

# Creation of pronunciation Lexicon in W3C PLS Standards for Indian Languages (Version 1.0)

The present document is created following the discussion at the 1<sup>st</sup> meeting of “Creation of PLS for Indian Languages” on 5<sup>th</sup> May 2011 at JNU, New Delhi.

## 1. What is PLS of W3C?

The PLS is the standard format of the documents referenced by the <lexicon> element of SSML. The PLS engine will load the external PLS document and transparently apply the pronunciations during the processing of the SSML document. An application may contain several distinct PLS documents to be used in different points of the application. If a pronunciation lexicon is referenced by a SRGS grammar, it can allow multiple pronunciations of the word in the grammar to accommodate different speaking styles

## 2. Why do we need such a markup language?

In voice browsing applications there is often a need to use proper nouns or other unusual words within speech recognition grammars and in texts to be read out by Text-to-Speech processors. These words may not be present in the built-in lexicons of the platforms. In such cases voice browsers typically resort to automatic pronunciation generation algorithms, which may be improved by manually specified pronunciations. The goal of the pronunciation lexicon markup is to provide a mechanism for application developers to supply high quality additional pronunciations in a platform independent manner.

In many cases application developers will need to only provide one or two additional pronunciations inline within other voice markup languages. But there are other cases where an application may make use of large pronunciation lexicons that cannot conveniently be specified inline and have to be provided as separate documents. The pronunciation lexicon markup will address both communities.

The markup language for pronunciation lexicons will be developed within the following broad design criteria. They are ordered from higher to lower priority. In the event that two goals conflict, the higher priority goal takes precedence. Specific technical requirements are addressed in the following sections.

1. The pronunciation lexicon markup language will enable consistent, platform independent control of pronunciations for use by voice browsing applications.
2. The pronunciation lexicon markup language should be sufficient to cover the requirements of speech recognition and speech synthesis systems within a voice browser.
3. The pronunciation lexicon markup language will be an XML language and shall be interoperable with relevant W3C specifications (see section 2 Interoperability Requirements for details).
4. The pronunciation lexicon markup language will be usable in a large number of human languages (see the requirements 3.4 and 3.5).
5. It should be easy and computationally efficient to automatically generate and process documents using the pronunciation lexicon markup language.
6. All features of the pronunciation lexicon markup language should be implementable with existing, generally available technology. Anticipated capabilities should be considered to ensure future extensibility (but are not required to be covered in the specification).
7. The pronunciation lexicon markup language should be easy to author, where appropriate, deriving from existing pronunciation lexicon formats and using existing pronunciation alphabets.

### 3. Background of PLS for Indian Languages

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 a 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. To provide for a maximum coverage of the words, multi-word expressions or even phrases, which commonly occur in a given application.-domain, application-specific word or phrase pronunciations may be required, especially for application-specific proper nouns, such as personal names or location names.

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.

The ISO-TC37 initiative, which started at LREC 2002, initiated work on a family of ISO standards related to natural language processing (Romary et al., 2006). Currently these standards are available in working drafts of high-level specifications for word segmentation, feature structures, annotations, and also for lexicons. The high-level specifications build on lower-level specifications in form of language and country codes, data categories, code scripts, and Unicode. Lexicon specifications are covered by the "Lexical Markup Framework" under ISO 24613 (Romary et al., 2006). The same description structure in terms of morphology, syntax and semantics (and translation) applies to monolingual up to multilingual lexicons. Multi-word expressions are given special attention.

Another initiative, the W3C Voice Browser Activity, has recently 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 language allows one or more pronunciations for a word or phrase to be specified using a standard pronunciation alphabet or if necessary using vendor specific alphabets. Pronunciations are grouped together into a PLS document which may be referenced from other markup languages, such as the Speech Recognition Grammar Specification SRGS and the Speech Synthesis Markup Language SSML.

