Proposal for an Oriya Script Root Zone Label Generation Ruleset (LGR)

*LGR Version:* 3.0

*Date:* 2018-08-08

*Document version:* 2.7

*Authors:* Neo-Brahmi Generation Panel [NBGP]

# General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed Root Zone Level Generation Rules for the Oriya script. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document:

Proposal-LGR-Orya-20180808.xml

Labels for testing can be found in the accompanying text document:

Oriya-Test-Labels-20180808.txt

# Script for which the LGR is proposed

ISO 15924 Code:  Orya

ISO 15924 Key N°: 327

ISO 15924 English Name: Oriya (Odia)

Latin transliteration of native script name: oḍiā

Native name of the script: ଓଡ଼ିଆ

Maximal Starting Repertoire (MSR) version: MSR-3

# Background on Script and Principal Languages using it

Odia (known in Unicode as Oriya) is an Eastern Indic language spoken by about 40 million people mainly in the Indian state of Odisha (Orissa), and also in parts of West Bengal, Jharkhand, Chhattisgarh and Andhra Pradesh. Oriya (Odia) is one of the many official languages of India. It is the official language of Odisha, and the second official language of Jharkhand. Eminent Linguists like John Beames, G. A. Grierson, L.S.S. O’Malley, Suniti Kumar Chatterjee, S. N. Rajaguru, John Boulton and others consider Odia as one of the most ancient languages of India. In Indic family of languages, Oriya (Odia) is closest to Sanskrit and least influenced by foreign languages. Only these two Indic languages (viz. Sanskrit and Odia) have got classical tag due to their rich, uninfluenced and long literary history. According to National Mission for Manuscripts, after Sanskrit (11,66,743), Odia (2,13,088) has the largest number of documented manuscripts in the India.

Odia was previously spelt as Oriya, and Odisha as Orissa. However, Odia and Odisha are now the preferred names officially in English as they are closer to their native names: ଓଡ଼ିଆ (oḍiā) [ɔɖiaː] and ଓଡ଼ିଶା (oḍiśā) [ɔɖisaː].

With reference to [Wikipedia](https://en.wikipedia.org/wiki/Odia_language) (https://en.wikipedia.org/wiki/Odia\_language) Oriya (Odia) language is also used by minority populations of the neighboring states of Jharkhand, West Bengal, Chhattisgarh and Andhra Pradesh. The region has been known at different stages of history as Kalinga, Udra, Utkala or Koshala. Odisha was a vast empire in ancient and medieval times, extending from the Ganges in the north to the Godavari in the south. During British rule, however, Odisha lost its political identity and formed parts of the Bengal and Madras Presidencies. The present state of Odisha was formed in 1936.

The modern Oriya (Odia) language is formed mostly from Pali words with significant Sanskrit influence. About 28% of modern Oriya (Odia) words have Adivasi origins, and about 2% have Hindustani (Hindi/Urdu), Persian, or Arabic origins. The earliest written texts in the language are about thousand years old. The first Oriya (Odia) newspaper was Utkala Deepika first published on 4 August 1866.

Among the Indo-European languages of India, only Oriya (Odia) and Sanskrit have been recognized as classical languages; and of the six Indian languages that have been conferred classical language status Oriya (Odia) was recognized most recently (in 2014).[[1]](#footnote-1) It forms the basis of Odissi dance and Odissi music. [[2]](#footnote-2)

Oriya script seems to be a variant of Devanāgarī, the main difference being the absence of the shirorekha or the line above the character and also its more rounded shapes. Since initially it was used for commercial ends, it has been referred to as śarāphi (banker's) or mahājani (trader's) script.

The Oriya (Odia) script is used to write Oriya (Odia) language and a number of other languages spoken in Odisha such as Munda, Santali, Kui, Ho and Sanskrit.

## The Evolution of the Script

The Oriya (Odia) script developed from the Kalinga script, one of the many descendants of the Brahmi script of ancient India. (Rajaguru, S.N., Odia Lipira Kramabikash, Odia Sahitya Akademi, page 2). The earliest known inscription in the Oriya (Odia) language, in the Kalinga script, dates from 1051. It descends from Odra-Magadhi Prakrit similar to Ardha Magadhi, prevalent in eastern India over 1,500 years ago.

The curved appearance of the Oriya script is a result of the practice of writing on palm leaves, which have a tendency to tear if written in straight lines.

The diagram below shows the major stages in the evolution of Oriya attesting its late divergence from Devanāgarī.



Figure 1: Pictorial depiction of Evolution of Oriya



Literature in the Oriya (Odia) language (Odia: ) is the predominant literature of the state of Odisha in India.

## Periods of Odia History



Oriya (Odia) language literature (Odia: ) is the predominant literature of the state of Odisha in India.

Historians have divided the history of the Oriya (Odia) literature into five main stages: Old Oriya (Odia) (8th century to 1300), Early Middle Oriya (Odia) (1300 to 1500), Middle Oriya (Odia) (1500 to 1700), Late Middle Oriya (Odia) (1700 to 1850) and Modern Oriya (Odia) (1850 to present).

## Use of Oriya language beyond India

According to [Wikipedia](https://en.wikipedia.org/wiki/Odia_language), <https://en.wikipedia.org/wiki/Odia_language> the Oriya (Odia) diaspora constitutes a sizeable number of speakers in several countries around the world, pushing the number of Oriya (Odia) speakers globally to 55 million.

It has a significant presence in eastern countries such as Bangladesh, Indonesia, mainly carried by the sadhaba, ancient traders from Odisha, who carried the language along with the culture during the old-day trading, and in western countries such as the United States, Canada, Australia and England as well. The language has also spread to Burma, Malaysia, Fiji, Sri Lanka and countries of the Middle East. Written Oriya (Odia) (or the standard Oriya (Odia) is used for official purpose. It has elements from different local Oriya (Odia) dialects but it usually avoids words of foreign origin such as Arabic and Persian. It has also assimilated many tribal words prevalent in Odisha.

## Notable features

The Oriya script is a syllabic alphabet written left to right in horizontal lines, in which all consonants have an inherent vowel. Diacritics, which can appear above, below, before or after the consonant they belong to, are used to change the inherent vowel.

When they appear at the beginning of a syllable, vowels are written as independent letters.

