

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

LGR Version: 3.0
Date: 2018-05-29
Document version: 2.3
Authors: Neo-Brahmi Generation Panel [NBGP]

1 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-20180529.xml

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

oriya-test-labels-20180529.txt

2 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

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) [ɔd̪iɑ:] and ଓଡ଼ିଶା (oḍiśā) [ɔd̪iʃɑ:].

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 It forms the basis of Odissidance and Odissi music.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

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;

²The variety of Oriya dialects etc. is reviewed in Appendix B.

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 etc, and Sanskrit.

3.1 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.

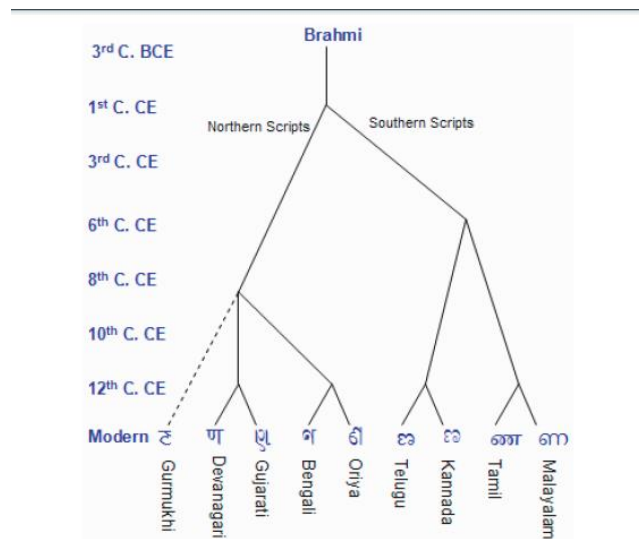


Figure 1: Pictorial depiction of Evolution of Oriya

Literature in Oriya (Odia) language (Odia: ଓଡ଼ିଆସାହିତ୍ୟ) is the predominant literature of the state of Odisha in India. The actual Oriya script closely resembled Bengali and Assamese scripts, but the one adopted for the printed typesets was completely different, leaning more towards the Tamil script.

3.2 Periods of Odia History

Oriya (Odia) language literature (Odia: ଓଡ଼ିଆସାହିତ୍ୟ) 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).

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

3.3 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.

3.4 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		Avargya consonants		
IPA	Oriya (Odia)	IPA	Oriya (Odia)	
<u>b</u>	ବ		ଘ	
<u>b^h</u>	ଭ	<u>i</u>	ଈ	
<u>d</u>	ଦ	<u>r</u>	ର	
<u>d^h</u>	ଢ	<u>l</u>	ଲ	
<u>ɖ</u>	ଡ	<u>l</u>	ଳ	
<u>d^h</u>	ଢ	<u>u</u>	ୂ	
<u>ɖʒ</u>	ଢ଼	<u>w</u>	ଓ	
<u>ɖʒ^h</u>	ଢ଼	<u>s</u>	ଷ	
<u>g</u>	ଗ	<u>ʂ</u>	ଞ	
<u>g^h</u>	ଘ	<u>ɕ</u>	ଞ	
<u>h</u>	ହ	<u>h</u>	ହ	
<u>k</u>	କ	<u>ŋ, ɲ, ɳ, ɹ, ɻ, ɽ, ʈ, ʡ</u>	ଂ	
<u>k^h</u>	ଖ	<u>̃</u>	ଂ	
<u>ɳ</u>	ଣ	<u>h</u>	ଃ	
<u>m</u>	ମ	Vowels and Matras		
<u>n</u>	ନ	IPA	Vowels	Matras
<u>n</u>	ଣ	<u>ə</u>	ଅ	
<u>n</u>	ଣ	<u>a:</u>	ଆ	।
<u>p</u>	ପ	<u>ɪ</u>	ଇ	।
<u>p^h</u>	ଫ	<u>i:</u>	ଈ	।
<u>r</u>	ର	<u>ʊ</u>	ଉ	।
<u>r</u>	ୠ	<u>u:</u>	ଊ	।
<u>r^h</u>	ୡ	<u>r</u>	ୠ	।
<u>s</u>	ସ	<u>e:</u>	ଏ	।
<u>t</u>	ଟ	<u>ɛ:</u>	ଐ	।
<u>t^h</u>	ଠ	<u>o:</u>	ଓ	।
<u>t</u>	ଡ	<u>ɔ:</u>	ଔ	।
<u>t^h</u>	ଢ			
<u>tʃ</u>	ଚ			
<u>tʃ^h</u>	ଢ଼			

