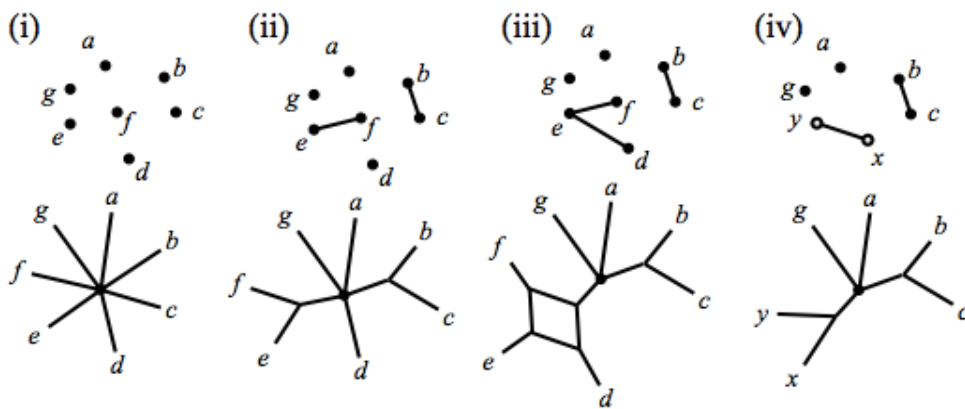
Figure 2: The Neighbor-Net method. (i) In the beginning, each taxon is represented by a single node. (ii) The program uses selection criteria to identify *b* and *c* and *e* and *f* as neighbours, respectively. (iii) *d* is also identified as a neighbour of *e*. (iv) The three nodes *d, e* and *f* are thus replaced by two nodes, *x* and *y*. Thus both of the possible splits *e f* | *abcdg* and *de* | *abcfg* are represented in the network.

When this process has been completed the amalgamation process is reversed, and the nodes that have been collapsed resurface as in Figure 3 (adapted from Figure 3, Bryant & Moulton, 2004:256).
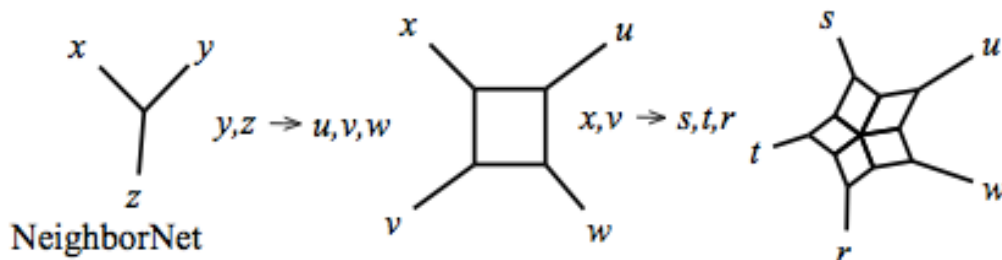
Figure 3: The reversal of the agglomeration process: nodes *y* and *z* are expanded to the original *u, v* and *w*. Nodes *x* and *v* are then expanded to the original *s, t* and *r*.

Thus the algorithm operates in such a way that the resulting network represents conflicting splits among the taxa. Such networks are informative in their

14

representations of linguistic phylogenies based on lexical data in a number of respects. Firstly, the presence of conflicting splits in the network gives an indication of borrowing or contact influences among the languages (McMahon & McMahon, 2003) and the degree of conflict in the networks gives an indication of the complexity of these contact patterns. The lengths of the edges separating the languages of the network are also informative since edge lengths are proportionate to the weight of the splits, or the degree of lexical distance between the languages separated by the split (Bryant, Filimon & Gray, 2005). Thus long divergent edges associated with certain taxa are indicative of large amounts of vocabulary not shared with the other languages of the data set. Additionally, the method of bootstrapping gives a measure of the confidence associated with each of the splits. In bootstrapping, the results of a certain number of random subsets of the data are compared in order to test the robustness of the evidence on which the splits of the network are based. While high confidence splits are based on more robust patterns in the data, low confidence splits are more likely to result from just a small number of lexical items (Huson & Bryant, 2006).

As well as representing conflicting splits, Neighbor-Net networks allow analysis of the linguistic and historical processes that may have caused the conflict. Networks containing complex nets of conflicting splits involving several languages may be an indication of dialect continua (Holden & Gray, 2006). Star-shaped splits with low levels of conflict but poorly marked hierarchical structure may represent processes such as rapid radiation, where a proto-language rapidly splits into several daughter languages, resulting in weak split signals (Holden & Gray, 2006).

2. Research Questions

In the current study, the contact and genealogical patterns of the Northern Aslian varieties are investigated, with a specific focus on the hitherto undescribed Menriq Rual variety. Following preliminary investigations into the lexical and phonological aspects of the data, Neighbor-Nets of more and less retentive lists of vocabulary from the Northern Aslian varieties are compared. The main research questions are threefold:

1) Does the data show traces of the genealogical and contact-based aspects of the relationships among the Northern Aslian varieties? While patterns evident in any identifiable regular sound changes and in the most basic vocabulary are likely to indicate genealogical aspects of the relationships, patterns in the less basic vocabulary are more likely to indicate relationships based on contact. In networks based on more and less basic subsets of vocabulary, the extent of contact between the Northern Aslian varieties is expected to be evident in the amount of conflicting splits in the networks. Traces of contact are expected to be seen in any differences in the positions of the varieties in the two networks. The direction of movement of varieties in the network representing the less basic vocabulary is expected to indicate the source/s of contact influences.

2) The second major research question of the current study concerns the connection of Menriq Rual to the Menriq varieties – is this a connection based primarily on contact or genealogy? In terms of the comparison of networks based on more and less basic vocabulary subsets, if the connection is one of genealogy rather than contact the analysis should show a closer connection between Menriq Rual and the Menriq varieties on the basis of the more basic vocabulary. If the connection is based primarily on contact the varieties should show a closer connection on the basis of the less basic vocabulary.

3) The connection of Menriq Rual with the other Northern Aslian varieties is also of importance in the current study. Any genealogical connections should show themselves in patterns of regular sound change as well as shared basic vocabulary, whereas connections based on contact should show themselves in the analysis of less basic vocabulary.

These questions are investigated using a range of methods. The analyses begin with a closer look at the phonological and lexical aspects of the Menriq Rual variety, followed by a traditional lexicostatistical comparison of the rates of shared vocabulary among the Northern Aslian varieties. Finally, Neighbor-Nets based on more and less basic subsets of vocabulary are compared. The combination of this range of methods is designed to strengthen the weight of the findings of the analyses through triangulation.

3. Method

3.1 The Menriq and Menriq Rual Data

Four wordlist recordings formed the major point of focus of the present study: one recording from each of the varieties Menriq Rual, Menriq Lah, Batek Teh Lebir and Batek Teh Taku. The recordings were made by Niclas Burenhult in connection with the DoBeS-sponsored 'Tongues of the Semang' survey. The recordings are of a 146-item wordlist based on Swadesh's (1952) 200-item list, which has been adapted by Geoffrey Benjamin to be more appropriate in the Aslian context (see Appendix 1 for the full list, and Benjamin, 1976, for information about the compiling of the modified list).

3.2 The Northern Aslian Data

The data for the wider Northern Aslian analysis was a subset of the data used in Dunn et al. (2011), with the addition of two language varieties: Maniq and Tea-De. The data was recorded between 1998 and 2008 by Niclas Burenhult and Nicole Kruspe, with the exception of the Maniq data which was collected by Ewelina Wnuk between 2009 and 2011, and the Tea-De data collected by Phaiboon Duangchan (Phaiboon, 2006).

3.3 Transcription and Cognate Coding

In the interests of gaining familiarity with the phonological aspects of the Menriq and Menriq Rual data, the Menriq Rual recording and the three Menriq recordings were listened to and transcribed. Potential cognates were then identified and coded. Since it is difficult to determine the actual cognacy of list items in the absence of detailed knowledge of regular sound changes among the Northern Aslian varieties, criteria were needed for the coding of potential cognates. The criteria were, as in Dunn et al. (2011), as follows: Forms were considered to be potential cognates where the initial and final consonants of the final syllable of forms matched in place of articulation. In some cases forms which do not meet the criteria have been coded as potential cognates, such as where forms failed to fulfil the criteria due to identifiable systematic sound changes or phonotactic changes, or to morphological operations. Similarly, forms that can be seen to fulfil the criteria due to chance have been analysed as non-cognates. These criteria are well suited to the material since the last syllable of words almost invariably has the structure /CVC/, and is most often part of the root. Suffixes are rare in the Aslian languages and the final syllable of words is not usually affected by morphophonemic processes (Dunn et al., 2011). Loans from Malay were coded as such so as to be excluded from the analyses.

3.4 Cognates vs potential cognates

Note the use of the term 'potential cognates' in the above description. In the current study, as in Dunn et al. (2011), the emphasis is on *shared lexicon*, rather than on true cognates. Since proto-forms for the items of the data set have not as yet been reconstructed, the shared lexical items can only be assumed to be apparent, or potential cognates, not true established cognates. These 'potential cognates' may be a result of genealogy or of intra-Aslian contact. They may turn out to be true cognates or loans. In the current study the term 'shared vocabulary' is used so as to avoid the issue of true cognacy.

3.5 The Analyses

The first part of the study involved a preliminary analysis of the phonological and lexical aspects of the data, with particular emphasis on the Menriq Rual and Menriq data. In the absence of established proto-language reconstructions for the Aslian

languages, and due to the limited amount of data available for many of the varieties, phonological analyses were necessarily preliminary. However, a comparison of the data set of the present study with Diffloth's (1975) outline of the patterns of sound changes in the Northern Aslian varieties gave several insights into the patterns in the data. Furthermore, since the current study included data from several varieties not included in Diffloth's study, this method allowed these varieties to be placed within the framework of the splits among the varieties proposed by Diffloth. The patterns of /r/ realisation in the Northern Aslian varieties were also investigated, since the realisation of /r/ in Menriq Rual is one of the most prominent features of this variety that set it apart from the Menriq varieties.

After an analysis of the Malay loan rates in the Menriq and Menriq Rual varieties, preliminary lexical analysis of the data focussed on the divergent forms of these four varieties. The rates of shared vocabulary among the Northern Aslian varieties were then calculated in an attempt to quantify the lexical similarity of Menriq Rual to the Menriq varieties and to the other Northern Aslian varieties. The final part of the analysis involved the use of Neighbor-Net network drawing software, which is part of the SplitsTree4 package (Huson & Bryant, 2006), combined with the comparison of more and less basic subsets of the lexical data.

3.6 The Sublists

The sublists used in the current study were based on those used by McMahon and McMahon (2003) and McMahon et al. (2005). However, due to differences in the makeup of the wordlist on which the data set of the current study is based compared to those of the two previous studies using the method, some modifications were necessary. While the full data set of the current study was based on a modified version of Swadesh's (1952) list adapted for use in an Aslian context, the lists used by McMahon and McMahon (2003) were based on Dyen, Kruskal and Black's (1992) version of the Swadesh list, and the lists used by McMahon et al. (2005) were based on a modified version of this list adapted for use with the languages of the Andes. Because of this, the number of items in the data set of the current study that were found in the lists of either of these two previous studies was somewhat limited, especially in the case of the two previous 'lolo' lists. Thus a hybrid version of the two earlier sets of lists was created, including all of the items from both previous versions that were found in the data set of the present study. This hybrid version of the two lists is shown in (1) a and b. The items found in both previous versions of the lists are shown in bold, those found in the McMahon and McMahon (2003) version only are shown in italics, and those found in the McMahon et al. (2005) version only are underlined.

(1)a. The hihi list of the current study (26 items)

day	**ear**	to eat	fingernail (claw)	**foot**
to give	**I**	to live	*long*	*mother*
name	**new**	**night**	**not**	**one**
salt	**to sleep**	*to spit*	*to stand*	*thin*
thou	**three**	**tongue**	**tooth**	**two**
wind				

(1)b. The lolo list of the current study (20 items)

back	**bird**	breast	far	*to flow*
heavy	left	**man**	**mouth**	*near*
neck	red	skin	*smooth*	**stone**
straight	**tail**	*to throw*	**to walk**	**wing**

4. Analyses and Results

4.1 Phonological Analyses

4.1.1 Analysis of Regular Sound Changes
While in-depth research into the regular sound changes of the Aslian varieties has yet to be conducted, Diffloth (1975) proposed a preliminary framework of the patterns of Aslian sound changes based on the data available to him. Several varieties of the current study did not form part of Diffloth's data set: Maniq, Tea-De, Menriq Rual, Batek Teq and the two Batek Teh varieties. Thus the patterns of sound change evident in the data set of the current study can be compared with Diffloth's findings in order to place these newly recorded varieties within this framework.

The ability to observe patterns was, as expected, limited by the small number of lexical items of the data set. However, a few important issues must be noted. Firstly, in terms of the split of Ceq Wong from the remainder of the Northern Aslian varieties, the additional varieties of the current study appear to share the phonological innovations that mark this split. Take for example the sound change *a > ɛ, shown in (2) a and b:

(2) a) Ceq Wong: /sac/ 'meat'; Batek and Menriq varieties, Menriq Rual, Jahai, Kensiw Perak: /sɛc/; Kensiw Kedah, Kintaq, Maniq: /sec/;

    b) Ceq Wong: /ʔak/ 'to give'; Batek and Menriq varieties, Menriq Rual, Jahai, Kensiw, Kintaq, Maniq, Tea-De: /ʔɛk/

Importantly, Maniq appears to have innovated (as should be expected) along with the remainder of the Northern Aslian varieties, despite the existence of a number of lexical retentions shared by Maniq and Ceq Wong.

