ISSN 2250-3153

Dialect variation in Boro Language and Grapheme-to-Phoneme conversion rules to handle lexical lookup fails in Boro TTS System

Chandan Sarma, Prof. P.H Talukdar

Department of Instrumentation & USIC, Gauhati University, Guwahati, Assam, India.

Abstract- It is not possible to include all the words in a natural language for general text-to-speech system. Grapheme-to-phoneme conversion system is essential to pronounce a word which is out of vocabulary. Grapheme-to-phoneme rules play a vital role where lexical lookup fails. Though basic Grapheme-to-phoneme rules system is very simple yet it is very powerful for naturalness of a TTS system. Letter-to-sound rules may be hand written or maybe automatic depending on the language. We have worked on Bodo language. After a systematic study of Boro language we found that there is a systematic relationship between the written form of a Bodo word and its pronunciation. So, it is fairly easy to build letter-to-sound rules by hand for Bodo language. We have used a Bodo corpora of 5000 words and built letter-to-sound rules. These rules have been tested using Festival, a most popular speech synthesizer and applying these rules, we were able to produce correct pronunciations for approximately 89% of the words. Again, dialect variation also influences grapheme-to-phoneme conversion rules. This paper gives overview of Boro dialect variation and grapheme-to-phoneme conversion rules developed for Boro TTS system.

Index Terms- Letter-to-sound rule, Text-to-speech, lexicon, NLP

I. Introduction

The prime cause of building grapheme-to-phoneme(G2P) rules system in a text-to-speech system(TTS) is to treat the words which are not available in the lexicon. Second cause, LTS system drastically reduces the memory amount required by a big lexicon. The rule set can be viewed as a sort of compression algorithm that captures language regularities [1].Complex word morphology as well as accentuation pattern in stress may disrupt regularities.

Several frameworks have been proposed to build a grapheme-to-phoneme system of a text-to-speech system, among which can be mentioned : information theoretic systems such as decision tree [2], automatically-trained decision tree [3], table look-up models [4], dictionary-based approaches [5],linguistically rule-base modules [6], hybrid systems [7], neural networks approaches [8], Finite State Transducers [9], statistical approaches [10] and HMMs [11] .A comparison of different frameworks with their results was made by Damper et al [12]. One of the most popular approaches is dictionary–based, which uses a large dictionary containing the phonetic transcription of a given number of words. This technique has been widely applied to languages whose orthography is roughly phonetically based, such as English. The main drawback of this approach is it fails to treat the new words that are not in the dictionary [13].

There are two basic logical stages of text-to-speech synthesis. The first stage accepts raw text as input, processes it and converts it into precise phonetic string to be spoken, appropriately annotated with prosodic markers (stress and intonation). This stage may be referred to as Natural Language Processing (NLP).The second stage which may be referred as Speech Synthesis (SS) , accepts this phonetic presentation of speech and generates the appropriate digital signal using a particular synthesis technique [14]. For SS, formant based techniques [15], or diphone based techniques are normally employed. These techniques are generally script independent. However, NLP is totally dependent on cultural and linguistic specific usage of script. NLP contains different components. The first component is dedicated to pre-processing and normalizing input text. After normalization, the second component performs phonological processing to generate a more precise phonetic string to be spoken. A major task in the Phonological Processing Component is to convert the input text into phonemic string using letter-to-sound rules [16].

For a natural language it is not possible to explicitly list all the words in that language. So, when a new word comes up which is not explicitly listed, the lexical lookup fails in that case. TTS uses letter-to-sound rules to treat the new words which are out of the lexicon. So, letter-to-sound acts as a backup. Letter-to-sound rules system makes TTS light as it saves memory.

We worked on Bodo language. Bodo shows a relatively regular behavior and thus Bodo pronunciation can be easily modeled from Bodo text by defining fairly regular rules. We have built letter-to-sound rules for Bodo and tested with Festival synthesizer. We were able to producing correct pronunciations for approximately 89% of the words. This paper overviews our works.

