









# (A Souvenir-Cum-Edited-Research Book) Importance of Ayurveda, Yoga and Sanskrit in the Present Scenario



Chief Editor

Dr. Rakesh Kumar Mishra

Head of Department

Sanskrit Government Women's College, Puri Odisha

Organized by

Department of Sanskrit, Government Women's College Puri
in collaboration with

Central Sanskrit University

Consistent College

Supplied College

# Dattowerl Boered

#### Chief Editor: -

# Dr. Rakesh Kumar Mishra.

(Awarded by Indian Prime Minister in 2008 Gold Medalist from B.H.U, Varanasi, M.A., PhD) Head of the Department Sanskrit, Govt. Women's College, puri.

### Editors: -

# Dr. Nihar Ranjan Mishra.

(M.Phil., PhD)

Lecturer In Sanskrit, Govt. Women's college, puri.

# Dr. Kishore Kumar Mishra.

(M.Phil., PhD)

Lecturer in Sanskrit, Govt. Women's College, Puri.

Title: जिजासा

(A Souvenir-Cum-Edited Research Book)

Importance of Ayurveda, Yoga and Sanskrit in the

© Chief Editor

First Edition: 2023

ISBN: 978-81-959162-7-6

Price : INR 600/-

# Publisher by: Kishor Vidya Niketan

B-2/236-A-1, Bhadaini, Varanasi-1 contact- 9415996512, Uttar Pradesh

email: kvnpublisher@gmail.com website: www.kishorvidyaniketan.com

किशोर विद्या निकेतन का मोनो बीसवी सदी के मध्य में तेल पोट्ट बनाने वाले मराहूर कलाकार **काञ्चिलाल** की कल्पना का एक ख़रूप है, उनके समकालीन केदारशर्मा, कंटाले आदि थे।