In the case of Indian Languages, grapheme to phoneme conversion is not much of a problem as there is very little discrepancy between the written text and its pronunciation. Phonological rules are thus comparatively less complex in most of the Indian Languages. But it is not true for Bangla language. The main problem of Bangla, Hindi and other Indian languages lies in the deletion of the neutral vowel /ə/ or /ɔ/ (in the case of Bangla). Bangla, like other Indian languages, is a semi-syllabic script. Each independent consonant letter in Bangla is associated

with an “inherent” neutral vowel, also known as ‘schwa’. If a consonant without a vowel has to be represented, then it should be marked with a subscript called ‘halant’ (to be interpreted as ‘consonant-ending’). But in practice, the neutral vowel is deleted in certain contexts, even if a ‘halant’ is not specified, and pronounced in others. For example, in the case of the word আচমকা ‘AchamakA’ the schwa with the medial ‘m’ is deleted whereas it is retained with ‘ch’, although there is no halant diacritic present for both the consonants. This is also true in the case of word final consonants. In the word final position, if a single consonant without any ligature is present then the neutral vowel is not pronounced, as in the Bangla word কঠোর ‘kaThora’ ‘hard’, which is pronounced as [kɔtʰor]) instead of [kɔtʰorɔ]. The same is the case with Hindi, where the lexical form ‘kaThora’ कठोर is pronounced as [kəʈho:r].

### 3. 1 An Example of Bangla PL in PLS standard

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- This pronunciation lexicon is licensed under the GPL. -->
<lexiconversion="1.0"
xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2005/01/pronunciation-lexicon
http://www.w3.org/TR/2007/CR-pronunciation-lexicon-20071212/pls.xsd"
alphabet="ipa" xml:lang="bn">
<lexeme>
<grapheme>আমি</grapheme>
<phoneme>ɛmi</phoneme>
</lexeme>
<lexeme>
<grapheme>ছাড়া</grapheme>
<phoneme>tʰɛɾɛ</phoneme>
</lexeme>
<lexeme>
<grapheme>দ্বিতীয়</grapheme>
<phoneme>ditijo</phoneme>
</lexeme>
<lexeme>
<grapheme>কেউ</grapheme>
<phoneme>keu</phoneme>
</lexeme>
</lexicon>
```

### 3. 2 An Example of Hindi PL in PLS standard

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- This pronunciation lexicon is licensed under the GPL. -->
<lexicon version="1.0"
xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2005/01/pronunciation-lexicon
http://www.w3.org/TR/2007/CR-pronunciation-lexicon-20071212/pls.xsd"
alphabet="ipa" xml:lang="bn">
<lexeme>
<grapheme>कसरत</grapheme>
<phoneme>kəsrət</phoneme>
</lexeme>
<lexeme>
<grapheme>एक</grapheme>
<phoneme>ek</phoneme>
</lexeme>
<lexeme>
<grapheme>कहानी</grapheme>
<phoneme>kəhani</phoneme>
</lexeme>
<lexeme>
<grapheme>सुंदर</grapheme>
<phoneme>sundər</phoneme>
</lexeme>
</lexicon>
```

### 4. Best practices for the creation PLS in W3C standard

- The pronunciation of the words in the lexicon will be represented using IPA symbols. Towards this end, a complete list of IPA symbols for all the phonemes in the Indic languages will be prepared (see Annexure-B).
- The pronunciation in the lexicon will be represented at the phonemic level.
- The pronunciation lexicons will be of the standard spoken variety.
- If an orthographic representation has multiple pronunciations, they may be labelled differently using the optional attribute POS, along with the phonemic elements, to indicate the detailed information for obtaining the pronunciation for speech recognition and speech synthesis. POS tags may have different morphological markers (optional) like finite/ non-finite, honorific/ non honorific (see Annexure A).

- Homophones, that is, different orthographic forms having identical pronunciation will be entered separately as different lexemes.
- If PLS standards of W3C are modified to include the representation of stress, then the lexical entries in the pronunciation lexicons will include stress marks.

### Proposed PLS structure for Indian Languages

```
<?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> G1</grapheme>
<pos> POS1.s1.s2.s3.s4.s5.s6.s7.s8</pos>
<phoneme: Standard="true" > P1</phoneme>
<phoneme: Standard="false" > P2</phoneme>
<pos> POS2.s1.s2.s3.s4.s5.s6.s7.s8 </pos>
<phoneme: Standard="true" > p3</phoneme>
<pos> null </pos>
<phoneme: Standard="true" > P4 </phoneme>
</lexeme>
<lexeme>
<grapheme> G2</grapheme>
<pos> POS1.s1.s2.s3.s4.s5.s6.s7.s8</pos>
<phoneme: Standard="true" > P1</phoneme>
<phoneme: Standard="false" > P2</phoneme>
<pos> null </pos>
<phoneme: Standard="true" > P4 </phoneme>
</lexeme>
</lexicon>
```