II. LANGUAGE OVERVIEW

Bodo belongs to the Bodo sub-section of Bodo-Naga section under the Assam-Burmese group of the Tibeto-Burman branch of the Tibeto-Chinese family. Before 1953, the Bodo language had no standard form of writing. Although Roman script and Assamese script were used in the past, recently Bodos adopted the Devanagiri script. According to some scholars, the Bodo language had a script of its own called 'Deodhai'. This language has a total of 22 phonemes: 6 vowels and 16 consonants. Use of the high back unrounded vowel phoneme /w/ is very frequent in Bodo language. The Bodo language has different special characteristics such as: It has intonation pattern, juncture and two types of tones. The words in Bodo are highly monosyllabic. It has agglutinative features also. The vowels and the consonants of Bodo language is given below in the TABLE I

Pure Vowels	***Consonants***
/i/	/pʰ/
/e/	/b/
/a/	/tʰ/
/ɔ/	/d/
/u/	/kʰ/
/ɯ/	/g/
	/m/
	/n/
	/ŋ/
	/s/
	/z/
	/h/
	/ɾ/
	/l/
	/w/
	/j/

Table 1

III. DIALECT AND ITS VARIATION IN BORO LANGUAGE

The term dialect was first coined in 1577 from the Latin dialectus, *way of speaking*. Dialectal variation is present in most language areas and often has important social implications. In sociolinguistics a variety, also called a lect, is a specific form of a language or language cluster. A dialect can be defined as a variant of a given language which is spoken by a specific group or in a particular location, but whose distinctive features – e.g. vocabulary, grammar, pronunciation – are "not distinctive enough" to be recognized as a separate language. In addition, though, there are dialect varieties associated with particular ethnic groups (sometimes called ethnolects), socioeconomic classes (sometimes called sociolects), or other social or cultural groups, speakers can understand each other although they don´t recognize the use of some words or their pronunciation

Before 1953 Bodo had no standard form of writing. It has dialect areas extending in the plains of Assam from Dhubri in the west to Sadiya in the east. Among the dialects that are currently outside of Assam, the prominent one is the Mech dialect prevailing in the northern area of West Bengal. Dr. Pramod Chandra Bhattacharjya, in his doctoral thesis, "A Descriptive Analysis of the Boro Language" (1977), mention that there are at least four dialect areas of the Boro language (1) North-West dialect area having sub-dialects of North Kamrup and North Goalpara district. (2) South-West dialect area comprising South Goalpara and Garo Hills Districts. (3) North-Central Assam dialect area comprising Darrang, Lakhimpur district and a few places of Arunachal. (4)Southern Assam dialect area comprising Nowgong, North-Cachar, Mikir Hills, Cachar and adjacent districts (the area and the name of the districts are as in then Assam) [17].

The Boro dialect areas divided by Dr. Promod Ch Bhattacharjya may be re-structured according to the change brought out due to further divisions of the prevailing districts. According to [18], the dialects spoken in Assam could broadly sub- divided into three main groups-

1) **The Western Bodo dialect (Swnabari) (WBD)**: The Western Boro dialects are spoken in the districts of Kokrajhar, Bongaigaon, Chirang and Dhubri.
2) **The Eastern Bodo dialect (Sanzari) (EBD):** The Eastern Bodo dialects are found mainly in the districts of Barpeta, Nalbari,Baksa, Kamrup and some parts of Darrang as well.
3) **The Southern Bodo dialect, (Hazari) (SBD):** The Southern Bodo dialects are found mainly in the district of Goalpara including Rani, Krishnai the southern part of Brahmaputra River.

There are similarities in western Bodo dialect and eastern Bodo dialect though slight phonological variation is there. But they are greatly differing from southern Bodo dialect in phonological sense. These dialects vary in the levels of structures of phonology, grammar and vocabulary.

There is lexical variation in dialects current in Kokrajhar and Dhubri districts from those current in Nabari, Barpeta, Baksa and Kamrup districts areas likes –

/leech/- /bedlao/p^{h}ansu/p^{h}ansuk^{h}u/

/the Moon/- /ok^{h}ap^h ɯr/uk^{h}umbri/ uk^{h}umbrɯi/

Again, the lexical variation in between the dialect current in Goalpara and Kamrup district areas also may be compared as –

/guest/-/bunda/alasi/

/milk/- /bundi/gakhir/gakher/

The lexical variation in between the dialect current in Kokrajhar, Dhubri and Darrang district areas also may be compared as –

/wife/- /bisi/bizi/

/eat/- /zado/zaio.

There is phonological variation in dialects current in Kokrajhar and Dhubri districts from those current in Nabari, Barpeta, Baksa and Kamrup districts areas likes –

/to come/- /phɯi/phai/

Again, the phonological variation in between the dialect current in Goalpara and Kamrup district areas also may be compared as –

/you/- /noŋ/nɯŋ/

The phonological variation in between the dialect current in Kokrajhar, Dhubri and Darrang district areas also may be compared as –

/creeper/- /bendɯŋ/bɯndɯŋ/

IV. . ALIGNMENT ISSUE IN LETTER-TO-SOUND CONVERSION

Alignment is a major consideration in Grapheme-to-phoneme conversion that is letter-to-sound conversion. Mainly we considered two types of alignments 1) one-to-one and 2) many-to-many.