Another notable example is the sound change *ə > e which separates the varieties of the Menraq and Maniq groups from the Batek varieties, shown in (3):

(3) Batek Deq Terengganu: /ʔntəŋ/ 'ear'; Batek Deq Koh: /ʔəntəŋ/; Batek Teq: /ʔntəŋ/; Menriq Lah, Batek Teh Taku, Jahai, Kensiw, Kintaq: /ʔnteŋ/; Maniq, Tea-De: /ʔantɛŋ/; Menriq Rual, Batek Teh Lebir: /ʔntiŋ/

The pattern seen here is in accordance with what should be expected: while all three Batek varieties have retained *ə/ə, all other varieties (notably, including Batek Teh) have innovated to e, ɛ or i. This suggests that the Batek Teh varieties are indeed not varieties of Batek. Two sound changes, *ɟC- > ʔiC- and *-r- > -j- are found by Diffloth to separate the Kensiw/Kintaq varieties from the remaining varieties. These sound changes are also found in the data set of the current study, illustrated in (4):

(4) a) Ceq Wong: /ɟʔeŋ/ 'bone'; Menriq, Menriq Rual: /ɟʔiŋ/, Jahai: /ɟʔeŋ/; Kensiw Kedah, Kintaq: /ʔiʔeŋ/; Kensiw Perak, Maniq: /ʔijeŋ/

    b) Menriq Lah, Jahai Banun: /krɔʔ/ 'back'; Batek Teh: /krəʔ/; Batek Deq Koh: /kʁɔʔ/; Batek Deq Terengganu: /kəɣɔʔ/; Kensiw, Kintaq: /kjɔʔ/; Maniq: /kaʔɔʔ/

Importantly, Maniq and Tea-De (compare Menriq, Menriq Rual: /ɟkɔp/ 'snake'; Kensiw, **Tea-De: /ʔikɔp/;** Jahai, Menriq: /braʔ/ 'not'; Kensiw, Kintaq, **Tea-De: /bjaʔ/**)

appear to have shared the changes of Kensiw and Kintaq while the Jahai, Menriq, Menriq Rual, Batek and Ceq Wong varieties of the data set have not. This lends support to the grouping of the Maniq variety together with the Kensiw and Kintaq varieties, and suggests that Tea-De also forms part of this group (Dunn et al.'s (2011) 'Maniq' subgroup). There is also evidence in the data that Maniq may have undergone a further sound change in which word-internal syllable-initial *-j- (corresponding to *-r- in the Menraq-Batek and Ceq Wong varieties) has become a glottal stop (*-j- > -ʔ-).

On the basis of the patterns found in the data it appears that the newly analysed varieties of the present study broadly follow the patterns outlined by Diffloth, and are generally in accordance with other previous classifications. Specifically, Batek Teh and Menriq Rual appear to share the innovation that sets the varieties of the Menraq and Maniq subgroups apart from the Batek subgroup, but do not share the innovations that characterise the further split of the Maniq subgroup. Meanwhile, Maniq and Tea-De appear to innovate along with Kensiw and Kintaq, and Batek Teq groups with the Batek Deq varieties. This gives support to the subgrouping proposed on the basis of lexical data in Dunn et al.'s (2011) study, the only other study to include these varieties (excluding Tea-De and Maniq): Their division of the Northern Aslian varieties into the Maniq (Maniq, Kensiw and Kintaq), Menraq (Jahai, Menriq, Batek Teh and Menriq Rual) and Batek (Batek Deq and Teq) subgroups is supported. Furthermore, the analysis suggests that Tea-De, not included in Dunn et al.'s (2011) analysis, is to be included in the Maniq group.

One pattern of sound change which is not discussed by Diffloth but which is noticeable in the data set of the current study is the seemingly regular changes in the Jahai nasal vowels of *ã͘ > ɛ̃ and *ɛ̃ > ĩ, shown in (5) a and b:

(5) a) Batek and Menriq varieties, Menriq Rual, Kensiw, Kintaq, Tea-De: /ʔãm/ 'breast'; Jahai: /ʔɛ̃m/;

   b) Ceq Wong, Batek and Menriq varieties, Kensiw, Kintaq, Maniq, Tea-De: /mɛ̃t/ 'eye'; Jahai: /mĩt/

The position of Menriq Rual in relation to these patterns appears to correspond to that of the Menriq varieties: while Menriq Rual and the three Menriq varieties appear to share the innovations of the Menraq group, they do not appear to share these innovations undergone by Jahai. Thus while Diffloth (1975) does not discuss any sound changes separating Menriq from Jahai, it appears that such changes are to be found. This also seems to argue against the subgrouping shown in some of the analyses of Dunn et al. (2011) which placed Menriq Rual as an outlier either to the Menraq subgroup, or to the combined Menraq-Batek subgroup, suggesting that on the basis of phonology, Menriq Rual is to be placed with the Menriq varieties within the Menraq subgroup. Assuming that these patterns are in fact indications of the historical splits among the varieties, we might conclude that Menriq Rual is, in terms of genealogy, more closely related to the Menriq varieties than to any other variety. However, analysis of a considerably larger data set is necessary before we can be sure that this is the case. On the basis of these very preliminary analyses, the newly recorded varieties of the current study may be placed into Diffloth's (1975) framework as in Figure 4.

Figure 4. The Northern Aslian Varieties of the Current Study within Diffloth's (1975) Framework
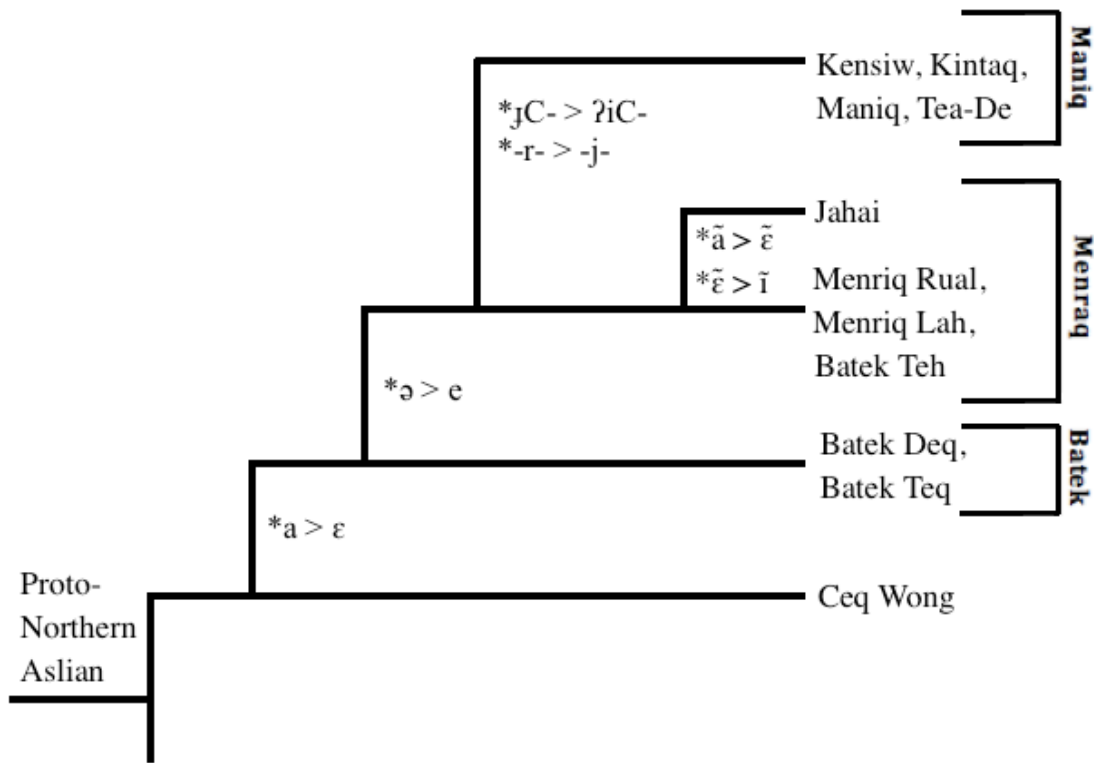
4.1.2 Analysis of /r/ Realisation

One of the most noticeable features that sets Menriq Rual apart from the other varieties of the Menraq subgroup is its characteristic realisation of /r/. A closer look at the realisation of /r/ in the Northern Aslian varieties should thus reveal whether this feature holds any clues as to the place of Menriq Rual among the Northern Aslian varieties. While Diffloth (1975) writes that the /r/ of the Northern Aslian varieties is invariably (save certain exceptions) realised as a voiced velar fricative, this is now known to be inaccurate. /r/ is realised as an alveolar trill in all positions in Ceq Wong, Batek Teq, Jahai and Menriq, this is in other words the norm among the Northern Aslian varieties that have not undergone the sound change *-r- > -j- (see 4.1.1 above). There are however two exceptions to this: Batek Deq and Menriq Rual.

In the two varieties of Batek Deq included in the current study, syllable-initial /r/ is realised as a uvular [ʁ] or velar fricative [ɣ], and word-final /r/ as a velar lateral approximant [ʟ]. In Batek Deq Terengganu, word-initial /r/ may be realised as an alveolar fricative [z] (as also noted by Diffloth, 1975). Meanwhile, Menriq Rual stands out among the members of the Menraq group as the only variety in which /r/ is not realised as a trill. In Menriq Rual, syllable-initial /r/ is realised as a uvular fricative [ʁ] while syllable-final /r/ is elided.

As noted above in section 4.1.1, the varieties of the Maniq group have undergone the sound change *-r- > -j-. However, more complex patterns can be seen among these varieties with respect to this change. Syllable finally, original *r is elided in syllable-final position in Kensiw, Kintaq and Maniq (and possibly also in Tea-De). While the Kensiw Kedah and Kintaq varieties have [j] word initially and between vowels, the Kensiw variety of Perak appears to have word-initial [ɣ] in some cases (although this

may only affect Malay loans, Benjamin, in press). Word-initial *r in Maniq is realised variably as a velar or uvular fricative or a velar or palatal approximant (Ewelina Wnuk, personal communication).

The /r/ realisations of Tea-De cannot be discussed with any certainty, since the data used in the current study is based solely on the transcriptions made by Phaiboon (2006). However, there appears to be a degree of variation in Tea-De /r/ realisations: in forms corresponding to those with original *r we find word-internal '-r-' as well as '-y-' (most likely representing [j]), and word-final elision as well as what is transcribed as 'l'.

The examples in Table 2 illustrate the general patterns of /r/ realisation among the Northern Aslian varieties of the data set of the current study:

Table 2. /r/ realisations among the Northern Aslian varieties

	'new'	'back'	'younger sibling'
Ceq Wong	rɛʔ		
Batek Deq Koh	ʁeʔ	kʁɔʔ	bɛʟ
Batek Deq Terengganu	zeʔ	kəɣɔʔ	bɛr
Batek Teq	reʔ		
Batek Teh Lebir	baroʔ	krəʔ	
Batek Teh Taku	baruʔ	krəʔ	bɛr
Menriq Lah	baroʔ	krɔʔ	bɛr
Menriq Rual	baʁoʔ		bɛ
Jahai Rual	baruh		bɛr
Jahai Banun	baruʔ	krɔʔ	bɛr
Kintaq	bajuʔ	kjɔʔ	bɛ
Kensiw Kedah	bajuʔ	kjɔʔ	bɛ
Kensiw Perak	bajuʔ	kjɔʔ	bɛh
Maniq		kaʔɔʔ	bɛ
Tea-De			bɛ

It is important to note that in the majority of the Central and Southern Aslian languages /r/ is realised as an alveolar trill [r]. Thus the non-trill /r/ realisations of the Batek Deq varieties and Menriq Rual, as well as the *-r- > -j- sound change of the varieties of the Maniq group are most likely innovations that have taken place after the split which separated Ceq Wong from the remaining Northern Aslian varieties. Assuming that the /r/ realisations of the Northern Aslian varieties are the result of historical sound changes and are not based solely on areal influences, the /r/ realisations of the Batek Deq varieties should be seen as the result of a change that has occurred after the split separating the Menraq and Batek groups. On the other hand, the uvular fricative [ʁ] of Menriq Rual should more likely be seen as a result of contact, since Menriq Rual shares other sound changes with Menriq. Thus the patterns observed among the Northern Aslian varieties in these two phonological analyses appear to suggest that while Menriq Rual is genealogically most closely related to the Menriq varieties, it may have been influenced by contact with Batek Deq. While the possibility exists that the /r/ realisations of Menriq Rual and Batek Deq are the result of other contact influences, such as that of local Malay dialects, this is nevertheless a useful finding that can serve as a hypothesis to be tested in the lexical analyses to come.

4.2 Preliminary Lexical Analysis

Preliminary lexical analyses included calculation of the Malay loan rates in the data from Menriq Rual and the three Menriq varieties as well as an analysis of the divergent forms of these varieties. Based on these analyses it would appear that the lexical differences between Menriq Rual and the Menriq varieties are not the result of differential influence from Malay. Malay loan rates for the four varieties were as follows: 23% for Menriq Rual, 18% for Menriq Lah, 21% for Batek Teh Lebir and 23% for Batek Teh Taku.

While a great number of the lexical items of the data set are shared by all four of these varieties (67 of 146 items) and a large number are shared by two or more of the varieties, a look at the items for which each of the varieties diverge from the others should reveal any obvious signs of influence from other Northern Aslian varieties. Firstly, Menriq Rual has a larger number of divergent forms compared to the three Menriq varieties – Menriq Rual has 24 divergent forms compared to Menriq Lah's 11, Batek Teh Lebir's 16 and Batek Teh Taku's 7. Analysis of these divergent forms also revealed some patterns. Table 3 shows the numbers of these divergent forms which are shared with at least one variety of the Batek group as well as at least one variety of the Maniq group, those forms which are shared with varieties of the Maniq group only, and those which are shared with varieties of Batek only, as well as those forms which are not shared with any other Northern Aslian variety in the sample.

Table 3. The Divergent Forms of Menriq Rual and the Menriq Varieties

	Not shared	Shared with Batek & Maniq	Shared with Maniq only	Shared with Batek only	Other varieties
Menriq Rual (22)	7	5	7	1	2 (Jahai, Maniq+Ceq Wong)
Menriq Lah (11)	7	2	1	-	1 (Jahai)
Batek Teh Lebir (16)	3	5	-	6	2 (Jahai, Ceq Wong)
Batek Teh Taku (4)	1	-	-	3	-

Firstly, in this analysis it can be seen that Menriq Rual and Menriq Lah have a greater proportion of divergent forms not shared with any other Northern Aslian variety in the sample. These two varieties also tend to share a greater proportion of their divergent forms with the varieties of the Maniq group and a lower proportion with the Batek varieties when compared to Batek Teh Lebir and Batek Teh Taku. This second pattern would seem to suggest that whether by genealogy or by contact, Menriq Rual and Menriq Lah may have a closer connection with the Maniq varieties than they do with the Batek varieties, and that the opposite may be true of Batek Teh Lebir and Batek Teh Taku.

