ADVANCEMENTS ON NLP APPLICATIONS FOR MANIPURI LANGUAGE

Maibam Indika Devi[1] and Prof. Bipul Syam Purkayastha[2]

[1]Department of Computer Science, Indira Gandhi National Tribal University, RCM, Makhan, Manipur
[2]Department of Computer Science, Assam University,Silchar, Assam

ABSTRACT

Manipuri is both a minority and morphologically rich language with genetic features similar to Tibeto Burman languages. It has Subject-Object-Verb (SOV) order, agglutinative verb morphology and is monosyllabic. Morphology and syntax are not clearly distinguished in this language. Natural Language Processing (NLP) is a useful research field of computer science that deals with processing of a large amount of natural language corpus. The NLP applications encompass E-Dictionary, Morphological Analyzer, Reduplicated Multi-Word Expression (RMWE), Named Entity Recognition (NER), Part of Speech (POS) Tagging, Machine Translation (MT), Word Net, Word Sense Disambiguation (WSD) etc. In this paper, we present a study on the advancements in NLP applications for Manipuri language, at the same time presenting a comparison table of the approaches and techniques adopted and the results obtained of each of the applications followed by a detail discussion of each work.

KEYWORDS

Manipuri Language, NLP, NER, MT, WSD.

1. INTRODUCTION

Human beings communicate in many ways such as through abstract languages, body gesture, facial expressions etc. Communication in human is unique because of the extensive use of abstract languages. Despite the diversity of languages that exist over the world, there is very much the need for communication and information exchange amongst people. Almost all of the day to day activities are done through natural languages whether it may be communicated directly in person or in a document. With the ever increasing technologies, the time where English was used as the language of the world is long gone. People nowadays prefer documents or works written in their own language which is easily understandable. The concept of NLP hinges on various disciplines namely computer science, computational linguistics, A.I etc. NLP comprises of two components, Natural Language Understanding (NLU) and Natural Language Generation (NLG). NLU forms the difficult component of NLP. NLP has various applications namely Morphological Analysis, POS Tagging, Parsing, WSD, NER, Automatic Summarization, MT, Co-reference resolution, Text-to-Speech, Speech segmentation, Speech-to-Speech, Question-Answering systems and many others.

2. MANIPURI LANGUAGE

Manipuri or Meitei-lon, meaning the language of Meiteis is spoken mainly as the first language in Manipur, a state situated in north-east India. Manipuri belongs to the Tibeto-Burman languages, a sub family of Sino-Tibetan languages. Amongst Tibeto-Burman languages, it is the first language

to be included in the Indian Constitution. The writing system of Manipuri language has two scripts- Meitei Mayek and Bengali script. An example sentence of both the scripts is given below. Example sentence in Bengali script:

নিঙোল চাক্কোবা মনিপুরগী অচৌবা কুন্ম্বৈ অমনি।

Example sentence in Meitei Mayek script:

ꯑꯤꯉꯣꯜ ꯆꯥꯛꯀꯧꯕ ꯃꯅꯤꯄꯨꯔꯒꯤ ꯑꯆꯧꯕ ꯀꯨꯝꯁꯤꯐ ꯑꯃꯅꯤ॥

Meitei Mayek is the original script and was used until 18th century. Bengali script was introduced in between 1709 and the middle of 20th century. Later local organizations and Meitei scholars put an effort in bringing back and adopting Meitei Mayek, thereby trying to replace Bengali script as the writing system. Manipuri language is a less computerized and there is not much work available on the web as compared to the language. Manipuri language being less computerized language, a very few works on NLP applications has been covered so far. This paper presents a detailed survey of the NLP applications developed for Manipuri language till now. Some of the works on NLP applications are described below.

