6.0 Introduction

This chapter attempts to provide a comparative study of the lexical items of the dialects of Khasi. In establishing genetic affinity between languages, many linguists have emphasized the importance of comparative lexical study. Lexical information is of paramount importance especially in the initial stages of classification because (i) it provides a larger number of independent and potentially differing items, and (ii) it offers a point of comparison among all languages in the sense that certain fundamental vocabulary items may be expected to occur in all languages. The comparison of various languages led to the assumption that some languages are related and they developed from a common source. In languages where documentations are available over a long period of time, there would be enough data for linguists to establish the genealogical relationship between those languages. However, not all languages will have documented records that cover a long period of history of a language. In such cases, linguists working on language reconstruction would be able reconstruct a proto-language only by comparing existing features of related languages or dialects. For example, in Khasi, the written scripts was developed in less than one hundred and seventy one years (171 years), hence, it will not be possible to conduct historical study based on written records. Scholars or linguists who want to work on this area will have to depend a great deal on systematic comparative study of the dialects and related languages to be able to reconstruct the Proto-Khasi language.

It is to be mentioned here that though comparative lexical study is important in genealogical classification, yet, this present work does not focus in establishing the genetic relatedness of the dialects. A major task of this study is to bring out the similarities and differences of the lexical items in the dialects undertaken for the study. The outcome expected from this study is the degree of relatedness between the dialects. This study focuses only on some important lexical items within the basic vocabulary list such as body parts, flora and fauna, natural elements, numerals, kinship, color terms, and
grammatical categories- verbs, adjectives, pronouns, prepositions and interrogatives are considered for comparative study. Before starting the discussion on comparative lexical study of the Khasi dialects, it is important to discuss the earlier works done in this area.

6.1 Previous work on comparative lexical study of Khasi dialects

Apart from Nagaraja (1996, 2004), no work has been done relating to comparative lexical study between Khasi dialects. In both of these works, the comparative study is between Standard Khasi and Lyngngam. According to Nagaraja, Grierson (1904) has classified Lyngngam, as one of the dialects of Khasi. He mentioned that in the late seventies, Bareh (1977), a Khasi scholar, doubted the appropriateness of Grierson’s classification by providing a few lexical items to make his point. It is in the light of Bareh’s observation that Nagaraja’s work (1996) has been undertaken to ascertain the position of Lyngngam in relation to Standard Khasi.

Nagaraja (ibid,38) observes that ‘nearly 40 percent of Lyngngam’s vocabulary is unrelated to that of Khasi. Though similarities may be found in some words, sound correspondences are not very regular’. With reference to the lexical items analyzed by him, Nagaraja found that out of 58 nouns, 38 items are related, 18 are unrelated, and the rest are problematic. Out of the 58 verbs, 21 are unrelated, 35 are related and two problematic. Out of 27 adjectives, 11 are unrelated, three are problematic and the rest are related. He further noted that with reference to numerals, there are great similarities between the two.

In addition to the lexical items, Nagaraja also presented a comparative study of the morphological features-pronoun marking, verb construction, word-formation, interrogatives, number and gender. The comparative phonological study has not been dealt with in detail.

Regarding the status of Lyngngam within Khasi, Nagaraja(ibid, 50) remarks that ‘the foregoing statements show clearly that the Lyngngam speech form is quite distinct from that of Standard Khasi. However, the question of whether this form should be considered as a dialect of Khasi or as a related language is a complicated one...’. He concluded by stating that future work is needed so as to be able to establish the status of Lyngngam within the Khasi language.
Nagaraja (2004) attempted to provide the possible time of separation from common/parent speech form of Khasi (Standard) and Lyngngam by adopting Swadesh’s method, for which a 200 word list was taken. In this work, Nagaraja provided a list of comparative vocabulary, followed by a discussion as to the determination of the number of possible cognates and non-cognates along with the problems involved in it. Nagaraja (ibid,44) mentioned that the most difficult part of this work is in determining whether an item is a cognate or non-cognates. He noted that as change is natural for all natural languages (speech forms), one can observe many changes between Standard Khasi and Lyngngam. On this basis, he said that the absence or presence of identical forms in both these speech forms is considered to be a sign of later development. Nagaraja further mentioned that the word list provided posed various problems in determining cognates. There are basically two types of entries: (i) cognates, that is, those having certain differences them numbering 87, and (ii) non cognates numbering 111. Within (ii), he stated that there are three sub-types a) those which are phonetically related, numbering 71; b) those which have identical phonetic shape in both the speech forms, numbering 30; and c) those having partial similarity, numbering 10 entries. In some of these entries, he mentioned that the similarity is in the first part of the word and in some in the second part. As it is difficult to decide whether these are really cognates or not, they have been treated as non-cognates. After analyzing the number of cognates and non-cognates, he applied a formula (based on lexicostatistics technique), so as to come up with the time separation between these two speech forms.

