

## Typography and Education

<http://www.typoday.in>

## Bringing Ol Chiki to the digital world

Saxena, Pooja, [anexasajoop@gmail.com](mailto:anexasajoop@gmail.com)

Panigrahi, Subhashish, Programme Officer, Access to Knowledge, Center for Internet and Society, Bengaluru (India), [psubhashish@gmail.com](mailto:psubhashish@gmail.com)

**Abstract:** Can a typeface turn the fate of an indigenous language around by making communication possible on digital platforms and driving digital activism? In 2014, a project was initiated with financial support from the Access to Knowledge programme at The Center for Internet and Society, Bangalore to look for answers to this question. The project's goal was to design a typeface family supporting Ol Chiki script, which is used to write Santali, along with input methods that would make typing in Ol Chiki possible. It was planned that these resources would be released under a free license, with the hope to provide tools to Santali speakers to read and write in their own script online.

**Key words:** *Ol Chiki script, typeface design, minority script, Santali language*

### 1. Introduction

The main aim of this paper is to share the experiences and knowledge gained by working on a typeface and input method design project for a minority script from India, in this case Ol Chiki, which is used to write Santali. This project was initiated by the Access to Knowledge programme at the Center for Internet and Society (CIS-A2K, whose mandate is to work towards catalysing the growth of the free and open knowledge movement in South

Asia and in Indic languages. From September 2012, CIS has been actively involved in growing the open knowledge movement in India through a grant received from the Wikimedia Foundation (WMF). The current focus of the CIS-A2K team spans over five language areas; Kannada, Konkani, Marathi, Odia, and Telugu, two community strengthening initiatives, and six stand-alone Wikimedia projects.

Today, there are over 6.2 million people living in four South Asian countries—India, Bangladesh, Nepal and Bhutan—that speak Santali. Even though Santali has been included in the eighth schedule of the Indian constitution, there has been little effort made to bring it to mobile phones, internet and mainstream media. Santali is written using several scripts including Bengali, Devanagari, Oriya, Latin, and OI Chiki. OI Chiki, the newest of these scripts, was developed between 1920-1940 by Pandit Raghunath Murmu to create a writing system that met all the orthographical needs of the Santali language and could unify its literature under one script. Pandit Murmu is said to have devised the shapes of OI Chiki letters based on objects commonly occurring in nature. The OI Chiki script has thirty letters, out of which six are vowels and the remaining twenty-four are consonants, along with six diacritics. A freely available, Unicode-compliant typeface can be the vehicle that takes the OI Chiki script to millions of people, enabling them to access information in their own language and contribute back to their shared repository of knowledge. In a project spanning about two years, an attempt was made to design an OI Chiki typeface that is consistent with the expectations of native users.

As a part of this project, the history and evolution of OI Chiki was studied and handwritten material in the script was collected to build a better understanding of how the letters look and interact with each other. Along with this, native readers and language experts were consulted throughout the design process to keep the project on the right track. The project set out to design a Unicode 7.0.0 compliant typeface family that supported the OI Chiki script in regular and bold weights, and an italic style. This font would be released under the SIL Open Font License. Along with the typeface family, input methods and keyboard layouts would be designed and a user manual for both would be created. This user manual would be complemented with a short research document describing the Santali language and OI Chiki script. Releasing this work as part of open standards will bring more power to the work. And we hope that it will help our aim of encouraging the Santali community to use the typeface and contribute to the Santali Wikipedia. This way, the next generation coming from the Santal homes would have the world's information on their fingertips, in their own language, and with open access.

The following sections contain an introduction to Santali and Ol Chiki, a brief account of the progress made in the design and input work so far, and the challenges faced in accomplishing them.

## 2. Santali and Ol Chiki Script

### 2.1 Santali

Santali belongs neither to the Indo-Aryan branch of the Indo-European language family (which includes Hindustani, or Hindi-Urdu, Bengali and Punjabi among others), nor to the Dravidian language family (which includes Tamil, Telugu, Malayalam and Kannada among others). Instead, Santali is an Austroasiatic language. It belongs to the Munda branch of the Austroasiatic family, along with languages such as Mundari, Ho and Sora Sompeng. The name Santali derives from the word Santal, used for the people who speak the language, and further Santal is believed to be a corruption of the common Bengali name for their tribe, *saotal* or *saotar*.<sup>1</sup> Santali was first studied by outsiders only in the middle of the 19th century, and foreign missionaries working in India were the first to do so.<sup>2</sup> Scholars who pioneered the study of Santali during that time were Rev. J. Phillips, who wrote *A Santali Primer* (1845), *Sequel to A Santali Primer* (1850) and *An Introduction to the Santal Language* (1852); Rev. L. O. Skrefsrud, who wrote *A Grammer [sic] of the Santal Language* (1873); and Rev. F. T. Cole, the author of *Santali Primer* (1969). Research and documentation of the script continued into the 20th century. Santali was discussed in detail in G. A. Grierson's mammoth publication, *Linguistic Survey of India* (1906) was the subject of P.O. Boddington's *Materials For A Santali Grammar Pt. I and II* (1922 and 1929 respectively).

