

An In-depth Study on POS Tagging for Assamese Language

Dr. Karabi Kherkatary Boro¹, Dr. Uzzal Sharma²

¹Assistant Professor, Department of Assamese,
Pragiyotish College, Guwahati, India
karabiboro1185@gmail.com

²Assistant Professor, Department of Computer Applications,
Assam Don Bosco University, Guwahati, India
druzzalsharma@gmail.com

Abstract: An automatic POS tagger is an essential component of any Natural Language Processing (NLP) work. It is one of the important steps towards the processing of Natural Language. There are various challenges in the tagging of POS and most of the time these are language-dependent. Assamese is a morphologically rich and free word order language. Because of this, the challenges are even more. In the present paper, the basic concept of the POS tagger and its importance in the NLP is discussed. In the later part of the paper, the overall characteristics of the Assamese language are discussed in short and its various challenge that may arise towards the tagging of POS is discussed. The paper also discusses about the various POS tagging techniques that are commonly used in the tagging of POS.

Keywords: POS Tagging, MRL, Free Word order Language, Word stripping, tagset.

(Article history: Received: 4th August 2020 and Accepted 7th December 2020)

I. INTRODUCTION: WHAT IS PARTS OF SPEECH (POS) AND ITS TAGGING

Words are the constituents of a sentence. Each word in a sentence is classified into different word classes. According to the use and its placement in a sentence, words are divided into different classes called Parts-of-Speech (POS) or word classes or syntactic categories. It is very much necessary to have different classes of words in a sentence to form a sentence and to give its appropriate meaning or semantics. In fact, the parts-of-speech works as the building blocks of the grammar of any language, more specifically, in the linguistic part of the syntax[1]. The grammar of any language is very important and it depends heavily on the parts-of-speech to produce required semantic expressions, both verbal as well as written[2][3]. It can also be said that parts-of-speech may be considered as the life-blood of any language in that, a lack of a sequence of appropriate words representing parts-of-speech would result in the spoken or written words meaningless[4].

Parts-of-Speech are very important as it reveals all the information about a word and its neighbor. In fact, POS helps us to get information about neighboring words knowing the class of a word. For example, if a word is a noun it is preceded by either a determiner or an adjective, if the word is a verb, it is preceded by a noun.

Parts-of-Speech(POS) tagging is a process of classifying each word in a sentence into one of the Parts-of-Speech belongs to that language. In simple words, we can say that tagging is a task, where each word in a sentence is labeled into the appropriate Parts-of-Speech. POS tagging is very much necessary for parsing that is the process of determining

the syntactic structure of a sentence. It means parsing is followed by POS tagging. Parsing is one of the very important steps towards the processing of Natural Language. But as a preliminary step, tagging of Parts-of-Speech is carried out, and this is generally performed by the language experts of the language under consideration in manual process or automatically by a program in computer based on the already prepared tagset for that specific language.

II. POS TAGGING AND ITS IMPORTANCE IN NLP

Natural Language Processing (NLP) is the ability of a computer system to process Natural Languages. Natural Languages are languages, which are spoken by human beings. Although, there are grammars associated with each language having different syntactic structures which are all grammatically correct can result in the same semantic meaning[5][6]. Because of not having a one-to-one relationship between syntactic and semantic, the processing of natural languages by a computer system is considered as a real challenge [7]. This is because, to understand and process different syntactic structures, the system is also required to be that much versatile.



Figure1: Natural Language Processing

In NLP, we develop a program, whose purpose is to tag each and every word of the corpus under study. The process of tagging is based on the Parts-of-Speech (POS) of the language under study[8]. It is known as POS tagging which

is a process of automatic assignment of lexical categories or word class markers such as Noun, Pronoun, Verb, Adjective, Adverb, etc. for each word in a sentence according to the grammar of that natural language. In NLP, POS tagging is carried out to remove any kind of disambiguates among the words.