## Annexure-B

The "POS" can be an optional attribute under the phoneme element which indicates the detailed information for obtaining the pronunciation for speech recognition and speech synthesis. The possible values for the POS information of the respective words of the language (i.e "verb","noun", "adjective" etc) are illustrated below.

```
<?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>___</grapheme>
<phoneme pos= "adjective">সরল</phoneme>
<!-- IPA string is: "সরল" -->
<!--Itrans is: "sarala" -->
<!--Meaning is : "easy" -->
<phoneme pos= "verb" >সরল</phoneme>
<!-- IPA string is: "সরল" -->
<!--Itrans is: "sarala" -->
<!--Meaning is: "moved" -->
<phoneme pos= "null">সরল</phoneme>
<!-- IPA string is: " সরল" -->
<!--Itrans is: "sarala" -->
<!--Meaning is: " easy" -->
</lexeme>
```

## A. Essential Requirement

1. Same orthography can produce different pronunciations based on the POS

S/L	Language	Orthography	POS	Pronunciation in IPA	Remarks
1					

### Example for Bengali

S/L	Language	Orthography	POS	Pronunciation in IPA	Remarks
1	Bengali	সরল	adjective	/ʃɔɾol/	Itrans is: "sarala" Meaning is : "easy"
			verb	/ʃorlo/	Itrans is: "sarala" Meaning is: "moved"

## B. Futuristic Requirement

### 1. The usage of morphological information for resolving multiple pronunciations

S/L	Language	Orthography	POS with morphological information	Pronunciation in IPA	Remarks
1			POS.s1.s2.s3.s4.s5.s6.s7.s8		

POS→ Part of Speech Marker; s1→ person marker; s2→ tense marker; s3→ aspect marker; s4→ mood marker; s5→ finite/nonfinite; s6→ emphatic/non emphatic; s7→ negative/ non negative; s8→ honorific/ non honorific

### Example for Bengali

S/L	Language	Orthography	POS with morphological information	Pronunciation in IPA	Remarks
1	Bengali		VM.3.Prs.smp.dcl.fin.n.n.n	/kore/	Itrans is: "kare" Meaning is: "do/does"
			VM.0.0.0.0.nfn.n.n.n	/kore/	Itrans is: "kare" Meaning is: "having done"
2	Bengali	ধর	VM.0.0.0.0.nfn.n.n.y	/dʰoro/	Itrans is: "dhar" Meaning is: "to hold"
			VM.0.0.0.0.nfn.n.n.n	/dʰor/	Itrans is: "dhar" Meaning is: "to hold"

### Where

VM→main verb; s1=3 (3<sup>rd</sup> person); s1=0 (not required) ; s8=y(honorific); s8=n(non-honorific)

### 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 for describing the morphological information of a particular word.



## Example from Bengali

```
<?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>ঢ়ত্ব</grapheme>
      <phoneme>tʃɐ</phoneme>
      <!-- IPA string is: "tʃɐ" -->
    </lexeme>
    <lexeme>
      <grapheme>কণ</grapheme>
      <phoneme>kɔɳ</phoneme>
      <!-- IPA string is: "kɔɳ " -->
    </lexeme>
    <lexeme>
      <grapheme>ঢ়ত্ব</grapheme>
      <phoneme>tʃɐɳ</phoneme>
      <!-- IPA string is: "tʃɐɳ" -->
    </lexeme>
  </stem>
  <suffix>
    <lexeme>
      <grapheme>খণ</grapheme>
      <phoneme>kʰɔɳ</phoneme>
      <!-- IPA string is: "kʰɔɳ" -->
    </lexeme>
    <lexeme>
      <grapheme>ত</grapheme>
      <phoneme>tɐ</phoneme>
      <!-- IPA string is: "tɐ " -->
    </lexeme>
    <lexeme>
      <grapheme>তৈ</grapheme>
      <phoneme>tɛ</phoneme>
      <!-- IPA string is: "tɛ" -->
    </lexeme>
    <lexeme>
      <grapheme>তি</grapheme>
      <phoneme>ti</phoneme>
      <!-- IPA string is: "ti" -->
    </lexeme>
  </suffix>
</lexicon>
```

## Annexure-1

### IPA representation of the phonemes of Indian languages

#### Abstract

This document is used for preparing the IPA representation of all the phonemes of Indian languages. It is proposed as a guide for the preparation of the phonemes in the Indian languages with their articulatory descriptions, example uses and sound samples as in Table 1.