According to the 2001 census conducted in India, a little under six million people in the country, i.e. 0.578% of the total, identify Santali as their mother tongue.<sup>3</sup> Another quarter of a million speakers of Santali live in neighbouring countries of Bangladesh and Nepal. Within India, most speakers live in Orissa, Bihar, Jharkhand, West Bengal, Assam and

---

<sup>1</sup> Grierson, G. A. (1906) *Linguistic Survey of India*, Vol IV. Munda & Dravidian Languages

<sup>2</sup> Ghose, Arun (2010) *Santali: A look into Santal Morphology*, Gyan Publishing House (New Delhi)

<sup>3</sup> [http://www.censusindia.gov.in/Census\\_Data\\_2001/Census\\_Data\\_Online/Language/Statement1.aspx](http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.aspx), last accessed on February 8, 2016

Tripura.<sup>4</sup> In 2003, Santali was added to the list of languages in the Eighth Schedule of the Indian Constitution, along with Dogri, Maithili and Bodo.

## 2.2 Scripts used for writing Santali

Up until Santali came to be studied in the 19th century, it was an oral language only, but it has been written using a number of scripts since. Rev. J. Phillips, for instance, used Bengali letters for his *An Introduction to the Santal Language* (1852). However, none of the Indic scripts used for writing Santali, whether Bengali or Devanagari or Odia, are considered adequate for representing all its sounds. One of the main drawbacks of using Indic scripts is that they are not equipped to represent checked consonants, which are peculiar to Munda languages, including Santali. In *The Munda Languages* (2008), Toshiki Osada phonetically describes checked consonants as—

‘first, the glottis is closed and the tongue or the lips simultaneously form an oral closure.

The tongue or lip position is the same as that of the corresponding normal stops. Then the glottal closure is released, which is optionally followed by nasal release and voicing.’

In addition, these scripts do not traditionally have a character for the glottal stop, which is used frequently in Santali. Finally, these scripts also fall short in their ability to represent Santali vowels. Santali uses eight or nine vowels that can be short or long, and nasalized, while Indic scripts provide only six vowels. Even with diacritic marks, the solution is not ideal. Latin, on the other hand, is able to represent checked consonants. But it, too, does not include a sign for the glottal stop in its basic or classical set of alphabets. This brings us to the Ol Chiki script, which was designed specifically to write Santali.

---

<sup>4</sup> <http://www.ethnologue.com/language/sat>, last accessed on February 8, 2016

### 2.3 OI Chiki Script

OI Chiki script was invented by Santali writer and scholar, Pandit Raghunath Murmu, who began work on its design in the 1920s and finished twenty years later. Even though it is known who created the OI Chiki script and when it was made, there are several myths in the community about its creation. One proclaims that the script came to be at the time when the Earth itself was created. Another says that the script was given as a divine gift to a learned man, i.e. Murmu.

The forms of the OI Chiki letters are said to be derived from nature, physical forms and everyday life of the Santals. The same is said about the pronounced sounds of the symbols. For instance, the pronounced sound *at* is depicted by a circle; its shape represents the earth and the meaning of the pronounced sound is the same. Similarly, the letter *ut* both looks and sounds like a mushroom. However, some discrepancies exist in the descriptions of these relationships provided by Pandit Raghunath Murmu himself and those by scholar, Rameshwar Murmu.

