Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study

Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)

==

1. Introduction

In this era of information technology explosion, translation is inevitable and has a major role to play in the development of languages and knowledge. The use of modern technology in Computer Aided Translation can help us translate quickly. A question arises as to whether a computer can fruitfully be utilized for translation. In this paper, an attempt is made to look into how far a computer can be used for translation. It seems, so far, no fruitful machine translation is done among Indian languages.

The number of structures in any language is finite to a greater extent. It is estimated that there may be around 40 structures and hence, they are very much rule based, and language-specific features may, sometimes, add a few to it. Similarities and differences between languages can be studied by Contrastive Analysis and it may help to evaluate the potentiality of computer in translation. This study brings out cross-linguistic similarities as well as differences between Manipuri and Tamil and tries to elicit the tools necessary for the

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study   397

translation. Further, a model of computer aided translation is also proposed. This will help us in producing translation tools as well for translation.

Scholars in the field of translation suggest a corpus-based translation to make the translation more successful and foolproof. This needs enormous size of corpus to make a successful translation; such an attempt is being made for corpus by LDC-IL, CIIL, Mysore. Furthermore, they try to make an automated POS tagging also, which will definitely help in machine translation.

This study stands on the view that computer can be used as an aid for very quick translation to support a human translator. In other words, the computer may translate quickly an enormous amount of texts with its limitations, such as giving equivalences, etc. for which a *bilingual dictionary (tool 1)* is necessary. Many of the decisions are to be taken by the human being involved in this kind of translation. Many of the structural-grammatical features are to be learned by practice only, for example, the PNG markers in Tamil. Computer may do the same after a long time practice. Only when these features are tagged to the words, a computer can translate them; otherwise only nonsensical translation would result in.

2. Computer based Translation

In this context, a design with a rule-based approach for Computer-Aided Translation is proposed. Since there is not much detailed information available to us, this paper depends on the translation theories and practices which have proposes five tasks for translation. The five tasks are:

1) Interpret the source language text, i.e., the pre-draft
2) Compose the translation
3) Conduct the research needed to complete the tasks (1) and (2)
4) Check the draft translation and correctness, and
5) Decide the implication of the communication, i.e., the post-draft

In the absence of clear-cut guidance, the work may begin with the basic question of how to make the computer work sensibly on the given design. Translators need equivalent sentence structures to translate. To start with, a comparative list of sentence structures and a bilingual dictionary may be prepared. Based on these a morphological analyzer and POS tagging have to be produced for computer aided translation.

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study   398

3. Manipuri and Tamil Structures

A few Manipuri and Tamil sentences are taken for analysis in this paper. Since both the languages belong to two different language families, there are many structural differences between them.

Normally, regular verbs take nominative subject and the PNG markers are added to the finite verb in Tamil. The defective verbs in Tamil take dative subjects; the person, number and gender markers (PNG) are not added to them, when conjugated for different subjects.

Example

1. enakku oru puttakam veeNDum.   'I want a book.'
2. enakku tamizh teriyum.   'I know Tamil.'

These sentences do not have equivalent structures with dative subject in Manipuri but have nominatives as the subject.

3. əi lairik əmə paammi.   'I want a book.'

4. əi tamil khəŋ-i.   'I know Tamil.'

Since there is no equivalent structure it may be difficult for the computer to translate. Whenever there are no equivalent structures available in both the languages, the translation of the sentences can be done using the *phrase structure tool (tool 2)* and *structure tool (tool 3).*

Another point of reference is the third person singular pronoun /məhak/ in Manipuri which can be equated to /avan 'he-non-hon.' or avaL 'she-non-hon' or avar 'he/she-hon.'/ in Tamil and the contrastive grammar should have tags to help the selection of the equivalents in such contexts. This *contrastive grammar (tool* 4) is the next one to be prepared for computer aided translation. Only a few examples are given here, but a detailed study of the features is to be carried out.