Based on the formula, Nagaraja (ibid,45) stated that these two speech forms became separated some 1500 years back from the common ancestor. He further stated that in spite of the limitation of this method, the study has thrown some light on the past stages of these two speech forms.

It is to be mentioned here that of all the scholars/linguists, who have worked on Khasi, Nagaraja, perhaps, is the only one, who have done some comparative works on Khasi dialects. Though, his works are not extensive in this area, yet, they have been of much importance to researchers working in this particular area.
6.2 Comparative lexical study of the dialects

For this study, 150 words are taken for comparative analysis. The lexical items are divided as under:

1. Body Parts: 20
2. Flora and Fauna: 30
3. Natural elements: 16
4. Kinship terms: 16
5. Numerals: 15
6. Color terminology: 5
7. Pronouns: 8
8. Verbs: 15
9. Adjectives: 16
10. Prepositions and Interrogatives: 9

The data referring to the above different lexical items are presented in Table 1-10, which are given at the end of this chapter.

6.2.1 Analysis of Body Parts: The data is presented in Table 6.1

Under body parts, 20 lexical items of the dialects have been selected. These lexical items are numbered from 1-20. Based on the data above, the following observations have been made:

With reference to body parts, it can be said that all the dialects seem to have similar/identical forms in most of the words. Their similarities are being based on phonetically related features/shapes. For example in data no.1, the term for ‘head’ is similar in all the eight dialects, where the variation is only in the vowel and in War-Jaintia, it is both in the vowel and consonant. However, there are some differences which have been identified. The difference is in certain lexical items which are phonetically unrelated in some of the dialects. These are seven in number. These are as under:

Data nos. 3, 4, 11 and 12 show that War-Jaintia has different forms, which are completely dissimilar compared to words from other dialects. In data no.13, War-Khasi shows a partially similar form, that is, the second part of the word has the same form,
when compared with the other dialects, and the first part is not related at all. In data no.10 and 17, Lyngngam shows a different unrelated form.

6.2.2 Analysis of flora and fauna: The data is presented Table 6.2

Out of the 15 lexical items taken for animals, 11 forms are found to be similar in all the dialects. 4 forms are found to be dissimilar and the dissimilarity is identified in the vocabulary of Lyngngam, War-Jaintia and War-Khasi. For example (data no.22), for the word ‘cat’ and ‘elephant’, all the dialects have similar forms, whereas in Lyngngam, the word is [miaŋgau] for ‘cat’ and [jawbaʔ] for ‘elephant’. Both the forms in Lyngngam are distinct from other dialects of Khasi. It is to be mentioned here that the word for ‘elephant’ found in the dialects is a borrowed term from Indo-Aryan source. This shows that Lyngngam is the only dialect which has a native term for ‘elephant’. Data no. 27 shows War-Jaintia has a different form for the word ‘horse’, whereas all the dialects have similar forms. War-Khasi shows a variant form in data no.35, which is unrelated with the others forms in the other dialects.

The analysis of the lexical items for birds shows that 6 out of 9 items are unrelated. For example, the word ‘duck’ (data no.38) is found to be similar in 6 dialects but in Pnar and War-Jaintia, the words are unrelated when compared with the other dialects. However, Pnar and War-Jaintia share similar forms, [rapasa] in Pnar and [rapasu] in War-Jaintia. The word shows that the variation is only in the final vowel.