The patterns seen here appear to be broadly in accord with what is known of the present-day contact patterns of these four varieties (discussed in sections 1.3 and 1.4). The relatively high number of forms shared by Batek Teh Lebir and the Batek varieties is congruent with the fact that the Batek Teh at Lebir live adjacent to a Batek Deq village. The overall smaller number of divergent forms of Batek Teh Taku could be a result of their present-day relative isolation. The fact that Menriq Rual has a higher proportion of shared forms with varieties of the Maniq group than with Batek varieties

is in accordance with the relative geographical proximity of this variety to the Maniq varieties compared to its distance from the Batek varieties. Thus the patterns observed in the analysis, however preliminary, suggest a connection between Batek Teh Lebir and the Batek varieties, as well as a connection between Menriq Rual and the Maniq varieties, whether this is based on (past or present) contact or genealogy. Although the number of divergent forms of Batek Teh Taku is small, the pattern suggests a closer connection with the Batek varieties than with the Maniq varieties. While Menriq Lah does not share any of its divergent forms with Batek varieties alone, it shares only a single divergent form with Maniq varieties alone. Thus the pattern regarding Menriq Lah is not clear enough to suggest a connection one way or the other.

One of the cases where Menriq Rual diverges from the Menriq varieties is of particular interest, and warrants individual discussion: the case of the first person singular. All Northern Aslian varieties except three share the form /jɛʔ/ - the three exceptions, Ceq Wong, Maniq and Menriq Rual, share the form /ʔiɲ/ (Ceq Wong /ʔiŋ/). Considering that the first person singular is generally considered to be an extremely basic and retentive item of vocabulary, and considering the geographical and genealogical distance separating the three varieties that share the form, this is of great interest. The form /ʔiɲ/ is thought to be an archaic form (Nicole Kruspe, personal communication). It is found in all Southern Aslian languages as well as several Central Aslian languages, and has cognates in other Mon-Khmer languages (Shorto, 2006). It is also found in Khmer inscriptions from the 8th century (Nicole Kruspe, personal communication). This suggests that the form is a retention rather than an innovation, that is, while the other Northern Aslian varieties have innovated away from this form, Ceq Wong, Maniq and Menriq Rual have retained it. The present-day isolation of the Ceq Wong and Maniq varieties suggests that this innovation from /ʔiɲ/ to /jɛʔ/ may have spread throughout the varieties of the Maniq, Menraq and Batek groups after the movements that resulted in the isolation of Ceq Wong and Maniq. The retention of this form in Menriq Rual could be interpreted as suggesting that Menriq Rual speakers may too have been cut off from contact with the other varieties of the data set at the time of this innovation, and that their present-day contact with these other varieties is the result of subsequent movements.

Thus far three different ideas have been proposed to account for the patterns relating to Menriq Rual observed in these first analyses. While the sound change patterns identified in the varieties of the sample suggest that Menriq Rual is most closely related to Menriq, the /r/ realisations of Menriq Rual suggest a connection – perhaps one of contact – with the Batek Deq varieties. Meanwhile, the analysis of the divergent forms of the Menriq Rual and Menriq varieties suggests a closer connection of Menriq Rual with the varieties of the Maniq subgroup than with the Batek varieties. Thirdly, a past isolation of Menriq Rual from contact with the other varieties of the data set has been proposed in explanation of its archaic first person singular form. In a situation of such high levels of speaker movement and linguistic variation and change, the possibility exists that Menriq Rual speakers have indeed undergone periods of such isolation as well as periods of contact with both the Batek varieties and the Kensiw/Kintaq varieties. Alternatively, we may find that some of these patterns are better explained by genealogy while others are better explained by contact. Also, the possibility cannot be ruled out that the patterns are the product of a dialect continuum. While the analyses on which these ideas are based are undeniably exploratory and

preliminary, the patterns observed may serve to suggest a direction for the analyses that follow, as well as add support to certain findings.

4.3 Analysis of shared vocabulary rates
Table 4. Shared vocabulary rates among the Northern Aslian varieties

| Tea-De														
57,5	Maniq													
68,8	59,2	Kensiw P												
72,2	63,2	83,2	Kensiw K											
67,1	55,1	82,2	88,2	Kintaq										
55,7	45,3	56,5	61,9	61,1	Jahai Banun									
53,9	40,2	55,5	58,3	59,3	88,9	Jahai Rual								
58,7	50,5	61,7	63,8	63,5	71,0	68,2	Menriq Rual							
58,4	47,4	56,8	61,5	60,2	77,7	73,8	76,9	Menriq Lah						
56,8	44,5	55,0	55,7	52,8	72,2	70,2	74,5	82,9	Batek Teh L					
57,9	47,7	56,5	58,5	58,1	73,8	70,2	78,3	89,2	84,3	Batek Teh T				
66,7	53,4	62,4	65,1	61,0	65,5	58,5	64,6	68,6	70,9	73,5	Batek Teq			
59,4	48,0	56,4	61,3	56,0	64,1	58,1	69,3	70,8	68,1	74,4	82,5	Batek Deq T		
59,2	45,8	57,1	62,5	60,8	67,0	59,0	68,7	72,4	71,8	76,2	83,5	89,6	Batek Deq K	
41,3	38,1	41,9	45,6	46,1	47,9	46,9	47,2	43,0	44,5	43,9	48,3	47,4	46,2	Ceq Wong

The next step in the analysis involved a traditional lexicostatistical comparison of the overall rates of shared vocabulary among the Northern Aslian varieties. This comparison somewhat crudely represents the degrees of similarity among the varieties. A comparison of the rate of vocabulary shared by Menriq Rual and the Menriq varieties with the rates found among varieties of the distinct Northern Aslian 'languages' gives an indication of the status of Menriq Rual as a variety separate from Menriq. According to current official (Malay government) and scientific praxis, the data set contains four languages for which more than one variety is represented: the Kensiw/Kintaq language represented by Kensiw Perak, Kensiw Kedah and Kintaq; the Jahai language represented by Jahai Banun and Jahai Rual; the Menriq language represented by Menriq Lah, Batek Teh Lebir and Batek Teh Taku; and the Batek language represented by Batek Teq, Batek Deq Terengganu and Batek Deq Koh. If the rate of vocabulary shared by Menriq Rual and the Menriq varieties is comparable to the shared vocabulary rates within these distinct languages of the data set, Menriq Rual might quantitatively be considered a variety of Menriq.

The analysis shows that the mean rate of shared vocabulary among the varieties of the data set considered to belong to distinct languages is 86%, and for each of these languages this rate was at least 82%. This is in itself an interesting result as it suggests that the current official division into distinct languages is quantifiable in terms of shared lexicon. Menriq Rual however shares an average of 76.6% of its vocabulary with the three Menriq varieties. Thus, in relation to its particular context, Menriq Rual should not be considered a variety of Menriq, at least not in a straightforward manner. Menriq Rual's shared vocabulary rates do however place it within the Menraq subgroup and within this subgroup it shares a higher rate of vocabulary with the three Menriq varieties (mean 76.6%) than it does with the two Jahai varieties (mean 69.6%). Thus while the analysis suggests that Menriq Rual should be considered to be a variety separate from the three Menriq varieties, it suggests a closer connection with these varieties than with any of the other varieties of the sample.

Patterns can also be seen in the rates of vocabulary shared by Menriq Rual and the remaining Northern Aslian varieties. The areas marked in green in Table 4 show the shared vocabulary rates of Menriq Rual and the Menriq varieties with Maniq, Kensiw and Kintaq on the one hand and with the Batek varieties on the other. Here it can be seen that Menriq Rual shares a lower rate of vocabulary with the three Batek varieties than do the Menriq varieties (a mean of 67.5% for Menriq Rual compared to 70.6% for Menriq Lah, 70.3% for Batek Teh Lebir and 74% for Batek Teh Taku). On the other hand, Menriq Rual has higher shared vocabulary rates with Maniq, Kensiw and Kintaq (63%) than do the three Menriq varieties (mean 59.5%, 54.5% and 57.7%, respectively). Thus the patterns observed in the analysis of shared vocabulary rates are in agreement with those observed in the preliminary analyses of the divergent lexical forms of Menriq Rual and the Menriq varieties of section 5.1 above.

The analysis also allows us to resolve some issues regarding the divergence of Menriq Rual from the Menriq varieties. Firstly, it appears that this divergence is not caused by a greater influence of Jahai Rual on Menriq Rual, as could be expected on the basis of the cohabitation of the speakers of these varieties. In fact, Menriq Rual shares a lower rate of vocabulary with both Jahai varieties than do the three Menriq varieties. While a connection of Menriq Rual to the Tea-De (or Jdek) variety has been proposed (Niclas Burenhult, personal communication; see section 1.3 above), shared vocabulary rates do not suggest a connection of Menriq Rual to Tea-De. Furthermore, although Menriq Rual appears to have a higher rate of shared vocabulary with Ceq Wong than do the Menriq varieties, this rate is not the highest among the remainder of the Northern Aslian varieties. On the basis of the shared vocabulary rates shown above, a tree-type phylogeny of the Northern Aslian varieties might be constructed as shown in Figure 5.

Figure 5. Northern Aslian subgroupings on the basis of shared vocabulary rates
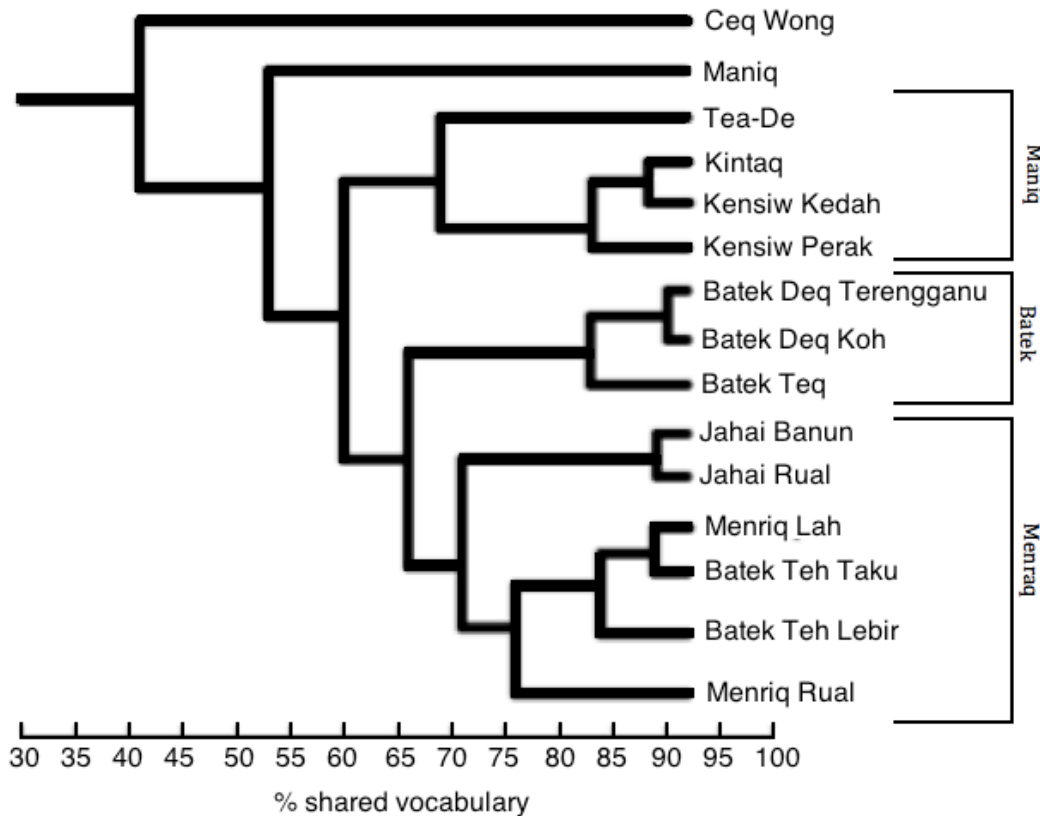

While the tree model is a visually useful tool in the representation of relationships among varieties of a language family, comparison of the distance matrix in Table 4 with the information represented in the tree in Figure 5 suggests that there are numerous aspects of the relationships among the Northern Aslian varieties which are not represented in this kind of tree diagram. A great degree of the complexity of the patterns is not represented in the tree above, for example the higher rate of vocabulary shared by Menriq Rual and the Kensiw/Kintaq varieties is lost in the tree, as is its lower rate shared with the Batek varieties. Furthermore, certain aspects of this tree are in all likelihood inaccurate: Maniq is shown as an outlier to the branch formed by the Maniq, Menraq and Batek groups, rather than as a member of the Maniq group. The remainder of the analysis will thus make use of the Neighbor-Net method in order to allow for a representation that takes into account the ambiguous and conflicting patterns that have been seen in the data.

4.4 Neighbour-Net Analysis

Using the Neighbour-Net method included in the SplitsTree4 software package (Huson & Bryant, 2006), the full Northern Aslian data set produced the network shown in Figure 6. This network demonstrates some major differences between traditional tree diagrams such as that in Figure 5 and the Neighbor-Net network. First of all, since Neighbor-Net does not assume a tree structure, we get a measure of how treelike the relationships between the varieties are. Second, the network shows not only the splits for which there is most evidence (that is, the most heavily weighted splits), but also conflicting splits of lower weight, allowing for the representation of weaker patterns or secondary relationships in the data. This gives a representation not only of patterns that are likely to be based on genealogy but also those that may be the result of contact. Thus, in cases where varieties are not products of clean splits between speaker groups, but where a degree of contact has been maintained among the groups (such as dialect continua), conflicting patterns in the data will result in a network with a non-treelike structure.