3. RELATED WORK

Many applications on NLP has been developed all around the world as well as for major Indian languages such as Hindi, Bengali, Malayalam. Very few works has been reported for northeastern languages of India. [1] has reported a survey on applications of NLP developed for north-eastern languages of India. Most of the works done are for Hindi, Bengali, Assamese, and Nepali with a few for Manipuri, Bodo and Kokborok. [2] reported a survey on NER and classification covering fifteen years of work from 1991 to 2006. [3] reported MT for low resource languages where a shallow analysis of source language , a translation dictionary and a mapping system are all that is needed for translation. They also presented many approaches to achieving it. [4] has reported a paper on WSD considering different approaches and a comparison of the results obtained. [5] presented a survey paper on POS Tagging for Indian languages covering the developments on POS tagset and POS tagging for various Indian languages. [6] has given a short review on language processing on Manipuri Language. They have highlighted the challenges imposed in processing the language and briefly described the language processing tools reported till date. There are many NLP works on Indian languages that are developed till now, most of the work carried out on major Indian languages. A little or no work has been done for minority languages of India. Manipuri language, though included in the Indian Constitution is lagging much behind in the field of NLP and its applications. A few applications that have been done so far are addressed here in this paper.

4. E-DICTIONARY

Given below is a table showing the e- dictionaries developed for Manipuri language.

Table 1: E-dictionaries developed for Manipuri Language.

Title	Developers	Technique	Year
Building Manipuri-English Machine Readable Dictionary by Implementing Ontology	S Ningombam, SP Meitei, BS Purkayastha	Using Ontology	2011
A Manipuri-English Bilingual Electronic Dictionary -Design and Implementation	S. Poireiton Meitei, Shantikumar Ningombam, H.Mamata Devi, Bipul Syam Purkayastha	Using Binary Search Technique	2012
Development Of English To Manipuri Electronic Dictionary:A database approach	S.Poireiton Meitei, H. Mamata Devi	Using Database Approach and Binary Search Technique	2017

In the work of [7] the lexicon provides only lexical information, ontology may provide concepts, relations between words; more specifically ontology adds sense to the lexical entries in the lexicon. The implementation of ontology here provides the e-dictionary an extra feature of WSD when polysemous words or sentences are encountered. [8] adopted the binary search algorithm for word searching mechanism. The database of the dictionary contains information about root word, lexical item, transliteration, meaning in English, examples in Manipuri and examples in English.   The E-dictionary of [9] takes English as the input language and gives output in Manipuri language. So far, this is the only English-Manipuri E-dictionary developed using database approach. However, Manipuri language being agglutinative and tone language, it has been reported that the use of ontology facilitate better results.

5. Morphological Analysis

The table below gives the existing morphological analyzer developed for Manipuri language.

Title	Developers	Technique	Performance	Year
Morphological Analyzer for Manipuri:Design and Implementation	Sirajul Islam Choudhury, Leihaorambam Sarbajit Singh,Samir Borgohain, and Pradip Kumar Das	Grammatical rules and dictionaries	75% accuracy	2004
Word Class and Sentence Type Identification in Manipuri "Morphological Analyzer"	Thoudam Doren Singh and Sivaji Bandyopadhyay	Root driven approach	-	2006
Morphological Analysis for Manipuri Nominal Category Words with Finite State Techniques	Ksh. Krishna B. Singha and Bipul Syam Purkyastha	Right to left suffix stripping approach	-	2012
Manipuri Morpheme Identification	Kishorjit Nongmeikapam, Vidya Raj RK, Yumnam Nirmal and Sivaji Bandyopadhyay.	Combined technique of bigram and standard deviation	Recall-59.80%precision-83.02% and f score- 69.52%.	2012

Table 2: Morphological Analyzers developed for Manipuri Language.

In [10] work, the root dictionary containing 3000 root entries as a model and an affix dictionary. The morphological analyzer comprises of three modules- segmentation, morph syntactic analyzer and tagging modules. The segmentation module employs two approaches; first it applies the left-to-right longest matching method to extract the longest root. If matching root isn't found in root dictionary then it employs the second approach, which involves suffix stripping from right-to-left. [11] make use of Manipuri English bilingual dictionary which has a collection of 500 root words, 15 prefixes, and 150 suffix collections. The suffixes form the basis for determining the word classes and sentence types. This morphological analyzer handles different types of words.[12] have devised a right to left suffix stripping approach for analyzing the nominal category words. They have also developed FSM (Finite State Machine) for the nominal category to represent the morph tactics of the language and have also converted the FSM from NFA (Nondeterministic Finite Automata) to DFA(Deterministic Finite Automata).

[13] Work includes segmenting words into syllables and then the syllables are identified as morphemes or not. The segmentation of a word into syllables make use of handcrafted rules. A combined technique of bigram and standard deviation is used for identifying the morphemes. The work has been carried out on Meitei Mayek scipt of the Manipuri language. An input file of 13045 words has been used for the system. The system performance varies with the change of corpus size as well as domain of the corpus.