Table 1: International Phonetic Alphabet Oriya Pronunciations

3.5 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) transcriptions.

	voiceless	voiceless aspirate	voiced	voiced aspirate	nasal
Velars	କ (ka)	ଖ (kha)	ଗ (ga)	ଘ (gha)	ଙ (ṅa)
Palatals	ଚ (ca)	ଛ (cha)	ଜ (ja)	ଝ (jha)	ଞ (ña)
Retroflex	ଟ (ṭa)	ଠ (ṭha)	ଡ (ḍa)	ଢ (ḍha)	ଣ (ṇa)
Dentals	ତ (ta)	ଥ (tha)	ଦ (da)	ଧ (dha)	ନ (na)
Labials	ପ (pa)	ଫ (pha)	ବ (ba)	ଭ (bha)	ମ (ma)

Table 2:structured consonants

3.6 Unstructured consonants

The unstructured consonants are consonants that do not fall into any of the above categories:

ଯ (ja), ଐ (ia), ର (ra), ଲ (la), ଳ (ḷa), ବ (va), ଷ (wa), ଶ (sa), ଷ (sa), ସ (sa), ହ (ha)

3.7 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).

3.8 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 “ୱ” U+0B21, “ୱ” U+0B22, “ୱ” U+0B15, “ୱ” U+0B16, “ୱ” U+0B17, “ୱ” U+0B1A, “ୱ” U+0B1C, and “ୱ” U+0B2B to show that words having these consonants with a nukta are to be pronounced in the Perso-Arabic style.

3.9 Visarga “ୱ” (U+0B03) and Avagraha “ୱ” (U+0B3D):

The Visarga (“ୱ” (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“ୱ”(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.

3.10 Candrabindu (ँ - U+0B01):

Candrabindu denotes nasalization of the preceding vowel and consonants as in ଞ୍ଜାଳା /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.

3.11 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: ଶୱଂ (0B0F+0B2C+0B02), ଶୱଂ (0B38+0B02+0B16+0B4D+0B5F+0B3E), etc.

3.12 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	Glyph	Unicode	Name
ଅ	U+0B05	ORIYA LETTER A			
ଅା	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
ଇି	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
ଉୁ	U+0B0A	ORIYA VOWEL LETTER UU	ୁୁ	U+0B42	ORIYA VOWEL SIGN UU
ଋ	U+0B0B	ORIYA VOWEL LETTER VOCALIC R	ୠ	U+0B43	ORIYA VOWEL SIGN VOCALIC R
ଏ	U+0B0F	ORIYA VOWEL LETTER E	େ	U+0B47	ORIYA VOWEL SIGN E
ଏୱି	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
ଓୱି	U+0B14	ORIYA VOWEL LETTER AU	ୌି	U+0B4C	ORIYA VOWEL SIGN AU
ଌ	U+0B0C	ORIYA LETTER VOCALIC L	ୡ	U+0B62	ORIYA VOWEL SIGN VOCALIC L
ଌି	U+0B61	ORIYA LETTER VOCALIC LL	ୡି	U+0B63	ORIYA VOWEL SIGN VOCALIC LL

Table 3: Vowel and its Matra Sign

“୫” U+0B0C, “୫” U+0B61, “୫” U+0B62 and “୫” U+0B63 are hardly in use in modern days.

4 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.

4.1 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.