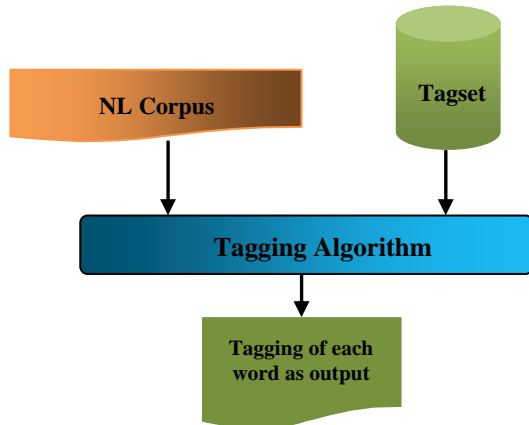


Figure 2: POS tagging Process

Any NLP related work typically have the steps as shown in Figure3:

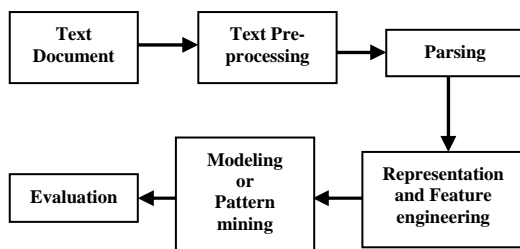


Figure3: NLP standard steps

Part-of-speech tagging is the process of marking up words and punctuation characters present in a corpus with the appropriate POS labels. The main purpose of the POS tagging is to determine the lexical property of word under study in the context and situation in which it was used. The grammatical or the syntactic category of a word is determined by two factors, namely, "lexical" and "contextual" information. The lexical information is associated with the type or category of the word whereas the contextual information is associated with the context or environment in which it is used. There are many problems associated with POS tagging[26]. They are:

1. In natural languages there are many words that are not listed in the lexical database. These words are known as "out of vocabulary" (OOV) words.
2. As far as the lexical category is concerned, there may be many words, which shows ambiguity regarding

the placement of that word into a lexical category. As an example, we can talk about the Assamese word কৰ /kɔr/, which is used as a verb as well as a noun. When it is used as a verb, it means "to do" and when it is used as a noun, it means "hand".

3. The training data used to train the system plays a vital role. If it is small in size, the POS tagger will not be considered as efficient.

III. CLASSIFICATION OF POS Tagger

There are several approaches for POS tagging. The Figure4 describes the different POS tagging models normally used.

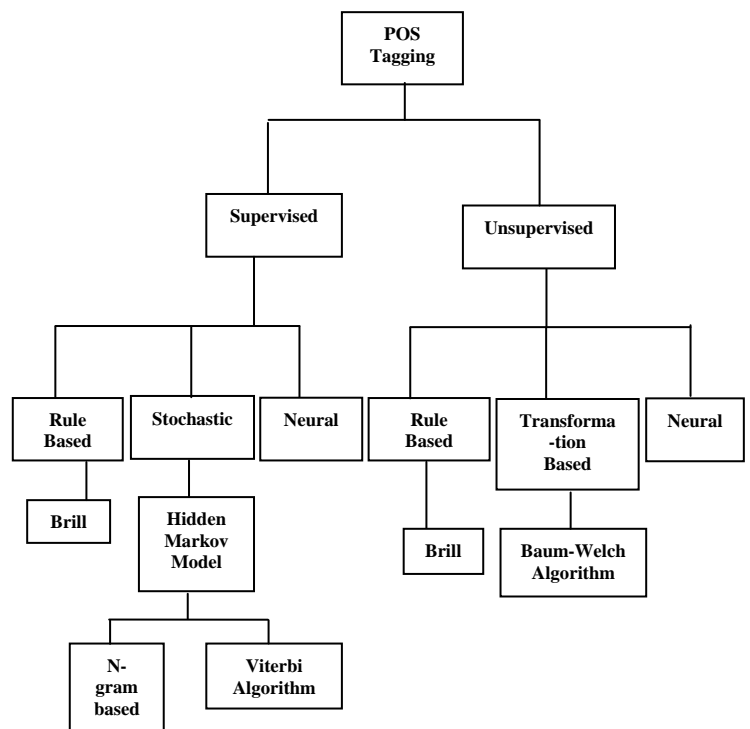


Figure4: Classification of POS Tagger

Although there are many models which can be used for the classification of POS taggers [9][10][11], in the current paper, some of the commonly used are discussed below which are most commonly used in case the Assamese Language:

Supervised models

In these models, a pre-annotated corpus is required which is used to train the model to learn information about the tagset, word-tag frequencies, rule sets, etc. [12][25]. The accuracy of the model depends on the size of the corpus. When the size of the corpus is large, it gives better accuracy and at the same time takes a long time for training and prediction.