6. RMWE

Existing works on RMWE for Manipuri language are given in the table below.

Title	Developers	Technique	Performance	Year
Identification of Reduplicated MWEs in Manipuri: A Rule based Approach	Kishorjit Nongmeikapam and Sivaji Banyopadhyay,	Rule based approach	Recall-94.24, Precision-82.27 F-score-87.68	2010
Web Based Manipuri Corpus for Multiword NER and Reduplicated MWEs Identification using SVM	Thoudam Doren Singh and Sivaji Bandyopadhyay	Support Vector Machine (SVM)	-	2010
Identification of Reduplicated Multiword Expressions Using CRF	Kishorjit Nongmeikapam , Dhiraj Laishram , Naorem Bikramjit Singh, Ngariyanbam Mayekleima Chanu, Sivaji Bandyopadhyay	Conditional Random Field (CRF)	Recall-92.91%, precision-91.90% F-score - 92.40%.	2011
Improvement of CRF based Manipuri POS tagger using RMWE	Kishorjit Nongmeikapam, Lairenlakpam Nonglenjaoba, Yumnam Nirmal and Sivaji Bandhyopadhyay	Incorporated RMWE in CRF based POS Tagger	Recall-80.20%, precision-74.31% F-score-77.14%	2012
Genetic Algorithm (GA) In Feature Selection For CRF Based Manipuri Multiword Expression (MWE) Identification	Kishorjit Nongmeikapam and Sivaji Bandyopadhyay	CRF including Genetic Algorithm	Recall - 64.08% precision-86.84% F-score - 73.74%.	2011

Table 3: RMWEs developed for Manipuri Language.

In the work of [14], it comprises of 2 models. The first model comprises of tokenizer, reduplication MWE identifier, valid inflection list and dictionary as its components which function for identification of the four RMWEs. The second model has an extra component, the semantic comparator which identifies similarities in semantics. The system uses a corpus size of 74,936 tokens having 20887word forms. [15] uses corpus size of 4649016 different word forms. The corpus is then used to identify multi-word NER and RMWE using SVM and their results in the form of recall, precision and F-score are evaluated.

The processes involved in [16] include feature selection, preprocessing, feature extraction, training, and testing. The system incorporates best feature selection step and used a corpus size of 55000 tokens.

[17] incorporated the concept of RMWE in an attempt to improve the efficiency of the Manipuri POS tagger using CRF[22]. The algorithm as given [14] is used for identifying the RMWE.[18] used Genetic Algorithm is used for the feature listing and feature selection step. The feature listing is done similar to that of chromosome population and the feature selection is done based on gene values. The table above shows the recall, precision and f-score of various approaches. The values shown depend on the number of words forms used. The overall performance increases with the increase in word forms.

7. NER

The table below presents the existing NER for Manipuri language.

Title	Developers	Technique	Performance	Year
Name Entity Recognition in Manipuri Using SVM	T. Doren Singh, N. Kishorjit, A. Ekbal and B. Sivaji	Active learning technique and SVM	Precision-93.91% Recall-95.32% F-score-94.59%	2009
CRF Name Entity Recognition (NER) in manipuri: A highly agglutinative indian language	Nongmeikapam, Kishorjit, Tontang Shangkhunem, Ngariyanbam Mayekleima Chanu, Laisuhram Newton Singh, Bishworjit Salam, and Sivaji Bandyopadhyay	CRF	Precision-85.67% Recall-81.12% F-score of 83.33%,	2011

Table 4: NERdeveloped for Manipuri Language.

[19] gave the first report on NER for Manipuri languages. The active learning technique based NER system is used as the baseline system. It makes use of unlabeled corpus with 174,921-word forms and generates lexical context patterns. The unlabeled corpus is then annotated with four NE tags to use by SVM. The SVM system makes use of contextual and orthographic information.
[20] takes into account many sets of features such as current word, surrounding stems, prefixes, suffixes and so on. Feature extraction and best feature selection form the vital part of the system. The system performs lower than the SVM based NER of [19].As per the results obtained, SVM performs better than CRF model of NER for Manipuri language.