4.1.1 Inclusion principles:

4.1.1.1 *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.

4.1.2 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.

4.1.2.1 *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 “ଓ” (U+0B21), “ଔ” (U+0B22) as well as their respective allophones “ଓ̃”, and “ଔ̃”. In Oriya (Odia) script, these differ in use of nukta. Thus “ଓ̃” and “ଔ̃” as distinct letters are not allowed but their decomposed form i.e. “ଓ”, “ଔ” followed by Oriya (Odia) sign nukta (U+0B3C) can be used. Similarly, for allophones of other consonants like ଩ (U+0B15), ଶ (U+0B16), ଷ (U+0B17), ଠ (U+0B1A), ଡ (U+0B1C), ଢ (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 "୪" (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.

4.1.2.2 *No Fraction Marks:*

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

Fraction signs		
0B72		ORIYA FRACTION ONE QUARTER
0B73	୩	ORIYA FRACTION ONE HALF
0B74	୩୩	ORIYA FRACTION THREE QUARTERS
0B75	୪	ORIYA FRACTION ONE SIXTEENTH
0B76	୫	ORIYA FRACTION ONE EIGHTH
0B77	୬	ORIYA FRACTION THREE SIXTEENTHS

Figure 2: Fraction Marks in Oriya

4.1.2.3 *No Symbols and Abbreviations:*

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

4.1.2.4 *No Rare and Obsolete Characters:*

There are characters which have been added to Unicode to accommodate rare forms especially like Oriya LETTER VOCALIC RR"ଠ" (U+0B60) and Oriya LETTER VOCALIC LL"ଢ" (U+0B61) as well as their Matra forms) "ଠ" (U+0B44) and "ଢ" (U+0B63). All such characters will not be included. This is in compliance with the Conservatism principle as laid down in the Root Zone LGR procedure.

5 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.

5.1 5.1 Oriya Section of Maximal Starting Repertoire [MSR] Version 3

	OB0	OB1	OB2	OB3	OB4	OB5	OB6	OB7
0		ଂ OB10	ଠ OB20	ଠ OB30	ଠ OB40		ଠ OB60	✓ OB70
1	ଠ OB01		ଠ OB21		ଠ OB41		ଠ OB61	ଠ OB71
2	ଠ OB02		ଠ OB22	ଠ OB32	ଠ OB42		ଠ OB62	ଠ OB72
3	ଠ OB03	ଠ OB13	ଠ OB23	ଠ OB33	ଠ OB43		ଠ OB63	ଠ OB73
4		ଠ OB14	ଠ OB24		ଠ OB44			ଠ OB74
5	ଠ OB05	ଠ OB15	ଠ OB25	ଠ OB35				ଠ OB75
6	ଠ OB06	ଠ OB16	ଠ OB26	ଠ OB36		ଠ OB56	ଠ OB66	ଠ OB76
7	ଠ OB07	ଠ OB17	ଠ OB27	ଠ OB37	ଠ OB47	ଠ OB57	ଠ OB67	ଠ OB77
8	ଠ OB08	ଠ OB18	ଠ OB28	ଠ OB38	ଠ OB48		ଠ OB68	
9	ଠ OB09	ଠ OB19		ଠ OB39			ଠ OB69	
A	ଠ OB0A	ଠ OB1A	ଠ OB2A				ଠ OB6A	
B	ଠ OB0B	ଠ OB1B	ଠ OB2B		ଠ OB4B		ଠ OB6B	
C	ଠ OB0C	ଠ OB1C	ଠ OB2C	ଠ OB3C	ଠ OB4C	ଠ OB5C	ଠ OB6C	
D		ଠ OB1D	ଠ OB2D	ଠ OB3D	ଠ OB4D	ଠ OB5D	ଠ OB6D	
E		ଠ OB1E	ଠ OB2E	ଠ OB3E			ଠ OB6E	
F	ଠ OB0F	ଠ OB1F	ଠ OB2F	ଠ OB3F		ଠ OB5F	ଠ OB6F	