Unsupervised models

In comparison to the supervised POS Tagging model, the unsupervised POS Tagging models do not require a pre-annotated corpus. Instead, they are more sophisticated and uses advanced computational techniques such as the Baum-Welch algorithm to automatically induce tagsets, transformation rules, etc. Once the tagsets and information rule is set, they either calculate the probabilistic information which is required by the stochastic taggers or apply contextual rules required by rule-based systems or transformation-based systems [12], [13].

Both the supervised and unsupervised models can be further classified into the following categories.

Rule-based and transformation-based models

In case of the rule-based POS Tagging models, there is a set of handwritten rules and uses contextual information to assign POS tags to words. These rules are often known as context frame rules set by linguistic experts. On the other hand, in the case of the transformation-based model a predefined set of handcrafted rules as well as automatically-induced rules that are generated during training is applied. Apart from it, some model also used morphological rules [14], capitalization and punctuation, etc. [13].

Stochastic models

In case of the stochastic models the frequency, probability, or statistics are used. It is based on different

methods such as Hidden Markov Models (HMM), maximum-likelihood estimation (MLE) or n-grams. The HMM-based methods require the evaluation of the "argmax" formula, which is very expensive in terms of time complexity as in this case all the possible tag sequences need to be checked, in order to find the sequence that maximizes the probability. So, in HMM-based approached, a dynamic programming model known as the Viterbi Algorithm is used to find the optimal tag sequence [15].

And lastly, the neural networks [16][17] is applied for both supervised and unsupervised POS Tagging models.

IV. POS IN ASSAMESE LANGUAGE

The grammatical information of any language is determined by the tags. All the tags belongs to a language is represented as a set known as tagset. The size of a tagset is very important as far as the versatility of the tagset is concerned. The efficiency and quality of the research depends on its tagset size. In the table (Table1), the Assamese version of the BIS Tagset is shown along with example of each tag type. Although the Table1 shows a tagset based on BIS tagset, it is not the complete one. There are many tags still need to be added to it and it is an ever-growing process as any language become richer and richer when it come in contact with other languages. In the present research, the authors tries to identify the most important and frequently occurring tags to represent the tagset.

TABLE1: THE ASSAMESE TAGSET AS PER THE BIS

S/no	POS	Category	Label	Annotation Convension	Assamese	Example
1	Noun(N) বিশেষ্য	Common	NN	N_NN	জাতিবাচক বিশেষ্য	মানুহ /manuh/, তৰা /tura/
		Proper	NNP	N_NNP	ব্যক্তিবাচক বিশেষ্য	ৰাম /ram/, ব্ৰহ্মপুত্ৰ /brohmoputro/
		Collective	SN	N_SN	সমষ্টিবাচক বিশেষ্য	সভা /xobha/, পৰিষদ /porixpd/
2	Pronoun(PP) সৰ্বনাম	Personal	PRP	P_PRP	ব্যক্তিবাচক সৰ্বনাম	মই /mni/, তই /tvi/, তুমি /tumi/, আপুনি /apuni/, সি /xi/, তাই /tai/
		Reflexive	PRF	P_PRF	আত্মবাচক সৰ্বনাম	নিজ /nij/, স্বয়ং /swopyng/, আপোন /apon/
		Relative	PRL	P_PRL	সম্বন্ধবাচক সৰ্বনাম	যি /ji/, যিহ /jih/
		Interogative	PRQ	P_PRQ	প্ৰশ্নবাচক সৰ্বনাম	কি /ki/, কিহ /kih/, কোন /kon/
		Inclusive	INC	P_INC	সাকল্যবাচক	সকলো /xokolo/,