### 3. Some useful definitions

#### 3.1 Consonants

**Place of articulation:** During the articulation the air-stream passing through the vocal tract is obstructed in some way. The place where the obstruction takes place is called the place of articulation

**Bilabial:** Bilabial sounds are produced when the two lips make the constriction

**Labiodentals:** These sounds are produced by contacting the lower lip with the upper teeth.

**Dental:** Dental sounds are produced by the constriction of the tip or blade of the tongue with the upper teeth.

**Alveolar:** The sound made by the tip or the blade of the tongue in contact against the alveolar ridge, which is the bony prominence immediately behind the upper teeth.

**Post alveolar:** The sound is articulated by the tip or the blade of the tongue with the back area of the alveolar ridge.

**Retroflex<sup>1</sup>:** Retroflex sounds are made when the tip of the tongue is curled back in the direction of the front part of the hard palate- in other words, just behind the alveolar ridge. Depending on how far the tongue curls back, retroflex consonants could be apico-postalveolar or apico-palatal.

**Palatal:** This sound is produced when the constriction is made by the front part of the tongue with the hard palate.

**Velar:** It refers to a sound made by the back of the tongue against the soft palate.

**Uvular:** This sound is produced when the back of the tongue touches the uvula.

**Pharyngeal:** It refers to a sound produced in the pharynx, the tubular cavity, which constitutes the throat above the larynx.

---

<sup>1</sup>Retroflexion is actually a manner of articulation where the tongue is curled back. It can have different places of articulation. If it is post alveolar or a palatal IPA symbols should be able to distinguish between them. Retroflexion can be indicated in the usual manner.

**Glottal:** These are the sounds which are made in the larynx due to the closure or narrowing of the glottis.

### **Manner of Articulation**

Manner of articulation is concerned with the airflow; the paths it takes and the nature of obstruction in the pathway.

**Plosive, or oral stop:** It is a type of consonant segment produced by a complete closure accompanied simultaneously by a velic closure. Because of these two closures the egressive air-stream is momentarily completely dammed up and unable to get through the vocal tract. The air is therefore compressed behind the point of articulation and will escape with a small explosion if the consonant is immediately followed by a vowel. But if the consonant is in the position before a pause then only the active articulator touches the passive articulator but no explosion sound is produced. In case of pulmonic ingressive stops, called implosives, the air behind the point of articulation is rarified and there will be a sudden rush of air inwards instead of outwards.

**Nasal stop: In the production of a Nasal Stop,** usually shortened to Nasal, there is complete occlusion of the oral cavity, and the air passes instead through the nose. The shape and the position of the tongue determine the resonant cavity that gives different nasal stops their characteristic sounds. Examples include English /m, n/. Nearly all languages have nasals.

**Fricative:** In the production of Fricative (sometimes called spirant) sounds two articulators come so close that the airstream moving between them produces an audible friction.

**Affricate:** An Affricate begins like a plosive, but releases into a fricative rather than having a separate release of its own.

**Lateral:** To produce a lateral sound, the air is obstructed by the tongue at a point along the center of the mouth but the sides of the tongue are left low so that the air can escape over its sides. In fact, the tongue is strongly flexed and the air is forced through a narrow oval cavity, producing a hushing sound.

**Approximant:** An approximant is a consonant that makes very little obstruction to the airflow. Approximants are divided into two main groups: semivowels (also called glides) and liquids.

**Tap:** A tap is produced with the tip of the tongue striking the hard palate at a place once and returning to its place of rest.

**Trill:** A trill is a sound in the production of which the articulator (usually the tip of the tongue) is held in place, and the airstream causes it to vibrate. More than one tap makes a trill. Taps, Trills and flaps, where there are one or more brief occlusions, constitute a class of 'r' like consonants called rhotics.