Figure 6. Neighbor-Net network based on the entire data set
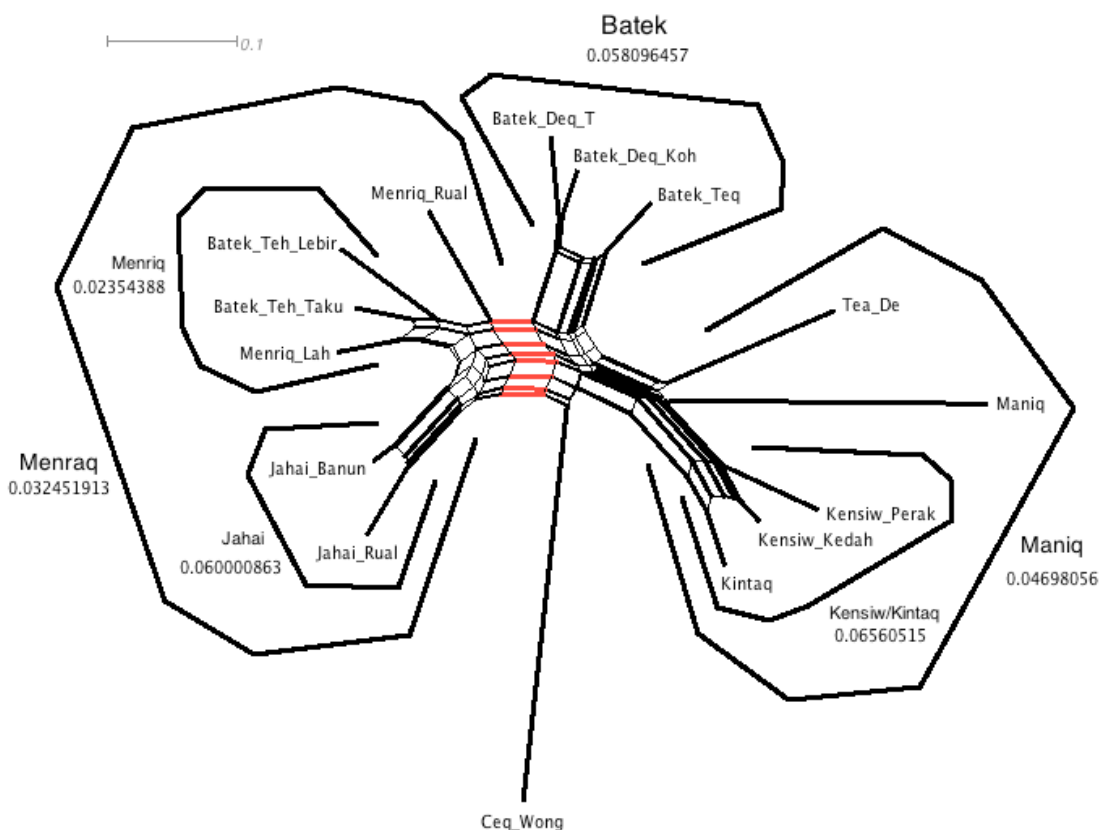
It is clear from the network in Figure 6 that the relationships among the Northern Aslian varieties are far from treelike. A complex net of reticulations in the centre of the network joins the majority of the varieties, suggesting the possibility that the relationships between these varieties may be akin to a dialect continuum-type situation. The network does however clearly delimit certain subgroups, corresponding to the major splits identified in earlier studies. To the left, the Menraq group can be seen, containing Jahai and Menriq as well as Menriq Rual. To the right we see the Maniq group containing outliers Maniq and Tea-De as well as a more close-knit group of

Kensiw/Kintaq varieties. The Batek group is also clearly shown, containing Batek Deq and Teq. While all of the varieties (except Ceq Wong) are involved in conflicting splits to some extent, some of these splits have more weight than others. The conflicting split connecting the Batek Deq varieties with the varieties of the Menraq group is comparatively robust – this is in accordance with the closer connection between the Batek and Menraq subgroups found by Dunn et al. (2011), and suggests that a higher degree of contact may exist between these groups than that of either group with the varieties of the Maniq group.

In Neighbor-Net networks, edge lengths are proportional to the weight of the associated splits. In Figure 6, the weight of the major splits is also shown numerically below the labels of the respective subgroups. These weights suggest that the evidence for the subgrouping of the Menraq group is less robust than that of the other groups. Furthermore, the split connecting the three Menriq varieties has a considerably lower weight compared to the splits connecting the varieties of the other distinct languages of the data set. In this network the confidence of the splits (obtained through bootstrapping with 1000 replicates) is also shown, by way of the differing widths of the edges of the network. Thus we see that while the network shows conflicting splits that connect the Batek varieties with Menriq Rual and the Menriq varieties on the one hand, and others which connect Batek Teq with the varieties of the Maniq subgroup (excluding Kintaq) on the other, as well as splits that connect Jahai, Ceq Wong and Kintaq, these splits are not robust.

Worthy of note are the lengths of the divergent edges of Ceq Wong, Tea-De and Maniq. Ceq Wong is clearly shown as an outlier to the entire Northern Aslian group. It makes a clean break from the remainder of the varieties and is only connected by a low weight and low confidence split to Jahai and Kintaq. Tea-De and Maniq on the other hand are clearly placed within the Maniq subgroup – a relatively high weight, high confidence split makes this clear. However, the lengths of the divergent edges of these varieties suggest that they diverge sharply from the other varieties of the subgroup. Thus the three varieties known to have little contact with the other groups of the data set are all shown to diverge sharply from the remaining varieties. After these three varieties, the next most divergent variety is Menriq Rual, with a comparatively low weight split connecting it with the other varieties of its subgroup as well as a comparatively long divergent edge. This represents the high rate of Menriq Rual vocabulary not shared with the other varieties of the data set. In contrast to the present-day isolation of the other more divergent varieties, however, Menriq Rual speakers are involved in extensive contact with other groups, in particular with Menriq and Jahai speakers. This divergence of the Menriq Rual variety thus raises questions as to the historical contact patterns of the Menriq Rual speakers. One possible explanation is the idea (proposed in section 4.2 above) that the Menriq Rual speakers have undergone a past period of isolation from contact with other groups, resulting in a higher number of divergent forms. Alternatively, the divergent forms that give rise to the long diverging edge of Menriq Rual may be remnants of a connection with a variety (existing or extinct) not included in the data set of the current study.

While the analyses thus far have suggested certain patterns in the connections of Menriq Rual with the other Northern Aslian varieties, it is not clear whether these patterns are the result of genealogy or contact. The method of comparing more and less

retentive subsets of vocabulary is thought to be well suited to precisely such situations. The final part of the analysis is thus based on this method.

4.5 The hihi and lolo sublists

Figure 7 shows Neighbor-Net networks constructed on the basis of the hihi (Figure 7a) and lolo (Figure 7b) sublists presented in section 3.6. The networks in Figure 7a and 7b differ in several respects from the network in Figure 6 constructed on the basis of the entire 146-item wordlist. Firstly, both networks are considerably less treelike than the network based on the entire 146-item wordlist, containing a larger degree of conflicting splits as well as a larger proportion of low confidence splits. Secondly, the clustering of the varieties in both networks is much less clear-cut than in the network of Figure 6. The connections among the varieties of the Menraq subgroup appear particularly affected in the two networks – there are no clear-cut splits connecting the varieties of this group in either network.

Comparison of the hihi and lolo networks also reveals several patterns. The network based on the more basic vocabulary (Figure 7a) contains comparatively fewer conflicting splits, and is thus more treelike than the network based on the less basic vocabulary (Figure 7b). This is to be expected since this vocabulary should be less prone to contact influences that give rise to conflicting splits. In addition, a greater proportion of the splits of Figure 7b are of low confidence, in particular those in the centre of the network. Another difference worth noting between the two networks is the lengths of the diverging edges of the three outlier varieties Ceq Wong, Maniq and Tea-De – all are considerably longer in Figure 7b than in Figure 7a. This is to be expected since the more basic vocabulary is thought to be more resistant to change and should thus give rise to less divergence, whereas the rates of replacement among the less basic vocabulary are thought to be higher. However, Menriq Rual is shown to diverge marginally *less* in Figure 7b than in Figure 7a, not behaving like the three outlier varieties in this respect. This suggests that the less basic vocabulary of Menriq Rual has converged with that of other varieties in the data set, since this less basic vocabulary is more prone to the influences of the present-day contact that the Menriq at Rual have with speakers of the other varieties of the data set. This pattern might be seen as lending support to the idea that Menriq Rual may have previously undergone periods of isolation, only subsequently coming into contact with the other groups of the data set.

Figure 7a. Neighbor-Net network based on the hihi list
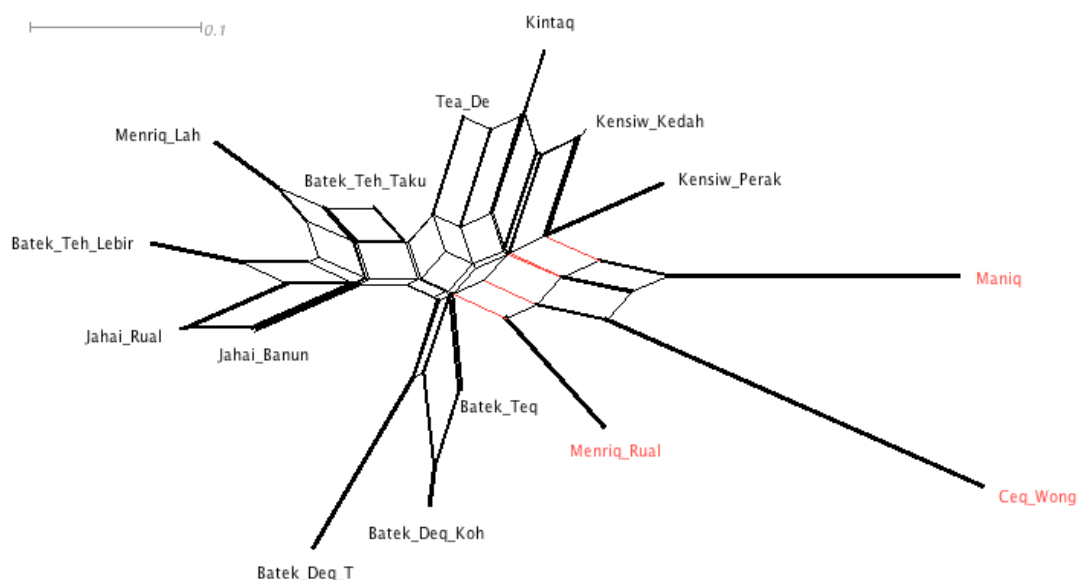
Figure 7b. Neighbor-Net network based on the lolo list
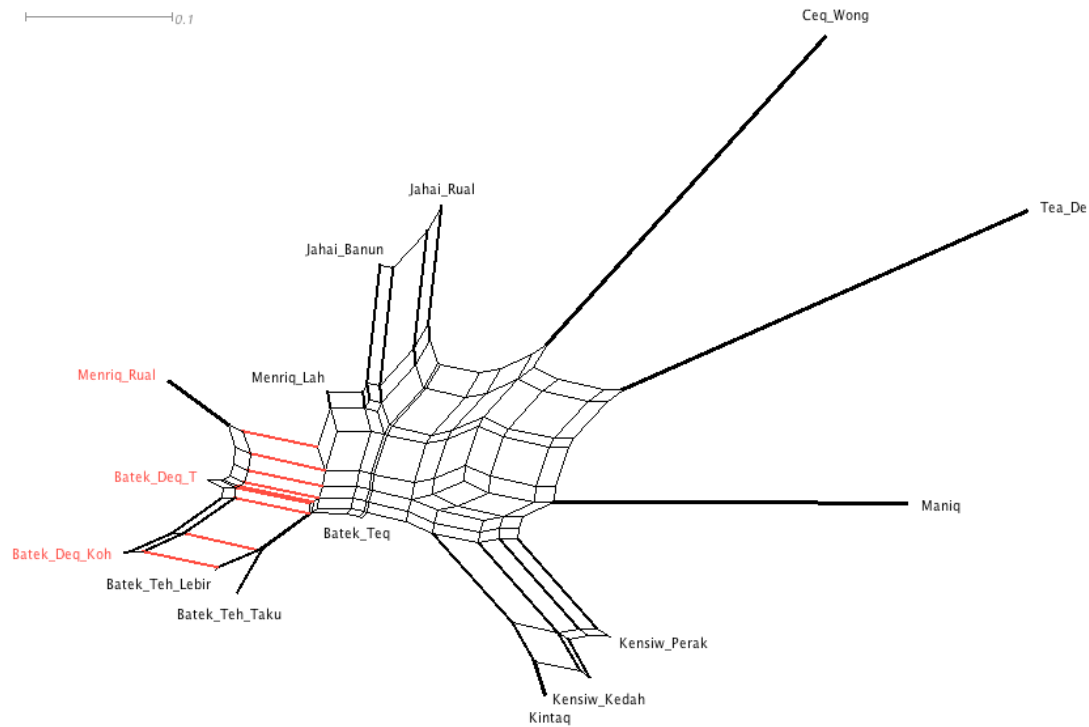
Apart from the relative divergence of Menriq Rual in the two networks, its position in relation to the other Northern Aslian varieties is also worthy of note. The position of Menriq Rual changes dramatically in the two networks: In Figure 7a it is connected by a relatively high-weight but low-confidence split to Maniq and Ceq Wong, whereas in Figure 7b it is connected by a high-weight, high-confidence split to the two Batek Deq varieties on the one hand, and a low-weight, low-confidence split with Menriq Lah and the two Jahai varieties on the other. This suggests that Menriq Rual shares a relatively large amount of basic vocabulary with Ceq Wong and Maniq, and a large amount of less basic vocabulary with Batek Deq along with a smaller amount with Menriq Lah and Jahai. It is interesting that neither network shows the major split connecting Menriq Rual with the varieties of the Menraq group that is shown in Figure 6 above. This is unexpected as earlier analyses suggest that Menriq Rual shares the greatest proportion of its lexicon with the varieties of the Menraq group, and in particular with the Menriq varieties. In addition, the analysis of regular sound change patterns in 4.1.1 above suggested that the connection between Menriq Rual and the remaining varieties of the Menraq group is one of genealogy. The reason for the absence of this connection from the networks of 7a and 7b is unclear, however it is possible that the bulk of vocabulary shared by Menriq Rual and the Menriq and Jahai varieties belongs to a subset not as basic as that of the hihi list, yet not as prone to borrowing as that of the lolo list. It may be that these varieties are indeed connected genealogically, but that the conservative aspects of the most basic vocabulary of Menriq Rual (perhaps caused by a period of isolation from contact with the other varieties of the data set) combined with contact influences from Batek Deq serves to conceal this connection in the networks.