In one-to-one alignment each letter or grapheme maps only to one phoneme and vice versa. There are several problems with this approach. Grapheme strings and phoneme strings are not typically the same length, so null phonemes and null graphemes must be introduced to make one-to-one possible. Again, two letters frequently combine to produce a single phoneme (double letters) and a single grapheme can sometimes generate two phonemes (double phonemes). So, following are the main problems with one-to-one alignments:

1.Double letters: two letters map to one phoneme (e.g. sh- [ʃ] , ph – [f])
2.Double phoneme: one letter maps to two phonemes (e.g. x-[ks] , u- [ju])

First we considered the double letter problem in case of Bodo language. In most cases when the grapheme sequence is longer than the phoneme sequence, it is due to some silent letters. For example, in the Boro word khra, pronounced [k r a], the letter h generates a null phoneme (€). This is well captured by one-to-one aligners. However, the longer grapheme sequence can be generated by double letters; for example, in the word bong, pronounced [b o η], the letters ng together produce the phoneme [η].In this case one-to-one aligners using null phonemes will produce an incorrect alignment. This can cause problems for the phoneme prediction model by training it to produce a null phoneme from either of the letters n or g.

A new phoneme is introduced to represents two or more phonemes in case of double phoneme. For example in the word zwj with phoneme sequence [z o w j], the letter w generates both [o] and [w] phonemes. According to [19], there are two possible solutions for constructing a one-to-one alignment in this case. The first is to create a new phoneme by merging the phonemes [o] and [w].This requires constructing a fixed list of new phonemes before beginning the alignment process. The second solution is to add a null letter in the grapheme sequence.

Many-to-many alignments overcome the problems of one-to-one. We have considered Boro words and their phonemes and alignments are made across graphemes and phonemes. For example, the word gwzwng , with phonemes [guzang], is aligned as :

g	w	zw	n	g
|	|	|	|	|
g	u	z	an	g

The letters zw are an example of the double letter problem (mapping to the single phoneme [z]), while the letter n is an example of the double phoneme problem (mapping to both [a] and [n] in the phoneme sequence).These alignments provide more accurate grapheme-to-phoneme relationship for a phoneme prediction model.

Once many-to-many alignments are built across graphemes and phonemes, each word contains a set of letter chunks, each consisting of one or more letters aligned with phonemes. Each letter chunk can be considered as a grapheme unit that contains either one or two letters. In the same way, each phoneme chunk can be considered as a phoneme unit consisting of one or two phonemes.

V. BUILDING GRAPHEME-TO-PHONEME RULES

Bodo language shows a very regular mapping from graphemes-to-phonemes. It has a systematic relationship between the written form of a word and its pronunciation. So, we preferred to write down letter-to-sound rules by hand. We used Festival to test the rules. In Festival there is a grapheme-to-phoneme rules system that allows rules to be written, but it also provided a method for building rule sets automatically which will often be more useful [20]. Letter-to-sound rules written by hand are totally context dependent. We re-write rules which are applied in sequence mapping string of letters to string of phones.

Before starting to build the rules, we first declare the set. The symbols in the rules are treated as set names if they are declared as such or as symbols in the input/output alphabets. The symbols may be more than one character long and the names are case sensitive. For example C denotes the set of all consonants, V denotes set of all vowels, # denotes word boundary etc.

The basic form that we have followed as per [21] to build letter-to-sound rules is as follows:

```
( LEFTCONTEXT [ ITEMS ] RIGHTCONTEXT =
NEWITEMS )
```

This interpretation is that if *ITEMS* appear in the specified right and left context then the output string is to contain *NEWITEMS*. Any of *LEFTCONTEXT, RIGHTCONTEXT* or *NEWITEMS* may be empty. Note that *NEWITEMS* is written to a different "tape" and hence cannot feed further rules (within this rule set). Some of the grapheme-to-phoneme rules for Bodo language developed by us are given in the Table 2.

G2p Rule	*Example*
(C[u] = o)	(khr[u]=khro)
(C[w] = oo)	(kh[w] =khoo)
(C[Ay] = i)	(n[Ay]=nai)
(C[i]C = ee)	(ph[i]d=pheed)
([u]C = o)	([u]n=on)
(C[g]A= ow)	(n[g]A=now

Table 2

The rules are tried in order until one matches the first (or more) symbol of the tape. The rule is applied adding the right hand side to the output tape. The rules are again applied from the start of the list of rules.

VI. CONCLUSION

We had used Bodo corpora of 5000 words. The words were selected from continuous paragraph of Bodo news bulletin. As an experimental basis we had developed 72 G2P rules. Due to limitation we have mention few of the rules only. All the rules were tested in festival speech synthesizer. We were able to produce correct pronunciations for approximately 89% of the words. Among the words name of persons, places, objects were there. These rules will helpful to develop a high quality Bodo TTS system.