					সর্বনাম	সমূহ /xɔmuh/, উভয় /ubhɔy/, আটাই /atai/
		Demonstrative	DRM	P_DRM	নির্দেশবোধক সর্বনাম	ই /i/, এই /ei/, সি /xi/, সেইয়া /xaiya/, সৌ /xou/
3	Verb(V) ক্রিয়া	Main	VM	V_VM	মূখ্য ক্রিয়া	হয় /hɔy/, আছে /ase/, যা /ja/, যাম /jam/, গ'ল /gɔl/
		Auxiliary	VAUX	V_VAUX	সহায়কাৰী ক্রিয়া	(ই) খাই /khai/, গৈ /goi/, বহি (বহ-ই) /bɔhi/, পড়িব (খুজিছোঁ) /pɔrhibɔ/ (/khujisɔ/), কোৱা (কথা) /kɔva/(/kɔtha/)
4	Adjective (JJ) বিশেষণ	--	JJ	JJ	বিশেষণ	গৰম /gɔrɔm/, দয়ালু /dɔyalu/, কিচকিচীয়া /kiskisia/
5	Adverb (RB) ক্রিয়া বিশেষণ	--	RB	RB	ক্রিয়া বিশেষণ	খৰকৈ /kɔrkoɪ/, সোনকালে /xɔnkale/, ওচৰে- পাজৰে /osɔrɛ - pajɔrɛ/, তালৈ /talɪ/
6	Postposition(PSP) অনুপদ	--	PSP	PSP	অনুপদ	অৰ্থে /ɔrthɛ/, দ্বাৰা /dwara/, চোন /son/, পাই /pai/, হবলা /hɔbɔla/, চাগৈ /sagɔɪ/
7	Conjunction(CC) সংযোজক	Co-ordinator	CCD	CC_CCD	সংযোজক	আৰু /aru/, অথচ /ɔthɔsɔ/, তেন্তে /tente/
		Subordinator	CCS	CC_CCS	সমন্বয়ক	যদিও /jɔdɪɔ/, তথাপি /tɔthapi/, কিন্তু /kintu/, অথবা /ɔthɔbɔ/
8	Indeclinable (RP) অব্যয়	Interjection / Exclamatory	INJ	RP_INJ	ভাৱবোধক অব্যয়	বাহ: (কি সুন্দৰ) /bah/ (/ki xundɔr/), হয় হায়! (অকালতে ঢুকাল) /hai hai! / (ɔkɔlɔtɛ dhukal/), ছি: (বৰ লেতেৰা) /sih/ (/bɔr leteɔ/), আও! (ইয়াৰ বৰ সাহ) /ao!/ (/iyar bɔr xah/)

		Vocative	VOC	RP_VOC	সম্বোধনবোধক অব্যয়	অ' (অ' মোৰ আপোন দেশ) /ʌ/(/ʌ mor apon dex/), ঐ (ল'ৰা) /ʌi/ (/lɔra/), হে (হৰি) /hɛ/ (/hɔri/), হেৰি /heri/, হেৰা (বন্ধু) /hɛra/ (/bɔndhu/)
9	Quantifiers(QT) পৰিমাণ নিদেৰ্শক	Cardinals	QTC	QT_QTC	অংকবাচক	এক /ɛk/, দুই /dui/, তিনি /tini/
		Ordinals	QTO	QT_QTO	পূৰ্ণবাচক)	প্রথম /prɔthɔm/, দ্বিতীয় /ditiyɔ/, পহিলা /pɔhila/, নৱমী /nɔbɔmi/, দহোকুৰি /dɔhɔkuri/, বাৰেকুৰি /barekuri/, ষোড়শী /xɔɔɔxi/, শতিকা /xɔtika/, হেজাৰী /hejari/, কুৰি /kuri/, গন্দা /gɔnda/,
10	Residuals(RD) অৱশিষ্ট	Foreign Word	RDF	RD_RDF	বিদেশী শব্দ	পেন /pen/, কি-ব'ৰ্দ /ki bɔrd/
		Symbol	SYM	RD_SYM	চিহ্ন	any symbol
		Punctuation	PUNC	RD_PUNC	যতি চিন	!, ?, !
		Unknown	UNK	RD_UNK	অজ্ঞাত	--
		Echowords	ECH	RD_ECH	ধ্বন্যাত্মক শব্দ	কা-কা /ka-ka/, ভৌ- ভৌ /bhɔu-bhɔu/

V. POS TAGGING AND ITS CHALLENGES IN ASSAMESE LANGUAGE

The Assamese language is one of the rich languages in terms of the types of word class or POS present in it. At the same time, Assamese is also a Morphologically Rich Languages (MRL). It belongs to the Indo-Aryan class of Languages along with other predominant languages such as Bengali, Hindi, etc[24].