Networks 7a and 7b show traces of some of the patterns suggested in earlier parts of the analyses of this study, but not others. The analysis of the phonological aspects of

the data suggested that while Menriq Rual shares sound changes with the other varieties of the Menraq group that separate it from the Batek varieties, subsequent contact with the Batek Deq varieties is suggested by the /r/ realisations of Menriq Rual. This idea of contact between Menriq Rual and the Batek Deq varieties is supported in the networks above – while no connection between these varieties is shown on the basis of the more basic vocabulary, the less basic vocabulary results in a strong connection of Menriq Rual to Batek Deq. On the other hand, while Menriq Rual appeared in earlier lexical analyses to share a larger amount of vocabulary with the Kensiw and Kintaq varieties than with the Batek varieties, no traces of this pattern can be seen in the networks of Figures 7a and 7b. The splits of the networks do not show any evidence of a connection of Menriq Rual with Kensiw and Kintaq. However, the connection of Menriq Rual with Maniq and Ceq Wong in Figure 7a is in line with the pattern suggested by Menriq Rual's conservative first person singular form. The patterns shown in Figures 7a and 7b thus appear to lend support to the idea that Menriq Rual may be a relatively conservative variety, at least in terms of its most basic vocabulary. These patterns also support the idea that this conservative variety has, following earlier periods of separation from the other varieties of the data set, undergone periods of contact with the Batek Deq varieties, as well as with Jahai and Menriq Lah. While the 'separation' proposed here may entail isolation of the Menriq Rual speakers, it may also be the result of contact with other groups not included in the data set. Importantly, the patterns seen in the networks of Figures 7a and 7b suggest that the words of the hihi list used in the current study are, as hoped, comparatively less prone to borrowing, and that even the intense contact known to exist between the Northern Aslian varieties has not had an effect on this vocabulary. This is clear as contact between Menriq Rual speakers and Ceq Wong and Maniq speakers is highly unlikely, given the geographical distance between the speakers of these varieties as well as the known isolation of these groups.

Since Neighbor-Net networks compare the lexical distances between varieties, the distance of Menriq Rual from each of the other Northern Aslian varieties in the two networks can be compared. The relative distance of Menriq Rual to the other Northern Aslian varieties in the two networks is compared in Figure 8. In terms of the connection of Menriq Rual to the other varieties of the Menraq subgroup, a few different patterns can be seen. Large differences are seen in the relative distances of Menriq Rual from Menriq Lah and from Batek Teh Lebir in the two networks, as well as marginal differences in the distance from Batek Teh Taku and the two Jahai varieties. While the distance of Menriq Rual from Batek Teh Taku is smaller based on the more basic vocabulary, the pattern for the other varieties of the Menraq group is the opposite. In other words, Menriq Rual shares a larger amount of more basic vocabulary than less basic vocabulary with Batek Teh Taku, but a larger amount of less basic vocabulary with Menriq Lah and Batek Teh Lebir, as well as (marginally) with the Jahai varieties. This would seem to suggest that while Menriq Rual is connected to Batek Teh Taku by genealogy rather than by contact, its relationship with the other varieties of the Menraq group is better explained by contact. This would seem to be particularly true of the connection of Menriq Rual with Menriq Lah, where the difference in distance between the two networks is substantial.

Figure 8. Comparison of lexical distances between Menriq Rual and the remaining Northern Aslian varieties in the two networks
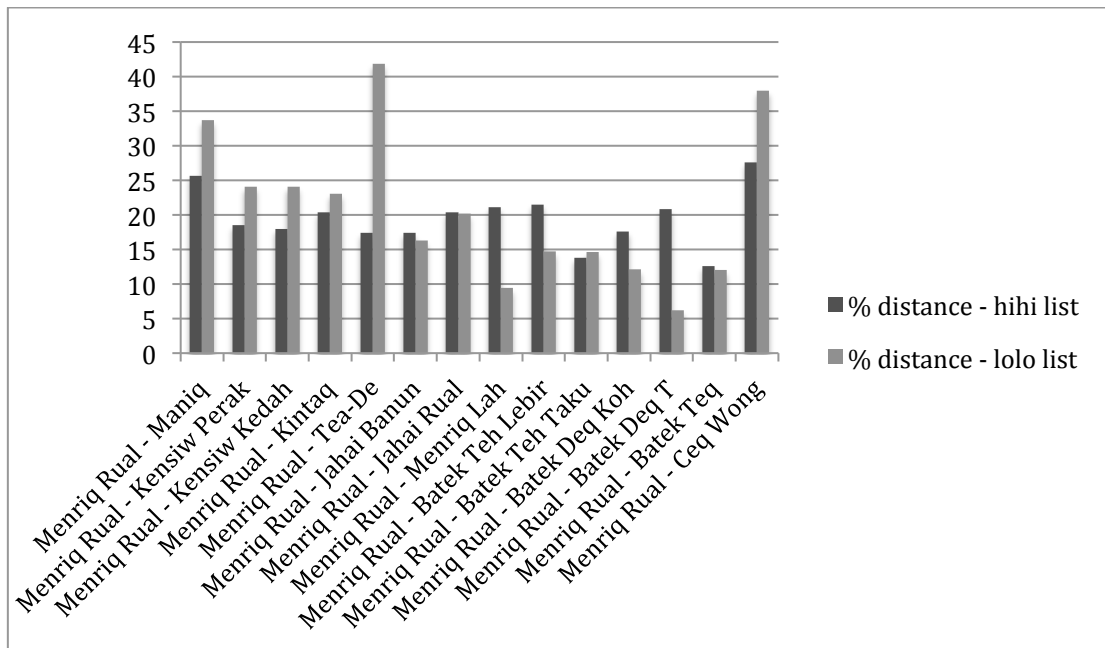
In terms of its connection with the varieties of the data set outside the Menraq subgroup, it can be seen that Menriq Rual has a closer connection with the Batek varieties on the basis of the less basic vocabulary, whereas its connection with the varieties of the Maniq subgroup, as well as with Ceq Wong, is closer on the basis of the more basic vocabulary. This pattern suggests that Menriq Rual has been affected to a greater extent by contact with the Batek varieties than with the Maniq varieties.

However, since the situation of the Northern Aslian varieties is so thoroughly marked by contact, it is difficult to draw straightforward conclusions from these results without a look at the patterns evident in the wider Northern Aslian context. Figures 9 through 22 (see Appendix 2) show distance comparisons for each of the Northern Aslian varieties. In these comparisons a clear pattern emerges. Only three Northern Aslian varieties consistently show closer connections with the other Northern Aslian varieties on the basis of the more retentive vocabulary: these are the three outlier varieties Ceq Wong, Maniq and Tea-De. This suggests an absence of contact influences from the varieties of the data set on these varieties, a finding that is in accordance with the present-day isolation of these varieties from the remainder of the group.

The remainder of the varieties show some signs of contact, in that they have a higher rate of shared vocabulary with certain other varieties based on the less basic vocabulary than on the basis of the more basic items. However, some varieties show signs of more wide-ranging contact than others. Importantly, we find suggestions of contact between all of the members of a group including the three Kensiw/Kintaq varieties, Menriq Lah, Batek Teh Lebir, Batek Deq Terengganu and Batek Teq. This finding is important since the Kensiw/Kintaq varieties are not known to be involved in any present-day contact with the Batek varieties. This pattern may be a sign that this group of varieties are members of a (present-day or past) dialect continuum stretching from Kensiw/Kintaq in the north to Batek in the south.

However, several varieties of the data set do not appear to form part of this continuum, showing signs of contact only with certain varieties. Batek Teh Taku shows signs of contact only with Batek Teh Lebir and the Batek varieties. While this is in agreement with the present-day contact of the Batek Teh at Sungai Taku with the Batek Teh at Pos Lebir, past contact with varieties of Batek are also not unlikely, given the geographical proximity of the Batek Teh at Sungai Taku to these varieties. In addition, the seeming lack of contact influences on Batek Teh Taku from the remainder of the Northern Aslian varieties is in accordance with its present-day relative isolation. While Batek Deq Koh shows signs of contact with the three Menriq varieties as well as with Menriq Rual and Batek Deq Terengganu, it appears not to be part of the patterns of contact stretching further north to Kensiw and Kintaq.

While the two Jahai varieties show signs of contact with certain varieties (Menriq Rual, Menriq Lah, Batek Teq and Batek Deq Terengganu), they appear to lack evidence of contact with the Kensiw/Kintaq and Batek Teh varieties. Thus the Jahai varieties also appear to be separate from the proposed dialect continuum. This pattern is in accordance with the idea proposed by Benjamin (1976; discussed in 1.3 above) that the arrival of Jahai speakers between the once-contiguous Kensiw/Kintaq and Menriq groups is relatively recent. The position of Menriq Rual in relation to the proposed dialect continuum is however somewhat difficult to discern. While Menriq Rual shows signs of contact with Jahai, Menriq Lah, Batek Teh Lebir and the Batek varieties (that is, all non-isolated Menraq-Batek groups), it does not show signs of contact with the Kensiw and Kintaq varieties.

Another aspect of the comparisons shown in Figures 9 through 22 is the connections shown between the varieties of distinct languages. The small distances between the two Jahai varieties and between the varieties of Kensiw/Kintaq on the basis of both sublists leave no doubt as to the close connection among these varieties thought to belong to distinct languages. Likewise, the large distances between Ceq Wong, Maniq and Tea-De and the remaining varieties of the data set on the basis of both lists leave no doubt as to the lack of close connections of these varieties with the other Northern Aslian varieties. On the other hand, the connections within the Menriq and Batek languages appear less straightforward, the patterns differing on the basis of the two sublists. It may be that this is a result of the intensity of the contact that takes place among these varieties. It may also be a sign that the connections among the Menriq and Batek varieties are more accurately represented as a dialect continuum than as distinct languages, while the cutting off of the Kensiw and Kintaq varieties from this continuum due to the movements of Jahai speakers has resulted in the formation of a more clear-cut and distinct Kensiw/Kintaq language.

5. Concluding Discussions

5.1 Summarising the Findings of the Analyses

5.1.1 Findings regarding the Northern Aslian varieties
In the current study, the contact and genealogical patterns of the Northern Aslian varieties have been explored, with a specific focus on the Menriq Rual variety. Traces of both genealogical and contact patterns among the Northern Aslian varieties were found in the analyses. Analysis of the phonological aspects of the data showed support for the genealogical groupings of the Northern Aslian varieties set up by Diffloth (1975). Furthermore, the additional varieties included in the current study were seen to fit into Diffloth's subgrouping as follows: the two Batek Teh varieties and Menriq Rual group together with Menriq Lah (forming the Menraq group together with Jahai), Batek Teq groups together with the Batek Deq varieties (forming the Batek group), and Maniq and Tea-De group together with the Kensiw/Kintaq varieties (forming the Maniq group). In addition, Jahai was seen to have undergone sound changes that set it apart from the Menriq varieties and from Menriq Rual, and Maniq appears to have undergone sound changes setting it apart from the remainder of the varieties of the Maniq subgroup. Thus according to the patterns of identifiable regular sound changes the genealogical relationships of the Northern Aslian varieties are relatively straightforward.

However, analysis of the lexicon of the Northern Aslian varieties suggests a range of more complex dynamics. Neighbor-Nets of the lexical data revealed a high degree of contact among the majority of the Northern Aslian varieties, in particular among the varieties of the Menraq and Batek subgroups. This was evident in the complex net of conflicting splits connecting these varieties. Comparison of more and less retentive subsets of vocabulary showed traces of contact among the majority of the Northern Aslian varieties. On the basis of this analysis it has been suggested that a previously existing dialect continuum may have stretched from the Kensiw/Kintaq varieties in the north down to the Batek varieties in the south, incorporating the Menriq varieties in between. More recently, however, the arrival of Jahai speakers between the Kensiw/Kintaq and Menriq groups appears to have cut off Kensiw/Kintaq from contact with the remaining varieties of the dialect continuum. These findings support ideas proposed by Benjamin (1976; discussed in section 1.3 above). Furthermore, those varieties known for their present-day isolation from the remainder of the group do not show signs of contact with the other Northern Aslian varieties: the Ceq Wong, Maniq and Tea-De varieties. Other varieties show signs of contact with some varieties but not with others, suggesting that they may not form part of the proposed dialect continuum.

5.1.2 Findings regarding Menriq Rual
Another major research question of the study related to the connection of Menriq Rual to the Menriq varieties. Analysis of the phonological aspects of the data suggested that Menriq Rual shares the phonological innovations of the Menriq varieties and has not undergone any regular sound changes that set it apart from these varieties. This suggests that the closest genealogical connection shared by Menriq Rual among the varieties of the study is that it shares with the Menriq varieties. Likewise, an analysis of rates of shared vocabulary suggested that Menriq Rual shares more lexicon with the Menriq varieties than it does with any other of the varieties of the data set. However, the lexical similarity of Menriq Rual to the Menriq varieties is not of the same order as the similarity found between other Northern Aslian varieties that are thought to belong

to distinct languages. Thus Menriq Rual cannot be seen as belonging to the Menriq 'language' in this sense. Indeed, comparisons of more and less basic subsets of vocabulary did not suggest a close connection of Menriq Rual with the Menriq varieties, whether on the basis of the most basic subset of vocabulary nor on the basis of the less basic subset.

Turning to the connection of Menriq Rual to the other Northern Aslian varieties, while Menriq Rual appears to share an overall higher rate of lexicon with the Kensiw, Kintaq and Maniq varieties, as well as a lower rate with the Batek varieties, than do the Menriq varieties, no traces of this are seen in the comparison of more and less basic vocabulary subsets. Instead, the comparison suggests a connection of Menriq Rual to Ceq Wong and Maniq on the basis of the most basic vocabulary and a connection to the Batek Deq varieties on the basis of the less basic vocabulary. It may be that the subset of vocabulary that Menriq Rual shares with the Kensiw, Kintaq and Maniq varieties, as with that shared with the varieties of the Menraq group, is of a nature more retentive than the items of the lolo list yet less retentive than those of the hihi list. These differing patterns concerning Menriq Rual evident in the analyses would seem to hint at a complexity in the history of this variety.

The patterns seen in the analysis of more and less basic sublists appear to offer possible explanations for the two most obvious aspects of Menriq Rual's divergence from the Menriq varieties: the /r/ realisations and first person singular form of Menriq Rual. While the possibility exists that the /r/ realisations of both Menriq Rual and Batek Deq are the results of influence from nearby Malay dialects, the patterns seen in the less basic lexical data strongly suggest a contact-based connection of these varieties, suggesting that the /r/ realisations of Menriq Rual may have come about through contact with Batek Deq speakers. Meanwhile, the archaic first person singular form Menriq Rual shares with Ceq Wong and Maniq may signal the otherwise conservative nature of Menriq Rual. One explanation of these patterns may be that Menriq Rual is a conservative variety, previously more isolated from the other varieties of the data set, whereas subsequent contact with Batek Deq-speaking groups has affected the less basic lexicon as well as the /r/ realisation of this variety. The geographical proximity of the Menriq at Rual to the Thai border suggests the possibility that this group may in the past have had contact with the Northern Aslian speaking groups of southern Thailand, with a possible connection to a group referred to as 'Jdek' by some Northern Aslian speakers. However, the analyses of the current study did not show any traces of a connection with Phaiboon's (2006) Tea-De group, who have been suggested to be synonymous with these 'Jdek' (Niclas Burenhult, personal communication). On the other hand, language consultants' reports of Menriq Rual origins further south in Sungai Taku would seem to allow for contact between Menriq Rual and Batek Deq, traces of which are apparent in the analyses of the current study.