**Flap:** A flap is produced when the tongue strikes the hard palate curled back and quickly returns to its place of rest. There are also lateral flaps.

### 3. 1 Vowels

Vowels are of two main types: Oral and Nasal or Nasalized.

**Oral vowels:** In the production of Oral vowels, there is velic closure, allowing the air to pass only through the oral cavity.

**Nasal vowels:** In the production of Nasal vowels, there is velic opening, allowing the air to pass through both the oral and the nasal cavities. Nasal vowels are symbolized using the diacritic '̃' above the symbols.

Each of the above types has further division into monophthongs or 'pure' vowels and 'diphthongs' or glides.

**Monophthongs:** are produced with the tongue being held in the same position throughout the production of the vowel.

**Diphthongs:** are produced with the tongue 'gliding' from one position to another in the production of a vowel. A diphthong is a single vowel.

Vowels are of three main types along the parameter of frontness-backness.

**Front:** Front vowels are produced with the front of the tongue raised against the roof of the mouth.

**Central:** Central vowels are produced with the Centre of the tongue raised against the roof of the mouth.

**Back:** Back vowels are produced with the back of the tongue raised against the roof of the mouth.

Vowels in languages can be described in terms of four **degrees of height**.

**Close:** the highest position to which the tongue is raised to produce a vowel like sound.

**Open:** the lowest position to which the tongue is brought to produce a vowel like sound.

**Close-mid:** the position closer to the Close position than the Open position to which the tongue is raised.

**Open-mid:** the position closer to the Open position than the Close position to which the tongue is brought.

#### 4. Labels for segmental sounds

Segmental sounds are identified using a three-term label, containing information about them being a) Voiced or Voiceless b) Place of Articulation and c) Manner of Articulation. In the case of stops, additional information is required about them being aspirated or unaspirated. A table of the phonemes of Bangla is produced below, containing the necessary details and examples. Similar tables need to be produced for other Indic languages.

Language								
Sl. no	IPA representation	Related Grapheme	Occurrence in Bangla Words			Place of Articulation	Manner of Articulation	Corresponding Sound Files
			Initially	Medially	Finally			
1.								
2.								

#### Example: Bangla Consonants and Vowels

Sl no.	Related Grapheme	IPA representation	Example Bangla Words			Place of Articulation	Manner of articulation
			Initial	Medial	Final		
1	क	/k/	/kobita/ 'Poetry'	/kaka/ 'Uncle'	/ɔbak/ 'Surprise'	Velar	Un-aspirated Un-voiced stop
2	ख	/k <sup>h</sup> /	/k <sup>h</sup> ɔtom/ 'End'	/ɔk <sup>h</sup> il/ 'Whole'	/akh/ 'Sugar cane'	Velar	Aspirated Un-voiced stop
3	ग	/g/	/gadha/ 'Donkey'	/ɔgad <sup>h</sup> / 'Plenty'	/tæg/ 'Abandonment'	Velar	Un-aspirated Voiced Stop
4	घ	/g <sup>h</sup> /	/ghɔr/ 'Home'	/ag <sup>h</sup> at/ 'Injury'	/bag <sup>h</sup> / 'Tiger'	Velar	Aspirated Voiced Stop