Festival has a facility to build LTS rules for bigger lexicon automatically. Our future aim is to build LTS rules for bigger lexicon automatically using festival.

REFERENCES

[1] Vincent Pagel, Kevin A. Lenzo, and Alan W. Black(1998).”Letter to Sound Rules for Accented Lexicon Compression”. In: CoRR, Vol. cmp-lg/9808010.
[2] J. M. Lucassen, R. L. Mercer(1984).”Discovering phonemic based forms automatically: an information theoretic Approach”. In IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 42.5.1-42.5.4, 1984.
[3] Black, A., Lenzo, K. and Pagel, V.(1988).”Issues in Building General Letter to Sound Rules”. In 3rd ESCA Workshop on Speech Synthesis, Jenolan Caves, Australia, pp. 77-80
[4] A. Bosch and W. Daelemans(1993).”Data-oriented methods for grapheme-to- phoneme conversion”. In Sixth Conference of the European Chapter of the Association forComputational Linguistics (EACL'93), pp. 45-53.
[5] Cecil H. Coker, Kenneth W. Church, and Mark Y.Liberman (1990).”Morphology and rhyming: Two powerful alternatives to letter-to-sound rules for speech synthesis”. In Conference on Speech Synthesis. European Speech Communication Association.
[6] R. M. Kaplan and M. Kay(1994). “Regular models of phonological rule systems”. In Computational Linguistics 20(3), pp. 331-378.
[7] H. M. Meng, S. Seneff, and V. Zue(1994).”Phonological parsing for bi-directional letter-to-sound/sound-to-letter Generation”. In ARPA Human Language Technology Workshop, Princeton, USA.
[8] T.J. Sejnowski and C. R. Rosenberg(1987).”Parallel networks that learn to pronounce English Text”. In Complex Systems, 1, pp. 145-168, 1987.
[9] E. Roche and Y. Schabes (1995). Exact Generalization of Finite-State Transductions: Application to Grapheme-to-Phoneme Transcription. Technical Report TR-95-08,Mitsubishi Electric Research Laboratories, Cambridge,USA.
[10] A. Chotimongkol, and A. Black (2000).”Statistically trained orthographic to sound models for Thai”.In ICSLP2000,Beijing, China.
[11] P. Taylor (2005).”Grapheme-to-Phoneme conversion using Hidden Markov models”. In Proc. Interspeech 2005,Lisbon, Portugal, September.
[12] R.I. Damper, Y. Marchand, M.J. Adamson, K. Gustafson (1998).”Comparative Evaluation of Letter-To-Sound Conversion Techniques For English Text-To-Speech Synthesis”. In Proceedings of 3rd ESCA/COCOSDA International Workshop on Speech Synthesis, Jenolan Caves, Australia, pp. 53-58.
[13] Braga D., Coelho L.(2006).”Letter—to—sound conversion for Galician TTS systems”. In: Actas de las IV Jornadas en Tecnologia del Habla, Zaragoza, pp. 171—l76.
[14] T. Dutoit(1997). An Introduction to Text-to-Speech Síntesis. Kluwer Academic Publishers,Dordrecht, The Netherlands.
[15] D. H. Klatt.(1980). Software for Cascade/Parallel Formant SynthesiZer. JASA 67: 971-995.
[16] Hussain, Sarmad(2004),” Letter-to-Sound Conversion for Urdu Text-to-Speech System”. In COLING 2004 Computational Approaches to Arabic Script-based Languages, Geneva, Switzerland
[17] Madhu Ram Baro, Structure of Boro language, 2008
[18] Phukan Basumatary, Introduction to Bodo Language.
[19] H. Lee, and C. Seong (1996).Experimental phonetic study of the syllable duration of Korean with respect to the positional effect. In The Fourth International Conference on Spoken Language Processing (ICSLP'96), Philadelphia (USA), pp. 1193-1196.
[20] A.W. Black and K.A. Lenzo (2003).Building synthetic voice. http://festvox.org/bsv/.
[21] www.cstr.ed.ac.uk/projects/festival/manual/festival-1.4.0.ps.gz

AUTHORS

First Author – Chandan Sarma, M.Sc(Computer Science), Gauhati University,chandan.rrc@gmail.com
Second Author – Prof.P.H Talukdar,M.Sc,PhD, M.Tech, Gauhati University,phtassam@gmail.com.

Correspondence Author – Chandan Sarma, chandan.rrc@gmail.com,sarmachandan14@rediffmail.com, 09864112523.