5.2 The exploratory nature of the study
While a number of patterns are evident on the basis of the analyses of the current study, a certain degree of caution is necessary in their interpretation. Firstly, the analyses reported here are based only on the limited data available to date for these varieties. However, this closer look at the available data, however limited, is considered to be an important step in the study of the Northern Aslian varieties, as it can point to useful directions for further research involving these varieties. Also, the

data that forms the basis of the current study is no more limited than that on which previous classifications of the Northern Aslian varieties have been based. One particular issue regarding the data of the study should however be noted, and that is that the different makeup of the published Tea-De data compared to the data set of the remaining varieties of the study resulted in a large number of missing forms for this variety (46 items were missing from the 146-item list, 11 from the hihi list and 9 from the lolo list). Thus while the overall patterns observed appear to be in accordance with what is known about the sociolinguistic situation of the Tea-De speakers (however limited this knowledge is), the findings regarding this variety must be treated with caution.

It is also possible that the particular sublists used for the comparison of more and less basic vocabulary are not ideally suited to the Northern Aslian context. In fact, a number of the list items chosen in other contexts to be resistant to borrowing (such as numbers two to five, salt, new) have been replaced with Malay loans in the majority of the Northern Aslian varieties. While the sublists used have been effective in representing likely scenarios for the history of the Northern Aslian varieties, adjustment of the makeup of the lists has the potential to increase their effectiveness in uncovering traces of genealogical and contact patterns among the varieties.

It is also important to note the exploratory nature of the methods used in the present study. While the method of comparing more and less retentive vocabulary subsets has given reliable results in previous studies, there are certain differences between the context of the current study and the contexts in which the method has been used previously. McMahon and McMahon (2003) found that the differences between networks based on more and less retentive vocabulary revealed traces of borrowing among certain Indo-European languages. Likewise, McMahon et al. (2005) found that while distances between groups of languages known to share ancestry decreased on the basis of the most basic vocabulary, the distance between the Quechua and Aymara groups *increased*, suggesting that the connection is one of contact. However, this method has not previously been tested at the within-family level. While this means that the findings must be treated with a certain degree of caution, it is clear that the method has generated interesting results and revealed clear patterns in the current study. The broad range of methods of analysis used in the study has allowed support for some findings to be strengthened and others to be discarded. In addition, the findings of the current study have lent support to previously proposed ideas about the history of the Northern Aslian varieties. Thus the analyses, in spite of their preliminary and exploratory nature, can be seen to contribute to the research into the Northern Aslian varieties, suggest patterns in the genealogical and contact relationships among the varieties, and point to directions for further study.

5.3 Directions for Further Study
The method of comparing phylogenetic relationships constructed on the basis of more and less basic vocabulary has been shown to be useful in differentiating the genealogical and contact-based aspects of language histories. The findings of the present study suggest that the method is likely to be successful in the context of other elusive language histories, even at the within-family level. Applied in the context of the entire Aslian language family, the method has the potential to offer important insights into the relationships between the Aslian varieties. Given the conservative nature of the phonology and lexicon of the Aslian languages (Benjamin, in press), a

better understanding of the classification and history of the Aslian languages is a crucial step in the reconstruction of proto-languages not only for the Aslian family but also for the Mon-Khmer family and in turn for Austro-Asiatic. Reconstruction of the history of the Aslian languages also has important implications for the speakers of these languages. While evidence from linguistics points to a long history of Aslian settlement in Peninsular Malaysia, the status of the speakers of Aslian languages as the indigenous people of the Peninsular is at present being challenged by the Malaysian government. Thus research into Aslian linguistic prehistory is of immediate practical use to the speakers of the Aslian languages in defending their right to occupy the land that they are thought to have occupied for millennia (see Benjamin, in press, for a discussion of local animal and plant names borrowed into Malay from Aslian).

The method also has potential for use in relation to issues of the correspondence of Aslian linguistic categories with ethnographic categories. In particular, the method has the potential to shed light on the genealogical and contact aspects of the relationships of the Northern Aslian varieties with the Central Aslian Lanoh and Southern Aslian Semaq Beri languages. While the way of life of the speakers of these languages resembles the Northern Aslian Semang, linguistically they are more closely connected with the non-forager speakers of the Central and Southern Aslian languages, respectively (Burenhult et al., 2011). Insights into the genealogical and contact-based patterns of the relationships of these varieties with the varieties of the Northern Aslian Semang have the potential to play an important part in our understandings of Aslian prehistory. Such insights not only have potential implications for the history of the Aslian languages but also more far-reaching implications for theories about human cultural development. In-depth study of the language varieties spoken by the Semang also has the potential to contribute to current understandings of the nature of language – while the history of human language is by far dominated by nomadic foraging populations like the Semang, linguistic theory is predominately based on knowledge of languages spoken in more sedentary circumstances (Benjamin, in press).

The analyses of the current study have suggested that the Menriq Rual variety, with its lexical divergence, its characteristic /r/ realisations, and its archaic first person singular form, is worthy of further study. The findings have suggested that this variety shows traces of an interesting relationship with the other Northern Aslian varieties as well as a complex history. Menriq Rual may be a remnant of a conservative Northern Aslian variety, which is genealogically related to the varieties of the Menraq group, but which has been affected by contact with the varieties of the Batek group. The findings of the present study have shown that the Menriq Rual variety is an interesting, urgent and thus high-priority target for future description and documentation, and that further study of this variety has the potential to contribute to our knowledge of the movements and histories of Northern Aslian speaking groups.

Furthermore, study of the varieties of Northern Aslian not included in the present study is likely to add greatly to our understandings of the connections among the Northern Aslian varieties, both those of genealogy and those of contact, as well as uncover clues as to the status and history of Menriq Rual. A thorough survey of the Northern Aslian varieties of Malaysia as well as the little-studied varieties of southern Thailand is crucial if the questions raised in the current study are to be answered. Indeed, the current study highlights the importance of linguistic surveys, a step that is often not prioritised in language documentation endeavours (Niclas Burenhult, personal

communication) – the existence of the Menriq Rual variety would not have been known to researchers had it not been for a linguistic survey. In particular, further investigations into the proposed existence of a Northern Aslian dialect continuum necessitate a much more detailed survey of the Northern Aslian varieties. Just as urgent is the issue of the scarcity of recorded data for many Northern (and other) Aslian varieties. This lack of data results in great difficulties in constructing reliable accounts of the classification and history of the Aslian languages. Thorough documentation of these fascinating languages should thus be a high-priority aim of future research.

References

Benjamin, G. (1976). Austroasiatic subgroupings and prehistory in the Malay Peninsula. In P. N. Jenner, L. C. Thompson, & S. Starosta (Eds.), *Austroasiatic Studies, Part I* (pp. 37−128). Honolulu: The University of Hawai'i Press.

Benjamin, G. (2009, March). *The Current Situation of the Aslian Languages*. Paper presented at Symposium on Language and Diversity, Nanyang Technological University, Singapore.

Benjamin, G. (in press). The Aslian languages of Malaysia and Thailand: an assessment. In P. K. Austin & S. McGill (Eds.), *Language Documentation and Description.* Vol. 11. London: Endangered Languages Project, School of Oriental and African Studies.

Bishop, N. M., & Peterson, M. M. (1993). *Maniq language survey report.* Bangkok: TUSIL- LRDP, Thammasat University.

Bryant, D., Filimon, F., & Gray, R. (2005). Untangling our past: Languages, trees, splits and networks. In R. Mace, C. Holden, & S. Shennan (Eds.), *The Evolution of Cultural Diversity: Phylogenetic approaches* (pp. 67−84). London: UCL Press.

Bryant, D., & Moulton, V. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution, 21(2),* 255−265.

Burenhult (n.d.) Unpublished field notes, DoBeS archive.

Burenhult, N., Kruspe, N., & Dunn, M. (2011). Language history and culture groups among Austroasiatic-speaking foragers of the Malay Peninsula. In N. J. Enfield (Ed.), *Dynamics of human diversity: The case of mainland Southeast Asia* (pp. 257-277). Canberra: Pacific Linguistics.

Campbell, L. (2004). *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Diffloth, G. (1975). Les langues mon-khmer de Malaisie: Classification historique et innovations. *Asie du Sud-Est et Monde Insulinde*, 6(4), 1−19.

Dunn, M., Burenhult, N., Kruspe, N., Tufvesson, S., & Becker, N. (2011). Aslian Linguistic Prehistory: A Case Study in Computational Phylogenetics. *Diachronica,* 28(3), 291-323.

Dyen, I., Kruskal, J. B., & Black, P. (1992). An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society,* 82, 1-132.

Heggarty, P., Maguire, W., & McMahon, A. (2010). Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Philosophical Transactions of the Royal Society B,* 365, 3829-3843.

Holden, C. J., & Gray, R. D. (2006). Rapid radiation, borrowing and dialect continua in the Bantu languages. In P. Forster, & C. Renfrew (Eds.), *Phylogenetic methods and the prehistory of languages* (pp. 19-39). Cambridge, UK: McDonald Institute for Archaeological Research.

Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution, 23*(2), 254-267.

Kruspe, N. (2009). Loanwords in Ceq Wong, an Austroasiatic language of Peninsular Malaysia. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the World's Languages: A Comparative Handbook* (pp. 659-686). Berlin, Germany: Mouton de Gruyter.

McMahon, A., & McMahon, R. (2003). Finding families: Quantitative methods in language classification. *Transactions of the Philological Society, 101*(1), 7–55.

McMahon, A., Heggarty, P., McMahon, R., & Slaska, N. (2005). Swadesh sublists and the benefits of borrowing: An Andean case study. *Transactions of the Philological Society, 103*(2), 147-170.

Phaiboon, D. (2006). Glossary of Aslian languages: The Northern Aslian languages of Southern Thailand. *Mon-Khmer Studies*, 36, 207–224.

Shorto, H. (2006). *A Mon-Khmer Comparative Dictionary*. Canberra: Pacific Linguistics.

Swadesh, M. (1952). Lexicostatistic dating of prehistoric ethnic contacts, *Proceedings of the American Philosophical Society, 96*, 452-63.

Appendix 1. The Wordlist Used for the Entire Northern Aslian Data Set

Table 5. Benjamin's (1976) modified Swadesh-list (146 items)

animal	back	bad	belly	big
bird	to bite	blood	to blow	bone
breast	to breathe	child	claw/fingernail	cloud
to cut	to dance	day	to die	to dig
dirty	dog	to drink	ear	earth
to eat	egg	eye	to fall	far
fat	father	to fear	fire	fish
to flow	flower	foot	fruit	full
to give	good	hair	hand	he
head	to hear	heavy	here	to hold
husband	I	knee	knife	to know
to laugh	leaf	left	to live	liver
long	louse	man	many	meat
moon	mother	mountain	mouth	name
near	neck	new	night	nose
not	old	person	to play	quiver
rain	red	rice	right	road
root	rotten	salt	to say	to scratch
to see	sharp	to shoot	short	sibling - elder
sibling - younger	to sing	to sit	skin	sky
to sleep	small	to smell	smoke	smooth
snake	spear	to spit	to squeeze	to stab
to stand	stick	stone	straight	to suck
to swell	tail	thin	this	thou
three	to throw	to tie	tongue	tooth
tree	to turn	two	to vomit	to walk
to wash	water	we	wet	what
when	white	who	wife	wind
wing	to wipe	woman	woods	you (pl)

Appendix 2. Comparisons of lexical distance among the Northern Aslian varieties on the basis of more and less basic vocabulary

Figure 9. Comparison of lexical distances between Tea-De and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b
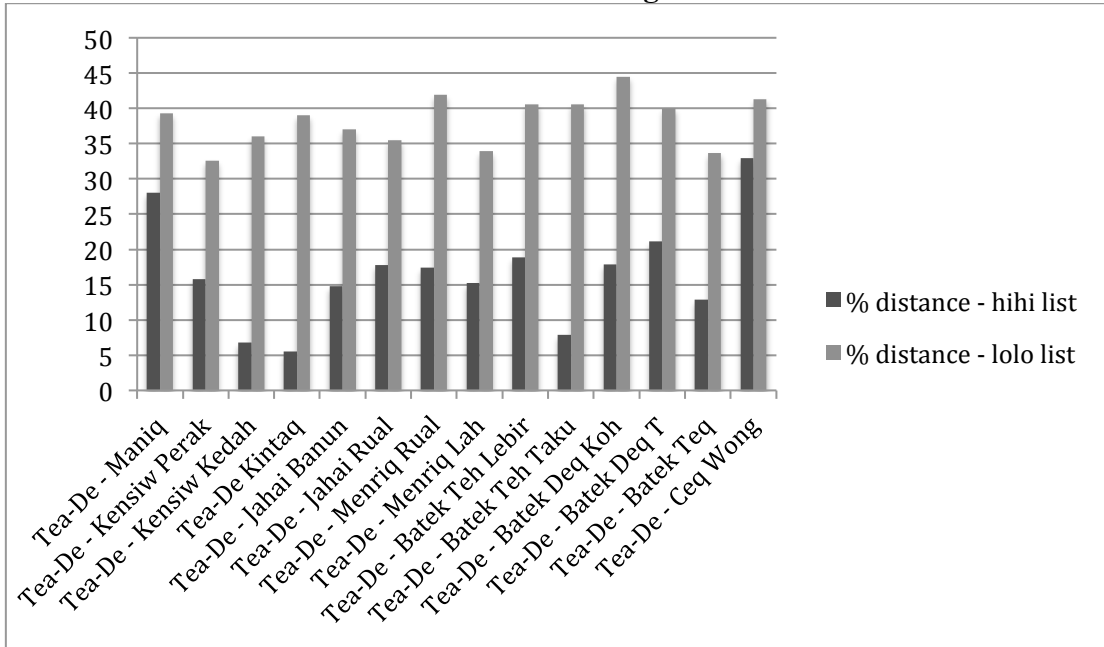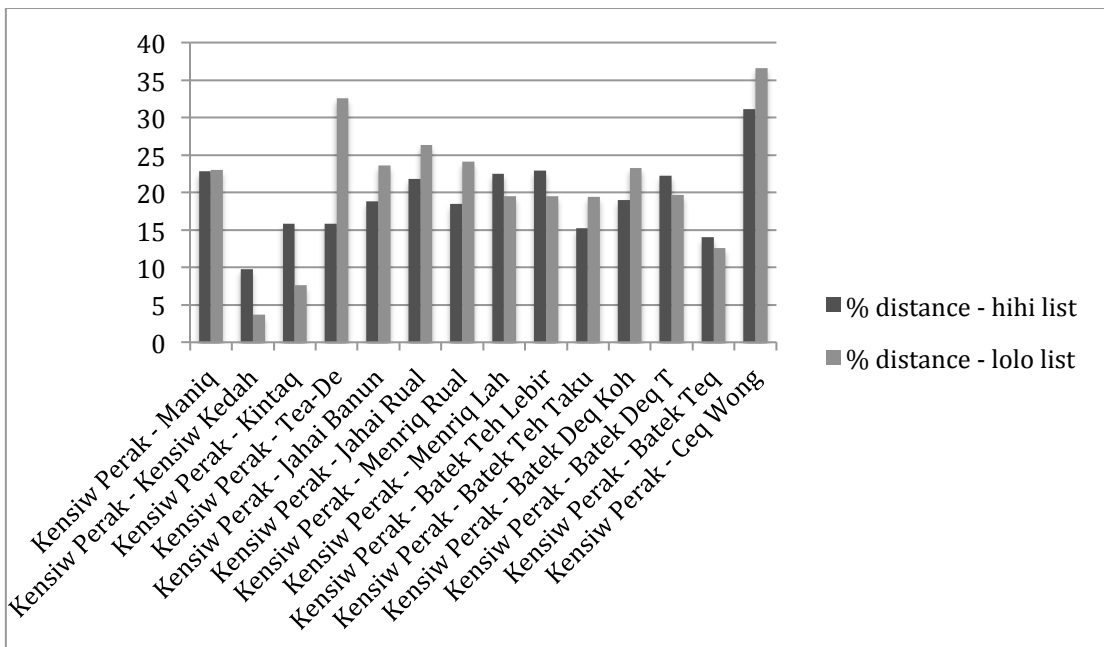

