International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Special Issue-1 E-ISSN: 2347-2693

Significance of the Study on Regional Dialects in Assamese Language

Jahnabi Borah[1*]
Department of CSE and IT
Assam Don Bosco University
Guwahati,India
jahnabiborah0@gmail.com

Dr. Ujjal Sharma[2]
Department of CSE and IT
Assam Don Bosco University
Guwahati,India
uzzal.sharma@dbuniversity.ac.in

***Abstract*- In any natural language there are two linguistic variations: dialect and accent. In any standard natural language, the presence of various dialects can be seen due to the pattern of pronunciation and the vocabularies used by a particular community because of geographical differences. Assamese Language also has several dialect variations based on the phonology and morphology. Recent studies show that there are four dialect groups in this language. Since very little work has been done in the regional dialect of Assamese language, it has become necessary to discriminate these dialects as it seems no significant and systematic study is carried out on the dialects of Assamese language. In this paper a study has been done to show the works that have been done in the area of Automatic dialect classification/recognition of some Foreign and Indian languages till date. Also the phonology and morphology of Assamese language and the need to study the dialects present in the language has been discussed.**

Keywords: Accent, Dialect, Phonology, Morphology.

I. INTRODUCTION

In any natural language there are two linguistic variations: dialect and accent. Accents are the pattern of pronunciations of a person's language while Dialects are the varieties of vocabularies, idioms, grammars and pronunciation within a specified language used by a particular community because of geographical differences. Social factor shows that members of a specific socioeconomic class such as working-class might have different dialects compared to high-class business man. So the way a person speaks his/her language is also influenced by both his/her social status[1].Dialects of a particular language differ from each other, still speakers of another dialect of the same language understands it. Automatic dialect classification has many applications: improvising the performance of the speech recognition and speaker recognition systems, dialect identification, human machine interaction and developing regional dialect spoken query system. These systems will be useful for the farmers in accessing information about the agricultural commodities; also automatic routing of the customer's call to the regional dialect desk in customer service center can be easily done.

Assam is located in the heart of northeast India spreading over an area of 78,438.00 square kilometers. It comprises of 27 districts. The inhabitants of Assam are an intermixture of the races like Mongolian, Indo-Burmese, Indo-Iranian and Aryan origin [3]. The principal language of Assam is Assamese and it is regarded as the lingua-franca of the whole northeast India [3]. Most of the natives of the state speak Assamese. Although Sanskrit is the basic of Assamese language, the vocabulary, phonology and grammar of this language have been influenced by the original inhabitants of Assam, such as the Bodos and the Kacharis. There is presence of various dialects in any

standard language due to the pattern of pronunciation and the vocabularies used by a particular community because of geographical differences.

It has become necessary to discriminate these dialects as it seems no significant and systematic study is carried out on the dialects of Assamese language. Section II discusses the related works that have been done in the area of Automatic dialect classification/recognition till date. Section III discusses about the phonology and morphology of Assamese language the need to study the dialects present in the language.

II. RELATED WORKS

"Automatic language identification is the process by which the language of a digitized speech utterance is recognized by a computer" [4]. Automatic dialect recognition studies were carried out for the languages like English (North and South dialect), Chinese (Mandarin and Taiwanese dialect), Spanish (Caribbean and Non-Caribbean dialect) [1], Japanese [2] and German [4]. Among Indian Languages few studies were carried out for Hindi [2, 7]. Dialects of Hindi Language such as Chattisgharhi (spoken in central India), Bengali (Bengali accented Hindi spoken in Eastern region), Marathi (Marathi accented Hindi spoken in Western region), General (Hindi spoken in Northern region) and Telugu (Telugu accented Hindi spoken in Southern region) were analyzed in [2]. While in [7] dialects like Hindi, Khariboli, Bhojpuri, Haryanvi and Bagheli were studied. Similar works were done for Sambalpuri Odia dialect in [8]. When we are concentrating on Assamese Language, in [9] Mech-Bodo dialect of North Bengal and standard Bodo language spoken in Assam has been analyzed. Nalbaria [10] and Barpetia [11] (Kamrupiya dialect group of Assam) has also been studied. But, it seems no significant and systematic study is carried out on theother dialect groups of Assamese language using the features derivedfrom speech. Different languages may vary not only in the structure, but in the type of structure- some may have elaborate rules of syntax while others may be richer in morphology [12].

According to [4] there are several Language identification Cues: Phonology, Prosody, Morphology and Syntax. In [5], some additional speech features such as Articulatory parameters, Spectral information, Phonotactic and Lexical knowledge has been discussed. Researchers have explored these above mentioned speech features to identify various dialects of a natural language. Hence, we are looking into the different features which can be derived from speech for identifying the dialects of Assamese. Grammar is the rules of a language, which is learned as one acquires a language [13]. These rules include: **phonology**: the sound of language, **morphology**: the structure of words, **syntax**: the combination of words into sentences, **semantics**: the relation of sounds and meanings, and the **lexicon**: mental dictionary of words.