When certain consonants occur together, special conjunct shapes are used which combine the essential parts of each letter.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| The chart below shows the way in which the International Phonetic Alphabet (IPA) represents Oriya (Odia) pronunciations. **Vargya Consonants** | | **Avargya consonants** | | |
| [IPA](https://en.wikipedia.org/wiki/Help:IPA) | Oriya (Odia) | [IPA](https://en.wikipedia.org/wiki/Help:IPA) | Oriya (Odia) | |
| [b](https://en.wikipedia.org/wiki/Voiced_bilabial_stop) | ବ |  | ଯ | |
| [bʱ](https://en.wikipedia.org/wiki/Breathy_voice) | ଭ | [j](https://en.wikipedia.org/wiki/Palatal_approximant#Palatal) | ୟ | |
| d̪ | ଦ | [ɾ](https://en.wikipedia.org/wiki/Alveolar_flap) | ର | |
| d̪ʱ | ଧ | [l](https://en.wikipedia.org/wiki/Alveolar_lateral_approximant) | ଲ | |
| [ɖ](https://en.wikipedia.org/wiki/Voiced_retroflex_stop) | ଡ | [l̪](https://en.wikipedia.org/wiki/Alveolar_lateral_approximant) | ଳ | |
| [ɖʱ](https://en.wikipedia.org/wiki/Breathy_voice) | ଢ | [ʋ](https://en.wikipedia.org/wiki/Labiodental_approximant) | ଵ | |
| [d͡ʒ](https://en.wikipedia.org/wiki/Voiced_palato-alveolar_affricate) | ଜ | [w](https://en.wikipedia.org/wiki/Voiced_labio-velar_approximant) | ୱ | |
| [d͡ʒʱ](https://en.wikipedia.org/wiki/Breathy_voice) | ଝ | [s](https://en.wikipedia.org/wiki/Voiceless_alveolar_fricative#Voiceless_alveolar_sibilant) | ସ | |
| [ɡ](https://en.wikipedia.org/wiki/Voiced_velar_stop) | ଗ | [ʂ](https://en.wikipedia.org/wiki/Voiceless_retroflex_fricative) | ଷ | |
| [ɡʱ](https://en.wikipedia.org/wiki/Breathy_voice) | ଘ | [ɕ](https://en.wikipedia.org/wiki/Voiceless_alveolo-palatal_fricative) | ଶ | |
| [h](https://en.wikipedia.org/wiki/Voiceless_glottal_fricative) | ହ | [ɦ](https://en.wikipedia.org/wiki/Voiced_glottal_fricative) | ହ | |
| [k](https://en.wikipedia.org/wiki/Voiceless_velar_stop) | କ | [ŋ](https://en.wikipedia.org/wiki/Velar_nasal),[ɲ](https://en.wikipedia.org/wiki/Palatal_nasal#Palatal_or_alveolo-palatal),[ɳ](https://en.wikipedia.org/wiki/Retroflex_nasal),[n](https://en.wikipedia.org/wiki/Dental,_alveolar_and_postalveolar_nasals#Alveolar),[m](https://en.wikipedia.org/wiki/Bilabial_nasal),[◌̃](https://en.wikipedia.org/wiki/Nasalization) | ଂ | |
| [kʰ](https://en.wikipedia.org/wiki/Aspirated_consonant) | ଖ | [◌̃](https://en.wikipedia.org/wiki/Nasalization) | ଁ | |
| [ɲ](https://en.wikipedia.org/wiki/Palatal_nasal#Palatal_or_alveolo-palatal) | ଞ | [ɦ](https://en.wikipedia.org/wiki/Voiced_glottal_fricative) | ଃ | |
| [m](https://en.wikipedia.org/wiki/Bilabial_nasal) | ମ | **Vowels and Matras** | | |
| [n](https://en.wikipedia.org/wiki/Dental,_alveolar_and_postalveolar_nasals#Alveolar) | ନ | **IPA** | **Vowels** | **Matras** |
| [ɳ](https://en.wikipedia.org/wiki/Retroflex_nasal) | ଣ | [ə](https://en.wikipedia.org/wiki/Mid_central_vowel) | ଅ |  |
| [ŋ](https://en.wikipedia.org/wiki/Velar_nasal) | ଙ | [aː](https://en.wikipedia.org/wiki/Open_front_unrounded_vowel) | ଆ | ା |
| [p](https://en.wikipedia.org/wiki/Voiceless_bilabial_stop) | ପ | [ɪ](https://en.wikipedia.org/wiki/Near-close_front_unrounded_vowel) | ଇ | ି |
| [pʰ](https://en.wikipedia.org/wiki/Aspirated_consonant) | ଫ | [iː](https://en.wikipedia.org/wiki/Close_front_unrounded_vowel) | ଈ | ୀ |
| [ɾ](https://en.wikipedia.org/wiki/Alveolar_flap) | ର | [ʊ](https://en.wikipedia.org/wiki/Near-close_back_rounded_vowel) | ଉ | ୁ |
| [ɽ](https://en.wikipedia.org/wiki/Retroflex_flap) | ଡ଼ | [uː](https://en.wikipedia.org/wiki/Close_back_rounded_vowel) | ଊ | **ୂ** |
| [ɽʱ](https://en.wikipedia.org/wiki/Breathy_voice) | ଢ଼ | r̩ | ଋ | ୃ |
| [s](https://en.wikipedia.org/wiki/Voiceless_alveolar_fricative#Voiceless_alveolar_sibilant) | ସ | [eː](https://en.wikipedia.org/wiki/Close-mid_front_unrounded_vowel) | ଏ | େ |
| t̪ | ତ | [ɛː](https://en.wikipedia.org/wiki/Open-mid_front_unrounded_vowel) | ଐ | ୈ |
| t̪ʰ | ଥ | [oː](https://en.wikipedia.org/wiki/Close-mid_back_rounded_vowel) | ଓ | ୋ |
| [ʈ](https://en.wikipedia.org/wiki/Voiceless_retroflex_stop) | ଟ | [ɔː](https://en.wikipedia.org/wiki/Open-mid_back_rounded_vowel) | ଔ | ୌ |
| [ʈʰ](https://en.wikipedia.org/wiki/Aspirated_consonant) | ଠ |
| [t͡ʃ](https://en.wikipedia.org/wiki/Voiceless_palato-alveolar_affricate) | ଚ |
| [t͡ʃʰ](https://en.wikipedia.org/wiki/Aspirated_consonant) | ଛ |

Table 1: International Phonetic Alphabet Oriya Pronunciations

## Structured consonants

The structured consonants are classified according to the place of articulation and are classified accordingly into five structured groups. These consonants are shown here with their IAST (International Alphabet of Sanskrit Transliteration)[[3]](#footnote-3) transcriptions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [**voiceless**](https://wikivisually.com/wiki/Voiceless) | [**voiceless  aspirate**](https://wikivisually.com/wiki/Aspiration_(phonetics)) | [**voiced**](https://wikivisually.com/wiki/Voice_(phonetics)) | [**voiced  aspirate**](https://wikivisually.com/wiki/Aspiration_(phonetics)) | [**nasal**](https://wikivisually.com/wiki/Nasal_stop) |
| [**Velars**](https://wikivisually.com/wiki/Velars) | କ (ka) | ଖ (kha) | ଗ (ga) | ଘ (gha) | ଙ (ṅa) |
| [**Palatals**](https://wikivisually.com/wiki/Palatals) | ଚ (ca) | ଛ (cha) | ଜ (ja) | ଝ (jha) | ଞ (ña) |
| [**Retroflex**](https://wikivisually.com/wiki/Retroflex) | ଟ (ṭa) | ଠ (ṭha) | ଡ (ḍa) | ଢ (ḍha) | ଣ (ṇa) |
| [**Dentals**](https://wikivisually.com/wiki/Dental_consonant) | ତ (ta) | ଥ (tha) | ଦ (da) | ଧ (dha) | ନ (na) |
| [**Labials**](https://wikivisually.com/wiki/Labial_consonant) | ପ (pa) | ଫ (pha) | ବ (ba) | ଭ (bha) | ମ (ma) |

Table 2:structured consonants

## Unstructured consonants

The unstructured consonants are consonants that do not fall into any of the above categories: ଯ (ja), ୟ (ia), ର (ra), ଲ (la), ଳ (ḷa), ଵ(va),ୱ (wa), ଶ (sa), ଷ (sa), ସ (sa), ହ (ha)

## The implicit vowel killer Halant (Virama)

Halant character is used after a consonant to "strip" it of its inherent vowel. A consonant syllable cannot end with halant. With a few exceptions, most of the Oriya words are svaranta (i.e ending with a vowel).

A syllable containing halant characters may be shaped with no visible halant signs, as the halants enable different consonants to form conjuncts.

Halant form of consonants - The form produced by adding the halant to the nominal shape. The halant form is used in syllables that have no vowel or as the half form when no distinct shape for the half form exists.