As far as the Assamese language is concerned, there are many challenges, that make the tagging of POS very difficult and complex. Like the other language of Indo-Aryan class of Languages, the Assamese is also a free word order language. A language is said to be a free word order language if the placement of words in the formation of a sentence has negligible influence as far as the semantics of the sentence is concerned. In the Assamese, although, the

predominant word order is subject(S), verb (V), and object (O) that is SVO, but we can see some other word orders as well in the construction of sentences. They are: SVO, SOV, VSO, VOS, OVS, and OSV. Out of these six, the first three that is SVO, SOV, VSO are mostly dominant in the formation of the sentence. For example, in the sentence

"মই ভাত খাওঁ।" /Moi bhat khao/ and
"ভাত মই খাওঁ।" /Bhat moi khao/,

there is hardly any difference in the semantic. This issue may create a problem as far as the tagging of POS is concerned [19][20].

Apart from it, there are some other issues, which makes, the POS tagging a very critical task. Firstly, we can talk about "Ambiguous words". Ambiguous words create a major problem as they can be categorized in different word classes[21]. There are many words, which can have more

than one tag. This incidence is known as "Lexical ambiguity". For example, sometimes the same word such as "mark" in English represents as a 'noun' and 'verb', in a sentence like:

'Mark was marked as a criminal'

Similarly, in Assamese, sentences such as "সুন্দৰ এজন সুন্দৰ ল'ৰা।" /Xundor ejon xundor lora/, the word "সুন্দৰ" /Xundor/ is categorized into proper-noun (the first "সুন্দৰ" /Xundor/) and adjective (the second "সুন্দৰ" /Xundor/).

In the same line, the word "কলা" /kɔla/, may function as Adjective(JJ) as well as Noun(N). For example,

- কলা /kɔla/ (Adjective(JJ)) - মানুহজন কলা। /manuhjon kɔla/- The man is deaf.
- কলা-কৌশল /kɔla-kɔuxɔl/ (Adjective(JJ)) - সি তাৰ কলা-কৌশল দেখাই আছে। /xi tar kɔla-kɔuxɔl dekhai ase/- He is showing his skills.
- কলা- সংস্কৃতি(N) /kɔla xɔŋskriti/- অসমৰ কলা- সংস্কৃতি(N) বাৰে বহনীয়া /ɔxɔmɔr kɔla-xɔŋskriti barerohoniya/- The Assamese culture is versatile.

in the above example, in the first two sentences, the word 'কলা' /kɔla/, the word is used as Adjective(JJ) and in the next two sentences, the same word 'কলা' /kɔla/, is functioning as a noun. So, during the tagging process, the automatic POS tagger, may try to place the word, either in the noun category or in the adjective category.

Similarly, for the word "কৰ" /kɔr/

- কৰ /kɔr/ (flower bud)(N) - ফুলৰ কৰটো বৰ ধুনীয়া। /phulor kɔrto bor dhuniya/- The flower bud is very beautiful.
- কৰ /kɔr/ (tax) (N) - কৰ সময়মতে পৰিশোধ কৰা উচিত। /kɔr xɔmɔymote pɔrixodh kɔra usit/- Tax should be paid on time.
- কষ্ট-কৰ (Troublesome) (Adjective(JJ)) - পাহাৰ বগোৱা বৰ কষ্ট-কৰ। /pahar bogowa bor kɔsɔkɔr/ - The climbing of hill is very difficult.

in the first two sentence, the 'কৰ' /kɔr/, is functioning as noun and in the third sentence, it is functioning as adjective(JJ).

In the following sentences, the word "বিষ" (bix),

- বিষ - (Noun(NN)) - poison - pain - তেওঁৰ হাতৰ বিষ হৈছে। /teor hator bix hoise/ - He is having pain in his hand.
- বিষ-ধৰ (Adjective (JJ)) - One which has poison(eg. snake) - সাপডাল বৰ বিষ-ধৰ আছিল। /xapdal bor bixɔdhɔr asil/- The snake was very poisonous.

is functioning as a noun whereas in the second sentence, the same word is functioning as Adjective(JJ).

In handling the above situations, we need to consider the individual words rather than the sentence, while POS tagging is carried out.