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

Figure 3: Oriya Code Page from MSR 3

5.2 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	OB01	◌◌̣	ORIYA SIGN CANDRABINDU	2-Oriya	Candrabinu	[0], [101], [102], [103], [104], [105]
2	OB02	◌◌̣◌◌̣	ORIYA SIGN ANUSVARA	2-Oriya	Anusvara	[0], [101], [102], [103], [104], [105]
3	OB03	◌◌̣◌◌̣◌◌̣	ORIYA SIGN VISARGA	2-Oriya	Visarga	[0], [101], [102], [103], [104], [105]
4	OB05	ଅ	ORIYA LETTER A	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
5	OB06	ଆ	ORIYA LETTER AA	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
6	OB07	ଇ	ORIYA LETTER I	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
7	OB08	ଈ	ORIYA LETTER II	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
8	OB09	ଉ	ORIYA LETTER U	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
9	OBOA	ଊ	ORIYA LETTER UU	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
10	OB0B	ଋ	ORIYA LETTER VOCALIC R	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
11	OBOF	ଏ	ORIYA LETTER E	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
12	OB10	ଐ	ORIYA LETTER AI	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]

Sr. No.	Unicode Code Point	Glyph	Character Name	Language with EGIDS	Indic Syllabic Category	References
13	OB13	ଓ	ORIYA LETTER O	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
14	OB14	ଔ	ORIYA LETTER AU	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]
15	OB15	କ	ORIYA LETTER KA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
16	OB16	ଖ	ORIYA LETTER KHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
17	OB17	ଗ	ORIYA LETTER GA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
18	OB18	ଘ	ORIYA LETTER GHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
19	OB19	ଙ	ORIYA LETTER NGA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
20	OB1A	ଚ	ORIYA LETTER CA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
21	OB1B	ଛ	ORIYA LETTER CHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
22	OB1C	ଜ	ORIYA LETTER JA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
23	OB1D	ଝ	ORIYA LETTER JHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
24	OB1E	ଞ	ORIYA LETTER NYA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
25	OB1F	ଟ	ORIYA LETTER TTA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
26	OB20	ଠ	ORIYA LETTER TTHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]

Sr. No.	Unicode Code Point	Glyph	Character Name	Language with EGIDS	Indic Syllabic Category	References
27	0B21	ୱ	ORIYA LETTER DDA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
28	0B22	ୱ	ORIYA LETTER DDHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
29	0B23	ୱ	ORIYA LETTER NNA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
30	0B24	ୱ	ORIYA LETTER TA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
31	0B25	ୱ	ORIYA LETTER THA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
32	0B26	ୱ	ORIYA LETTER DA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
33	0B27	ୱ	ORIYA LETTER DHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
34	0B28	ୱ	ORIYA LETTER NA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
35	0B2A	ୱ	ORIYA LETTER PA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
36	0B2B	ୱ	ORIYA LETTER PHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
37	0B2C	ୱ	ORIYA LETTER BA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
38	0B2D	ୱ	ORIYA LETTER BHA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
39	0B2E	ୱ	ORIYA LETTER MA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
40	0B2F	ୱ	ORIYA LETTER YA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]

Sr. No.	Unicode Code Point	Glyph	Character Name	Language with EGIDS	Indic Syllabic Category	References
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	ବ	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	ୠ	ORIYA SIGN NUKTA	2-Oriya	Nukta	[0], [101], [102], [103], [104], [105]
50	0B3E	ୡ	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]

Sr. No.	Unicode Code Point	Glyph	Character Name	Language with EGIDS	Indic Syllabic Category	References
55	0B43	ୠ	ORIYA VOWEL SIGN VOCALIC R	2-Oriya	Matra	[0], [101], [102], [103], [104], [105]
56	0B47	ୡ	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	୦	ORIYA SIGN VIRAMA	2-Oriya	Halant	[0], [101], [102], [103], [104], [105]
61	0B56	ୠ	ORIYA AI LENGTH MARK	2-Oriya	Matra	[2], [101], [102], [103], [104], [105]
62	0B5F	ୡ	ORIYA LETTER YYA	2-Oriya	Consonant	[0], [101], [102], [103], [104], [105]
63	0B71	ୢ	ORIYA LETTER WA	2-Oriya	Consonant	[6], [101], [102], [103], [104], [105]