Half form of consonants (pre-base form) - A variant form of consonants which appear to the left of the base consonant, if they do not participate in a ligature. Consonants in their half form precede the ones forming the base glyph. Some Indic scripts, like Devanagari have distinctly shaped half forms for most of the consonants. If no distinct shape exists, the full form will display with an explicit Virama (same shape as the halant form).

## Nukta (଼ – U+0B3C):

The nukta sign is used in Oriya language just like any other Indian scripts. It is used with a few consonants to represent sounds found only in words borrowed from Perso-Arabic. It can be commonly used with “ଡ” U+0B21, “ଢ” U+0B22, “କ” U+0B15, “ଖ” U+0B16, “ଗ” U+0B17, “ଚ” U+0B1A, “ଜ” U+0B1C, and “ଫ” U+0B2B to show that words having these consonants with a nukta are to be pronounced in the Perso-Arabic style.

## Visarga “ଃ” (U+0B03) and Avagraha “ଽ” (U+0B3D):

The Visarga (“ଃ” (U+0B03) is frequently used in Sanskrit and represents a sound very close to /h/. Example, ଦୁଃଖ /du:kh/ sorrow (U+0B26 U+0B41 U+0B03 U+0B16).

The Avagraha"ଽ"(U+0B3D) creates an extra stress on the preceding vowel and is used in Sanskrit texts. It is rarely used in other languages using Oriya. In case of LGR, the Avagraha is not part of the repertoire as it is barred in the Maximal Starting Repertoire.

## Candrabindu (ँ - U+0B01):

Candrabindu denotes nasalization of the preceding vowel and consonants as in ଅଁଳା /ãala/name of seasonal fruit (U+0B05 U+0B01 U+0B33 U+0B3E). Oriya users commonly use it for writing the words and sounds of Sanskrit language.

## Anusvara (ଂ - U+0B02):

Anusvara replaces a conjunct group of a Nasal Consonant + Halant + Consonant belonging to that particular varga. The Anusvara represents a homorganic nasal. Before a non-varga consonant the Anusvara represents a nasal sound. For example: ଏବଂ (0B0F+0B2C+0B02), ସଂଖ୍ୟା (0B38+0B02+0B16+0B4D+0B5F+0B3E), etc.

## Matra sign (Dependent Vowel)

It is used to represent a vowel sound that is not inherent to the consonant. Dependent vowels are referred to as "matras". They are always depicted in combination with a single consonant, or with a consonant cluster. The greatest variation among different Indian scripts is found in the rules for attaching dependent vowels to base characters. The rules specific to Oriya are mentioned in Section 6 (Variants) and Section 7 (WLE Rules).

Following table explains the correlation between a vowel and its matra sign.

| **Vowel and its matra sign** | | | | | |
| --- | --- | --- | --- | --- | --- |
| Glyph | Unicode | [Name](https://www.compart.com/en/unicode/block/U+0B00?sort=name) | Glyph | Unicode | [Name](https://www.compart.com/en/unicode/block/U+0B00?sort=name) |
| ଅ | U+0B05 | Oriya Letter A |  |  |  |
| ଆ | U+0B06 | Oriya Vowel Letter Aa | ା | U+0B3E | Oriya Vowel Sign Aa |
| ଇ | U+0B07 | Oriya Vowel Letter I | ି◌ | U+0B3F | Oriya Vowel Sign I |
| ଈ | U+0B08 | Oriya Vowel Letter Ii | ୀ | U+0B40 | Oriya Vowel Sign Ii |
| ଉ | U+0B09 | Oriya Vowel Letter U | ୁ◌ | U+0B41 | Oriya Vowel Sign U |
| ଊ | U+0B0A | Oriya Vowel Letter Uu | ୂ◌ | U+0B42 | Oriya Vowel Sign Uu |
| ଋ | U+0B0B | Oriya Vowel Letter Vocalic R | ୃ◌ | U+0B43 | Oriya Vowel Sign Vocalic R |
| ଏ | U+0B0F | Oriya Vowel Letter E | େ | U+0B47 | Oriya Vowel Sign E |
| ଐ | U+0B10 | Oriya Vowel Letter Ai | ୈ | U+0B48 | Oriya Vowel Sign Ai |
| ଓ | U+0B13 | Oriya Vowel Letter O | ୋ | U+0B4B | Oriya Vowel Sign O |
| ଔ | U+0B14 | Oriya Vowel Letter Au | ୌ | U+0B4C | Oriya Vowel Sign Au |
| ଌ | U+0B0C | ORIYA LETTER VOCALIC L | ୢ◌ | U+0B62 | Oriya Vowel Sign Vocalic L |
| ୡ | U+0B61 | ORIYA LETTER VOCALIC LL | ୣ◌ | U+0B63 | Oriya Vowel Sign Vocalic Ll |

Table 3: Vowel and its Matra Sign

“ଌ” U+0B0C, “ୡ” U+0B61, “ୢ” U+0B62 and ”ୣ” U+0B63 are hardly in use in modern days.

# Overall Development Process and Methodology

Under the Neo-Brahmi Generation Panel, there are many different scripts belonging to separate Unicode blocks. Each of these scripts is the basis for a separate LGR proposal; however Neo-Brahmi GP ensures that the fundamental philosophy behind building these various script LGRs is same across the different scripts being considered. This is the Oriya (Odia) LGR, which caters to Oriya (Odia) languages written using Oriya (Odia) belonging to EGIDS scale 1 to 4.

## Guiding Principles

The NBGP adopts the following broad principles for the selection of code-points in the repertoire across the board for all the scripts within its scope.

### Inclusion principles:

#### Modern usage:

Every character proposed should be in the everyday usage of a particular linguistic community. The characters which have been encoded in the Unicode for transcription purposes only or for archival purposes will not be considered for inclusion in the code-point repertoire.

# 4.1.1.2. Unambiguous use:

Every character proposed should have unambiguous understanding among the linguistic about its usage in the language.

### Exclusion principles:

The main exclusion principle is that of External Limits on Scope. These comprise of protocols or standards which are pre-requisites to the Label Generation Rules etc. All further principles are in fact subsumed under these limitations but have been spelt out separately for the sake of clarity.

#### External Limits on Scope:

The code point repertoire for root zone being a very special case, up the ladder in the protocol hierarchies, the canvas of available characters for selection as a part of the Root Zone code point repertoire is already constrained by various protocol layers beneath it. The following three main protocols/standards act as successive filters:

i. The Unicode Chart:

Out of all the characters that are needed by the given script, if the character in question is not encoded in Unicode, it cannot be incorporated in the code point repertoire. Such cases are quite rare, given the elaborate and exhaustive character inclusion efforts made by Unicode consortium.

ii. IDNA Protocol:

Unicode being the character encoding standard for providing the maximum possible representation of a given script/language, it has encoded as far as possible all the possible characters needed by the script. However, the Domain name being a specialized case, it is governed by an additional protocol known as IDNA (Internationalized Domain Names in Applications). The IDNA protocol excludes some characters out of Unicode repertoire from being part of the domain names.

Example: Oriya script frequently uses “ଡ” (U+0B21), “ଢ” (U+0B22) as well as their respective allophones “ଡ଼”, and “ଢ଼”. In Oriya (Odia) script, these differ in use of nukta. Thus “ଡ଼” and “ଢ଼” as distinct letters are not allowed but their decomposed form i.e. “ଡ”, “ଢ” followed by Oriya (Odia) sign nukta (U+0B3C) can be used. Similarly, for allophones of other consonants like କ (U+0B15), ଖ (U+0B16), ଗ (U+0B17), ଚ (U+0B1A), ଜ (U+0B1C), ଫ (U+0B2B) nukta can be used.

iii. Maximal Starting Repertoire:

As the Root Zone LGR is used for creation of the root zone TLDs, which in turn are an even more specialized case of domain name labels, the Root Zone LGR procedure introduces additional exclusions for characters allowed by IDNA.