Features like subject verb agreement, case frames and the verbs, use of /-nə/ as both nominative and instrumental case marker, etc., in Manipuri are to be well defined for the use of the computer. This would clearly mean that *computational grammars (tool 5)* of both the languages are to be produced as a tool to help.

Some of the problematic contrasting structures are mentioned here as examples.

The first person plural /əikhoi/ 'we' in Manipuri has two equivalents /naaŋkaL/ and /naam/ in Tamil.

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil
Case Study   399

For example, for the Manipuri sentence,

     5. əikhoi yumdə cətli.   'We go home.'

there is a possibility of translating the sentence into Tamil either as

     6. (a) naaŋkaL viiTTukku pookiRoom. (excluding the hearer)

     or

     6. b) naam viiTTukku pookiRoom. (including the hearer)

Only the context can give us the clue for the choice. This is true of the third person singular pronoun /məhak/ also.

Example

     7. məhak-nə həi cai.   'He eats fruit.'

 This may be translated as

     8. (a) avan pazham caappiTukiRaan. (indicating masculine singular-non-hon.)

         He eats fruit.

     or

     8. (b) avar pazham caappiTukiRaar.   (indicating masculine singular-hon.)

        'He/she eats fruit.'

     or

     8. (c) avaL pazham caappiTukiRaaL. (Indicating feminine singular-non-hon.)

        'She eats fruit.'

     9. makhoi(-nə) həi cai.   'They eat fruit.'

This sentence may be translated as

     10. (a) avarkaL pazham caappiTukiRaarkaL.

        'They eat fruit.' (they- human)

     or

     10. (b) avaikaL pazham caappiTukinRana.

        'They eat fruit.' (they-neuter)

This problem arises because of equating /məkhoi/ as

    /avarkaL/ 'they- human' or /avaikaL/ 'they- non-human'

    Out of context translation will lead to such problems. But sentences of the following types would definitely solve the problem to a greater extent only when we mark every word for POS tagging for its features and the equivalents in Tamil.

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study   400

11. tom ŋəsi əyuk əikhoigi yumdə lak-i, məhak məNipurigi nobel əmə pammi.

'Tom came to our house today morning, he wants a Manipuri novel.'

Here, /məhak/ definitely represents /tom/, a masculine singular pronoun which may be equated with /avan-he/ (avar 'he/she-hon.' or avaL 'she-non-hon.' is also possible) in Tamil. But the solution to the problem is not that easy. How will the computer recognize this anaphoric reference? A human mind can easily bring out this by mere observation. The immediate first draft translation would be

12. *tom inRu kaalai eŋkaLuTaiya viiTu vantaan.

*tom inRu kaalai eŋkaLuTaiya viiTu vantatu.

avanukku (/ avaLukku / atukku) oru (/ onRu) maNippuri naaval veeNTum.

But the sentence should have the bracketed features also to make a good sentence.

13. tom inRu kaalai(yil) eŋkaLuTaiya viiTu(kku) vantaan.

avanukku oru maNippuri naaval veeNTum.

14. tom eŋkaLuTaiya viiTTukku inRu kaalaiyil vantaan.

avanukku oru maNippuri naaval veeNTum.

As discussed earlier, only the context of occurrence of a sentence can lead to correct translation.

4. Gender and Number

Deciding the gender and number of a noun in Manipuri is not that easy for translation. This is due to the fact that there is no subject-verb agreement in Manipuri. When a problem of subject verb agreement comes after a sentence is translated, only a translator can do this in the absence of *a grammar checker (tool 6) and a morphological analyzer (tool 7)*. Example

15. nupiməcadu phəjəi.

ciRumi azhakaaka irukkiRaaL.

'The girl is beautiful.'

16. cəukidu phəjəi.

*naaRkaali azhakaaka irukkiRaaL.

'The chair is beautiful.'

But in Tamil, the second sentence above is not acceptable. It should be

17. naaRkaali azhakaaka irukkiRatu.