# विषय सूपा

| <ol> <li>व्याकरणस्य महत्त्वं प्रयोजनाः</li> <li>यज्ञानां रोगोपशामकता</li> </ol>  | डॉ० प्रसेनणिन् माइति<br>मधुस्मिता डेका  | 7  |
|--|---|--|
| ्र आध्निकयुर्ग कथाराहरू ।  | अमित कुमार पण्डा  | 15   |
| मूल्यानां प्रासिद्धता  |   | 21   |
| मूल्याना जाराज ।<br>4. संस्कृतसाहित्यस्य सुक्तिनां अधुनातन समाजे आवश्यकता  |   | 3 3  |
| 4. संस्थारणायाः विवेचनम्<br>5. वैदिकवाङ्मये यज्ञस्य अवधारणायाः विवेचनम्  |   | 41   |
| 6. वेदेषु नैतिकशिक्षा  |   | 44   |
| 7. योगशास्त्रदिशा ध्यानासनप्राणायामानाम् आवश्यकता  | अन्नपूर्णा आचार्यः  | 46   |
| <ul> <li>क्ष्माजस्य हितं संस्कृते निहितम्</li> </ul>   | रू<br>खोनकदासः  |  |
| <ol> <li>रामायणस्य बालाख्य काण्डयोः पर्यावरणसचेतनता</li> </ol>   | बलभद्र कर्णः  |  |
| 10. वैदिकवाङ्मये पर्यावरणसचेतनताः एकमध्ययनम्   | सोम्यरञ्जन वडपण्डा  | 64   |
| 11. वैदिकवाक्येषु विश्वएकतायाः परिशीलनम्   | डॉ० किशोरकुमारमिश्रः  | 67   |
| 12. दार्शनिकं सामाजीकञ्च चित्रं महाभाष्ये  | विनोदः  |  |
| 13. साम्प्रतिकयुगे योगस्य आवश्यकता   | दीपकः साहुः   | 8.0  |
| क्ष्मिल्या उपदर्शी   |   |  |
| ज्या जोगटणेनाक्रमक्षिस्य उपपरे   | डॉ० ओमनारायण मिश्र  | 89   |
| 14. साम्प्रतिकयुगे योगदर्शनाक्रमाक्षस्य उपदर्शा  | डॉ० ओमनारायण मिश्र<br>डॉ० सुनेली देई 1  | 89   |
| <ul><li>14. साम्प्रतिकयुगे योगदर्शनाक्रमक्षिस्य उपदर्शा</li><li>15. समग्रविकासाय योगः सिवद्वयशिक्षानीतिश्च</li><li>16. साम्प्रतिककाले पाणिनीयव्याकरणस्य उपयोगिता</li></ul>   | डॉ० ओमनारायण मिश्र  | 89   |
| <ul> <li>14. साम्प्रतिकयुगे योगदर्शनाक्रमाक्षस्य उपदरता</li> <li>15. समग्रविकासाय योगः सिवद्वयिशक्षानीतिश्च</li> <li>16. साम्प्रतिककाले पाणिनीयव्याकरणस्य उपयोगिता</li> <li>17. योगदर्शने कर्मानुशीलनम्  English Articles</li> </ul>   | डॉ० ओमनारायण मिश्र<br>डॉ० सुनेली देई 1  | 89   |
| <ol> <li>समप्रतिकयुगे योगदर्शनाक्रमक्षिस्य उपदेशा</li> <li>समप्रविकासाय योगः सिवद्वयिशक्षानीतिश्च</li> <li>साम्प्रतिककाले पाणिनीयव्याकरणस्य उपयोगिता</li> <li>योगदर्शने कर्मानुशीलनम्         English Articles     </li> <li>Medical Science in Ancient India</li> <li>Utility of Yoga in Practical Life</li> <li>Emotional Intelligance: with reference to</li> <li>Emotional Practical Life</li> </ol>   | डॉ० ओमनारायण मिश्र<br>डॉ० सुनेली देई 1<br>डोली पण्डा 1<br>Prof. Rayi Prakash Arya   | 89<br>03<br>111<br>119   |
| 14. साम्प्रतिकयुगे योगदर्शनाक्रमाक्षस्य उपयस्ता  15. समग्रविकासाय योगः सिवद्वयशिक्षानीतिश्च  16. साम्प्रतिककाले पाणिनीयव्याकरणस्य उपयोगिता  17. योगदर्शने कर्मानुशीलनम्  English Articles  1. Medical Science in Ancient India 2. Utility of Yoga in Practical Life 3. Emotional Intelligance: with reference to some basic test of Sanskrit 4. Significance of Yoga in Daily Life 4. Significance of Yoga in Daily Life 5. Impact and Importance of TRATAKA in Present Scenario  Allen Glispong: The Poef as Yogi  Allen Glispong: The Poef as Yogi  Yoga for Physical, Mental and Social Wellbeing | डॉ० ओमनारायण मिश्र<br>डॉ० सुनेली देई 1<br>डोली पण्डा 1<br>Prof. Ravi Prakash Arya<br>Smt. Prabhati Nalini Devi<br>Dr. Manisha Mishra<br>Nishigandha Panda<br>Akarshit Mourya<br>Netranand Das | 89<br>03<br>111<br>119<br>137<br>150<br>159<br>167<br>178        |
| 14. साम्प्रतिकयुगे योगदर्शनाक्रमाक्षस्य उपदेशा  15. समग्रविकासाय योगः सिवद्वयशिक्षानीतिश्च  16. साम्प्रतिककाले पाणिनीयव्याकरणस्य उपयोगिता  17. योगदर्शने कर्मानुशीलनम्  English Articles  1. Medical Science in Ancient India  2. Utility of Yoga in Practical Life  2. Utility of Yoga in Practical Life  3. Emotional Intelligance: with reference to some basic test of Sanskrit  4. Significance of Yoga in Daily Life  5. Impact and Importance of TRATAKA in Present Scenario  Present Scenario  Allen Glispong: The Poef as Yogi  | डॉ० ओमनारायण मिश्र<br>डॉ० सुनेली देई 1<br>डोली पण्डा 1<br>Prof. Ravi Prakash Arya<br>Smt. Prabhati Nalini Devi<br>Dr. Manisha Mishra<br>Nishigandha Panda<br>Akarshit Mourya<br>Netranand Das | 89<br>03<br>111<br>119<br>137<br>150<br>159<br>167<br>178<br>182 |