Example: Oriya Sign Avagraha "ଽ" (U+0B3D) even if allowed by IDNA protocol, is not permitted in the Root Zone Repertoire as per the [MSR].

Maximal Starting Repertoire also excludes invisible characters Zero Width Non-Joiner (U+200C) and Zero Width Joiner (U+200D). These are required in certain cases where a typical visual shape of an akshar is desired.

To sum up, the restrictions start off with admitting only such characters as are part of the code-block of the given script/language. This is further narrowed down by the IDNA Protocol and finally an additional filter in the form of Maximal Starting Repertoire restricts the character set associated with the given language even more.

#### No Fraction Marks:

The TLDs being identifiers, fraction markers present in Brahmi based languages such as given below will not be included.

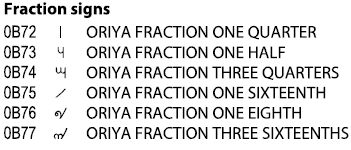


Figure 2: Fraction Marks in Oriya

#### No Symbols and Abbreviations:

Abbreviations, weights and measures and other such iconic characters like Isshar"୰" (U+0B70) will not be included.

#### No Rare and Obsolete Characters:

There are characters which have been added to Unicode to accommodate rare forms especially like Oriya LETTER VOCALIC RR"ୠ" (U+0B60) and Oriya LETTER VOCALIC LL"ୡ" (U+0B61) as well as their Matra forms ) “ୄ“ (U+0B44) and "ୣ" (U+0B63). All such characters will be excluded. This is in compliance with the Conservatism principle as laid down in the Root Zone LGR procedure.

# Repertoire

This section provides the relevant section of [MSR] applicable to the Oriya script on which Oriya code point repertoire for the Root Zone LGR is based on. Section 5.1 details the code-point repertoire that the Neo-Brahmi Generation Panel [NBGP] proposes to be included in the Oriya Root Zone LGR.

## Oriya Section of Maximal Starting Repertoire [MSR] Version 3

|  |  |
| --- | --- |
| Figure 3: Oriya Code Page from MSR-3 | **Color convention:**  All characters that are included in the [MSR]- Yellow background  **PVALID** in IDNA2008 but excluded from the [MSR]- Pinkish background  **Not PVALID** in IDNA2008 - White background |

## Code Point Repertoire

The table below lists all the code points included in the repertoire for Oriya script. For each of the code points, language references have been given in the last column.

| **Sr. No.** | **Unicode Code Point** | **Glyph** | **Character Name** | **Language with EGIDS** | **Indic Syllabic Category** | **References** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 0B01 | ଁ | ORIYA SIGN CANDRABINDU | 2-Oriya | Candrabindu | [0], [101], [102], [103], [104], [105] |
| 2 | 0B02 | ଂ | ORIYA SIGN ANUSVARA | 2-Oriya | Anusvara | [0], [101], [102], [103], [104], [105] |
| 3 | 0B03 | ଃ | ORIYA SIGN VISARGA | 2-Oriya | Visarga | [0], [101], [102], [103], [104], [105] |
| 4 | 0B05 | ଅ | ORIYA LETTER A | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 5 | 0B06 | ଆ | ORIYA LETTER AA | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 6 | 0B07 | ଇ | ORIYA LETTER I | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 7 | 0B08 | ଈ | ORIYA LETTER II | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 8 | 0B09 | ଉ | ORIYA LETTER U | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 9 | 0B0A | ଊ | ORIYA LETTER UU | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 10 | 0B0B | ଋ | ORIYA LETTER VOCALIC R | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 11 | 0B0F | ଏ | ORIYA LETTER E | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 12 | 0B10 | ଐ | ORIYA LETTER AI | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 13 | 0B13 | ଓ | ORIYA LETTER O | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 14 | 0B14 | ଔ | ORIYA LETTER AU | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 15 | 0B15 | କ | ORIYA LETTER KA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 16 | 0B16 | ଖ | ORIYA LETTER KHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 17 | 0B17 | ଗ | ORIYA LETTER GA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 18 | 0B18 | ଘ | ORIYA LETTER GHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 19 | 0B19 | ଙ | ORIYA LETTER NGA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 20 | 0B1A | ଚ | ORIYA LETTER CA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 21 | 0B1B | ଛ | ORIYA LETTER CHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 22 | 0B1C | ଜ | ORIYA LETTER JA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 23 | 0B1D | ଝ | ORIYA LETTER JHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 24 | 0B1E | ଞ | ORIYA LETTER NYA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 25 | 0B1F | ଟ | ORIYA LETTER TTA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 26 | 0B20 | ଠ | ORIYA LETTER TTHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 27 | 0B21 | ଡ | ORIYA LETTER DDA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 28 | 0B22 | ଢ | ORIYA LETTER DDHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 29 | 0B23 | ଣ | ORIYA LETTER NNA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 30 | 0B24 | ତ | ORIYA LETTER TA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 31 | 0B25 | ଥ | ORIYA LETTER THA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 32 | 0B26 | ଦ | ORIYA LETTER DA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 33 | 0B27 | ଧ | ORIYA LETTER DHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 34 | 0B28 | ନ | ORIYA LETTER NA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 35 | 0B2A | ପ | ORIYA LETTER PA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 36 | 0B2B | ଫ | ORIYA LETTER PHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 37 | 0B2C | ବ | ORIYA LETTER BA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 38 | 0B2D | ଭ | ORIYA LETTER BHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 39 | 0B2E | ମ | ORIYA LETTER MA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 40 | 0B2F | ଯ | ORIYA LETTER YA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 41 | 0B30 | ର | ORIYA LETTER RA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 42 | 0B32 | ଲ | ORIYA LETTER LA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 43 | 0B33 | ଳ | ORIYA LETTER LLA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 44 | 0B35 | ଵଵ | ORIYA LETTER VA | 2-Oriya | Consonant | [6], [101], [102], [103], [104], [105] |
| 45 | 0B36 | ଶ | ORIYA LETTER SHA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 46 | 0B37 | ଷ | ORIYA LETTER SSA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 47 | 0B38 | ସ | ORIYA LETTER SA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 48 | 0B39 | ହ | ORIYA LETTER HA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 49 | 0B3C | ଼ | ORIYA SIGN NUKTA | 2-Oriya | Nukta | [0], [101], [102], [103], [104], [105] |
| 50 | 0B3E | ା | ORIYA VOWEL SIGN AA | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 51 | 0B3F | ି | ORIYA VOWEL SIGN I | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 52 | 0B40 | ୀ | ORIYA VOWEL SIGN II | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 53 | 0B41 | ୁ | ORIYA VOWEL SIGN U | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 54 | 0B42 | ୂ | ORIYA VOWEL SIGN UU | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 55 | 0B43 | ୃ | ORIYA VOWEL SIGN VOCALIC R | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 56 | 0B47 | େ | ORIYA VOWEL SIGN E | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 57 | 0B48 | ୈ | ORIYA VOWEL SIGN AI | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 58 | 0B4B | ୋ | ORIYA VOWEL SIGN O | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 59 | 0B4C | ୌ | ORIYA VOWEL SIGN AU | 2-Oriya | Matra | [0], [101], [102], [103], [104], [105] |
| 60 | 0B4D | ୍ | ORIYA SIGN VIRAMA | 2-Oriya | Halant | [0], [101], [102], [103], [104], [105] |
| 61 | 0B56 | ୖ | ORIYA AI LENGTH MARK | 2-Oriya | Matra | [2], [101], [102], [103], [104], [105] |
| 62 | 0B5F | ୟ | ORIYA LETTER YYA | 2-Oriya | Consonant | [0], [101], [102], [103], [104], [105] |
| 63 | 0B71 | ୱୱ | ORIYA LETTER WA | 2-Oriya | Consonant | [6], [101], [102], [103], [104], [105] |