The data provided in data no. 45-50 show that most of the words are related with an exception in War-Jaintia. For example, the word ‘tree’ (data no. 25) can be treated as similar forms in all the dialects, but in War-Jaintia, the term used is [ɔt], which is not related with the other dialects. The word ‘crow’ (data no. 40) is similar in all the dialects, but in Lyngngam and War-Jaintia, the forms are different. In Lyngngam, the form is [kawa] and in War-Jaintia, it is [lətua]. Both the forms are phonetically unrelated compared with the forms of other dialects. Lyngngam has a different form for the word ‘crane’ as shown in data no. 42. For the word ‘pigeon’, Pnar and War-Jaintia seem to share similar forms, which are different from others.
6.2.3 Analysis of Natural Objects: The data is presented in Table 6.3

Out of 16 items (data 51-66), 9 are found to be unrelated. The unrelated forms are found to be more in War-Jaintia. The examples of these unrelated forms in War-Jaintia are [pnuai], ‘moon’ [khloʃmin] ‘star’, [juːɾɛ] ‘sun’, [pdeŋ] ‘mountain’, [ɕəbliaŋ] ‘thunder’, [pəɾhut] ‘wind’. Data no. 52 [khloʃmin] ‘star’ in War-Jaintia shows partial similarity (similarity in the first part of the word) with the other dialects.

In Lyngngam, there are three forms which are distinct from the other dialects. This is seen in data nos. 53, 56 and 60 in words [jawɾɛj] ‘sun’, [gmiaŋ] ‘earth’, and [naw] ‘lightning’.

It has been observed that War-Jaintia and Bhoi (Tyrso) show similar forms for the word ‘lightning’, Bhoi (Tyrso) [clabaŋ] and War-Jaintia [ɕəbliaŋ].

The data nos. 62 and 64 show that all the dialects share the same form for the words ‘ice’ and ‘cloud’. All the forms have the same phonetic shapes.

6.2.4 Analysis of Kinship terms: The data is presented in Table 6.4

The data under kinship terms show that most of the lexical items are similar in all the dialects, with some exceptions in War-Jaintia and Lyngngam. In Standard Khasi, the word for ‘child’[khun], for ‘male’ [ʃnranŋ] and ‘female’ is [kinthɛj]. Other dialects (other than War-Jaintia and Lyngngam) have similar forms, though there may be slight variation. In Standard Khasi, the marker for feminine is ‘ka’ and ‘u’ for masculine, which is the same in other dialects. Thus, for the word ‘son’ and ‘daughter’, the lexical items referring to male and female can be omitted, and it can be understood simply by using the term [khun] preceded by gender markers ‘u’ and ‘ka’. This is shown in data nos. 70 and 71. But this is not the same in Lyngngam, as it has been found that no gender marker is employed in Lyngngam. Hence, for the terms ‘son’ and ‘daughter’, the form for ‘child’ is used along with the forms for ‘male’ and ‘female’ as shown in data nos. 71 and 72. Interestingly, in War-Jaintia, though there is the presence of gender markers, yet, it is similar to Lyngngam in the use of the terms for referring to ‘son’ and ‘daughter.
6.2.5 **Analysis of numerals**: Data presented in Table 6.5

Under numerals, most of the items are related with the exception of some of the lexical items in War-Jaintia. Data nos. 89 [hənthɛ] ‘seven’ and 91 [honʃʔɛ] show that War-Jaintia has a different form. The words are partially similar when compared with the other varieties. The similarity is only in the first part of the word, whereas the second part is totally different. Further, the same is applied to data nos. 93 and 94. It is to be mentioned here that data nos. 96 and 97 show similar forms in all the dialects. The reason for this can be attributed to borrowing, as these words are from Indo-Aryan source.

6.2.6 **Analysis of Color terms**: Data presented in Table 6.6

The lexical items taken for color terminology are only the basic color terms, numbering 5 of them (data nos. 98-102). With reference to color terms, similar forms are found in all the dialects, with exception in War-Jaintia for the words ‘white’ and ‘yellow’. Nongtrai and Lyngngam show different form in the word ‘green’ as shown in data no. 101. In both of these dialects, the forms referring to ‘green’ are related.