Table 4: Code Point Repertoire

5.2.1 Code Points Excluded

Sr. No.	Unicode Code Point	Glyph	Character Name	Language with EGIDS	Indic Syllabic Category	Reference
1	0B0C	ୠ	ORIYA LETTER VOCALIC L	2-Oriya	Vowel	[0], [101], [102], [103], [104], [105]

2	0B44	ୠ	ORIYA VOWEL SIGN VOCALIC RR	2-Oriya	Matra	[9], [101], [102] [103] [104] [105]
3	0B57	ୠ	ORIYA AU LENTH MARK	2-Oriya	Matra	[0], [101], [102] [103] [104] [105]

Table 5: Code Point Excluded from Repertoire

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

5.2.2 Variables involved

C	→	Consonant
M	→	Matra
V	→	Vowel
B	→	Anusvara
H	→	Halant / Virama
N	→	Nukta
C1	→	{ ୫ 0B15, ୬ 0B16, ୭ 0B17, ୮ 0B21, ୯ 0B22, ୧ 0B1A, ୨ 0B1C, ୩ 0B2B}
X	→	Visarga
D	→	Candrabindu

6 Variants

6.1 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.

6.2 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 provide a set of cross-script variant code points it identifies with all 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 only with the Malayalam script. Cases listed in Table 6 are cross-script variants between Oriya and Malayalam.

It is to be noted that none of the combinations listed in Table 6 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	୦	0D20	୦

Table 6: 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.

7 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, ଢ 0B21 DDA, ଢ 0B22 DDHA, ଚ 0B1A CA, ଜ 0B1C JA, ଫ 0B2B PHA }
X	→	Visarga
D	→	Candrabindu

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

For example:

ଡ(0B21)+ଠ(0B3C) = ଢ଼

ଢ(0B22)+ଠ(0B3C) = ଢ଼

Rule2: B (ୡ) 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: ଢ଼ଗା)
- 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: ପ୍ରାୟଃ, କ୍ରମଶଃ)
- ii) X may be preceded by N (example: ଢ଼ଃ)
- iii) X may be preceded by M, (examples: ଦୁଃଖ, ଦୁଃଖତ୍ତ୍ୱ,)
- iv) X may be preceded by V, (examples: ଅଃ, ଆଃ, ଇଃ, ଊଃ) commonly used when writing Sanskrit or when there is religious requirement

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

- i) D may be preceded by V (examples: ପାଌ, ଯେଌ, ନିଆଁ)
- ii) D may be preceded by C (example: ମୁହଁ, ପହଁରା, ନୁହଁ)
- iii) D may be preceded by N (example: ଢାଁଶ)
- 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: ଢୁକ୍ତି, ଢ୍ରାଗନ)

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: ଢ୍ରାଗନିଚର)

8 Contributors

NBGP Co-chairs: Dr. Udaya Narayan Singh, Mr. Mahesh D Kulkarni and Dr. Ajay Data

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

Name	Language Expertise
Udaya Narayana Singh	Bengali, Maithili, Hindi, English
Ajay Data	Hindi
Mahesh D. Kulkarni	Marathi, Hindi
Anupam Agrawal	Hindi, Bengali
Akshat S. Joshi	Hindi, Marathi
Abhijit Dutta	Bengali, Hindi
Neha Gupta	Hindi
Nishit Jain	Hindi
Prabhakar Pandey	Hindi
Raiomond Doctor	English, Hindi, Marathi, Gujarati
N. DeivaSundaram	Tamil
Shantaram S. WardeWalawalikar	Konkani
Bal Krishna Bal	Nepali
Ganesh Murmu	Santali
BalaramPrasain	Nepali
Rajib Chakraborty	Bangla (Bengali)
Gurpreet Singh Lehal	Panjabi
SarojaBhate	Sanskrit