8. POS TAGGING

The existing POS tagging systems for Manipuri languages along with the techniques adopted are given in the following table.

Table 5: POS Tagging developed for Manipuri Language.

Title	Developers	Technique	Performance	Year
Morphology Driven Manipuri POS Tagger	Thoudam Doren Singh and Sivaji Bandyopadhyay	Dictionaries along with a tagset	69% accuracy	2008
Manipuri POS Tagging using CRF and SVM: A Language Independent Approach	Thoudam Doren Singh, Asif Ekbal and Sivaji Bandyopadhyay	CRF and SVM	CRF based model-72.04% accuracy whereas SVM based model-74.38% accuracy	2008
Developing a Tagset for Manipuri Part of Speech Tagging	Kh Raju Singha, Bipul Syam Purkayastha, Kh Dhiren Singha and Arindam Roy	Tagset based on ILPOST (Indian Language POS Tagset) framework	-	2011
A Rule-based Approach International Journal of Computer Applications	Kh Raju Singha, Bipul Syam Purkayastha, Kh Dhiren Singha:Part of Speech Tagging in Manipuri	Rule-based Approach	Increases with increase in number of rules and number of words in lexicon	2012
Transliterated SVM Based Manipuri POS Tagging	Nongmeikapam, Kishorjit, Lairenlakpam Nonglenjaoba, Asem Roshan, Tongbram Shenson Singh, Thokchom Naongo Singh, and Sivaji Bandyopadhyay	SVM	86.04% accuracy	2012
Part of Speech Tagging in Manipuri with Hidden Markov Model	Singha, Kh Raju, Bipul Syam Purkayastha and Dhiren Singha	Hidden Markov Model (HMM)	92% accuracy	2012

[21] used 3 dictionaries- prefix, suffix and root dictionaries, with the root dictionary having 2051 entries. It also employs a basic Manipuri tagset having 13 categories. The tag generator generates the POS tag of each word. The tagger has been tested on 3784 sentences with 10917 unique words.[22] manually annotated 63000 tokens using the 26 tags defined for Indian Languages. The experimental results show a difference of 2.34% implying SVM model performs better than the CRF model.

The tagset of [23] includes generic as well as language specific attributes totaling to 97 tags. The tagset comprises of two tables. They also suggested 12 categories of word classes of Manipuri words and also list the typological features of the language giving a concise idea on phoneme, gender, number, case relation, word formation etc.

A set of 3 types of rules- orthographic, morphological and disambiguation has been applied along with the use of a lexicon by [24]. A 3-tier tagset comprising of major category, sub-category and the attributes has been designed.

[25] first tagged Bengali scripted Manipuri words and then the tagged words are transliterated into the Meitei Mayek script. The algorithm for transliteration scheme from Bengali having 52 consonants and 12 vowels mapped to Meitei Mayek having 27 alphabets and its supplement vowels is explained. This reduction as they claimed to be is due to difference in domain of the corpus, their work confined to article domain as compared to newspaper domain of the latter.
[26] used the tagset designed in the work as in [2] This tagger is a sentence based approach where the probabilities of tagged words in sequence are combined and the maximum probability for the sentence is selected. The maximum probability is determined using the Viterbi algorithm in the HMM.

Various approaches to POS tagging are adopted by various researchers. Each of the approaches has its pros and cons. The rule based approach requires hand written rules but doesn't require a large set of data. While stochastic approach saves the time and effort of applying linguistic rules but requiring a large dataset of statistical data. As per the accuracy percentage reported, while SVM performs better than CRF, the HMM based POS tagging serves the best approach from that of SVM, CRF and rule based approaches.

9. MT SYSTEMS

The MT systems developed for Manipuri language till dates are presented in the following table.

Table 6: MT systems developed for Manipuri Language.