6.2.7 **Analysis of Pronouns**: Data presented in Table 6.7

The pronominal markers are found in all the dialects of Khasi. The pronominal markers are phonetically similar in all the dialects, with exception of Lyngngam and War-Jaintia. For example, data nos. 104, 107, 108, 109 (forms marking IPPI, IIPFem, IIPMas and IIPPl), show that Lyngngam has different forms which are not related with the forms present in other dialects of Khasi. It is to be noted here that in Lyngngam, forms in data nos. 108 and 109 are not used as markers of gender, while in other dialects of Khasi, the III P pronouns are also used to mark gender. War-Jaintia shows different forms of markers as shown in data nos. 105, 106, 107, 108, 109 (forms marking IIPSg, IIPPl, IIPFem, IIPMas, IIPPl) which are neither related with other dialects nor with Lyngngam. A careful examination of the data nos. 105 and 107 show that War-Jaintia has a same form to mark IIPSg(Fem) and IIPFem. When this particular question was asked to the informants, the reply was that the form can be understood according to the context.
6.2.8 Analysis of Verbs: Data presented in Table 8

Out of 15 verbs (include motion, static, psych verbs) taken for the analysis as shown in table 8 data nos. 111-125, 8 of them are found to be phonetically related in all the dialects. In phonetically related forms, the variation is seen in either the vowel or consonant. For example, the word ‘drink’ (data no. 112) shows that Standard Khasi, War-Khasi, Bhoi (Nongpoh) and Pnar have the same form that is, [diʔ] ‘drink’, whereas in Bhoi (Tyrso) and Nongtrai, the variation is in the final consonant, that is [ʔ] is replaced by [c], in Lyngngam, the variation is in the vowel and consonant, that is, [i] is replaced by [ɛ] and [ʔ] is replaced by [c]. In War-Jaintia, there is a variation in the vowel, where [i] is replaced by [e]. Data no.113 shows that Standard Khasi and Nongpoh share the same form for the word ‘walk’, that is [jac], Pnar and War-Jaintia have the same form [kap˺], War-Khasi and Bhoi(Tyrso) share similar form, and, Nongtrai and Lyngngam, show different forms which are unrelated. The word for ‘love’ (data no.118) shows that Nongtrai, Lyngngam, Pnar and War-Jaintia have different forms. It has been observed that Pnar and War-Jaintia have similar forms, whereas, in Nongtrai and Lyngngam, the forms are different, which are not related with other dialects of Khasi. With reference to verbs, it has been found that War-Jaintia and Lyngngam are the two dialects, which seem to have more different forms, when compared with other dialects. Interestingly, an important observation which has been made on Lyngngam is that, in most of the verb forms, there is the presence of [ə] in the initial syllable of the word, which supports the presence of sesquisyllabic structure in the dialect.

6.2.9 Analysis of Adjectives: Data presented in Table 6.9

A total number of 16 adjectives from the dialects are taken for comparison. Out of the 16 adjectives, 5 of them are found to be similar in all the dialects. These are shown in data nos. 131, 133, 134, 135 and 136. There are two forms which are similar in all the dialects but are different in Lyngngam and War-Jaintia (data nos. 126, 127). Lyngngam shows a different form in data no.128, while all the other dialects have similar forms. Though the other dialects seem to have similar form for the word ‘thin’(data 129),
Nongtrai and Lyngngam have same forms which are different from others. It is to be mentioned here that in some dialects of Khasi (not included here), \([k^h\text{rew}]\) is a word which is used to refer to a person who is thin and sickly, hence, it can be said here that the absence of lexical item \([\text{rajk}^h\text{r}?]\) meaning ‘thin’ in the vocabulary of Lyngngam and Nongtrai is compensated by \([k^h\text{ru}/k^h\text{ro}]\). In data no. 140, Standard Khasi has a different form compared with other dialects. Some scholars proposed that the word \([b^h\text{a}]\) meaning ‘good’ is borrowed from Indo-Aryan language, either Assamese or Bengali. The data shown in 140 also shows that in addition to Standard Khasi, War-Khasi and Lyngngam also have different forms. Bhoi (Nongpoh), one of the dialects, which commonly have similar forms with other dialects, seem to have a different form in data no. 141. In addition to Bhoi (Nongpoh), Nongtrai, Lyngngam and War-Jaintia also have forms which are distinct from other dialects.