Another challenging task that we can see in the following example for the word "কলা" /kɔla/,

- কল /kɔl/(Industry) (NN) - জাগিৰডোত কাগজ কল আছে। /jagirodot kagj kɔl ase/- There is a paper mill in Jagiroad.
- কল /kɔl/(fruit) (NN) - কল অতি পুষ্টিকৰ ফল। /kɔl ɔti pustikɔr phɔl/- Banana is very nutritious fruit.
- কল /kɔl/ (switch)(NN) - আকাশবাণীৰ কলটো টিপি দিয়ক। /akaxbanir kɔltu tipi diyɔk/- Switch on the radio.[23]

To, identify, the root word in a sentence, Suffix stripping is used[22]. The process basically, strips off the suffix from a word, to extract the root word. For example, the root word ভাৰত /bharot/ may have a different suffix such as

- 'ীয়' /-iyɔ/ (ভাৰতীয়) /bharotiya/,
- 'বাসী' /baxi/ (ভাৰতবাসী) /bharotbaxi/,
- 'ৰত্ন' /rotno/ (ভাৰতৰত্ন) /bharotrotno/,

suffix stripping is a step, which is carried out before the POS tagging is done so that the word can be placed in the appropriate class. But as it is done most of the time automatically, there is a chance of incorrect stripping output which is reported by the researchers[Q]. For example, the word 'মাজুলী' /majuli/ may be wrongly stripped of to মাজুল /majul/, same is the case with the word 'দিল্লী'; /dilli/ be stripped off as দিল্ল /dillo/ [18][22].

Based on the above circumstances, it can be said that the tagging of POS in the Assamese Language is one of the challenging task, which needs to be taken care of very carefully. Otherwise, the tagging process may result in an error.

In the above example, although, they all fall under the same POS tag, that is a noun, but gives different meanings in different contexts. So, while tagging is carried out, the context in which it is used needs to be considered.

VI. CONCLUSION

Although the POS tagging process is carried out automatically, there are many challenges that need attention and human intervention. Especially in languages like Assamese, which is morphologically rich and a free word order language, the challenges are even more. The need for the design of a versatile POS tagger is very much felt. Although, there are many automatic POS taggers for the Assamese language are there, and performing their job very well, but still, the need for a complete automatic POS tagger is always felt, which can handle all the aspects of the language. In the present paper, the authors try to showcase the different challenges that need to overcome, while designing an automatic POS tagger.

REFERENCES

- [1] Bahl, L. R. and Mercer, R. L. (1976). "Part of speech assignment by a statistical decision algorithm". In *Proceedings IEEE International Symposium on Information Theory*, 88–89.
- [2] Gil, D. (2000). Syntactic categories, cross-linguistic variation and universal grammar. In Vogel, P. M. and Comrie, B. (Eds.), *Approaches to the Typology of Word Classes*, 173– 216. Mouton.
- [3] Karlsson, F., Voutilainen, A., Heikkilä, J., and Anttila, A. (Eds.). (1995). *Constraint Grammar: A Language- Independent System for Parsing Unrestricted Text*. Mouton de Gruyter.
- [4] Voutilainen, A. (1995). Morphological disambiguation. In Karlsson, F., Voutilainen, A., Heikkilä, J., and Anttila, A. (Eds.), *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*, 165–284. Mouton de Gruyter.
- [5] Arora, Sanjeev, Ge, Rong, Halpern, Yonatan, Mimno, David, Moitra, Ankur, Sontag, David, Wu, Yichen, and Zhu, Michael. A practical algorithm for topic modeling with provable guarantees. In *Proceedings of The 30th International Conference on Machine Learning*, pp. 280– 288, 2013.
- [6] Lee, Moontae, Mimno, David, and Bindel, David. Robust spectral inference for joint stochastic matrix factorization. In *Advances in neural information processing systems*, 2015.
- [7] James Allen. *Natural Language Understanding*. Pearson Education, Singapore, second edition, 2004.
- [8] A.Chen and F.C.Grey, "Generating statistical Hindi stemmer from Parallel texts," *ACM Trans. Asian Language Inform Process*, vol. 2(3), 2003.
- [9] S. Dasgupta and V. Ng, "Unsupervised morphological parsing of Bengali," *Language Resources and Evaluation*, pp. 311–330, 2006.
- [10] Yair Halevi, "Part of Speech Tagging", Seminar in Natural Language Processing and Computational Linguistics (Prof. Nachum Dershowitz), School of Computer Science, Tel Aviv University, Israel, April 2006.
- [11] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*, Prentice Hall, 2001.
- [12] Fahim Muhammad Hasan, Naushad UzZaman, Mumit Khan, "Comparison of Different POS Tagging Techniques (n-grams, HMM and Brill's Tagger) for Bangla", *International Conference on Systems, Computing Sciences and Software Engineering (SCS2 06) of International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CIS2E 06)*, December 4-14, 2006.
- [13] Karthik Kumar G, Sudheer K, Avinesh Pvs, "Comparative Study of Various Machine Learning Methods for Telugu Part of Speech Tagging", In *Proceeding of the NLP AI Machine Learning Competition*, 2006.
- [14] Linda Van Guilder, "Automated Part of Speech Tagging: A Brief Overview", Handout for LING361, Georgetown University, Fall 1995.
- [15] Andrew MacKinlay, "The Effects of Part-of-Speech Tagsets on Tagger Performance", Undergraduate Thesis, University of Melbourne, 2005.
- [16] Manoj Kumar C, "Stochastic Models for POS Tagging", IIT Bombay, 2005.
- [17] L. E. Baum, "An Inequality and Associated Maximization Technique in Statistical Estimation on Probabilistic Functions of a Markov Process", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 22, Issue: 4, April 2000, pp. 371-377.