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil
Case Study   401

(and not irukkiRaaL, which is feminine singular conjugation)

This should have the correct subject and verb agreement, namely, the third person neuter singular PNG marker.

Let us see another example.

18. tombə waŋ-i.

   toompa uyaramaaka irukkiRaan.

   'Tomba is tall.'

19. kutubminaar waŋ-i.

   *kutubminaar uyaramaaka irukkiRaan.

   'Kutubminar is tall.'

The second sentence in Tamil should be

20. kutubminaar uyaramaaka irukkiRadu.

Hence, it may be concluded here that a mere tagging of the words for gender as feminine, masculine or neuter alone will not lead to acceptable translation. *POS Tagging tool (tool 8)* can solve gender problem of individual words only. Hence, it should be understood that the computer can retrieve the equivalent words with the greatest possible speed for the translator and the remaining has to be done manually by the translator.

In Tamil the gender is more of a natural one. The classification of gender and number in the third person is as follows:

21. (a) Human

Masculine Singular	avan	'he-non-hon.'
Feminine Singular	avaL	'she-non-hon.'
Honorific Singular	avar	'he/she-hon.'
Human Plural	avarkaL	'they-human'

21. (b) Non-human

Neuter Singular	atu	'it'
Neuter Plural	avai	'they- non-human'

A dictionary with these details of word usage alone is not sufficient for translation. The above type of classification is not there in Manipuri, hence there are problems in translating Manipuri sentences into Tamil and vice versa.

Take for example,

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study   402

22. sənbinə səŋgom pi.   'The cow gives milk.'

This sentence may be translated as

23. (a) *pasu paal tarukiRaaL.   'She gives milk.'

or

23. (b) *pasu paal tarukiRaan.   'He gives milk.'

Both are unacceptable sentences. The correct sentence should be

24. pasu paal tarukiRatu.   'The cow gives milk.'

Only when we have the grammar checker this correct sentence will be the out-come.

5. Numerals

The numerals in Manipuri are also problematic in translation. This is due to the fact that when the nouns precede the numerals, the nouns do not take the plural marker.

Example

25. əigi lairik tərə ləi.   'I have ten books.'

This sentence may be translated into Tamil as

26. enniTam pattu puttakaŋkaL irukkinRana.

27. yum əsidə mi tərə ləi.   'There are ten persons in this house.'

The equivalent sentence in Tamil is

28. inta viiTTil pattu aaTkaL irukkiRaarkaL.

In both the Tamil sentences, the plural marker /-kaL/ is added to the nouns. This plural marker is not necessary in Manipuri.

6. Case Markers

The case markers are always problematic since they are language specific. More often mere equivalents in the form of comparative statements alone will result in unaccepted sentences.

Example

29. əi-nə kələm-nə cithi əmə i.

      naan penaa-v-aal oru katitam ezhutukiReen.

      'I write a letter with a pen.'

Here, the marker /-nə/ functions both as nominative and instrumental marker with nouns. Dative and locatives markers are same as in the following sentences.

30. imanə əiŋondə phurit əmə pi.

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil
Case Study   403

en amma ennakku oru sattai koTukkiRaar.
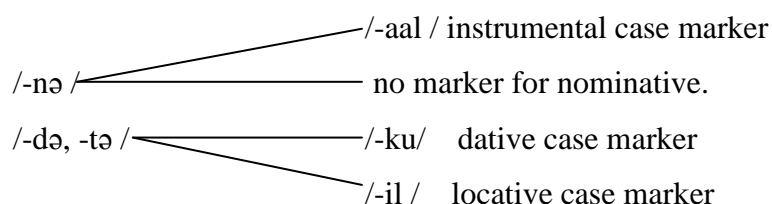
'My mother gives me a shirt.'

31. əi yumdə ləi.

naan viTTil irukkiReen.

'I am in the house.'

The use of the above mentioned markers depend on the grammatical context.

A study of the case markers reveals the following facts. This cross over is a problematic one when translation is taken up. Manipuri cases are equated here

/-nə / ⎯⎯⎯⎯⎯⎯⎯⎯ /-aal / instrumental case marker

⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯ no marker for nominative.