6.2.10 PREPOSITIONS AND INTERROGATIVES: The data are presented in Table 6.10

With reference to preposition, all the dialects seem to have similar forms, though there are some which have distinct forms, and which cannot be related with the others. The most notable forms are \([\text{pl}\text{len}]\) in Lyngngam and \([\text{p}\text{rli}]\) in Nongtrai meaning ‘in’ as shown in data no. 143. However, War-Jaintia has a related form, that is, having partial similarity, and this is seen in the second part of the word. Data no. 144 shows that War-Jaintia has a different form, which phonetically unrelated, when compared with other dialects.

In analyzing data nos. 148, 149 and 150, it has been found that the interrogative formation is quite complex. All the question markers (bound morphemes) are suffixed probably, to pronoun morphemes. This is a pattern in all the eight dialects. Though the pattern is the same, yet, there is a difference noted in the markers. This can be illustrated by taking data no. 148. Standard Khasi, Nongpoh, Tyrso, Nongtrai Pnar and Pdenghakap have similar forms. Their similarities can be traced phonetically. However, this is not the same with War-Khasi and Lyngngam, as they seem to have complete distinct forms. For the interrogative ‘why’ (data no. 149), Nongtrai, Lyngngam and War-Jaintia have different forms, whereas in other dialects, they seem to have identical phonetic shape.
Data no.150 shows that Standard Khasi, War-Khasi, Nongpoh and Tyrso have similar phonetic forms and they are can be easily identified. However, Nongtrai, Lyngngam, Pnar and War-Jaintia have different forms. It is noted that the forms in Pnar and War-Jaintia seem to be related, that is, $[ka-te/u-te]$ change to $[kəɛ/u-ɛ]$ in War-Jaintia. The variation in the two dialects is that $[t]$ in Pnar is deleted $[ø]$ in War-Jaintia.

The data and analysis of the eight dialects presented above give an idea of the amount of divergence between the dialects. Nongdaju (representing Lyngngam dialect) and Pdengshakap (representing War-Jaintia dialect) show more divergences when compared with other dialects. Interestingly, what has been noted is that, whenever any different/unrelated forms are identified in the data, in most cases, it is either Lyngngam or Pdengshakap (in some instances both the dialects), which show the presence of unrelated forms. However, the Bhoi dialects (represented by Nongpoh and Tyrso), War-Khasi dialect (represented by War-Khasi), Standard Khasi, and to some extent Nongtrai (represented by Nongstoin), seem to have close resemblance between them. This observation is made on the basis that many lexical items (as shown in data nos. 1-150) are phonetically similar in these dialects. However, the observation made may be not true in all instances, as there are some lexical items (as shown in the analysis presented above), in which the dialects exhibit unrelated forms. Another observation which can be made based on the analysis is that dialects spoken in contiguous areas display some similarities in their lexical items, when compared with dialects spoken in farther areas. For example, the amount of lexical items varies greatly in War-Jaintia and Lyngngam, yet, both the dialects seem to share some similar forms with dialects adjacent to them. In terms of geographical location, Pdengshakap (War-Jaintia) is nearer to Jowai (Pnar), and Nongdaju is nearer to Nongstoin, hence, one can find a number of lexical items in these dialects which are phonetically related. This same fact can be accounted for Nongpoh which shows closer resemblance to Standard Khasi. This is as expected, since people/villagers from Pdengshakap and Nongdaju have more communication with people from Jowai and Nongstoin respectively, as the latter are commercial places and district headquarters.
A comparative lexical study of the eight Khasi dialects shows that six of the dialects are more closely related to each other. This is based on the fact that these dialects share more related forms, when compared with the other two dialects- Lyngngam and Pdengshakap. However, based on the analysis, it can be said that all the dialects are related to a certain degree.