Use of Part Of Speech (POS) and morphological information for resolving Multiple Pronunciations in Pronunciation Lexicon Specification (PLS) for Indian Languages – Bengali as a Case Study

Shyamal Das Mandal[1], Somnath Chandra[2] and Swaran Lata[2]
[1]Centre for Development of Advanced Computing (C-DAC), Kolkata 70091
[2] Dept of Information Technology, Govt. of India and W3C India, 6 CGO Complex , New Delhi 110003

e-mail: [1]shyamal.dasmandal@kolkatacdac.in ; [2] schandra@mit.gov.in , slata@mit.gov.in

Abstract
Pronunciation dictionary is one of the important components for the speech technology development for a particular language. This is because it represents the interface between speech analysis on the acoustic level and speech interpretation. The W3C Voice Browser Activity has published a Pronunciation Lexicon Specification (PLS) Version 1.0 [1] for generation of PLS in different languages. This paper proposes some modification of the published PLS specification with respect to Indian languages with Bengali as a typical case study.

1. Introduction
Man machine communication in speech mode involves the integration of all technologies needed for both speech input, as well as output, as per all the attributes demanded by the discipline of associated language. In this context, as explained earlier for India, Speech Synthesis and Speech Recognition are considered to be of primary need not only to empower disabled people, but also to functionally literate population.

One of the important components for the speech technology development for the particular language is the pronunciation dictionary. This is because it represents the interface between speech analysis on the acoustic level and speech interpretation. For example in Automatic Speech Recognition (ASR), the search module relies on phonetic transcriptions to select appropriate acoustic models against which to score the input utterance. Likewise, in Text-to-Speech (TTS) synthesis, phonemic transcriptions are required for the selection of the proper units from which to generate the desired waveform.

Consistent specification of word pronunciation is critical to the success of many speech technology applications. Most state-of-the-art Automatic Speech Recognition (ASR) and Text-To-Speech (TTS) systems rely on lexicons, which contain pronunciation information for many words.
Several guidelines have been reported to define the structure of a pronunciation lexicon, ranging from simple two-column ASCII lexicons providing the mapping between graphemic and phonemic transcriptions, to more general de-facto standards and new standardization attempts, which are also handling multiple orthographies and multiple pronunciations. Another initiative, the W3C Voice Browser Activity, has published a Pronunciation Lexicon Specification (PLS) Version 1.0. PLS is designed to enable interoperable specification of pronunciation information for both speech recognition and speech synthesis engines within voice browsing applications. The language is intended to be easy to use by developers while supporting the accurate specification of pronunciation information for international use. The present Pronunciation Lexicon Specification (PLS) Version 1.0; W3C, may require some modification to make it more usable. In this paper we have proposed some modification of present PLS with typical example of an Indian language namely Bangla.

2. Text-to-phoneme conversion
One of the important components for the speech technology development for a particular language is the pronunciation dictionary which in turn is largely dependent on grapheme to phoneme conversion system. Each word must be assigned its lexical stress and should be converted into a phonetic representation. Rules and lexicons are designed for this purpose [2]. The letter-to-sound (grapheme-to-phoneme) conversion is relatively direct for some languages, while it can be highly unpredictable for others, like English and some of the Indian languages. In such languages same grapheme information may have different pronunciation based on its Part of Speech information and its semantics.

3. The usage of POS information for resolving multiple pronunciations
In many Indian languages (like Bangla, Hindi) Parts of Speech (POS) plays an important role in pronunciation. Based on the POS same orthography can produce different pronunciation. For example

the Bengali word সরল (“sarala”) is pronounced in two different ways when it is verb it is pronounced as */ʃorlo/* (moved) and */ʃɔrol/* (easy) when it behaves as adjective.
In PLS specification, multiple pronunciations can be shown with phoneme elements. And attribute “prefer” can be used to give one pronunciation high priority among many pronunciation candidates but this does not solves the above problem for TTS and ASR. The attribute “prefer” can be used for defining the ideolectal variation of the same orthographic information. As figure-1 represent an example use of the attribute “prefer” for ideolectal variation of a Bengali word.

