# भारतीय भाषाओं के समानांतर शब्दावली का सर्वेक्षण एवं समीक्षा

# (Survey and Review of Parallel Glossary of Indian Languages)

**Pratiksha kumari, Prof.Chandrakant S.Ragit**

Research Scholar, Pro-vice-chancellor and Director

Center for Information and Language Engineering

Mahatma Gandhi Antarrastriya Hindi Vishwavidhyalaya Wardha, Maharashtra India

**सारांश:**

प्रस्तुत शोध-पत्र में भारतीय भाषाओं के समानांतर शब्दावली के सर्वेक्षण समीक्षा पर प्रकाश डाला गया है जो आज के शिक्षा जगत के लिए उपयोगी हैं। इस शोध-पत्र का उदेश्य है-प्रकृतिक भाषा संसाधन को विभिन्न भाषाओं के शब्दों को संग्रहीत कर व्यापक बनाना तथा शिक्षा जगत में त्रि-भाषाई फॉर्मूला को सुदृढ़ बनाना ,विलुप्त हो रही भाषाओ को संरक्षित करना और मशीनी अनुवाद को सुगम बनाना इत्यादि । उक्त उदेश्यों को ध्यान में रखते हुए भारतीय भाषाओं के लिए बहुभाषिक शब्दावली का निर्माण किया जाएगा जिसकी मदद से उपरोक्त सभी शैक्षिक उदेश्यों को पूर्ण करने में सहायता प्राप्त होगी । यह शोध पत्र एक तरह का सर्वे है बहुभाषिक शब्दावली के उपर जिसके माध्यम से विभिन्न प्रकार की भारतीय बहुभाषिक शब्दावली का समीक्षा किया जाएगा जो ऑनलाइन अलग- अलग तरह के पद्धतियों का उपयोग कर तैयार किया गया हैं।

**मुख्य शब्द :** प्रकृतिक भाषा संसाधन,मशीनी अनुवाद ,शिक्षा जगत ,त्रि भाषाई फार्मूला।

## परिचय :

सभ्यता और संस्कृति के प्रारंभ मे ही मानव जान गया था कि भाव के सही सम्प्रेषण के लिए सही अभिव्यक्ति आवश्यक है। सही अभिव्यक्ति के लिए सही शब्द का चयन आवश्यक है। और सही शब्द के चयन के लिए शब्दों के संकलन आवश्यक हैं। शब्दों और भाषा के मानकीकरण कि आवश्यकता समझ कर इस शब्दावली का निर्माण करना आवश्यक लगा। जिससे कि कुछ निहित क्षेत्र का शब्द संग्रह तीन भाषाओं में किया जाएगा और इसके विभिन्न पहलुओं को ध्यान मे रखते हुये इस शब्दावली का निर्माण किया जा रहा हैं।

ऑनलाइन शब्दावली के लिए बहुत से संसाधन मौजूद है जिसमें स्काइनेट ,वर्डनेट ,हिंदी शब्दकोश आदि की समीक्षा करने पर पता चला ऑनलाइन शब्दावली में अलग –अलग पद्धतियाँ उपलब्ध हैं जैसे रूल आधारित अभिगम ,ज्ञान आधारित अभिगम ,मशीन लर्निंग आधारित सामील हैं।

प्राकृतिक भाषा संसाधनके लिए यह बहुत उपयोगी होती है। सूचना प्रौद्योगिकी के क्षेत्र में क्रांति आने के बाद मानव विकाश के लिए सूचनाओं को सही जगह पे सार्थक रूप से संप्रेषित करने का काम इस तरह के शब्दावली से करना आसान होगा, अभी भी बहुत सारी भाषा असंगठित और अर्द्ध-संरचित रूप में है। आज के इस तकनीकी वाले इस दौर में इन भाषाओं को संगठित रूप में विकसित करना अति आवश्यक है।