Shambhu Kumar Singh	Maithili
SwarnaPrabhaChainary	Bodo
Ghanashyam Nepal	Nepali
KalyanVasudeo Kale	Marathi
Shashi Pathania	Dogri
Santhosh Thottingal	Malayalam, Sourashtra, Tamil
Uma Maheshwar G	Telugu
Girish Chandra Mishra	Odia
K. C. Tikayat ray	Odia
Kuldeep Patnaik	Odia
Debajit Sharma	Assamese
DrDevasisaJethy	Oriya
Basanta Kumar Panda	Oriya
Arvind Bhandari	Gujarati
Harish Chowdhary	Hindi
Chitrita Chatterjee	Multiple languages represented by members of IAMA
U.B. Pavanaja	Kannada
Hempal Shrestha	Nepali, Newari
Suraj Adhikari	Nepali
Gangadhar Panday	Telugu
Vinay Murarka	Hindi
Mukesh Saini	Hindi
Jay Paudyal	Hindi

PawanChitrakar	Nepali
Nirajan Parajuli	Nepali
Uttam Shrestha Rana	Nepali
Dev Dass Manandhar	Nepali,Newari
Bhim Dhoj Shrestha	Nepali, Newari
Rajiv Kumar	Hindi
Shubham Saran	Hindi
Anivar A. Aravind	Malayalam
Shanmugam R	Tamil
Prasad PK	Malayalam
Sinnathambi Shanmugarajah	Tamil

Table 7: Contributors and NBG Panel

9 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 12th Jan. 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>
- [102] Odia (Oriya) alphabet - Wikipedia https://en.wikipedia.org/wiki/Odia_alphabet
- [103]Odia language - Wikipedia https://en.wikipedia.org/wiki/Odia_language
- [104] Oriya (Unicode block) -
Wikipedia,[https://en.wikipedia.org/wiki/Oriya_\(Unicode_block\)](https://en.wikipedia.org/wiki/Oriya_(Unicode_block))
- [105] Odisha State Govt. Primary School Grade 1 e-book “HasaKhela”: by [Odisha Primary Education Programme Authorityhttp://opepa.odisha.gov.in/website/Download/e-Text-Book/Class%20I/Hasa%20Khela%20Part%20II/Haso%20Khelo-II-Page-112.pdf](http://opepa.odisha.gov.in/website/Download/e-Text-Book/Class%20I/Hasa%20Khela%20Part%20II/Haso%20Khelo-II-Page-112.pdf)

ମୁଁ କ'ଣ ଶିଖିଲି ?

ଅ	ଆ	ଇ	ଈ	ଉ	ଊ	ଋ
ଏ	ଐ	ଓ	ଔ			
କ	ଖ	ଗ	ଘ	ଙ		
ଚ	ଛ	ଜ	ଝ	ଞ		
ଟ	ଠ	ଡ	ଢ	ଣ		
ତ	ଥ	ଦ	ଧ	ନ		
ପ	ଫ	ବ	ଭ	ମ		
ଯ	ର	ଲ	ଳ	ୱ	ଶ	ଷ
ସ	ହ	ୱ	ୱ			
ଂ	ଃ	ଌ	ୠ			

ମାତ୍ରା : | ̇ 1 ̣ ̤ ̥ ̦ ̧ ̨ ̩

ଯୁକ୍ତାକ୍ଷର : କ ଖ ଗ ଘ ଙ ଛ ଝ ଞ ଟ ଠ ଡ ଢ ଣ ତ ଥ ଦ ଧ ନ ପ ଫ ବ ଭ ମ ଯ ର ଲ ଳ ଳ ଶ ଷ ସ ହ ଣ ଣ ଣ

ଫଳା : । ଂ ଃ

୧୯୮

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

10 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.