OI Chiki is written from left to right. It contains six vowels and twenty-four consonants, along with five basic diacritics (a sixth is made by the combination of two existing ones). The letters are organized in a matrix of 6 by 5, and the first letter of each row, i.e. the first column, are vowels, and the rest consonants. An additional three vowels are generated by using the diacritic *gahla tudag*, which can follow the vowels *la*, *laa* and *le*. The diacritic *mu tudag* nasalizes vowels, and *mu-gahla tudag*, which is made by the combination of the *mu tudag* and *gahla tudag* nasalizes a newly created vowel. The other three diacritics—*rela*, *phaarkaa* and *ahad*—are used as a length mark, glottal protector and for deglottalizing respectively. OI Chiki also has two punctuation marks, *mucaad* and double *mucaad*. Both are used in poetry, and only the former in prose to mark the end of a sentence. Latin punctuations comma, question mark, exclamation mark, parenthesis and quotation marks are also. Finally, OI Chiki, which also uses the decimal number system, has its own form of numerals.

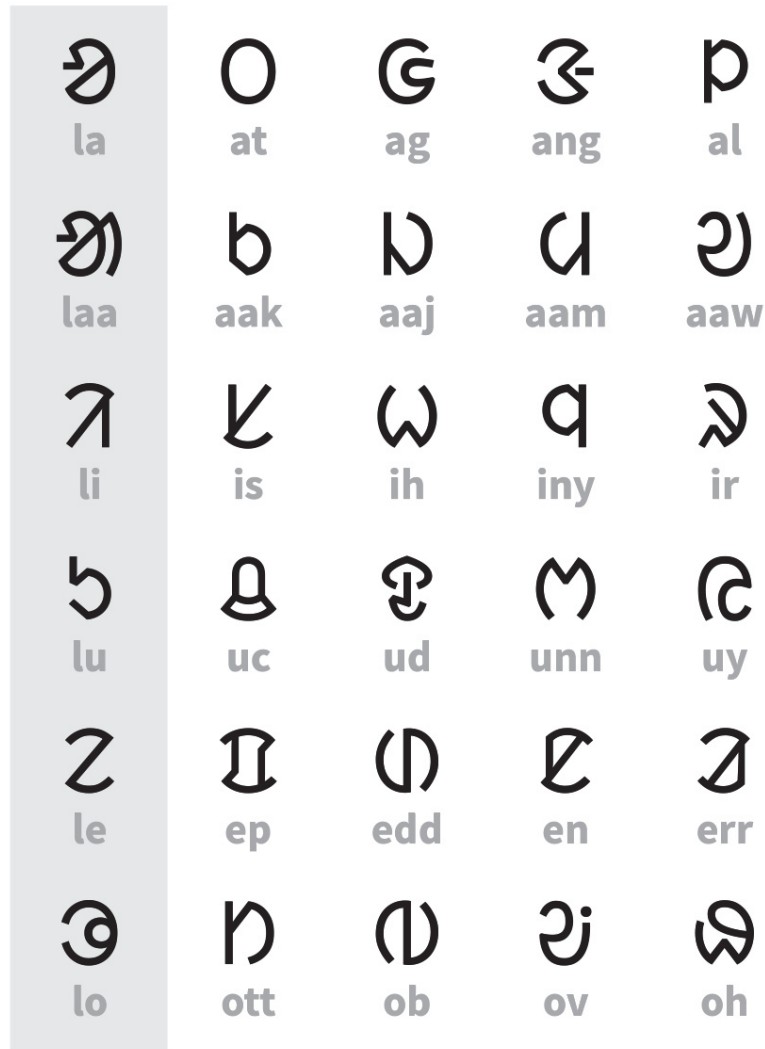


Figure.1 The OI Chiki alphabet set in the typeface being developed as a part of this project

OI Chiki letters have different forms for printing and handwriting. Unlike printing forms, handwriting forms cater to the need of writing fast. These forms are usually slightly slanted and narrower, and the shapes are simplified in a way that they can be written quickly and without lifting the writing instrument. In particular, in the handwriting form, the diacritic *ahad* is ligated with the four consonants it can follow, instead of being written separately.