Title	Developers	Technique	Performance	Year
Manipuri-English Bidirectional Statistical Machine Translation Systems using Morphology and Dependency Relations	Thoudam Doren Singh, Sivaji Bandyopadhyay	Using Morphology and Dependency Relations	BLEU score of 16.873 for English-Manipuri and 17.573 for Manipuri-English	2010
Integration of Reduplicated Multiword Expressions and Named Entities in a Phrase Based Statistical Machine Translation System	Thoudam Doren Singh, Sivaji Bandyopadhyay	SVM along with Phrase Based SMT(PBSMT) integrated with RMWEs	BLEU score of 15.023 and NIST score of 5.21	2011
"Statistical machine translation of English–Manipuri using morpho-syntactic and semantic information	Singh, Thoudam Doren, and Sivaji Bandyopadhyay	Factored approach of SMT	-	2010
Manipuri-English Example Based Machine Translation System	Thoudam Doren Singh, Sivaji Bandyopadhyay	Example Based Machine Translation (EBMT)	BLEU and NIST score of 0.137 and 3.361	2010

In the work of [27] suffix and dependency relations of English language and case markers for the Manipuri language forms the translation factors for English-Manipuri translation. While for Manipuri-English translation, case markers along with POS tags of Manipuri language, suffixes and dependency relation of English forms the important translation factors. Both the systems are tested on a news domain corpus of 10350 sentences; the testset is of 500 sentences.

Manipuri NEs has been identified and RMWEs are also identified and classified in this work of [28] using the SVM technique. It shows an increase in translation quality as compared to baseline system.

The use of factored approach in [29] tightens the integration of linguistic information into translation model, which overcomes the data sparseness problem and provides availability of many language aspects to the language model. They have employed 3 translation factors and a corpus size of 10350 sentences for the system. The system has also been evaluated using subjective and automatic scoring techniques.

[30] developed parallel corpora of 16919 sentences, dictionary with 12229 entries and 57629 aligned phrases. MT systems for Manipuri language has been reported using SMT, PBSMT, EBMT etc. The performance of an approach relies on the language pair chosen. Manipuri is a morphologically rich language. For language pair with same morphology, SMT proves to be better provided there is large corpus available. EBMT serves better for language pair with different morphology and for poor resource language. Manipuri language, as it is a low resource language, EBMT approach will provide better results.

10. WORDNET

[31] build the Indo-wordnet, a wordnet for Manipuri language. Here for each word being newly entered, a synonym set that represents one lexical concept is found out. Then entered synonyms are then linked to other synonym sets to form hypernymy, hyponymy, meronymy, and antonymy, by using the semantic relationships.

11. WSD

[32] develop a WSD system for Manipuri language based on decision tree model. The system used a corpus size of 672 sentences, containing 13,167 words and nearly 2,000 polysemous words. The systems preprocess the training data to select and extract features. The classification and regression tree (CART) based algorithm, a decision tree based algorithm are used as the learner's algorithm to train the classifier. The system which was trained using 1600 words, after testing on 400 words gave accuracy with 71.75% accuracy.

12. CONCLUSION

NLP is a field of computer research in which research applications is ever increasing day by day. Many of the rich resourced languages such as English, French etc. has a long history of developments and advancements in NLP applications and tools. For example Google Translate, we all know itself is an outcome of NLP. One can translate web pages from Chinese to English and vice versa very easily. This is due to the availability of huge amount of resources as well as the rapid and fast growing applications in the field of NLP. In India too, many research has been going on for a long time. Much of the work has been done on Hindi, Nepali, Malayalam, Bengali, and few some languages. There are multiples of languages where in terms of NLP applications, they are in the prime stage of its development. The main reason behind this is the scarcity of language resource. Manipuri language is one such language, with scarce resources. This paper has given a report on the NLP works for Manipuri with the hope of facilitating and aiding researchers in having a brief understanding of the status of the language in the NLP scenario. Besides, many researchers are working on Manipuri language for developing NLP applications for both scripts.The upcoming researchers need to do a lot of work in this field, to contribute towards the development of the language. In the effort of doing so, it is required to make use of many preprocessing tools and applications. This paper will serve to aid the researchers in providing knowledge and information about already existing tools and the yet to be developed ones towards developing the language.

REFERENCES

[1]   Saiful Islam, Maibam Indika Devi, Prof. Bipul Syam Purkayastha. (2017). A Study on Various Applications of NLP Developed for North-East Languages, International Journal o Computer Science and Engineering (IJCSE) Vol. 9 No.06 Jun 2017 ISSN : 0975-3397

[2]   D. Nadeau and S. Sekine.(2007).A survey of named entity recognition and classification, Linguisticae Investigationes, 30(7).