/-də, -tə / ⎯⎯⎯⎯⎯⎯ /-ku/   dative case marker

/-il /   locative case marker

7. Idioms

Idioms are always language specific and are very problematic in translation. Hence, there is a need for a list of equivalent idioms *(Idioms tool 9)* in Manipuri and Tamil before going in for translation. Take for example,

32. məhak əŋaŋ onhənkhre.

'He was given death punishment.'

* avan kuzhantai aanaan.

* He became a child.

In this sentence, 'əŋaŋ onbə' literally means 'becoming a child' but as an idiom, it means 'death punishment.' This idiom has no equivalent in Tamil. Without a dictionary of idioms in Tamil, the translation will be awful as given above.

This pilot research clearly shows that the computer can never give an acceptable, foolproof, cent percent sensible and satisfactory translation till all the features discussed in this paper and the other similar problems are incorporated in the tools. This paper suggests that the capacity of the computer may be fully utilized for a quick pre-draft translation and using this raw translation, translators can complete the job successfully to begin with. To achieve a better computer aided translation there is a need for some tools in both the languages for translation.

1. A bilingual dictionary (Vocabulary tool)

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study   404

2. A phrase tool (A list of equivalent phrases in both the languages)

3. A structure tool (A list of equivalent sentence structures)

4. A contrastive grammar

5. A computational grammar

6. A spelling and grammar checker

7. A morphological analyzer

8. POS Tagging tool

9. An idioms tool (A list of parallel idioms and proverbs)

Incorporating these tools in the computer, the translation work can be done by the following proposed scheme, however, this needs a lot of discussions and interpretations, as this is only a rough one. The scheme represents all the tools in both the languages.

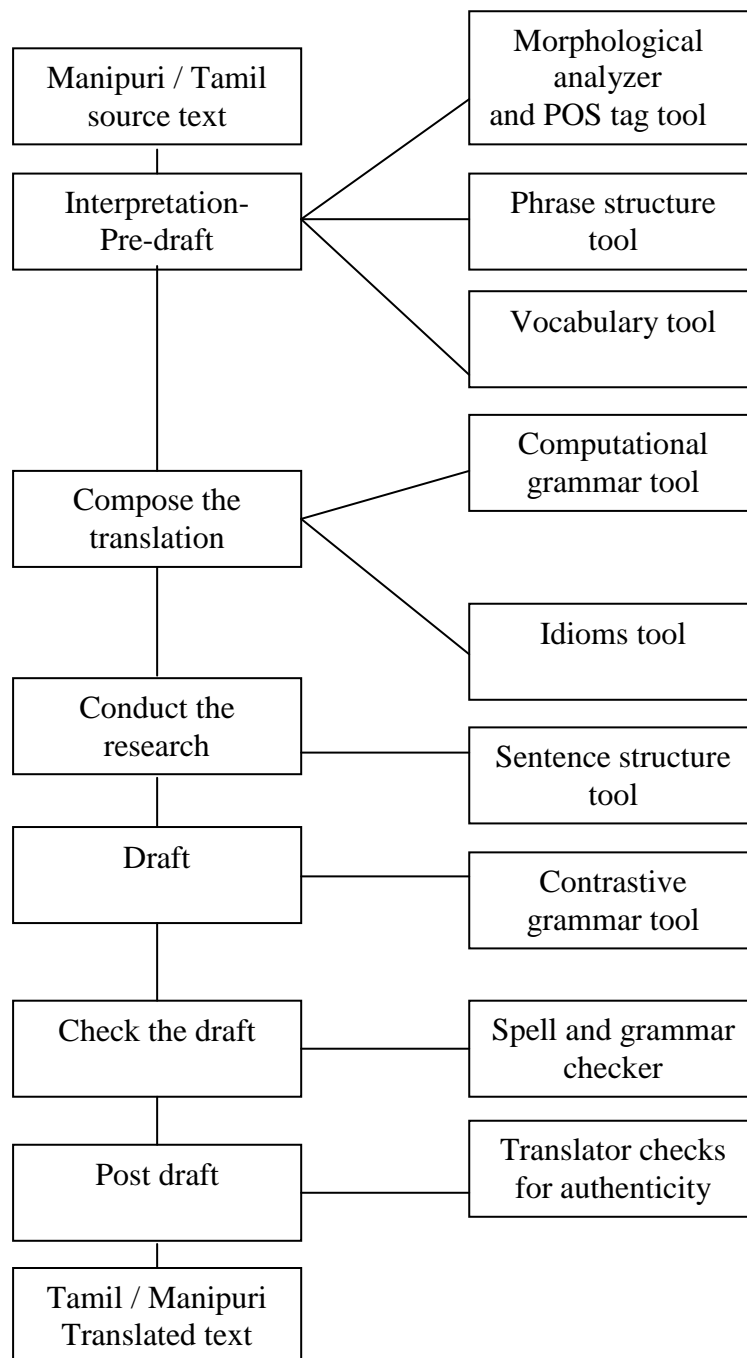Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil
Case Study   405