Table 4: Code Point Repertoire

### Code Points Excluded

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Unicode Code Point** | **Glyph** | **Character Name** | **Language with EGIDS** | **Indic Syllabic Category** | **Reference** |
| 1 | 0B0C | ଌ | ORIYA LETTER VOCALIC L | 2-Oriya | Vowel | [0], [101], [102], [103], [104], [105] |
| 2 | 0B44 | ୄ | ORIYA VOWEL SIGN VOCALIC RR | 2-Oriya | Matra | [9], [101], [102] [103] [104] [105] |
| 3 | 0B57 | ୗ | ORIYA AU LENTH MARK | 2-Oriya | Matra | [0], [101], [102] [103] [104] [105] |

Table 5: Code Point Excluded from Repertoire

Since the matra ୗ (U+0B57) ORIYA AU LENTH MARK is not in current use by the Oriya Community, it is decided by the NBGP to exclude it. Also, “ଌ” U+0B0C, “ୡ” U+0B61, “ୢ” U+0B62 and ”ୣ” U+0B63 are hardly in use in modern days.

### Variables involved

C → Consonant

M → Matra

V → Vowel

B → Anusvara

H → Halant / Virama

N → Nukta

C1 → { କ 0B15, ଖ 0B16, ଗ 0B17, ଚ 0B1A, ଜ 0B1C, ଡ0B21, ଢ0B22, ଫ 0B2B}

X → Visarga

D → Candrabindu

# Variants

## In-Script Variants

In Oriya script, there are no characters/character sequences which can be created by using the Oriya characters permitted as per the [MSR] and look identical. There are no in-script variants.

## Cross-Script Variants

A cross-script variant label, also sometimes referred to as "Whole Label confusable", is the variant case where one label in one script can be composed in such a way that it can resemble another entire label in a different script.

Every individual LGR under NBGP provides a set of cross-script variant code points that it identifies with members of other related scripts.

NBGP has ensured that not only the individual characters but also most of the akshar variations are taken into consideration during the cross-script variant analysis of Oriya with all the other scripts under NBGP. It was achieved by sharing a list of most of the akshar combinations with all the other script teams (‘most’ is used here as all the possible Consonant + Halant + Consonant +… cases cannot be practically covered. Case of all the Oriya “Consonant + Halant + Consonant” was included in the analysis).

Oriya script has a set of possible cross-script variants with the Malayalam script. Cases listed in Table 6 are cross-script variants between Oriya and Malayalam. This follows the NBGP Cross-script Variant inclusion policy available in Appendix D.

It is to be noted that none of the combinations listed in Table 6a are termed to be equivalents of each other semantically or otherwise. They are only grouped because they are considered visually same by the two script communities.

NBGP has ensured that Oriya and Malayalam LGR teams propose a same set of cross-script variants by meeting face-to-face on many occasions as well as through mail communications. The same set of cross-script variants (with Malayalam) is supposed to be found in the Malayalam LGR documents.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variant Set | Oriya | | Malayalam | |
| CP | Glyph | CP | Glyph |
| 1. | 0B20 | ଠ | 0D20 | ഠ |

Table 6a: Variant set between Oriya script and Malayalam script

Oriya script also has a set of possible cross-script variants with the Myanmar script. Cases listed in Table 6b are cross-script variants between Oriya and Myanmar. This follows the NBGP Cross-script Variant inclusion policy available in Appendix D.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variant Set | Oriya | | Myanmar | |
| CP | Glyph | CP | Glyph |
| 1. | 0B20 | ଠ | 101D | ဝ |

Table 6b: Variant set between Oriya script and Malayalam script

The cases listed in Appendix B are the visually confusable code points for reference, but they are not defined as variant code points.

# Whole Label Evaluation Rules (WLE)

This section provides the whole label evaluation rules for text written in Oriya script. The rules have been drafted in such a way that they can be easily translated into the LGR specification.

Below are the symbols used in the WLE rules for each of the "Indic Syllabic Category" as mentioned in Table 4: Code point repertoire

In addition, a few additional symbols define the appropriate subsets for various rules.

C → Consonant

M → Matra

V → Vowel

B → Anusvara

H → Halant / Virama

N → Nukta

C1 → {କ 0B15 KA, ଖ 0B16 KHA, ଗ 0B17 GA, ଚ 0B1A CA, ଜ 0B1C JA, ଡ0B21 DDA, ଢ 0B22 DDHA,

ଫ 0B2B PHA}

X → Visarga

D → Candrabindu

**Rule1**: N( ଼) must be preceded only by C1

For example:

ଡ(0B21)+଼(0B3C) = ଡ଼

ଢ(0B22)+଼ (0B3C) = ଢ଼

**Rule2**: B ( ଂ) must be preceded by V, C, N or M

i) B may be preceded by V (examples: ଅଂଶ,)

ii) B may be preceded by C, (example: ସଂସାର, ବଂଶ)

iii) B may be preceded by N (example: ଡ଼ଂଗା)

iv) B may be preceded by M, (examples: ସିଂହ, ମାଂସ, ବିଂଶ, ସୁତରାଂ)

**Rule3**: X (ଃ) must be preceded by C, V, N or M

i) X may be preceded by C, (example: ପ୍ରାୟତଃ, କ୍ରମଶଃ)

ii) X may be preceded by N (example: ଡ଼ଃ)

iii) X may be preceded by M, (examples: ଦୁଃଖ, ଦୁଃଖିତ)

iv) X may be preceded by V, (examples: ଅଃ, ଆଃ, ଇଃ, ଉଃ) commonly used when writing Sanskrit or when there is religious requirement

**Rule4**: D (ଁ) must be preceded by V, C, N or M

i) D may be preceded by V (examples: ପାଇଁ, ଯେଉଁ ,ନିଆଁ)

ii) D may be preceded by C (example: ମୁହଁ, ପହଁରା, ନୁହଁ)

iii) D may be preceded by N (example: ଡ଼ାଁଶ)

iv) D may be preceded by M (examples: ନାହିଁ, ନାଁ, ଗାଁରେ)

**Rule5**: H (୍) must be preceded by C or N

i) H may be preceded by C, (example: ଠିକ୍, ଭୁଲ୍)

ii) H may be preceded by N (example: ଡ଼୍ୟୁଟି, ଡ଼୍ରାଗନ)

**Rule6**: M must be preceded by C or N

i) M may be preceded by C (example: ମୁହଁ, ପହଁରା, ନୁହଁ)

ii) M may be preceded by N (example:ଡ଼ାଇମିଟର)

# Contributors

This proposal is prepared and submitted by Mr. Kuldeep Patnaik (Freelancer) and reviewed by Dr. Debashishya Jethy Oriya (Odia) linguistic analyst, translating medical science to Odia, coauthor of a book (Odia equivalents of scientific terms) from Central Institution of Indian Languages, Mysore.