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>ঊনত্রিশ</grapheme>
    <phoneme prefer="true">untriʃ</phoneme>
    <!-- IPA string is: "untriʃ" -->
    <phoneme>unotiriʃ</phoneme>
    <!-- IPA string is: "unotiriʃ" -->
  </lexeme>
</lexicon>
```

Figure-1

W3C PLS framework tackled the problem of POS using “role attribute” under the lexeme tag in the Pronunciation Lexicon Specification (PLS) version 1.0 14 October 2008. But if a separate XML element or an attribute under Phoneme tag is provided for POS then the number of entry of lexeme will be reduced.

3.1 Proposal 1: Pos as an attribute

The "pos" can be an optional attribute under the phoneme element which indicates the detail information for obtaining the pronunciation for speech recognition and speech synthesis. The possible values for the POS information of the respective word of the language (i.e "verb","noun", “adjective” etc.

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>সরল</grapheme>
    <phoneme pos= “adjective”> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol " -->
    <!—Itrans is: "sarala" -->
    <!—Meaning is : "easy" -- >
    <phoneme pos= “verb” > ʃorlo </phoneme>
    <!-- IPA string is: " ʃorlo" -->
    <!—Itrans is: "sarala" -->
    <!—Meaning is: “moved” -- >
     <phoneme pos= “null”> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol " -->
    <!—Itrans is: "sarala" -->
    <!—Meaning is: “ easy” -- >
  </lexeme>
 </lexicon>
```

Figure -2

This will be very effective for solving the problem of same grapheme having different pronunciation in most of the Indian languages

3.2 Proposal 2: Pos as an element

The <lexeme> element may contain optionally one or more <pos> element. Each <pos> element contains the pronunciation of the word depending on pos element information.

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>সরল</grapheme>
    <pos> adjective </pos>
    <phoneme> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol " -->
    <!—Itrans is: "sarala"  -->
    <!—Meaning is : "easy"  -- >
    <pos> verb </pos>
    <phoneme> ʃorlo </phoneme>
    <!-- IPA string is: " ʃorlo" -->
    <!—Itrans is: "sarala"  -->
    <!—Meaning is: “moved” -- >
    <pos> null </pos>
    <phoneme> ʃɔrol </phoneme>
    <!-- IPA string is: " ʃɔrol " -->
    <!—Itrans is: "sarala"  -->
    <!—Meaning is: “ easy”  -- >
  </lexeme>
 </lexicon>
```

Figure -3

4. The usage of morphological information for resolving multiple pronunciations

4.1 POS with morphological information

In some of the languages like Bengali not only POS information but also morphological information especially in case of verb finiteness and honorificity information are very crucial in determining the pronunciation of a homograph. This information can be defined in the same attribute or element used for POS using proper POS tag set for that language. The above problem is illustrated by the example below for Bengali language.

Example-1:

The POS and morphological information of Bengali word “করে” can be defined as *VM.3.prs.smp.dcl.fin.n.n.n or VM.0.0.0.0.nfn.n.n.n* depending on its context

where VM → main verb; 0 → person not required ; 0 → tense not required; 0 → aspect not required ; 0→ mood not required ; nfn→ non finite; n→ no emphatic; n→ no negative marker ; n→ no honorific marker. This can be handled in the PLS as in Figure-4 or Figure-3

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>করে</grapheme>
    <phoneme: pos= “VM.3.prs.smp.dcl.fin.n.n.n” >kɔre</phoneme>
    <!-- IPA string is: "kɔre" -->
    <!—Itrans is: “kare” -->
    <!—Meaning is: “do/does” -- >
    <phoneme: pos= “VM.0.0.0.0.nfn.n.n.n” >kore</phoneme>
    <!-- IPA string is: "kore" -->
    <!—Itrans is: “kare” -->
    <!—Meaning is: “having done” -- >
  </lexeme>
</lexicon>
```