साम्प्रदायिकयुगे योगस्य आवश्यकता

साम्प्रदाः युज् समाधौ, युज् संयमने, युजिन हरण्य गर्भः अनुशासनम्, असंप्रज्ञात समाधिः, अस्तेय, अपन् हिरण्य गर्भः अनुसार्धः, असंप्रज्ञात समाधिः, अस्तेय, अपिष्पिहः संप्रज्ञातसमाधिः, तन्त्रशास्त्रम्, ऋग्वेदः कैवल्यः न्यवृतिः चित्तभूमिः, सप्रशास्त्रम्, तन्त्रशास्त्रम्, त्रहग्वेदः कैवल्यप्राणिः, तन्त्रशास्त्रम्, त्रहग्वेदः कैवल्यप्राणिः, व्याप्यायः, ईश्वरप्रणिधानम्, समाधिः, तन्त्रशास्त्रम्, त्रहग्वेदः कैवल्यप्राणिः, यः, ईश्वरप्रणिषानगरः वित्रं जम्बुद्रीपं मनोरमं जीवितं मनुष्याणाम्'' इति गौतमबुद्धस्य के ंचित्रं जम्बुहाप पारतवर्षः नाम्नाः सुविदितः यत्र संस्कृतिः भारतवर्षः नाम्नाः सुविदितः यत्र संस्कृतिः भिक्षे वस्तुगत्या अयं भूखण्डः भारतवर्षः नाम्नाः सुविदितः यत्र संस्कृतिः भिक्षे वस्तुगत्या अयं भूखण्डः वस्तुगत्या अयं भूखण्डः वस्तुगत्या अयं भूषण्डः विविधाः उपायाः भिक्ता वस्तुगत्याः किताकर्षकी, मानवजीवनं सुशृङ्खचितं कर्तुं मनीषिणः विविधाः उपायाः किताकर्षकी, भानवजीवनं सुशृङ्खचितं अयं योगः न केवलं प्रमूप वित्ताकर्षकी, मानवजावन पुरुष्त । अयं योगः न केवलं शरीरसुर्यतायं किल्पतेषु उपायेषु 'योगः' अन्यतमः। अयं योगः न केवलं शरीरसुर्यतायं अ

इ।
संसारोऽयं दुःखपूर्णम् सांसारिकदुःखात् विमोचनाय तथा निःश्रेयसा षड्षुआस्तिक दर्शनेषु योगदर्शनस्य महत्त्वं नितरानुभूयते। अतः उच्यते-

''योगमात्रोत्परोमार्गो नास्ति नास्ति श्रुतौ स्मृतौ। शास्त्रेस्वन्येषु सर्वेषु शिवेन कथितः पुराः॥''

योगदर्शनस्य प्राणप्रतिष्ठा यद्यपि पतञ्जलिविरचित योगसूत्रादवगम्पेतः पतञ्जलिः योगदर्शनस्य आदिप्रवर्त्तकः न। अधिकन्तु ''योगियाज्ञवल्क्यः सिक् योगदर्शनस्य आदि प्रवक्ता उपदेष्टा वा स्वीक्रियते, पतञ्जलिनाऽपि स्वग्रन्यात्मेः

"अथ योगानुसासनम्<sup>2</sup>" इति।

अनुशासनम् इति शब्दप्रयोगात् योगभास्यटीकाकारेण वाचस्पतिमिश्रेण 🐲 निनु हिरण्यगर्भो योगस्य वक्ता नान्यः पुरातनः''। यद् वा भवतु आत्मक ोगसाधनम् अवश्यं करणीयम्।