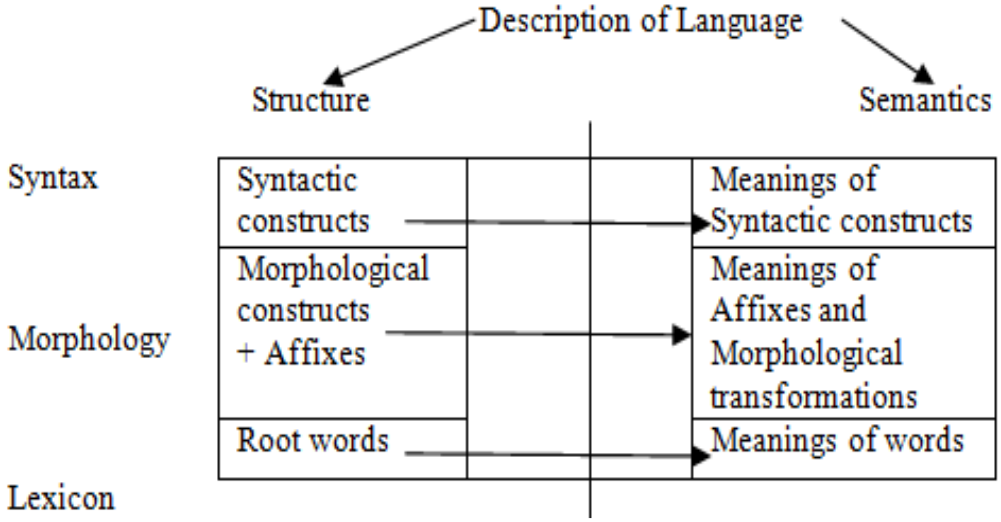

Fig: 1 Describes language with respect to structure and semantics [12]

Spectral features vary in languages due to dialect. According to [5] Spectral features, are easier to obtain but volatile because speech variations such as speaker or channel variations are present. Prosody is about duration,

pitch contours, and stress patterns that are present in the dialects and also in different languages. Prosodic features may be useful in applications such as recognition of language or speaker, where explicit phoneme/syllable boundaries are not easily available. A new approach for extracting and representing prosodic features directly from the speech signal has been introduced in[6] and a hypothesis was established that prosody is linked to linguistic units such as syllables, and it is manifested in terms of changes in measurable parameters such as fundamental frequency (F_0), duration and energy. Mel frequency cepstral coefficients (MFCC) can be used as spectral features and duration, pitch contour values can be used as prosodic features for the dialect recognition using 2-layer feed forward neural network [7]. Lexical/syntactic features rely on large vocabulary speech recognizer, which is language and domain dependant. They are therefore difficult to generalize across languages and domains [8]. In the development of natural language software, Morphological analyzer plays an important role. It takes the word as input and it gives the output in the form of root word, suffix and also some additional information. There are morphological analyzer for many Indian and foreign languages. They are based on standard language, but till now very few morphological analyzer has been developed for dialect of any Indian language [9]. Morphological analyzer was developed for Odia dialect [9], Assamese language [12, 14]. Applications of morphological analyzer can be found in grammar parser, checker, machine translation, query system, spell checker, speech recognition system and many more [9].

III. EMPHASIS ON THE STUDY OF ASSAMESE DIALECT

The rules discussed in Section II have been described below with respect to Assamese Language:
Phonology & Morphology: According to linguists, phone and phoneme set differ from one language to another.

A. Consonants:

Twenty-three consonant sounds and two semi-vowels are present in the Assamese Language:/b/,/p/,/pʰ/,/bʰ/, /t/, /tʰ/, /d/,/dʰ/,/k/,/kʰ/,/g/,/gʰ/,/m/, /n/ , /ŋ/,/s/,/z/,/ɦ/,/h/,/r/,/l/,/w/,/j/.[3]

B. Vowels:

- Monopthongs: Eight monopthongs are present as follows: /i/ , /e/, /ɛ/, /a/, /u/, /ʊ/, /o/, /ɔ/. The vowels in Assamese language occur namely in the three positions, word-initially, medially and finally [3].
- Diphthongs: Ten diphthongs are present in Assamese : /ai/, /ei/, /oi/, /ɔi/, /ui/, /iu/, /ou/, /au/, /eu/, /ua/.[3]

C. Nominal:

The nominal used in Assamese are: Noun, Pronoun, Interrogative, Demonstrative, Reflexive, Indefinite [3].