Following NBGP members helped Mr. Kuldeep Patnaik to take crucial decision while working together for nine Indian languages including Oriya (Odia).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Position** | **Name** | **Organization** | **Country** | **Language Expertise** |
| **Co-Chair** | Ajay Data | [Data Xgen Technologies](https://www.xgenplus.com/) | India | Hindi, English |
| **Co-Chair** | Mahesh D. Kulkarni | C-DAC | India | Marathi, Hindi |
| **Co-Chair** | Udaya Narayana Singh | Visva-Bharati, Santiniketan, West Bengal | India | Bengali, Maithili, Hindi, English |
| Member | Akshat S. Joshi | C-DAC | India | Hindi, Marathi |
| Member | Atiur Rahman Khan | C-DAC | India | Bangla |
| Member | Dr Debasishya Jethy | Oriya (Odia) linguistic analyst | India | Odia |
| Member | Jay Paudyal | Consultant | India | Hindi |
| Member | Neha Gupta | C-DAC | India | Hindi |
| Member | Shanmugam R | C-DAC | India | Tamil |
| Member | Veena Solomon | (freelancer) | India | Malayalam |

Following is the list of other NBGP members with their language expertise.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Position** | **Name** | **Organization** | **Country** | **Language Expertise** |
| Member | Abhijit Dutta | Wikimedia | India | Bengali, Hindi |
| Member | Anivar A. Aravind | Indic Project | India | Malayalam |
| Member | Anupam Agrawal | Tata Consultancy Service | India | Hindi, Bengali |
| Member | Arvind Bhandari | Gujarat University | India | Gujarati |
| Member | Ashish Modi | [Data Xgen Technologies](https://www.xgenplus.com/) | India | Hindi |
| Member | Bal Krishna Bal | Kathmandu University | Nepal | Nepali |
| Member | BalaramPrasain | Tribhuvan University | Nepal | Nepali |
| Member | BASANTA KUMAR PANDA | Regional Institute of Education (NCERT) | India | Odia |
| Member | BhimDhoj Shrestha | Consultant | Nepal | Nepali, Newar |
| Member | Chitrita Chatterjee | Internet and Mobile Association of India (IAMAI) | India | Multiple languages represented by members of IAMAI |
| Member | DEBAJIT SHARMA | AnundoramBorooah Institute of Language Art and Culture | India | Assamese |
| Member | Dev DassManandhar | Consultant | Nepal | Nepali,Newar |
| Member | Dhanalakshmi KT | Northern Trust | India | Kannada |
| Member | Ganesh Murmu | Ranchi University | India | Santali |
| Member | GangadharPanday | Babul Films Society | India | Telugu |
| Member | Ghanashyam Nepal | Benares Hindu University& University of North Bengal | India | Nepali |
| Member | Girish Chandra Mishra | Language Technology Centre, Ravenshaw University | India | Odia |
| Member | Gurpreet Singh Lehal | Punjabi University Patiala | India | Panjabi |
| Member | Harish Chowdhary | NIXI | India | Hindi |
| Member | Hempal Shrestha | Nepal Entrepreneurs' Hub (NEHUB) | Nepal | Nepali, Newar |
| Member | JijoPappachan | DN. Domains | India | Malayalam |
| Member | K. C. Tikayat ray | Odia BhasaPratisthan | India | Odia |
| Member | KalyanVasudeo Kale | Formerly affiliated with University of Pune | India | Marathi |
| Member | Kuldeep Patnaik (Editor) | Freelancer | India | Odia |
| Member | Mukesh Saini | Essel Group | India | Hindi |
| Member | N. DeivaSundaram | NDS Lingsoft Solutions Pvt Ltd | India | Tamil |
| Member | NirajanParajuli | NREN | Nepal | Nepali |
| Member | Nishit Jain | C-DAC | India | Hindi |
| Member | PawanChitrakar | Gapsco | Nepal | Nepali |
| Member | Prabhakar Pandey | C-DAC | India | Hindi |
| Member | Prasad PK | A-one Publishers | India | Malayalam |
| Member | Prateek Pathak | ISOC Mumbai | India | Devanagari |
| Member | Raiomond Doctor | NLP Consultant | India | English, Hindi, Marathi, Gujarati |
| Member | Rajib Chakraborty | Society for Natural Language Technology Research | India | Bangla (Bengali) |
| Member | Rajiv Kumar | NIXI | India |  |
| Member | S.Maniam | International Forum IT for Tamil | Singapore | Tamil |
| Member | Santhosh Thottingal | Wikimedia foundation | India | Malayalam, Sourashtra, Tamil |
| Member | SarojaBhate | University of Pune | India | Sanskrit |
| Member | Shambhu Kumar Singh | National Translation Misson, Mysore | India | Maithili |
| Member | Shantaram S. WardeWalawalikar | Independent Researcher | India | Konkani |
| Member | Shashi Pathania | P.G.D. of Dogri, University of Jammu | India | Dogri |
| Member | Shubham Saran | NIXI | India |  |
| Member | SinnathambiShanmugarajah | University of Colombo School of Computing | Sri Lanka | Tamil |
| Member | SujithKartha | [Digitalkz.com](http://digitalkz.com/) | India | Malayalam |
| Member | SurajAdhikari | Mercantile Communications (and .np ccTLD) | Nepal | Nepali |
| Member | SwarnaPrabhaChainary | Guwahati University | India | Bodo |
| Member | U.B. Pavanaja | <http://vishvakannada.com/> | India | Kannada |
| Member | Uma Maheshwar G | CALTS, Univ. of Hyderabad | India | Telugu |
| Member | Uttam Shrestha Rana | NPNOG | Nepal | Nepali |
| Member | Vinay Murarka | Consultant; https://मेरा.भारत | India | Hindi |

Table 7: Contributors and NBG Panel

# References

[MSR] Integration Panel, "Maximal Starting Repertoire — MSR-3 Overview and   
Rationale", 28 March2018  
<https://www.icann.org/sites/default/files/packages/lgr/msr/msr-3-wle-rules-28mar18-en.html>

[NBGP] Neo-Brahmi Generation Panel

[Code Charts] The Unicode Standard 10.0 Character Code Charts <http://www.unicode.org/charts/PDF/U0B00.pdf> (Accessed on 12 January 2018)

[0] The Unicode Standard 1.1  
 Any code point originally encoded in Unicode 1.1

[6] The Unicode Standard 4.0  
 Any code point originally encoded in Unicode 4.0

[9] The Unicode Standard 5.1  
 Any code point originally encoded in Unicode 5.1

[101] Omniglot, "Oriya", <https://www.omniglot.com/writing/oriya.htm>   
(Accessed on 12 January 2018)

[102] Odia (Oriya) alphabet - Wikipedia <https://en.wikipedia.org/wiki/Odia_alphabet> (Accessed on 12 January 2018)

[103] Odia language - Wikipedia <https://en.wikipedia.org/wiki/Odia_language>   
(Accessed on 12 January 2018)

[104] Oriya (Unicode block) - Wikipedia, <https://en.wikipedia.org/wiki/Oriya_(Unicode_block)>

[105] Odisha State Govt. Primary School Grade 1 e-book “HasaKhela”: by  
 [Odisha Primary Education Programme Authority](http://opepa.odisha.gov.in/website/e-textbook.aspx) http://opepa.odisha.gov.in/website/Download/e-Text-Book/CLass%20I/Hasa%20Khela%20Part%20II/Haso%20Khelo-II-Page-112.pdf

# Appendix A: Cross-script Confusable Code Points

Oriya script has a set of possible cross-script confusable code points with the Gujarati, Bengali, Telugu, and Kannada.