ोग शब्दस्यार्थ:-

'युज् समाधौं'' 'युजिर् योगे' 'युज् संयमने'' चेति धातूनां भावं छ्यू र्र्ल यं योगशब्दः निस्पद्यते।

तत्रादां युज् समाधौ धातोः निष्पत्रः योगशब्दः समाधौ स्थिरता अर्थ प्रति ायामुक्तमस्ति-"यत्रोपरम्यते चित्तं निरुद्धं योगसेवयाट"। 6/20

ततः ''यजिर् योगे'' धातोः निष्पन्नः योगशब्दः समस्वरूपः पर्पा सम्बन्धत्वं द्योतयित प्रकाशयित वा। प्रमाणं तावत्-"समत्वं योग उच्यो"।

अनुनारं "युज संयमने भनो (महा थार भूवयति। "प्रथमं योगमंश्वरम्" १/६१८ अस्सान् अपि तु पारलांकिकः एव।

योगस्य स्वरूपम्-

त्तिः

युगे युगे जनाः दुःग्धविनाशाय मृखप्रात्रः जनानां मनिस शान्तिप्रदानाय पत्रज्ञालना याग्य "योगः चित्तवृत्तिनिरोधः" इति, योग-1/ चित्तवृतीनां निरोधेनैव योगस्यिमिंडः भ प्रमाण-पिपर्यय-विकल्प-निद्रा-स्मृतयः

क- प्रत्यक्षानुमान शब्द प्रमाणानि। ख-मिथ्याज्ञानं विपर्ययः।

ग-शब्दज्ञानानुपाती वस्तुशून्यः विकल घ-अभाव प्रत्ययालम्बना वृत्तिः निद्रा ङ-अन्भूतविषया सम्प्रमोषः स्मृतिः

आत्मनः वृत्ति-अभावात् भ्रमवशतः चि वृत्तयः कार्याणां सिद्धौ कारणानि भवन्ति। मनः चित्तवृत्तिनिरोधः **बहुक्ले**शकरः। परन्तु अभ्य

चित्तस्यास्य पञ्चअवस्थाः भूमयः। वा निद्वद्धः⁴ इति।

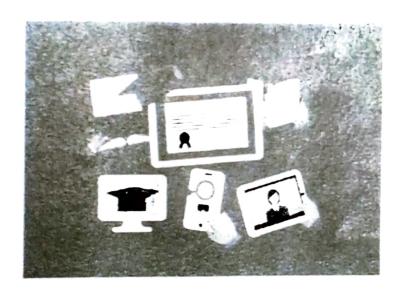
> \* तत्र चञ्चलत्वं क्षिप्तस्य स्वभावः सिद्धयति।

\* तमोगुणाधिक्यः मूढावस्था, अ अनन्तरं विक्षिप्तावस्था, अत्र

सत्वगुणाः विराजिताः तथापि \* एकामावस्था तु तादृशी अवस् किन्तु पूर्णतय न निरुद्धयते।

74 । जिज्ञासा





Editors
Prof. Ashutosh Biswal
Prof. R.C. Patel
Ms. Shama P. Ansari

Bore winds to sty

Published by

Geetanjali Nangia

Ashish Books

4435 367, Ansari Road, Darya Ganj,

New Delhi-110002 Phone: 011 23274050

e-mail; aphbooks a gmail.com

2022

Typeset by
Ideal Publishing Solutions
C-90, J.D. Cambridge School,
West Vinod Nagar, Delhi-110092

Primed at

BALAJI OFFSET

Navin Shahdara, Delhi-110032

#### Effectiveness of TLM in Munda Language for Developing Language Skills of MLE (Munda) Students in Sundargarh District, Odisha