5	ঙ	/ŋ/		/kɔŋkal/ 'Skeleton'	/bæŋ/ 'Frog'	Velar	Nasal murmur
6	ঢ	/ʈʃ/	/ʈʃal/ 'Rice'	/ɔʈʃɔ/ 'Still'	/ʈʃamotʃ/ 'Spoon'	Place of release Alveolar and place of friction Post alveolar	Un-aspirated Un-voiced Affricate
7	ছ	/tʃʰ/	/tʃʰagol/ 'Goat'	/bitʃʰana/ 'Bed'	/matʃʰ/ 'Fish'	Place of release Alveolar and place of friction Post alveolar	Aspirated Unvoiced Affricate
8	জ	/dʒ/	/dʒɔ/ 'Water'	/adʒkal/ 'Now a days'	/kadʒ/ 'Work'	Place of release Alveolar and place of friction Post alveolar	Un-aspirated Voiced Affricate
9	ঝ	/dʒʰ/	/dʒʰikmik/ 'Twinkle'	/madʒʰe/ 'In the middle of'	/sādʒʰ/ 'Evening'	Place of release Alveolar and place of friction Post alveolar	Aspirated Voiced Affricate
10	ঞ	/ɲ/		/goɲai/ 'A title of vaisnava guru'		Alveolar	Nasal murmur
11	ট	/ʈ/	/ʈɔk/ 'Sour'	/ʈatʃka/ 'Fresh'	/motʃ/ 'Total'	Post-alveolar	Un-aspirated Unvoiced stop (retroflex)
12	ঠ	/ʈʰ/	/ʈʰakuma/ 'Grandmoth er'	/kātʰal/ 'Jack fruit'	/katʰ/ 'Wood'	Post-alveolar	Aspirated Unvoiced stop (retroflex)
13	ড	/ɖ/	/ɖal/ 'Pulses'	/adɖa/ 'Gossip'		Post-alveolar	Un-aspirated Voiced stop (retroflex)
14	ঢ	/ɖʰ/	/ɖʰal/ 'Shield'	/ɔɖʰel/ 'Plenty'		Post-alveolar	Aspirated Voiced stop (retroflex)
15	ণ	/ɳ/		/tʰaɳɖa/ 'Cold'		Post-alveolar	Nasal murmur
16	ত	/t/	/tumi/ 'You'	/batasʰ/ 'Wind'	/otit/ 'Past'	Dental	Un-aspirated Unvoiced stop
17	থ	/tʰ/	/tʰaka/ 'Stay'	/kɔtʰa/ 'Talk'	/pɔtʰ/ 'Road'	Dental	Aspirated Unvoiced stop
18	দ	/d/	/doi/ 'Curd'	/kodal/ 'Spade'	/pɔd/ 'Parts of speech'	Dental	Un-aspirated Voiced stop

19	ধ	/d <sup>h</sup> /	/d <sup>h</sup> an/ 'Paddy'	/ad <sup>h</sup> ar/ 'Container'	/bɔd <sup>h</sup> / 'Kill'	Dental	Aspirated Voiced Stop
20	ন	/n/	/na/ 'No'	/ɔnek/ 'Many'	/bon/ 'Woods'	Dental	Nasal murmur
21	প	/p/	/paka/ 'Ripe'	/kɔpal/ 'Forehead'	/pap/ 'Sin'	Bilabial	Un-aspirated Unvoiced stop
22	ফ	/p <sup>h</sup> /	/p <sup>h</sup> ɔl/ 'Fruit'	/ap <sup>h</sup> ʃɔf/ 'Repentance'	/sap <sup>h</sup> / 'Clear'	Bilabial	Aspirated Unvoiced Stop
23	ব	/b/	/boi/ 'Book'	/kobita/ 'Poetry'	/ɔb/ 'Dead body'	Bilabial	Un-aspirated Voiced stop
24	ভ	/b <sup>h</sup> /	/b <sup>h</sup> ɔy/ 'Fear'	/ab <sup>h</sup> as <sup>h</sup> / 'Hint'	/lab <sup>h</sup> / 'Profit'	Bilabial	Aspirated Voiced stop
25	ম	/m/	/ma/ 'Mother'	/amar/ 'My'	/am/ 'Mango'	Bilabial	Nasal murmur
26	র	/r/	/rɔktɔ/ 'Blood'	/kɔra/ 'Do'	/tomar/ 'Your'	Alveolar	Trill
27	ল	/l/	/lal/ 'Red'	/balɔk/ 'Boy'	/kal/ 'Time'	Dental	Lateral
28	শ, ষ	/ʃ/	/ʃɔʃa/ 'Cucumber'	/aʃi/ 'Eighty'	/pɔlaʃ/ 'A flower'	Post alveolar	Fricative
29	স	/s/	/sat/ 'Seven'	/bæstɔ/ 'Busy'	/tas/ 'Playing cards'	Alveolar	Fricative
30	হ	/h/	/het/ 'Hand'	/biher/ 'To travel'	/beh/ 'An exclamati on'	Glottal	Fricative
31	ড়	/ɽ/		/bɔrɔ/ 'Big'	/ʃɔɽ/ 'Conspira cy'	Post alveolar	Un-aspirated flap (Retroflex)
32	ঢ	/t <sup>h</sup> /		/driɽ <sup>h</sup> / 'Rigid'	/baɽ <sup>h</sup> / 'Flood'	Post alveolar	Aspirated flap (Retroflex)
33	য়	/j/		/paja/ 'Leg of a table'	/hɔj/ 'Is'	Palatal	Approximant
34		/v/			/hava/ 'Wind'	Bilabial	Approximant
<p style="text-align: center;"><b>Vowels:</b> <b>Oral:</b> <b>Monophthongs</b></p>							
1	অ	/ɔ/	/ɔsim/ 'Infinity',	/kɔt <sup>h</sup> a/ 'Talk'		Open, rounded	Back vowel