## Oriya and Gujarati

The following characters are visually confusable. The NBGP discussed and concluded that they are similar code points but should not be considered as variant code points.

|  |  |
| --- | --- |
| Oriya | Gujarati |
| ଃ (0B03) | ઃ (0A83) |
| ପ (0B2A) | ઘ (0A98) |
| ଥ (0B25) | થ (0AA5) |

Table 8: Confusable code points between the Oriya and Gujarati scripts

## Oriya and Bengali

The following characters are visually confusable. The NBGP discussed and concluded that they are similar code points and should not be considered as variant code points.

|  |  |
| --- | --- |
| Bengali | Oriya |
| ও (0993) | ଓ (0B13) |

Table 9: Confusable code points between the Oriya and Bengali scripts

The following characters were discussed and the NBGP concluded that they are neither variant code points nor confusable code points.

|  |  |  |
| --- | --- | --- |
| Bengali | Oriya | Resolution |
| ঘ (0998) | ସ (0B38) | Distinguishable |

Table 10: Other resolutions between Oriya script and Bengali script

## Oriya and Telugu

The following characters were discussed and the NBGP concluded that they are not variant code points nor confusable code points

|  |  |  |
| --- | --- | --- |
| Oriya | Telugu | Resolution |
| ଠ (0B20) | ర (0C30) | Distinguishable |
| ଠ (0B20) | ఠ (0C20) | Distinguishable |

Table 11: Other resolutions between the Oriya and Telugu scripts

## Oriya and Kannada

The following characters were discussed and the NBGP concluded that they are not variant code points nor confusable code points

|  |  |  |
| --- | --- | --- |
| Oriya | Kannada | Resolution |
| ଠ (0B20) | ರ (0CB0) | Distinguishable |
| ଠ (0B20) | ಠ (0CA0) | Distinguishable |

Table 12: Other resolutions between the Oriya and Kannada scripts

# Appendix B: Oriya Dialects

There are different ways of speaking and meaning of words in local Oriya Language. However the script remains the same.**[[4]](#footnote-4)**

#### Standard Odia

**Kataki Odia** or *The Odia of Mughalbandi region* considered as Standard Odia due to literary traditions. It is spoken mainly in the eastern half of the state of Odisha, with little variation, in districts like Khurdha, Puri, Cuttack, Jajpur, Jagatsinghpur, Kendrapada, Dhenkanal, Angul and Nayagarh district.

#### Major forms, or dialects

Midnapori Odia:

Spoken in the undivided [Midnapore](https://en.wikipedia.org/wiki/Midnapore) District of West Bengal.

[***Singhbhumi Odia***](https://en.wikipedia.org/wiki/Singhbhumi_Odia):

Spoken in [East Singhbhum](https://en.wikipedia.org/wiki/East_Singhbhum), [West Singhbhum](https://en.wikipedia.org/wiki/West_Singhbhum) and [Saraikela-Kharsawan](https://en.wikipedia.org/wiki/Saraikela-Kharsawan) district of [Jharkhand](https://en.wikipedia.org/wiki/Jharkhand)

[***Baleswari Odia***](https://en.wikipedia.org/wiki/Baleswari_Odia):

Spoken in [Baleswar](https://en.wikipedia.org/wiki/Balasore_district), [Bhadrak](https://en.wikipedia.org/wiki/Bhadrak) and [Mayurbhanj](https://en.wikipedia.org/wiki/Mayurbhanj) district of Odisha.

[***Ganjami Odia***](https://en.wikipedia.org/wiki/Ganjami_Odia):

Spoken in [Ganjam](https://en.wikipedia.org/wiki/Ganjam) and [Gajapati](https://en.wikipedia.org/wiki/Gajapati_district) districts of Odisha and [Srikakulam](https://en.wikipedia.org/wiki/Srikakulam) district of Andhra Pradesh.

[***Sambalpuri***](https://en.wikipedia.org/wiki/Sambalpuri_language) ***Odia:***

Spoken in [Bargarh](https://en.wikipedia.org/wiki/Bargarh_district), [Bolangir](https://en.wikipedia.org/wiki/Bolangir_district), [Boudh](https://en.wikipedia.org/wiki/Boudh_district), [Debagarh](https://en.wikipedia.org/wiki/Debagarh_district), [Jharsuguda](https://en.wikipedia.org/wiki/Jharsuguda_district), [Kalahandi](https://en.wikipedia.org/wiki/Kalahandi), [Nuapada](https://en.wikipedia.org/wiki/Nuapada_district), [Sambalpur](https://en.wikipedia.org/wiki/Sambalpur_district) and [Subarnapur](https://en.wikipedia.org/wiki/Subarnapur_district) districts of Odisha and by some people in [Raigarh](https://en.wikipedia.org/wiki/Raigarh_district), [Mahasamund](https://en.wikipedia.org/wiki/Mahasamund_district), [Raipur](https://en.wikipedia.org/wiki/Raipur_district) districts of [Chhattisgarh](https://en.wikipedia.org/wiki/Chhattisgarh) state.

[***Desiya Odia***](https://www.ethnologue.com/language/dso):

Spoken in   
[Koraput](https://en.wikipedia.org/wiki/Koraput), [Rayagada](https://en.wikipedia.org/wiki/Rayagada), [Nowrangpur](https://en.wikipedia.org/wiki/Nowrangpur) and [Malkangiri](https://en.wikipedia.org/wiki/Malkangiri) Districts of Odisha and in the hilly regions of [Vishakhapatnam](https://en.wikipedia.org/wiki/Vishakhapatnam), [Vizianagaram](https://en.wikipedia.org/wiki/Vizianagaram) District of Andhra Pradesh.

[***Bhatri***](https://en.wikipedia.org/wiki/Bhatri_language):

Spoken in South-western Odisha and eastern-south Chhattisgarh.

[***Halbi***](https://en.wikipedia.org/wiki/Halbi_language):

Spoken in undivided Bastar district of [Chhattisgarh](https://en.wikipedia.org/wiki/Chhattisgarh). Halbi is a mixture of Odia and Marathi with influence of Chatishgarhi tribal languages.

[***Phulbani Odia***](https://en.wikipedia.org/wiki/Phulbani_Odia):

Spoken in [Phulbani](https://en.wikipedia.org/wiki/Phulbani), [Phulbani Town](https://en.wikipedia.org/wiki/Phulbani), Khajuripada block of [Kandhamal](https://en.wikipedia.org/wiki/Kandhamal_district), and in nearby areas bordering [Boudh district](https://en.wikipedia.org/wiki/Boudh_district) .This language gained momentum during the amalgamation of [Kandhamal(Phulbani)](https://en.wikipedia.org/wiki/Kandhamal), and [Boudh](https://en.wikipedia.org/wiki/Boudh), region into a single district [Phulabani](https://en.wikipedia.org/wiki/Phulbani),

#### Minor non-literary and tribal forms or dialects

[**Sundargadi**](https://en.wikipedia.org/wiki/Sundargadi)**Odia** :

Variation of Odia Spoken in Sundargarh district of [Odisha](https://en.wikipedia.org/wiki/Odisha) and in adjoining pockets of [Jharkhand](https://en.wikipedia.org/wiki/Jharkhand) and [Chhattisgarh](https://en.wikipedia.org/wiki/Chhattisgarh).

[**Kalahandia**](https://en.wikipedia.org/wiki/Kalahandia)**Odia** :

Variation of Odia spoken in undivided [Kalahandi District](https://en.wikipedia.org/wiki/Kalahandi_District) and neighboring districts of [Chhattisgarh](https://en.wikipedia.org/wiki/Chhattisgarh).

**Kurmi**: Spoken in Northern Odisha and South west Bengal.

**Sounti**: Spoken in Northern Odisha and South west Bengal.

**Bathudi**: Spoken in Northern Odisha and South west Bengal.

**Kondhan**: A tribal dialect spoken in Western Odisha..

[**Laria**](https://www.ethnologue.com/language/hne): Spoken in bordering areas of Chatishgarh and Western Odisha.