Figure 10. Comparison of lexical distances between Maniq and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b
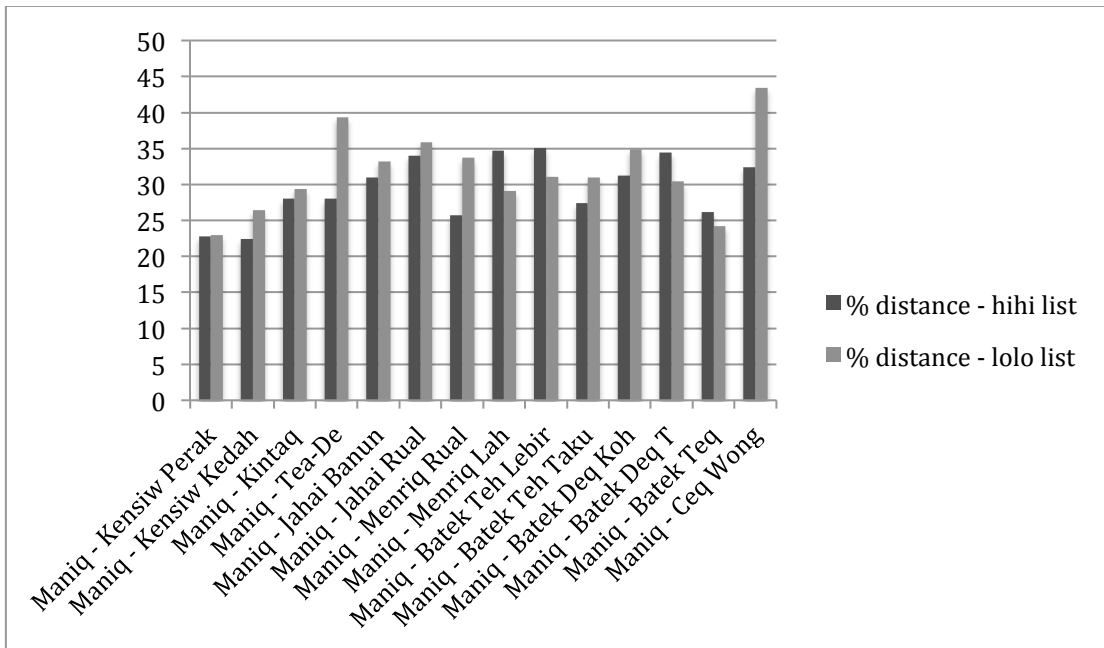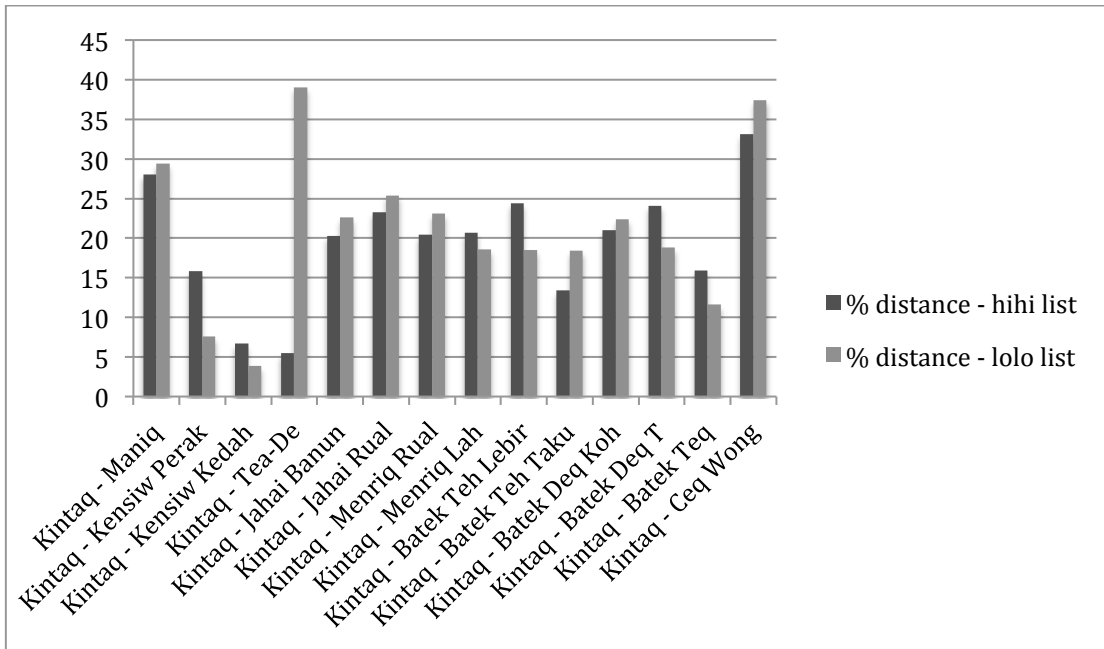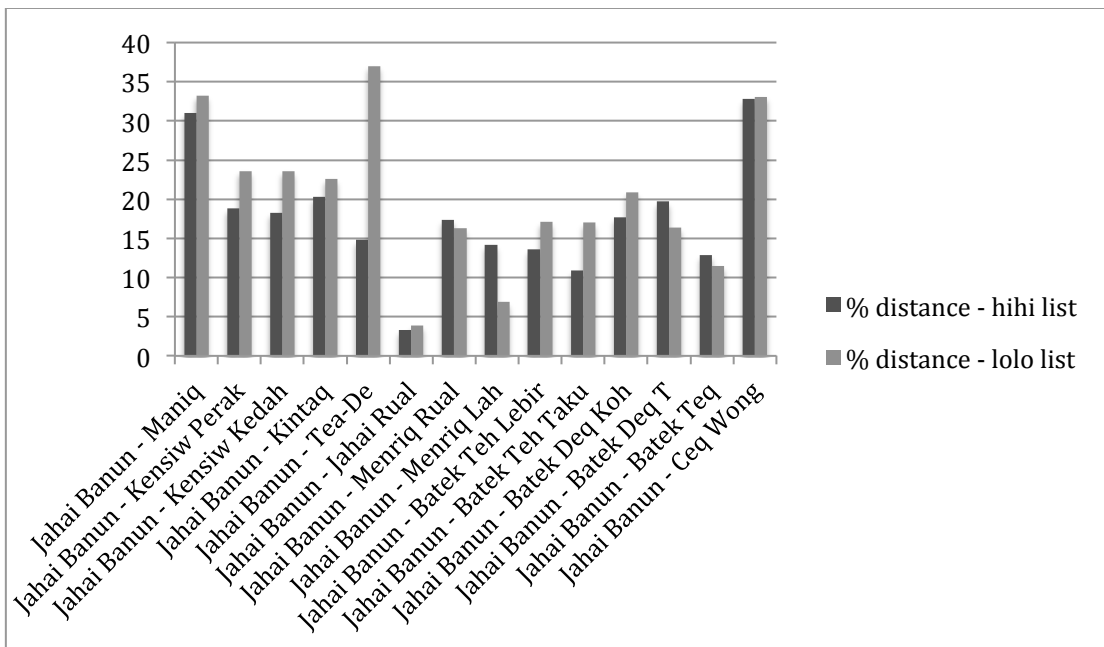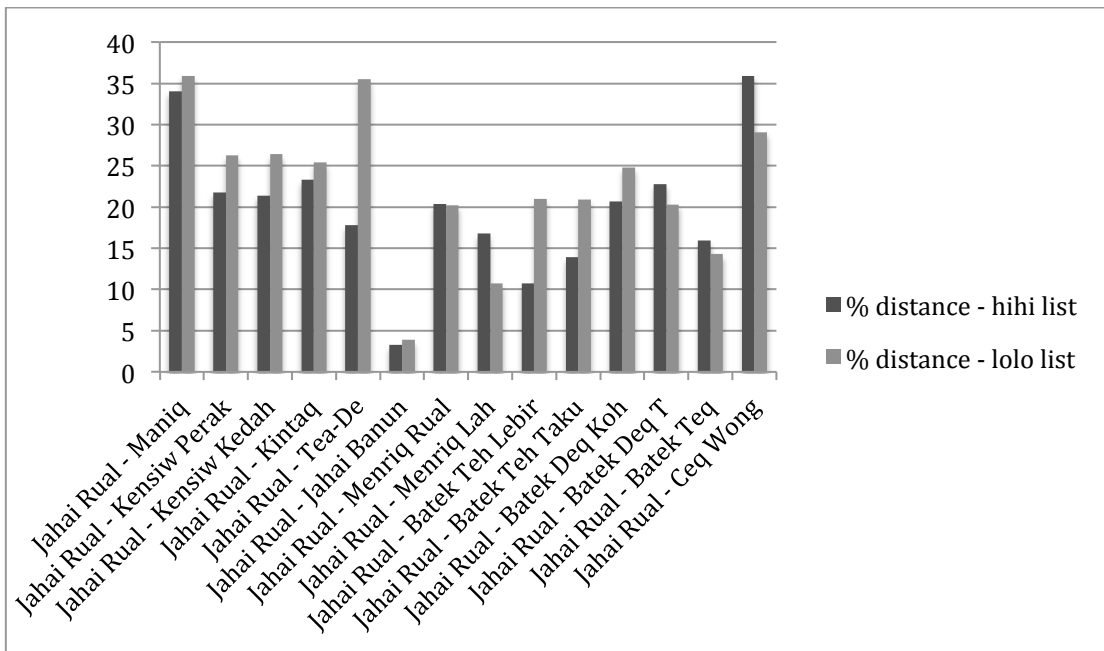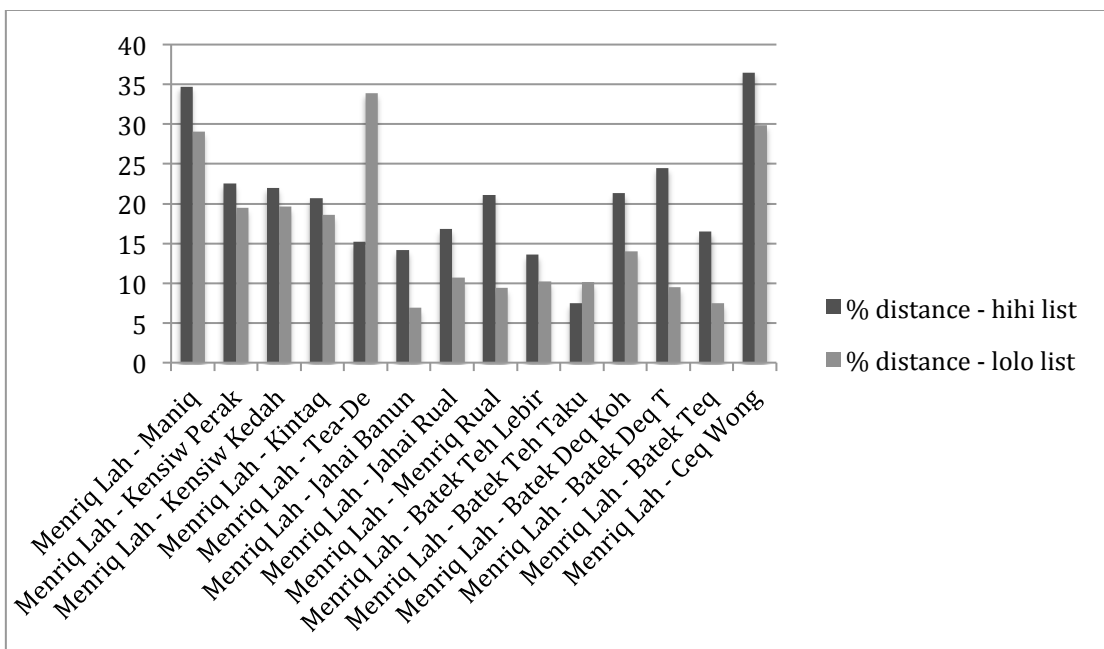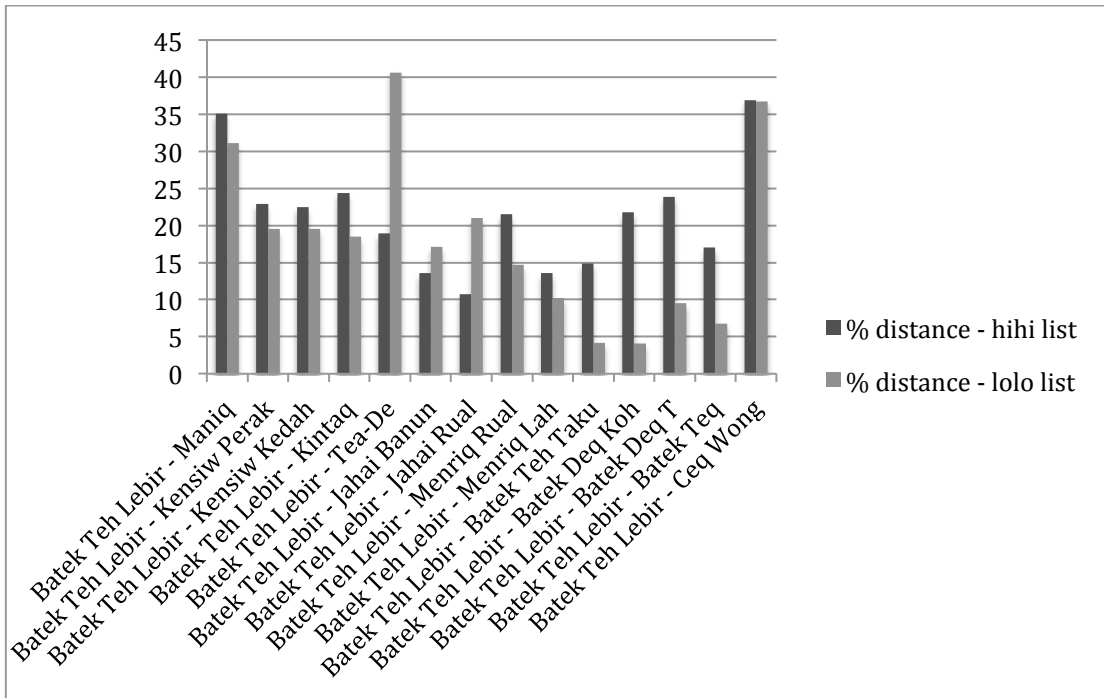
Figure 11. Comparison of lexical distances between Kensiw Kedah and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b
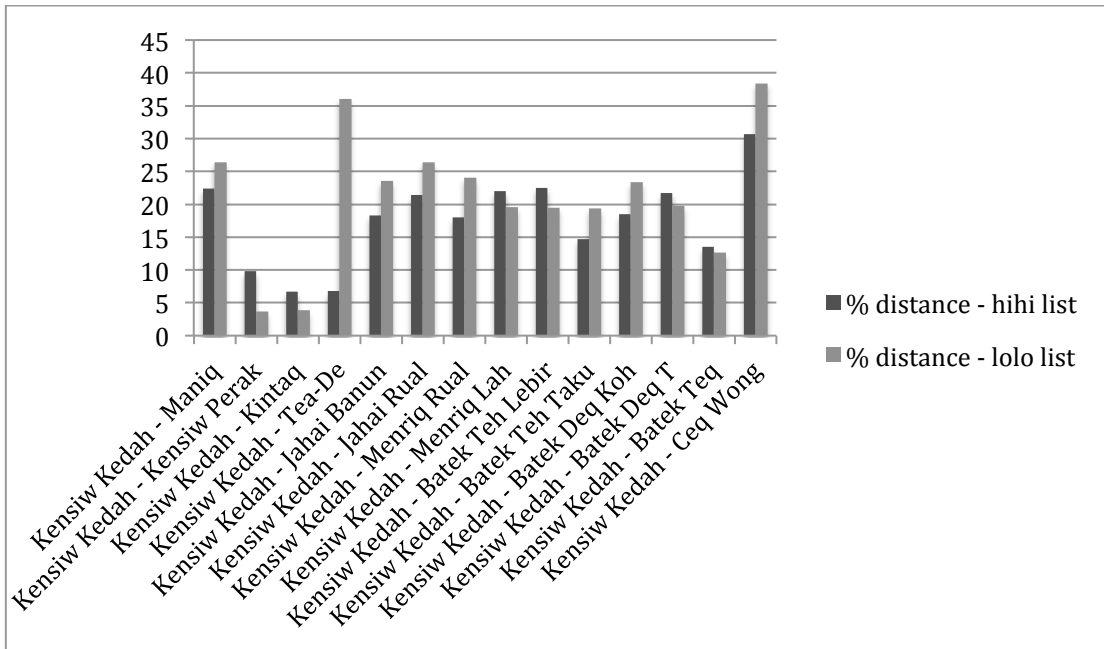
Figure 12. Comparison of lexical distances between Kensiw Perak and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b