[3]   Khan Md. Anwarus Salam†, Mumit Khan and Tetsuro Nishino Vandeghinste, V., Schuurman, I., Carl, M., Markantonatou, S. and Badia, T.(2006), METIS-II: Machine Translation for Low Resource Languages,In Proceedings of LREC 2006.

[4]   Alok Ranjan Pal and Diganta Saha. (2015).Word Sense Disambiguation: A Survey,International Journal of Control Theory and Computer Modeling (IJCTCM) Vol.5, No.3, July 2015.

[5]   Antony P J and Dr. Soman K P.(2011).Parts Of Speech Tagging for Indian Languages: A Literature Survey,International Journal of Computer Applications (0975 – 8887) Volume 34– No.8.

[6]   Surjit Singh R.K., Gunasekaran S., Anand Kumar M. and Soman K.P.(2014).A Short Review about Manipuri Language Processing,International Science Congress Association,Research Journal of Recent Sciences ISSN 2277-2502,Vol. 3(3), 99-103.

[7] Ningombam Shantikumar, Sagolsem Poireiton Meitei, and Bipul Syam Purkayastha. (2011). Building Manipuri-English machine readable dictionary by implementing ontology, International Journal of Engineering Science and Technology.

[8] S.Poireiton Meitei, Shantikumar Ningombam, H.Mamata Devi and Bipul Syam Purkayastha. (2012). A Manipuri-English Bilingual Electronic Dictionary -Design and Implementation, International Journal of Engineering and Innovative Technology (IJEIT).

[9] S.Poireiton Meitei and H. Mamata Devi. (2017). Development Of English To Manipuri Electronic Dictionary:A database approach, International Journal of Innovations & Advancement in Computer Science (IJIACS), Volume 6, Issue 3

[10] Sirajul Islam Choudhury, Leihaorambam Sarbajit Singh,Samir Borgohain, and Pradip Kumar Das. (2004). Morphological Analyzer for Manipuri:Design and Implementation, Proceedings of Second Asian Applied Computing Conference, AACC 2004, Kathmandu,Nep al.

[11] Thoudam Doren Singh, and Sivaji Bandyopadhyay. (2006). Word class and sentence type identification in manipuri morphological analyzer, Proceedings of MSPIL, Mumbai, India.

[12] Singha, Ksh Krishna B., and Bipul Syam Purkyastha. (2012). Morphological Analysis for Manipuri Nominal Category Words with Finite State Techniques, International Journal of Computer Applications 58.15.

[13] Kishorjit Nongmeikapam, Vidya Raj RK, Yumnam Nirmal and Sivaji Bandyopadhyay. (2012). Manipuri Morpheme Identification. Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP), pages 95-108, COLING.

[14] Nongmeikapam Kishorjit, and Sivaji Bandyopadhyay. (2010). Identification of Reduplicated MWEs in Manipuri: A Rule Based Approach,(ICCPOL-2010), Redwood City, San Francisco, pp49-54.

[15] Thoudam Doren Singh and Sivaji Bandyopadhyay. (2010).Web Based Manipuri Corpus for Multiword NER and Reduplicated MWEs Identification using SVM, Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pages 35–42,the 23rd International Conference on Computational Linguistics (COLING).

[16] Nongmeikapam Kishorjit, Ningombam Herojit Singh, Sonia Thoudam, and Sivaji Bandyopadhyay. (2011). Manipuri transliteration from Bengali script to Meitei Mayek: A rule based approach, In Information Systems for Indian Languages (pp. 195-198). Springer, Berlin, Heidelberg.

[17] Nongmeikapam Kishorjit, Lairenlakpam Nonglenjaoba, Yumnam Nirmal, and Sivaji Bandhyopadhyay. (2012).Improvement of CRF based Manipuri POS tagger by using Reduplicated MWE (RMWE), arXiv preprint.

[18] Nongmeikapam Kishorjit and Sivaji Bandyopadhyay. (2011). Genetic algorithm (GA) in feature selection for CRF based manipuri multiword expression (MWE) identification, arXiv.

[19] Thoudam Doren Singh, Kishorjit Nongmeikapam, Asif Ekbal, and Sivaji Bandyopadhyay. (2009). Named Entity Recognition for Manipuri using Support Vector Machine, 23rd PACLIC , Hong Kong, Vol.2 pp. 811-818.