D. Nominal Inflections :

- Number: Assamese has singular and plural. Plural numbers are expressed by adding suffixes to the singular forms of nouns, pronouns and sometimes also to adjectives. They are even expressed by adding qualifying words. [3]
- Gender: In Assamese Language, gender can be differentiated mainly by the following three ways[3]:
  - Separate noun bases are used for male and female.
  - To distinguish gender separate qualifying words are used before or after the common nouns.
  - "In many cases *tatsama* words puruɦ (Skt. purusa, male) and mohila 'female' are used before the terms belonging to common gender" [3].
  - Sometimes gender is indicated by addition of enclitic definitive 'tɔ' for masculine and 'zoni' for feminine after the nouns [3].
  - Some suffixes are present in Assamese which are generally added to the masculine noun bases to indicate feminine gender. These suffixes are known as feminine suffixes.
- Verb:

  Verbs in Assamese are classified to Main verbs and Auxiliary verbs. There are simple and derived main verbs. To negativize verbs in Assamese /n/ can be added before the verb [3].

- Tense:

  Assamese verbs have three tenses: present, past and future [3].

Assamese Language has also several dialect variation based on the phonology and morphology.

According to *Banikanta Kakati* the Assamese dialects has been divided into two major groups [3]. They are:
a) Eastern Assamese.
b) Western Assamese.

But according to recent studies there are four major dialect groups [3]:
1. *Eastern group:* It is spoken in Sibsagar district and nearby districts surrounding it.
2. *Central group:* It is spoken in Nagaon district and adjoining areas.
3. *Kamrupi group:* It is spoken in Barpeta, Darrang, Kamrup, Nalbari, Kokrajhar and Bongaigaon district.
4. *Goalparia group:* It is spoken in Bongaigaon, Dhubri, Goalpara and Kokrajhar.

IV. CONCLUSION

Studies have showed that most of the works in the automatic dialect detection have been done in languages of western and eastern countries and very little work has been done for Indian languages. The study of the various dialects of Assamese Language will contribute a lot in developing regional dialect spoken query system and speech recognition system .These systems will be useful for the farmers in accessing information about the agricultural commodities, also automatic routing of the customer's call to the regional dialect desk in customer service center will be easily done. In future work, the 4 dialect group of Assamese Language will be studied in details on the basis of Phonology, Morphology, Prosody and Lexical features and will develop an automatic dialect recognition system for the Language.

REFERENCES

[1] Behravan, Hamid. "Dialect and Accent Recognition", Master's Thesis, University of Eastern Finland, School of Computing, December, 2012

[2] Rao, K. Sreenivasa, Saurav Nandy, and Shashidhar G. Koolagudi. "Identification of Hindi dialects using speech." WMSCI-2010 (2010).

[3] www.iitg.ernet.in/rcilts/pdf/assamese.pdf

[4] Zissman, Marc A., and Kay M. Berkling. "Automatic language identification."Speech Communication 35, no. 1 (2001): 115-124.

[5] Tong, Rong, Bin Ma, Donglai Zhu, Haizhou Li, and EngSiongChng. "Integrating acoustic, prosodic and phonotactic features for spoken language identification." In Acoustics, Speech and Signal Processing, 2006.ICASSP 2006 Proceedings.2006 IEEE International Conference on, vol. 1, pp. I-I.IEEE, 2006.

[6] Mary, Leena, and BayyaYegnanarayana. "Extraction and representation of prosodic features for language and speaker recognition." Speech communication 50, no. 10 (2008): 782-796.

[7] Sinha Sweta, Jain Aruna, Agawam Sham."Speech processing for Hindi dialect recognition."Advances in Signal Processing and Intelligent Recognition Systems, Advances in Intelligent Systems and Computing Volume 264, 2014, pp 161-169

[8] Sethi Dhabal Prasad "Morphological Analyzer for Sambalpuri Odia Dialect Inflected Verbal Forms."International Journal of Advanced Research in Computer Science and Software Engineering: Volume 3, Issue 10, October 2013 ISSN: 2277 128X.

[9] Muchahary, Rujab. "Variation in the lexicon of the Mech (Boro) dialect of North Bengal and standard Boro language spoken in Assam." International Journal of Scientific and Research Publications: 616.

[10] Patgiri, Bipasha. "Stress in the Nalbaria Dialect of Assamese". PhD Student, JNU

[11] Das, Harish."A Regional Dialect of Assamese Language". IJCA SPECIAL ISSUE ON BASIC, APPLIED & SOCIAL SCIENCES, VOLUME II, JULY 2012 [ISSN: 2231-4946] 159

[12] Sharma, Utpal."Unsupervised Learning of Morphology of a Highly Inflectional Language",2006.Ph.DThesis,Supervisors:Rajib K Das and Jugal K Kalita

[13] http://ielanguages.com/linguist.html

[14] Saharia, Navanath, Utpal Sharma, and Jugal Kalita. "Analysis and evaluation of stemming algorithms: a case study with Assamese." In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, pp.842-846.ACM, 2012.