Sri Garudeswar Day\*

#### ABSTRACT

Appropriate language of instruction, perfect teacher-punit communication, suitable curriculum and textbook, compatible school environment plays a vital role in the learning of tribal children. Odisha being a tribal-dominated state has the distinct feature of linguistic diversity. Sundargarh district is unique and something different from another part of the state concerning language and culture. For children belonging to Munda tribal community, Peer group communication, and social communication are restricted to their home language. They seldom get the opportunity to speak a language other than their mother tongue. But the language of instruction at school is the state language Odia. The context of learning as given in the curriculum and textbook is not suitable for these marginalized tribal students. Examples cited in the general textbook are difficult for their comprehension. As a result, they struggle with the content given in the textbook. Mother tongue-based primary education can enhance their learning ability. To ensure equity and quality education for tribal children, the Government of Odisha has 184 - The large of the set of the section of Posterior has an freelige I consecutive.

takes as ther tengue-hased multilingual education (MLT) for these . Add from Mother tongue-hased MCF tries to heldge the gap between mustly at again and state language Orbit. The curriculum of the MLL proof its coverages the broad objective of multilinguist education cliematic approach. Theme webs were developed from ..... is deficing, based on curricular provisions for grade-specific ... bieve bolistic learning goals. Alphabet charts, number are based big and small books, bilingual textbooks, picture divine a . ( from mother tongue to Odia), and grammar book ined and implemented in class. Approaches adopted in MEL at To develop (r)BICS(Basic interpersonal communication skill, and (ii)CALPS(Cognitive Academic Language Proficiency Skills, BICS is learned through experiential knowledge by the child in a locial context while CALPS aims to leach the skill of reading. writing. The present study is based on primary data to evaluate the effectiveness of these materials for developing language skills of MLE (Munda) students in Sundargarh district. Odisha

Aeywords: Multilingual Education, theme web, thematic approach, LLM

#### INTRODUCTION

Children are born with the ability to learn a language and the mother tongue is the first language the child comes in contact with Language plays a vital role in delivering quality education. Instruction through a language that learners don't speak has been called "submersion" (Skutnabb-Kangas, 2000, P-64) because it is analogous to holding learners underwater without teaching them how to swim. It is key to communication and understanding the classroom. In elementary classes, mother tongue-based instruction is important due to the following grounds.

- · It lays a solid foundation for learning
- It facilitates learning both academic subjects as well as the second language (1.2). At school, children can concentrate on the subject matter instead of struggling to understand the language of instruction as well.

<sup>\*</sup> Leacher Educator, DH.T. Sundargarh, Odisha



# TEACHER EDUCATION

1

21<sup>ST</sup> CET





# Edited by

DR. BIMAL CHARAN SWAIN
PROF. SUREKHA SUNDARI SWAIN
DR. RAJALAKSHMI DAS

Government College

# **KUNAL BOOKS**

4648/21, 1st Floor, Ansari Road,

Daryaganj, New Delhi-110002

Phones: 011-23275069, 9811043697,9868071411

E-mail: kunalbooks@gmail.com

Website: www.kunalbooks.com

Teacher Education in 21" Century

© Editors

First Published 2022

ISBN: 978-93-91908-46-1

[All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publisher].

The opinions and views expressed are exclusively those of the authors and in no way the editors or the publisher is responsible for the same.

Published in India by Prem Singh Bisht for Kunal Books, and printed at **Trident Enterprises**, Noida, (U.P.).

# EFFECTIVENESS OF ALTERNATIVE INTERNSHIP PROGRAMME (AIP) IN ACHIEVING TEACHING SKILLS BY STUDENT TEACHERS

Garudeswar Das

## Introduction

The COVID-19 pandemic has resulted physical closure of educational institutions and so also teacher education institutions. Regular physical learning of the students paralyzed in schools, colleges, universities and Teacher Education Institutions. Many educational institutions adopted digital device based blended learning for learning of their students. Both children of Government primary schools and student-teachers of teacher education institutes suffered a lot due to COVID-19 pandemic.