Figure 13. Comparison of lexical distances between Kintaq and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b

Figure 14. Comparison of lexical distances between Jahai Banun and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b

Figure 15. Comparison of lexical distances between Jahai Rual and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b
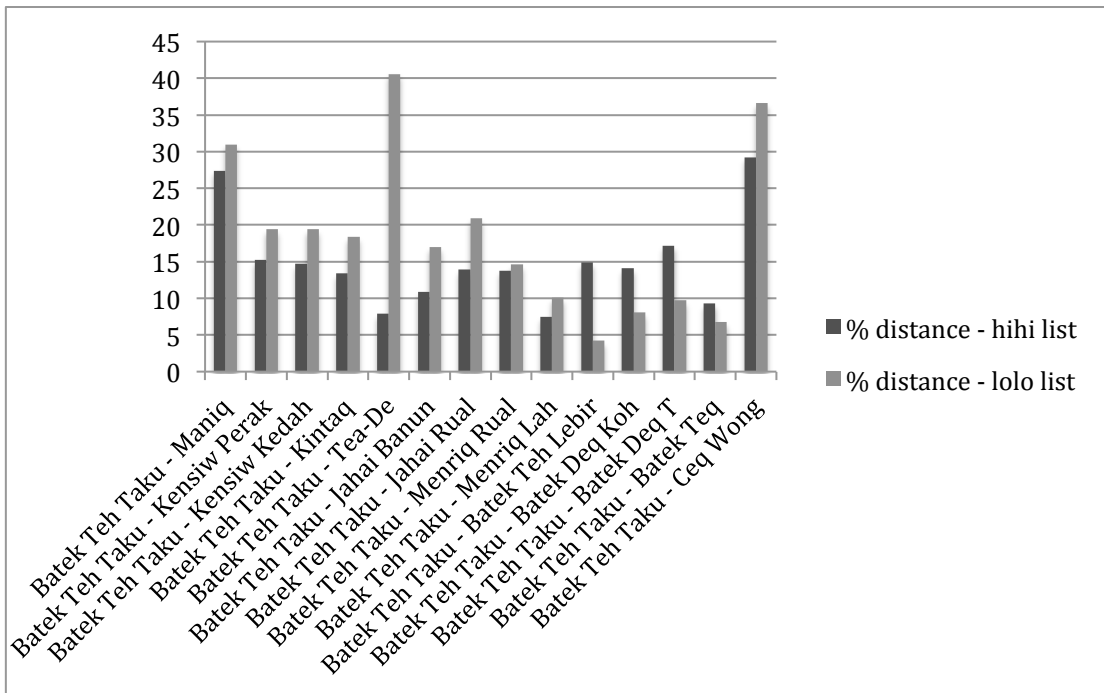

Figure 16. Comparison of lexical distances between Menriq Lah and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b

Figure 17. Comparison of lexical distances between Batek Teh Lebir and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b
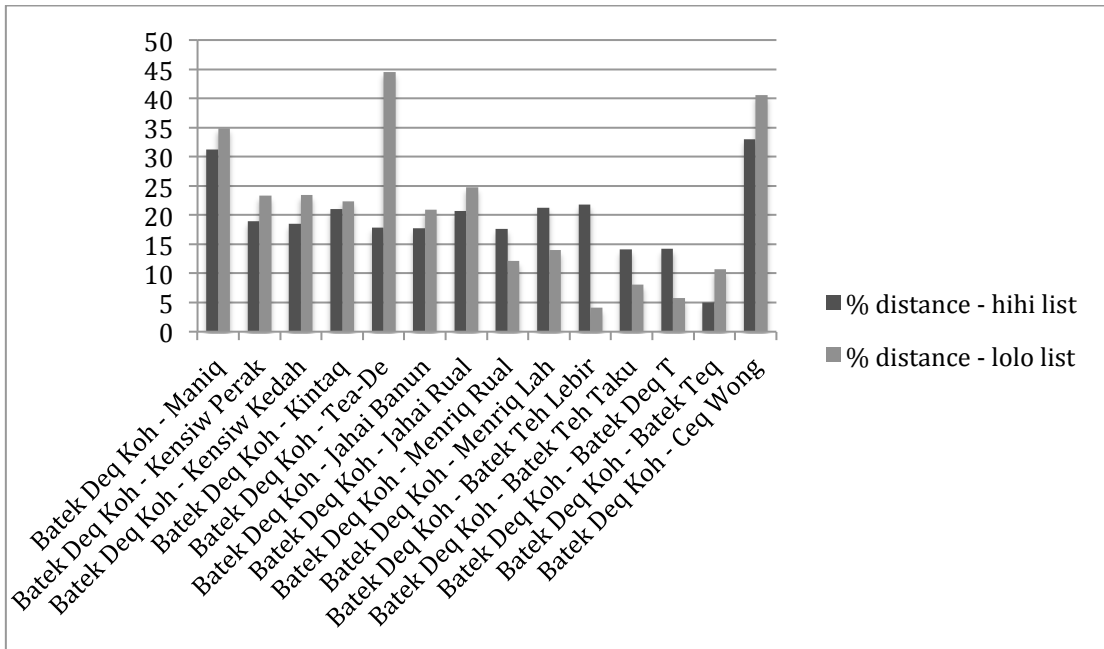

Figure 18. Comparison of lexical distances between Batek Teh Taku and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b
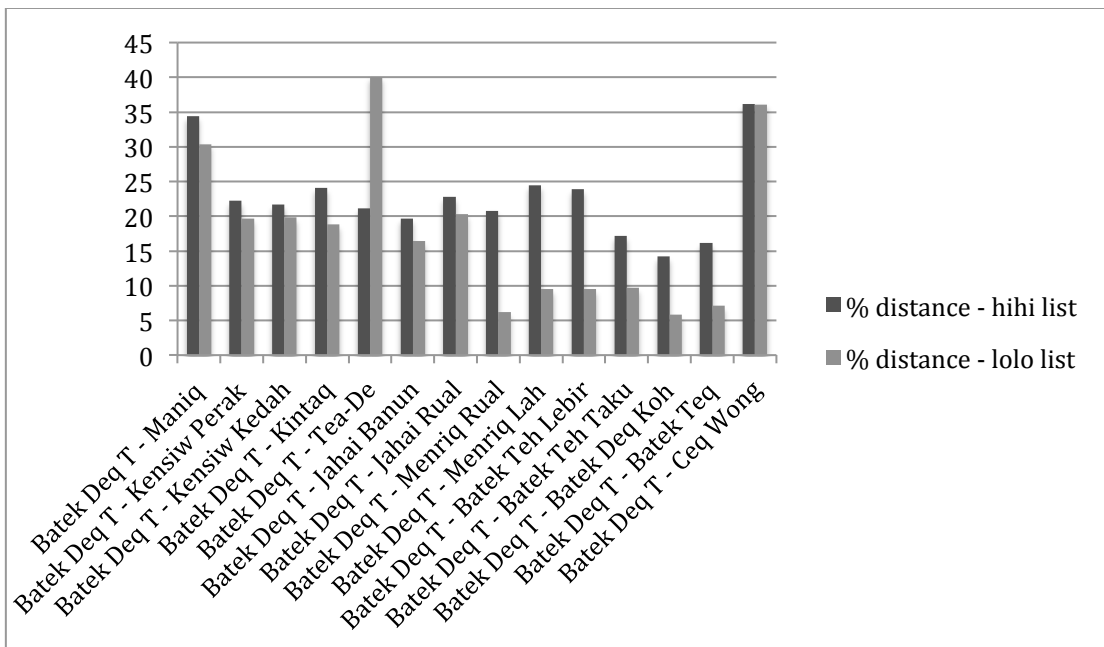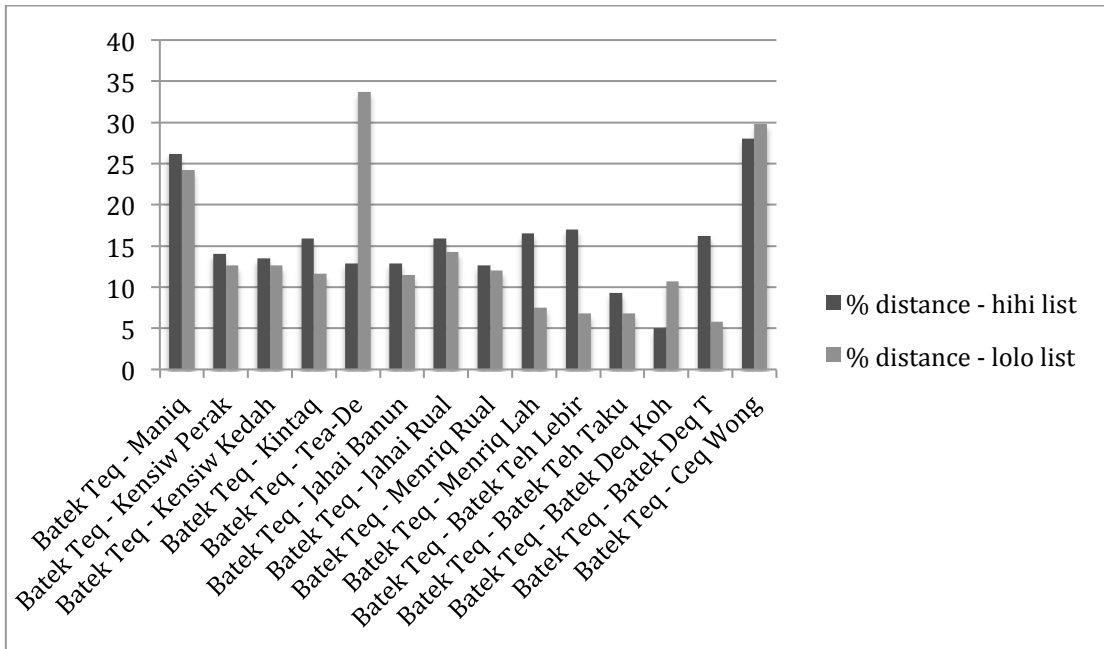
Figure 19. Comparison of lexical distances between Batek Deq Koh and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b

Figure 20. Comparison of lexical distances between Batek Deq Terengganu and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b

Figure 21. Comparison of lexical distances between Batek Teq and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b

Figure 22. Comparison of lexical distances between Ceq Wong and the remaining Northern Aslian varieties in the networks of Figure 7a and 7b