- [18] Juan Antonio P'erez-Ortiz and Mikel L. Forcada, "Part-of-Speech Tagging with Recurrent Neural Networks", Universitat d'Alacant, Spain, 2002.
- [19] Dhanalakshmi V, Anand Kumar1, Shivapratap G, Soman KP and Rajendran S, "Tamil POS Tagging using Linear Programming", International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.
- [19] Gurleen Kaur Sidhu, Navjot Kaur, "Role of Machine Translation and Word Sense Disambiguation in Natural Language Processing", IOSR Journal of Computer Engineering (IOSR-JCE), May. - Jun. 2013.
- [20] Hem Chandra Baruah. Assamiya Vyakaran. Hemkosh Prakashan, Guwahati, 2003.
- [21] D. Deka and B. Kalita. Adhunik Rasana Bisitra. Assam Book Dipot, Guwahati, 7th edition, 2007.
- [22] P. Sharma, U. Sharma and J. Kalita, "Suffix stripping based NER in Assamese for location names," *2012 2nd National Conference on Computational Intelligence and Signal Processing (CISP)*, Guwahati, Assam, 2012, pp. 91-94, doi: 10.1109/NCCISP.2012.6189684.
- [23] A collection of Linguistics and Phonological essays written by a number of eminent writers, edited by Naherandra Padum and published by Sri Ujjal Hazarika, Bani Mandir, Dibrugarh-786001, August, 2004, ISBN: 81-7206-189-4.
- [24] A. Dalal, K. Nagraj, U. Sawant, S. Shelke, P.Bhattacharya."Building Feature Rich POS Tagger for Morphologically Rich Languages: Experiences in Hindi", In ICON 2007: International Conference on Natural Language Processing 2007
- [25] Bipul Roy, Bipul Syam Purkayastha: A Study on Different Part of Speech (POS) Tagging Approaches in Assamese Language, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016.
- [26] Nitish Chandra, Sudhakar Kumawa, Vinayak Srivastava: Various Tagsets for Indian Languages and Their Performance in Part Of Speech Tagging, Department of Computer Science & Engineering, IIT (BHU), Varanasi, Proceedings of 5th IRF International Conference, Chennai, 23rd March. 2014, ISBN: 978-93-82702-67-2

AUTHOR PROFILE



Dr. Karabi Kherkatary Boro is presently working as an Assistant Professor in the Department of Assamese at Pargjyotish College, Guwahati, India. She has more than 12 years of teaching and research experience. She has published more than 15 papers in Books and Journals and presented more than 10 papers in International and National conferences. She has a book to her credit as an Author and also co-edited a book.



Dr. Uzzal Sharma is currently working as an Assistant Professor, Stage 2 in the Department of Computer Applications at Assam Don Bosco University, Guwahati, India. He has over 16 years of experience in the field of teaching and research. He has published more than 19 SCOPUS indexed papers and presented many papers in international conferences. He also has a patent to his credit. He has written 5 books as sole author, published by national and international publishers and contributed many book chapters. Dr. Sharma has produced 2 Ph.D. scholars and currently guiding a few scholars for their Ph.D. research.