The Proposed Development Model

```
┌─────────────────────┐            ┌─────────────────────┐
│  Manipuri / Tamil   │            │    Morphological    │
│    source text      │            │      analyzer       │
│                     │            │  and POS tag tool   │
└──────────┬──────────┘            └─────────────────────┘
           │                      ╱
┌──────────┴──────────┐          ╱ ┌─────────────────────┐
│   Interpretation-   │─────────────│  Phrase structure  │
│     Pre-draft       │          ╲ │        tool         │
└──────────┬──────────┘          ╲ └─────────────────────┘
           │                      ╲
           │                       ┌─────────────────────┐
           │                       │   Vocabulary tool   │
           │                       └─────────────────────┘
           │
           │                       ┌─────────────────────┐
           │                       │    Computational    │
┌──────────┴──────────┐          ╱ │    grammar tool     │
│   Compose the       │─────────────└─────────────────────┘
│   translation       │          ╲
└──────────┬──────────┘          ╲ ┌─────────────────────┐
           │                       │     Idioms tool     │
           │                       └─────────────────────┘
┌──────────┴──────────┐
│   Conduct the       │            ┌─────────────────────┐
│    research         │────────────│  Sentence structure │
└──────────┬──────────┘            │        tool         │
           │                       └─────────────────────┘
┌──────────┴──────────┐            ┌─────────────────────┐
│      Draft          │────────────│    Contrastive      │
│                     │            │   grammar tool      │
└──────────┬──────────┘            └─────────────────────┘
           │
┌──────────┴──────────┐            ┌─────────────────────┐
│  Check the draft    │────────────│  Spell and grammar  │
│                     │            │      checker        │
└──────────┬──────────┘            └─────────────────────┘
           │                       ┌─────────────────────┐
┌──────────┴──────────┐            │  Translator checks  │
│     Post draft      │────────────│   for authenticity  │
└──────────┬──────────┘            └─────────────────────┘
           │
┌──────────┴──────────┐
│  Tamil / Manipuri   │
│   Translated text   │
└─────────────────────┘
```

8. Conclusion

Machine translation (MT) systems are now omnipresent. This omnipresence is due to a combination of increased need for translation in this day of global marketing and an exponential growth in computing and exploding knowledge power. Under these

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil
Case Study   406

circumstances, MT system is a powerful tool. Unfortunately, despite the widespread accessibility of MT, it is clear that the purpose and limitations of such systems are frequently misunderstood, and their capability is widely overestimated. But as discussed above, the computer can never work fruitfully without the necessary tools mentioned above.

At present, there is still a mismatch between the performance of MT systems and the expectations of users. The responsibility for closing this gap is lying in the hands of linguists, users and developers. Linguists need to think more about making their grammars computer friendly and learn how to assess the output of MT systems. Language courses and grammars need to address these issues. Developers with the help of linguists should produce tools as suggested above. The users must think of their needs so that an understanding of the problems by all concerned will be tackled fruitfully.