[20] Nongmeikapam, Kishorjit, Tontang Shangkhunem, Ngariyanbam Mayekleima Chanu, Laisuhram Newton Singh, Bishworjit Salam, and Sivaji Bandyopadhyay. (2011). Crf based name entity recognition (ner) in manipuri: A highly agglutinative indian language, In Emerging Trends and Applications in Computer Science (NCETACS), 2nd National Conference on, pp. 1-6. IEEE.

[21] Thoudam Doren Singh and Sivaji Bandyopadhyay. (2008). Morphology Driven Manipuri POS Tagger, Proceedings of the Workshop IJCNLP on NLP for Less Privileged Languages, 11 January 2008 IIIT, Hyderabad, India.

[22] Thoudam Doren Singh, Asif Ekbal, and Sivaji Bandyopadhyay. (2008). Manipuri POS Tagging using CRF and SVM: A Language Independent Approach, Proceedings of ICON-2008: 6th International Conference on Natural Language Processing, Pune, India, pp 240-245.Volume 2, Issue 1.

[23] Kh Raju Singha, Bipul Syam Purkayastha, Kh Dhiren Singha, and Arindam Roy. (2011). Developing a Tagset for Manipuri Part of Speech Tagging, Journal Of Computer Science And Engineering, Volume 5.

[24] Kh Raju Singha,Bipul Syam Purkayastha, and Kh Dhiren Singha. (2012). Part of Speech Tagging in Manipuri: A Rule-based Approach, International Journal of Computer Applications (0975 – 8887) Volume 51– No.14.

[25] Nongmeikapam, Kishorjit, Lairenlakpam Nonglenjaoba, Asem Roshan, Tongbram Shenson Singh, Thokchom Naongo Singh, and Sivaji Bandyopadhyay. 2012. Transliterated SVM Based Manipuri POS Tagging, In Advances in Computer Science, Engineering & Applications (pp. 989-999). Springer, Berlin, Heidelberg.

[26] Kh. Raju Singha, Bipul Syam Purkayastha, and Dhiren Singha. (2012). Part of Speech Tagging in Manipuri with Hidden Markov Model. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, ISSN (Online): 1694-0814.

[27] Thoudam Doren Singh and Sivaji Bandyopadhyay. (2010). Manipuri-English Bidirectional Statistical Machine Translation Systems using Morphology and Dependency Relations, Proceedings of SSST-4, Fourth Workshop on Syntax and Structure in Statistical Translation, pages 83–91,COLING, Beijing.

[28] Thoudam Doren Singh and Sivaji Bandyopadhyay. (2011). Integration of Reduplicated Multiword Expressions and Named Entities in a Phrase Based Statistical Machine Translation System, Proceedings of the 5th International Joint Conference on Natural Language Processing, pages 1304–1312,Chiang Mai, Thailand.

[29] Thoudam Doren Singh and Sivaji Bandyopadhyay. (2010). Statistical machine translation of English–Manipuri using morpho-syntactic and semantic information, Proceedings of the Association for Machine Translation in the Americas (AMTA 2010).

[30] Thoudam Doren Singh and Sivaji Bandyopadhyay. (2010). Manipuri-English Example Based Machine Translation System, International Journal of Computational Linguistics and Applications (IJCLA), ISSN (2010): 0976-0962.

[31] Yumnam Bablu Singh and Bipul Syam Purkyastha.(2011). Experiences in building the indo wordnet- a wordnet for Manipuri, International Journal of Engineering Science and Technology (IJEST),ISSN : 0975-5462,Vol. 3 No. 5.

[32] Richard Laishram Singh, Krishnendu Ghosh, Kishorjit Nongmeikapam and Sivaji Bandyopadhyay. (2014). A Decision Tree Based Word Sense Disambiguation System In Manipuri Language, Advanced Computing: An International Journal (ACIJ), Vol.5, No.4.

AUTHORS

Maibam Indika Devi is an Assistant Professor at Indira Gandhi National Tribal University (IGNTU), Regional Campus Manipur, Makhan, Manipur. She is currently doing PhD in Assam University, Silchar.

Prof. Bipul Syam Purkayastha is working as a Professor, at Assam University, Silchar , Assam. He has published many papers in many fields – NLP, Networking, DataMining etc.