Parents at lower strata of economic ladder could not afford smart phone to their siblings in primary schools. Internet shadow areas were also a hurdle for this. Digital learning was not an easy task for small children. Primary school children were segregated from outer world and—confined to their home which resulted mental stress in them. The children deprived from digital or physical learning can tend to be out of school children or learning deficit children.

Teacher education institutions faced different type of problems. Though 80% of the student teachers had on-line accessibility to learning, learning of practical skills was not possible due to closure of schools. COVID-19 pandemic is an obstacle for internship of student-teachers undergoing teachers training programme.

Internship programme is compulsory for D.El.Ed. student-teachers. The



# **Domain Adaptation of Parts of Speech Annotators** in Hindi Biomedical Corpus: An NLP Approach

Pitambar Behera<sup>1\*</sup> and Om Prakash Jena<sup>2</sup>

Centre for Linguistics, SLL & CS, Jawaharlal Nehru University, New Delhi <sup>2</sup>Department of Computer Science, Ravenshaw University, Cuttack, India

#### Abstract

The envisaged research demonstrates the development of bio-medically annotated Parts of Speech (POS) corpus in Hindi. The study presents the adaptation of POS tagger trained in general domain corpus to automatically annotate the corpus of health domain. The tagger is trained with 200k word tokens applied from the ILCI (Indian Languages Corpora Initiative) data of mixed domains (in addition to 50k newswire tokens of biomedical data) which provides a satisfactory accuracy of 92%. When adapted and tested with the fresh data of the biomedical domain, the tagger registers an accuracy of 86.5%. In addition, the paper also focuses light on the resource-poor scenario of Hindi and other Indian regional languages in general domain and biomedical corpus in particular. Furthermore, the study provides a detailed account of the issues and challenges encountered pertaining to inter-rater reliability, domain adaptation of corpus, linguistics, and NLP (Natural Language Processing).

Keywords: Parts of Speech Annotation, NLP, ILCI, biomedical text processing, Hindi, resource-poor, domain adaptation

#### Overview 20.1

In the age of information revolution, the scientific research output of biomedical research is phenomenally and unprecedentedly developing in general and the knowledge base in particular is increasing and expanding rapidly [1]. The MEDLINE database in 2004 is one of the databases which contains 12.5 million records and receives 500k new and innovative citations each year [2]. Owing to the revolution of information, the biomedical research encounters information overload in the form of big data. To cater to this situation, NLP applications such as text mining, information retrieval, summarizing, and others have been developed. The boom in biomedical research in recent years has been excelled by the significant contribution of research and development in the fields of Artificial Intelligence (AI) and Natural Language Processing (NLP) in Computer Science. Therefore, one of the pivotal purposes of the text mining is to shift the responsibility of information overload from the

Om Prakash Jena, Alok Ranjan Tripathy, Ahmed A. Elngar and Zdzisław Polkowski (eds.) Computational Intelligence and Healthcare Informatics, (377-392) © 2021 Scrivener Publishing LLC

15BN - 978-11-19-8188-7

<sup>\*</sup>Corresponding author: pitambarbehera2@gmail.com



### A Qualitative Evaluation of Google's Translate: A Comparative Analysis of PBSMT and NMT Systems

#### Pitambar Behera

CL/SLL&CS, Jawaharlal Nehru University, New Delhi-110067, India

Abstract: The paper demonstrates the qualitative evaluation of the English to Indian languages Machine Translation Systems, namely PBSMT and NMT hosted on Google's Translate. This system is popularly known as Rosetta, formerly governed by the Phrase-based approach and is presently governed by the neural module of source and target languages. In this study, a model corpus set of 1k English sentences of cross-domain data has been applied considering various types of verbs as input text to evaluate the output of the online systems in Indian languages.