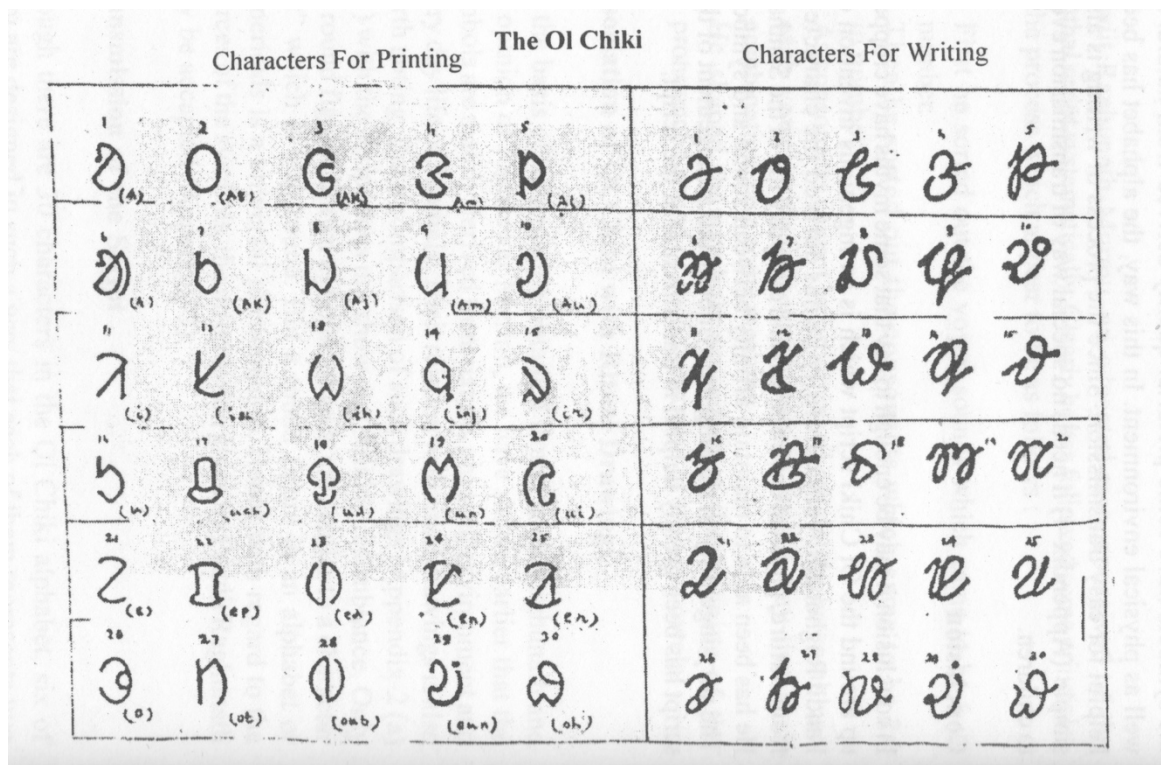


Figure.2 Printing and handwriting forms of the Ol Chiki alphabet (Source: Santhal Worldview (2001), Edited by Nita Mathur, Indira Gandhi National Center for the Arts)

### 3. Designing a typeface family

#### 3.1 Design Process

The process of designing this typeface family began with (a) the study of printed material or their images and (b) by practicing how to draw Ol Chiki letters. These two steps were essential in familiarizing one's self with an alphabet that was completely foreign. Ol Chiki is less than a century old, so there are no historical manuscripts to be consulted. Literacy in the script is quite limited, and the communities who speak Santali are marginalized. As a result, printed material in Ol Chiki is neither found in abundance, nor readily. Contact details for several publications, for instance, turned out to be dead-ends. With help of the limited material that could be collected, a first draft of the regular weight was designed. This design was then used as a starting point to understand how the script should be translated into a typeface. It was shared with native readers for their feedback on shapes and proportions. To make this critique more useful, inputs on specific matters were also sought. They were presented with, for instance, different positions of placement for diacritic marks *rela* and *phaarka*. Or how close or far they would expect to the space on the left of the *mucaad* to be. The *mucaad* had been found to be very close to the last

letter of the last word of a sentence in some of printed material that was consulted. Changes were made to the letter shapes according to this feedback.

### 3.2 OI Chiki Italic

An Italic style was needed for this typeface for emphasis within text, and for complying with Latin-based typographic conventions that have come to be expected. OI Chiki handwriting forms, as seen in writing manuals and in handwritten samples collected from native readers (during the feedback for the regular), have been used as reference to design this style. The basic letterforms have been designed and are awaiting feedback so the design can be improved.