===

References

1. ALPAC. 1966. Language and Machines: Computers in Translation and Linguistics. A Report by the Automatic Language Processing Advisory Committee (ALPAC). Division of Behavioral Sciences, National Academy of Sciences, National Research Council Publication 1416. Washington: NAS/NRC

2. Bar-Hillel, Y. 1960. The Present Status of Automatic Translation of Languages. In F. Alt (ed.) Advances in Computers (Vol. 1). New York: Academic Press.

3. ……………….1962. The Future of Machine Translation. Times Literary Supplement. London: Times Newspaper. April 20th.

4. Bhat, D.N.S. & Ningomba, M.S. 1997. Manipuri Grammar. Newcastle: Lincom Studies in Asian Linguistics.

5. Chelliah, Sobhana L. 1997. A Grammar of Meithei. New York: Mouton de Gruyter.

6. Chomsky, N. 1957. Syntactic Structures. Mouton: The Hague.

7. ……………. 1966. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

8. Farwell, D., Guthrie, L. & Wilks, Y. 1993. Automatically Creating Lexical Entries for ULTRA, A Multi-Lingual MT System, Journal of Machine Translation, 8(3).

9. Minsky, M. 1968. Semantic Information Processing. Cambridge, Mass: MIT Press.

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study   407

10. Nadaraja Pillai, N. 1986. A Guide for Advanced Learners of Tamil. Mysore: Central Institute of Indian Languages.

11.……………………1992. A Syntactic Study of Tamil Verbs. Mysore: Central Institute of Indian Languages.

12. Steiner, G. 1975. After Babel: Aspects of Language and Translation. London: Oxford University Press.

13. Singh, T. D. & Bandyopadhyay, S. 2006. *Word Class and Sentence Type Identification in Manipuri Morphological Anlyzer.* In Proceedings of 'Modeling and Shallow Parsing of Indian Languages' (MSPIL) 2006. (pp. 11-17), IIT Bombay, Mumbai, India.

14. …………………………………... 2008. *Morphology Driven Manipuri POS Tagger. In Proceedings of International Joint Conference on Natural Language Processing (IJCNLP-08).* Workshop on Natural Language Processing of Less Privileged Languages (NLPLPL) 2008 (pp. 91-98), Hyderabad, India.

15……………………………………. 2010a. *Manipuri-English Example Based Machine Translation System.* International Journal of Computational Linguistics and Applications (IJCLA), 1/1-2, 147-158.

16. Rebika Devi, Soibam. 2011. '*A Contrastive Study of Gender in Manipuri and Tamil'.* in Akattiyam. Eds. V.S. Arulraj and R. Muralidharan. Thanjavur: Thanjavur University. pp. 407-413.

17………………………... 2010. *'Teaching some Syntactical features of Tamil to Manipuri Speakers'.* World Classical Tamil Conference. Coimbatore.

18...…………………….....2009. '*A Contrastive Study of the Case Systems of Manipuri and Tamil'.* International Conference on Dravidian Linguistics in honour of Prof. S. Agesthialingom. Annamalainagar: Annamalai University.

19. Renuga Devi, V. 1997. Grammatical Comparison of Tamil & English: A Typological Study. Madurai: Devi Publications.

20. Yashwanta Singh, Ch. 2000. Manipuri Grammar. New Delhi: Rajesh Publications.

==

Colophon:

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil Case Study   408

Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Resource Person
National Translation Mission
Central Institute of Indian languages
Mysore 570 006
Karnataka
India

Language in India www.languageinindia.com
12 : 5 May 2012
Soibam Rebika Devi, M.Sc. (Biology), M.A. (Linguistics), M.A. (Translation)
Some Problems in Computer Aided Translation and the Need for Tools – Manipuri and Tamil
Case Study   409