Figure-4: POS with morphological information as an attribute

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="bn">
  <lexeme>
    <grapheme>করে</grapheme>
    <pos> VM.3.prs.smp.dcl.fin.n.n.n </pos>
    <phoneme>kɔre</phoneme>
    <!-- IPA string is: "kɔre" -->
    <!—Itrans is: “kare” -->
    <!—Meaning is: “do/does” -- >
    <pos> VM.0.0.0.0.nfn.n.n.n </pos>
    <phoneme>kore</phoneme>
    <!-- IPA string is: "kore" -->
    <!—Itrans is: “kare” -->
    <!—Meaning is: “having done” -- >
  </lexeme>
</lexicon>
```

Figure-5: POS with morphological information as an element

Similarly Bengali homograph ধর has two pronunciations depending upon its context which can be tackled by the morphological information. “ধর” can be defined as VM.2.prs.smp.dcl.fin.n.n.***y*** *or* VM.3.prs.smp.dcl.fin.n.n.***n*** where y→ with honorific marker; n→ no honorific marker

4.2 Morphological Analysis for reducing the size of the Lexicon

Morphological Analysis is also very helpful in reducing the size of the Lexicon. Most of the dictionary used a standard format [3] for describing the morphological information of a particular word. The example of such format is given below.

চার *a.* four.˜ইয়ারি *a.* elicited out of the meeting of four friends.˜কোনা *a.* square; qudrangular.˜খানা *n.* four pieces or bits or units.˜চালা *a.* having a thatched roof consisting of four parts.˜চৌকো *a.* square.˜টা (coll.)˜টে *a.* four. □ *n.* four o'clock.˜টি *a.* four; a few or a little. *n.* four˜টিখানি *a.* a little.˜দিক *n.* four directions or sides or quarters; all directions or sides or quarters or places.˜পায়া, (dial. corrup.)˜পাই *n.* a kind of four-legged cot (usu. made of bamboo and coconutcoir). ˜পেয়ে *a.* four-legged, four-footed, quadruped.˜পোয়া, (dial.)˜পো *a.* complete, full.˜ভিত same as চারদিক ।˜সন্ধ্যা four evenings; four parts or points of the day, namely, the morning, the noon, the evening and the midnight

This will reduce the multiple entry of the same word with different suffixes and word compounding. This can be implemented using the proposed xml format as in Appendix-1.

From Appendix-1it is observed that if the Bengali word চার is entered in the proposed PLS format then the derived word like চারকোণা, চারটা, চারখানা etc. are not required to enter separately. This will not only reduce the size of the lexicon but also reduce the search time for the pronunciation.

5. Conclusion

In this paper, we propose to use the POS along with morphological information for resolving multiple pronunciations and reducing the size of the lexicon. This can be used to choose the proper pronunciation among multiple pronunciations of the same orthography of many Indian Languages. Finally, this information can reduce the search time in a large vocabulary recognition and synthesis system.

Reference

[1] Pronunciation Lexicon Specification (PLS) Version 1.0; W3C Recommendation 14 October 2008

[2] Basu, J.; Basu, T.; Mitra, M.; Mandal, S, "Grapheme to Phoneme (G2P) conversion for Bangla," Speech Database and Assessments, 2009 Oriental COCOSDA International Conference on , vol., no., pp.66-71, 10-12 Aug. 2009.

[3] Biswas, Sailendra. Samsad Bengali-English dictionary. 3rd ed. Calcutta, Sahitya Samsad, 2000.

Appendix-1

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
      alphabet="ipa" xml:lang="bn">
 <stem>
   <lexeme>
    <grapheme>চার</grapheme>
    <phoneme>ʧɐr</phoneme>
    <!-- IPA string is: "ʧɐr" -->
  </lexeme>
   <lexeme>
    <grapheme>কোণা </grapheme>
    <phoneme>konɐ</phoneme>
    <!-- IPA string is: "konɐ " -->
   </lexeme>
   <lexeme>
    <grapheme>চালা</grapheme>
    <phoneme>ʧɐlɐ</phoneme>
    <!-- IPA string is: "ʧɐlɐ" -->
   </lexeme>
 </stem>
 <suffix>
  <lexeme>
    <grapheme>খানা</grapheme>
     <phoneme>kʰɐnɐ</phoneme>
    <!-- IPA string is: "kʰɐnɐ" -->
  </lexeme>
  <lexeme>
    <grapheme> টা</grapheme>
    <phoneme>ʈɐ</phoneme>
    <!-- IPA string is: "ʈɐ " -->
  </lexeme>
  <lexeme>
    <grapheme>টে</grapheme>
     <phoneme>ʈe</phoneme>
    <!-- IPA string is: "ʈe" -->
  </lexeme>
  <lexeme>
    <grapheme>টি</grapheme>
    <phoneme>ʈi</phoneme>
    <!-- IPA string is: "ʈi" -->
   </lexeme>
 </suffix>
</lexicon>
```