2	আ	/a/	/ami/ 'I'	/kal/ 'Time',	/kaka/ 'Uncle'	Open, Un- rounded	Front vowel
3	ই, ঐ	/i/	/ilif/ 'Hilsa'	/din/ 'Day'	/nodi/ 'River'	Close, Un- rounded	Front vowel
4	উ, ঊ	/u/	/upore/ 'Up'	/kukur/ 'Dog'	/ʃadhū/ 'Saint'	Close, rounded	Back vowel
5	এ	/e/	/ekʰane/ 'Here'	/ʃeba/ 'Nursing'	/kobe/ 'When'	Close-mid, Unrounded	Front vowel
6	ও	/o/	/otʰa/ 'To get up'	/lok/ 'Person'	/alo/ 'Light'	Close-mid, rounded	Back vowel
7		/æ/	/æk/ 'One'	/tæg/ 'Abandonment',		Open-mid, Unrounded	Front vowel



		<b>Vowels: Oral: Diphthongs</b>		
Sl no.	Diphthong	Examples Bangla Words		
		Initially	Medially	Finally
1.	/ui/	/uipoka/ ‘a kind of insect’	/tʃoruib <sup>h</sup> ati/ ‘feast’	/babui/ ‘A kind of bird’
2.	/ua/		/duar/ ‘door’	/dʒua/ ‘gamble’
3.	/ue/		/duek/ ‘two’	/ʃue/ ‘having laid’
4.	/uo/		/ʃuor/ ‘pig’	/kuo/ ‘well’
5.	/oi/	/oiṭa/ ‘that’	/ʃoilo/ ‘mountain’	/boi/ ‘book’
6.	/oa/		/dʒoar/ ‘tide’	/ʃoa/ ‘to lie down’
7.	/oe/		/koel/ ‘a river’	/k <sup>h</sup> oe/ ‘erosion’
8.	/ou/		/kouṭo/ ‘container’	/bou/ ‘wife’
9.	/ɔe/		/k <sup>h</sup> ɔer/ ‘catechu’	/b <sup>h</sup> ɔe/ ‘in fear’
10.	/ɔo/		/kɔoa/ ‘to say’	/kɔo/ ‘say’
11.	/ai/		/baire/ ‘outside’	/k <sup>h</sup> ai/ ‘eat’
12.	/ao/		/t <sup>h</sup> aor/ ‘understand’	/gao/ ‘sing’
13.	/au/	/auṣ/ ‘a kind of paddy’	/pauruṭi/ ‘bread’	/lau/ ‘gourd’
14.	/æe/		//	//
15.	/æo/		//	//
16.	/æi/		/mitobæita / ‘frugality’	/mitobæi / ‘frugal’
17.	/iɔ/		/dʒiɔ/ ‘alive’	
18.	/ia/		/ʃial/ ‘jackal’	/daraia/ ‘having stood up’
19.	/ie/		/b <sup>h</sup> ien/ ‘one who makes sweets’	/bie/ ‘wedding’
20.	/io/		/mrioman/ ‘sorrowful’	/prio/ ‘favourite’
21.	/iu/		/ʃiuli/ ‘a kind of flower’	/piu/ ‘a name’
22.	/ei/	/eiṭa/ ‘this’	/beiman/ ‘ungateful’	/ʃei/ ‘that’
23.	/ea/		/k <sup>h</sup> ea/ ‘attention’	/k <sup>h</sup> ea/ ‘a kind of boat’
24.	/eo/		/sreofi/ ‘a name’	/ʃreo/ ‘better’
25.	/eu/		/keuṭe/ ‘a kind of snake’	/keu/ ‘any one’