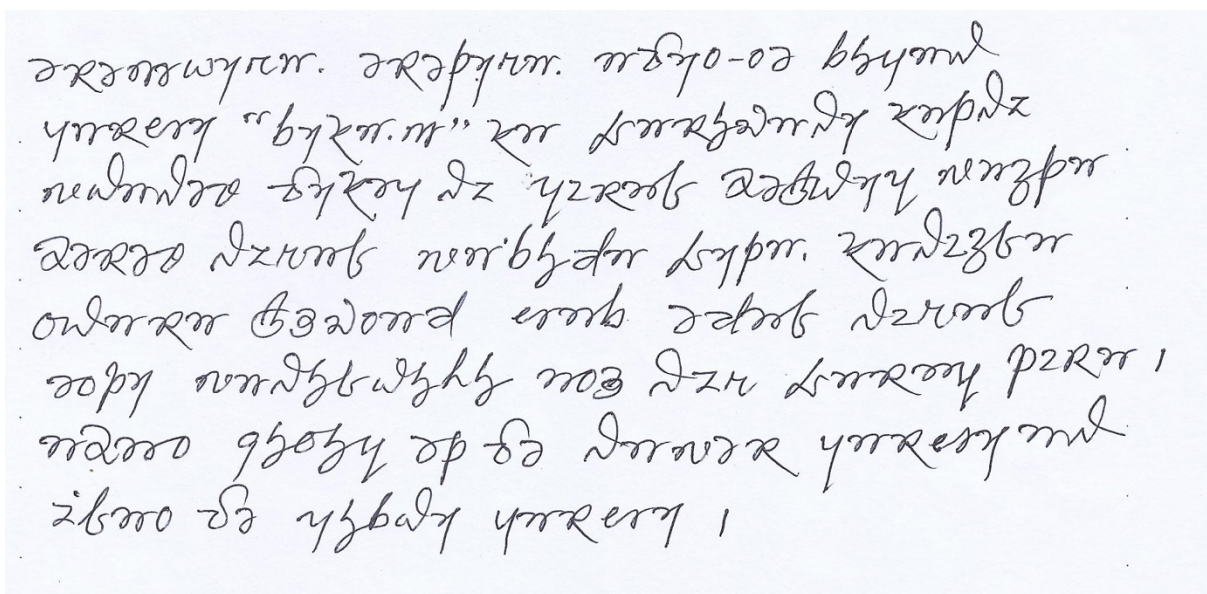


Figure.3 Handwriting sample collected from native readers of OI Chiki

### 4. Creating input methods

OI Chiki, so far, has been typed by the native speakers using two major input methods for computer; InScript, a standard keyboard for all the Indian languages<sup>5</sup>, and a phonetic method<sup>6</sup>. The first one being a universal method is freely-licensed and the second one is a freeware that is free to use but copyrighted with no source code open. There are a few

<sup>5</sup> [http://www.ilcd.in/SantaliO/GIST/santaliO\\_cd\\_2/linux/index.html](http://www.ilcd.in/SantaliO/GIST/santaliO_cd_2/linux/index.html) b) <https://fedorahosted.org/releases/i/n/inscript2/inscript2-20110721.tar.gz>, last accessed on February 8, 2016

<sup>6</sup> <http://wesanthals.tripod.com/id24.html>, last accessed on February 8, 2016



phonetic input tools available for mobile operating systems.<sup>7</sup> The non-availability of a free software and the favorability for phonetic over InScript by common users, as shared by the creators of the input method on We Santhals over email, pushed for working on both InScript and phonetic method “Sarjom baha”<sup>8</sup> and making the source code available under a free license for reuse and modification. The phonetic method *Sarjom baha* is inspired by another phonetic method designed by the We Santhals group.<sup>9</sup>



Figure.4 [OI Chiki InScript input layout](#) (Designed by [Pooja Saxena](#), modified by [Subhashish Panigrahi](#), [CC-BY-SA-4.0](#))

## 5. Challenges

The development of the typeface designing has been done in consultation with some of the language experts including a professor and a publisher. Their inputs in shaping the characters for better legibility and keeping the alphabet’s orthographic features intact has been taken into account fully. However, it has been challenging to have viewpoints of a larger community in both the development of the typeface and the input methods. Organizing consultations to review the typeface in a regular basis could not happen because lack of on-ground support of the native language community or any organization working on the language research. For the same reason the initial idea of building a Wikipedia editor community was also dropped.

<sup>7</sup> <https://play.google.com/store/apps/details?id=com.braingen.santalinotepad&hl=en>, last accessed on February 8, 2016

<sup>8</sup> [https://www.mediawiki.org/wiki/Help:Extension:UniversalLanguageSelector/Input\\_methods/sat-Sarjom\\_baha](https://www.mediawiki.org/wiki/Help:Extension:UniversalLanguageSelector/Input_methods/sat-Sarjom_baha), last accessed on February 8, 2016

<sup>9</sup> <http://wesanthals.tripod.com/id24.html>, last accessed on February 8, 2016