10.1 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
ଠ (0B03)	ઠ (0A83)
ଢ (0B2A)	ઢ (0A98)
ଢ (0B25)	ઢ (0AA5)

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

10.2 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
ଓU+0993	ଓU+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
ଐU+0998	ଐU+0B38	Distinguishable

10.3 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
ୠ (0B20)	ఁ (0C30)	Distinguishable
ୠ (0B20)	ం (0C20)	Distinguishable

Table 11: Other resolutions between the Oriya and Telugu scripts

10.4 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
ୠ (0B20)	ಁ (0CB0)	Distinguishable
ୠ (0B20)	ಂ (0CA0)	Distinguishable

Table 12: Other resolutions between the Oriya and Kannada scripts

11 Appendix B: Oriya Dialects

There are different ways of speaking and meaning of words in local Oriya Language. However the script remains the same.³

11.1.1.1 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.

³Extracted from [Wikipedia](https://en.wikipedia.org/wiki/Odia_language#Major_forms_or_dialects), https://en.wikipedia.org/wiki/Odia_language#Major_forms_or_dialects

11.1.1.2 Major forms, or dialects

Midnapori Odia:

Spoken in the undivided Midnapore District of West Bengal.

Singhbhumi Odia:

Spoken in East Singhbhum, West Singhbhum and Saraikela-Kharsawan district of Jharkhand

Baleswari Odia:

Spoken in Baleswar, Bhadrak and Mayurbhanj district of Odisha.

Ganjami Odia:

Spoken in [Ganjam](#) and [Gajapati](#) districts of Odisha and [Srikakulam](#) district of Andhra Pradesh.

Sambalpuri Odia:

Spoken in

[Bargarh](#), [Bolangir](#), [Boudh](#), [Debagarh](#), [Jharsuguda](#), [Kalahandi](#), [Nuapada](#), [Sambalpur](#) and [Subarnapur](#) districts of Odisha and by some people in [Raigarh](#), [Mahasamund](#), [Raipur](#) districts of [Chhattisgarh](#) state.

Desiya Odia:

Spoken in

[Koraput](#), [Rayagada](#), [Nowrangpur](#) and [Malkangiri](#) Districts of Odisha and in the hilly regions of [Vishakhapatnam](#), [Vizianagaram](#) District of Andhra Pradesh.

Bhatri:

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

Halbi:

Spoken in undivided Bastar district of [Chhattisgarh](#). Halbi is a mixture of Odia and Marathi with influence of Chatishgarhi tribal languages.

Phulbani Odia:

Spoken in [Phulbani](#), [Phulbani Town](#), Khajuripada block of [Kandhamal](#), and in nearby areas bordering [Boudh district](#). This language gained momentum during the amalgamation of [Kandhamal\(Phulbani\)](#), and [Boudh](#), region into a single district [Phulabani](#),

11.1.1.3 Minor non-literary and tribal forms or dialects

Sundargadi Odia :

Variation of Odia Spoken in Sundargarh district of [Odisha](#) and in adjoining pockets of [Jharkhand](#) and [Chhattisgarh](#).

[Kalahandia Odia](#) :

Variation of Odia spoken in undivided Kalahandi District and neighboring districts of 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: Spoken in bordering areas of Chhattisgarh and Western Odisha.

Aghria: Spoken mostly by the ingenious people of Aghria caste in Western Odisha.

Bhulia: Tribal form spoken in Western Odisha.

Sadri: A mixture of Odia and Hindi language with major regional tribal influence.

Bodo Parja / Jharia: 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: Spoken in Southern Odisha and bordering areas of Andhra Pradesh.

Kupia: Spoken by [Valmiki](#) caste people in the Indian state of [Telangana](#) and [Andhra Pradesh](#), mostly in [Hyderabad](#), [Mahabubnagar](#), [Srikakulam](#), [Vizianagaram](#), [East Godavari](#) and [Visakhapatnam](#) districts.