In order to evaluate the output text in a qualitative manner, the Intertranslator Agreement (IA) of three human translators has been considered with their scores on a five-point scale. The scores are calculated by the Fleiss' Kappa statistical measure with regard to comprehensibility and grammaticality on the basis of which error analysis and suggestions have been provided for improvement. In addition, the system has also been quantitatively evaluated on the basis of word error rate and sentence error rate. Furthermore, all the erroneous entities have been analyzed through computational typology. The strategy for evaluation is to evaluate the output text of Indian languages based on the five-point scale with scores that range from 0-4 where 0 refers to incomprehensible or ungrammatical, 1 = little meaning or disfluent, 2 = neutrality, 3 = comprehensible or grammatical and 4 suggests flawless in both cases.

**Keywords:** PBSMT, NMT, Google's Translate, MT, Indo-Aryan, NLP, Fleiss' Kappa

#### Overview

As discussed in Castilho et al. (2018), since the advent of the Machine Translation (MT) or automated translation, new methods, approaches and techniques have really created high expectations among researchers. On one hand, the qualitative approaches have

1SBN-978-81-942780-3-0

Cone 2 mugardary

Europe College



# lssues and Challenges in Developing Statistical POS Taggers for Sambalpuri

Pitambar Behera<sup>1\*</sup>, Atul Kr. Ojha<sup>2</sup> and Girish Nath Jha<sup>3</sup>

<sup>15</sup> Centre for Linguistics, <sup>23</sup>Special Centre for Sanskrit Studies Jawaharlal Nehru University, New Delhi, India {pitambarbehera2, shashwatup9k & girishjha}@gmail.com

Abstract. Low-density languages are also known as lesser-known, poorly-described, less-resourced, minority or less-computerized language because they have fewer resources available. Collection and annotation of a voluminous corpus for the purpose of NLP application for these languages prove to be quite challenging. For the development of any NLP application for a low-density language, one needs to have an annotated corpus and a standard scheme for annotation. Because of their non-standard usage in text and other linguistic nuances, they pose significant challenges that are of linguistic and technical in nature. The present paper highlights some of the underlying issues and challenges in developing statistical POS taggers applying SVM and CRF++ for Sambalpuri, a less-resourced Eastern Indo-Aryan language. A corpus of approximately 121k is collected from the web and converted into Unicode encoding. The whole corpus is annotated under the BIS (Bureau of Indian Standards) annotation scheme devised for Odia under the ILCI (Indian Languages Corpora Initiative) Project. Both the taggers are trained and tested with approximately 80k and 13k respectively. The SVM tagger provides 83% accuracy while the CRF++ has 71.56% which is less in comparison to the former.

Keywords: Low-density language: Parts of speech tagger: SVM: CRF++: Sambalpuri: Eastern IA language

#### 1 Introduction

Low-density languages have fewer resources in terms of the availability of voluminous corpus [1] for NLP applications. The unavailability of corpus for a low-density language proves to bear adverse impacts on its future NLP development. As rightly pointed out by Ostler [2], languages that lack active participation in the electronic media are doomed to be endangered in the forthcoming years. Most of these languages are either dialects or languages with no government recognition. As a result, the situations of these languages in South Asia in general and in Indic languages, in particular, are 'relatively bleak' [1]. Although India is a land of more than 6000 languages with five prominent diverse language families [3] only 22 are scheduled and the rest are fighting for their survival.

This paper is concerned with demonstrating the issues and challenges in developing statistical parts of speech taggers for a low-resource language nomenclatured as Sambalpuri or Kosli. The paper has broadly three objectives. Firstly, it highlights the issues in corpus collection with regard to non-uniform orthographic language standards and non-Unicode encodings of the written text. Secondly, it further attempts to bring out the issues in annotation having without any guideline. Finally, it demonstrates the challenges in developing statistical POS taggers for Sambalpuri owing to the typical, unobserved and language-specific linguistic nuances.

Human Language Technology. Challenges for Computer Science and Linguistics, Lecture Notes in Computer Science, vol. 10930. Springer, Cham

ISBN-978-3-319-9378)-6

Welling gige