[**Aghria**](https://en.wikipedia.org/wiki/Agharia): Spoken mostly by the ingenious people of Aghria caste in Western Odisha.

**Bhulia**: Tribal form spoken in Western Odisha.

[**Sadri**](https://www.ethnologue.com/language/sck): A mixture of Odia and Hindi language with major regional tribal influence.

[**Bodo Parja / Jharia**](https://www.ethnologue.com/language/bdv): Tribal dialect of Odia spoken mostly in Koraput district of Southern Odisha .

**Matia**: Tribal dialect of Odia spoken in Southern Odisha.

**Bhuyan**: Tribal dialect of Odia spoken in Southern Odisha.

[**Reli**](https://www.ethnologue.com/language/rei): Spoken in Southern Odisha and bordering areas of Andhra Pradesh.

[**Kupia**](https://www.ethnologue.com/language/key): Spoken by [Valmiki](https://en.wikipedia.org/wiki/Valmiki_(caste)) caste people in the Indian state of [Telangana](https://en.wikipedia.org/wiki/Telangana) and [Andhra Pradesh](https://en.wikipedia.org/wiki/Andhra_Pradesh), mostly in [Hyderabad](https://en.wikipedia.org/wiki/Hyderabad_district,_India), [Mahabubnagar](https://en.wikipedia.org/wiki/Mahabubnagar_District), [Srikakulam](https://en.wikipedia.org/wiki/Srikakulam), [Vizianagaram](https://en.wikipedia.org/wiki/Vizianagaram), [East Godavari](https://en.wikipedia.org/wiki/East_Godavari_district) and [Visakhapatnam](https://en.wikipedia.org/wiki/Visakhapatnam_district) districts.

# Appendix C: Oriya Characters

Odisha State Government Primary School Grade 1 e-book “HasaKhela” [105] page 112 lists all the Oriya characters as shown in Figure 4.

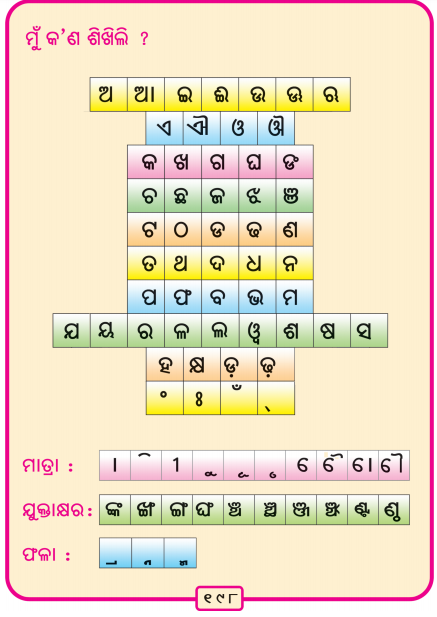


Figure 4: Odisha State Govt. Primary School Grade 1 e-book (Page 112)

# Appendix D: NBGP Cross-script Variant Inclusion Policy

If, in any two given scripts, all the potential cross-script variants consist of dependent (e.g. Vowel Signs, Anusvara, Visarga, Chandrabindu etc.) characters **ONLY**, then that entire set can be ignored and no cross-script variants be proposed between those two scripts.

If, in any two given scripts, there is **AT LEAST ONE** non-dependent (e.g. Consonant, Vowel etc.) cross-script variant character/sequence present, all the potential cross-script variants be considered and proposed between the two scripts.  
  
This cross-script analysis has been restricted to the scripts that have descended from the Brahmi as most of them share similar usage patterns. By and large, all of these scripts have a common set of characters that existed in Brahmi script and bear the same identities. However, as the scripts branched out from the Brahmi, depending on various factors, the shapes of the characters changed. This change in the shape was not uniform across all the characters and the scripts. Some characters shapes did change significantly whereas some of them still retained similarity. The cross-script similarity analysis also aims to identify such cases where the same character retained almost the same shape despite being part of the different scripts. These set of characters are variants of each other in true sense than merely of co-incidental visual similarity. 

**Case of Malayalam and Odia (Oriya) TTHA Consonant:**  
This is the case of "Consonant Ttha" which happened to retain the same shape despite being part of different scripts, i.e., Malayalam and Odia. These characters are:

ഠ - MALAYALAM LETTER TTHA (U+0D20)  
ଠ - ORIYA LETTER TTHA (U+0B20)

Both the characters, look exactly alike and belong to a "Consonant" category. As they are consonants, each of them, even in the simplest form i.e. the characters themselves, are valid labels. As per the NBGP cross-script variant inclusion policy, this is a valid case for inclusion. Also, even if they are single characters, when the same character combines, theoretically they can form infinite[[5]](#footnote-5) number of cross-script variant labels between the scripts involved. Here are some samples of some of those labels:

|  |  |
| --- | --- |
| **Malayalam** | **Oriya** |
| ഠഠഠ U+0D20 U+0D20U+0D20 | ଠଠଠ U+0B20 U+0B20U+0B20 |
| ഠഠഠഠ U+0D20 U+0D20U+0D20U+0D20 | ଠଠଠଠ U+0B20 U+0B20U+0B20U+0B20 |
| ഠഠഠഠഠ U+0D20 U+0D20U+0D20U+0D20U+0D20 | ଠଠଠଠଠ U+0B20 U+0B20U+0B20U+0B20U+0B20 |

Since, having such labels is a realistic possibility and the corresponding labels look almost exactly alike, NBGP has proposed them as blocked variants.

NBGP acknowledges the concern that this shape is quite generic and may have parallels in other scripts not under its ambit.  However, as NBGP does not have any exposure about actual usage of those characters in those particular scripts, NBGP desisted from including them in the analysis.  As NBGP has already considered all the related scripts under the cross-script variant analysis, the similarity of the characters belonging to NBGP scripts with other scripts not under the NBGP ambit, may be of a mere co-incidental visual nature.

Additionally, this concern is not limited to these two characters but for all the characters in all the scripts under the scope of the Root LGR procedure. Carrying out this analysis can practically be done only with the Generation Panels that exist while the NBGP is active. This still leaves out those scripts out of the scope which may not have a Generation Panel established yet. Hence, carrying out this exercise in entirety is quite impracticable. This conundrum can be resolved if all the such cases are handled by the "String Similarity Assessment Panel" of ICANN.

1. Criteria for this status include: high antiquity of its early texts/recorded history over a period of 1500–2000 years; a body of ancient literature/texts, which is considered a valuable heritage by generations of speakers; a literary tradition that is original and not borrowed from another speech community; [↑](#footnote-ref-1)
2. The variety of Oriya dialects etc. is reviewed in Appendix B. [↑](#footnote-ref-2)
3. International Alphabet of Sanskrit Transliteration (I.A.S.T.) is a transliteration scheme that allows the lossless romanization of Indic scripts as employed by Sanskrit and related Indic languages. IAST makes it possible for the reader to read the Indic text unambiguously, exactly as if it were in the original Indic script. Example: କ 0B15 (ka), ଖ 0B16 (kha) etc. [↑](#footnote-ref-3)
4. Extracted from [Wikipedia](https://en.wikipedia.org/wiki/Odia_language#Major_forms_or_dialects), <https://en.wikipedia.org/wiki/Odia_language#Major_forms_or_dialects> [↑](#footnote-ref-4)
5. Though theoretically infinite, this number would be limited to the number of such labels whose equivalent punycode string would not exceed 63 characters including the ACE prefix "xn--". [↑](#footnote-ref-5)