## साहित्य पुनरवलोकन :

1. Barneva P. et.al (2002) के द्वारा "Multilingual Multimedia Electronic Dictionary For Children" यह bilingual form मे बच्चों के लिए तैयार किया गया electronic dictionary है जिसमे शब्द अर्थ के साथ स्रोत,उच्चारण तथा व्याख्या भी है । यह prototype type तथा user friendly dictionary है इसको बनाने के लिए advanced database compression technology तथा text searching algorithms का उपयोग किया गया।
2. Tsvetkov &Wintner, (2010) ने लघु समानांतर कार्पोरा (Small Parallel Corpora) से बहुशब्दीय अभिव्यक्तियों के विभिन्न स्वरूपों के प्रत्ययन एवंलक्ष्य भाषा में उनके अनुवाद हेतु एक सामान्य प्रविधि प्रस्तुत की गई। उनके द्वारा लघु द्विभाषीय कार्पसलिया गया तथा दोषपूर्ण (noisy) शब्द अलाइनमेंट में से बहुशब्दीय अभिव्यक्तियों का प्रत्ययन किया। इसके पश्चात एक वृहद एकभाषीय कार्पस से बहुशब्दीय अभिव्यक्तियों का चयन करने एवं उन्हें वरीयता प्रदान करने हेतु सांखिकीय विधि का प्रयोग किया।
3. Dr. B. Kamakoti (university of Annamalai) (2015) के द्वारा developed डिक्शनरी "Multilingual electronic dictionary for Tamil, Hindi, English language" यह रिसर्च प्रोजेक्ट U.G.C का major approval electronic dictionary हैं। इसमे Hindi equivalent to Tamil word की feasibilityजानने के उदेश्य से बनाया गया। और एकत्रित किए गए डेटा को Microsoft excel sheet मे enter किया गया और Unicode का उपयोग करके बदल दिया गया जिसका electronic version available था।
4. Priya viji et.al (2017) ने व्युत्पन्न जरनल जिसका शीर्षक "study on development of multilingual dictionary software"है यह शोध एक से ज्यादा भाषा का उच्चारण, समानार्थी शब्द और इससे संबन्धित विचार भी उपलब्ध कराता है। यह एक python-based E.M Dictionary है।इसमे कुल पाँच सेक्शन हैं।

**5 . Burke G Abbot (21 December 2018) "A Brief Sanskrit Glossary: A Spiritual Student's Guide to Essential Sanskrit Terms"**

(Swami Nirmalananda Giri) यह संस्कृत की शब्दावली है जिसमे संस्कृत के अध्यात्मिक शब्द का पूर्ण अनुवाद और व्याख्या है।

6 . Angelov krasimir (2020) "A Parallel Word Net for English, Swedish and Bulgarian

नें तीन भाषाओं ( अंग्रेजी, स्वीडिश और बुल्गारिया) को लेकर एक समानांतर स्रोत तैयार किया है जो मुख्य रूप से स्वीडिश तथा बुल्गारियन भाषा के उथान के लिए बनाया गया है, इस वर्डनेट को प्रिंस्टन वर्डनेट के साथ जोड़कर तैयार किया गया है। यह शाब्दिक तथा सिनसेट दोनों स्तर के लिए तैयार किया गया है । जिसमें स्रोत भाषा के शब्दों को लक्ष्य भाषा के समानांतर टहता सन्निकट शब्दों का अनुवाद किए हुये शब्द के साथ मिलान किया गया है। यह मशीनी अनुवाद और प्रकृतिक भाषा निर्माण के लिए बहुत आवश्यक है । यह संसाधन एक कोश के साथ आता है जहाँ प्रिंस्टन वर्डनेट के उदाहरणों का स्वीडिश और बुलगोरियाई में अनुवाद किया जाता है इसमें मौजूद ओपन- सोर्स संसाधन का उपयोग किया गया है

7. **Redkar Hanumant et.al** (2015) के द्वारा निर्मित शिर्षक "IndoWord Net Dictionary: An Online Multilingual Dictionary Using IndoWord Net" यह एक बहुभाषिक IndoWordNet Dictionary है जिसमें अनेकों प्रकार से शब्दों का व्यवस्थित रूप में सूचना प्रदान किया गया है । जैसे –सेंस आधारित, thesaurus based, शब्द आधारित और भाषा आधारित है । तथा इसके टूल के निर्माण में frontend के लिए, PHP, CSS, JAVASCRIPT और BACKEND के लिए MYSQL का उपयोग किया गया है। यह एक प्रकार का ऑनलाइन रेसोर्स तैयार किया गया है।

8 . **Pirkola A. et.al** (2001) University of Tampre, Finland "Dictionary- Based Cross-Language Information Retrieval: Problems, Method, and Research Findings"

इस पेपर में CLIR में प्रयोग किये गए पद्धति का उपयोग किया गया ,जिसमे pirkola के द्वारा query मॉडल पर चार भाषाओं को लेकरquery पर आधारित रिसर्च का कार्य किया गयाहैं। इस पत्र में structured queries को आधार मानकर शोध किया गया हैं।

इस पत्र में टेक्स्ट मटेरियलको इंटरनेट और दूसरी इन्फॉर्मेशन देने वाली जगहों से लिया गया है। जिसके लिए क्रॉस-लैंगवेज इन्फॉर्मेशन रिट्रीवल मुख्य रूप में महत्वपूर्ण रिसर्च पर प्रयोग हुआ। इसमें रिट्रीवल क्वेरी , रिट्रीवल डाक्यूमेंट्स से अलग उपस्थिती थी। इस पेपर की चार मुख्य बिन्दु इस प्रकार हैं।

1 सामान्य शब्दकोशों के अनुवाद के लिए untranslatable search key की सीमाओं पर कार्य

2 विभक्ति शब्दों के संस्करण पर।

3 वाक्यांश पहचान और अनुवादके लिए।

4 स्रोत भाषा और लक्ष्य भाषा में शाब्दिक अस्पष्टता।

इस पेपर में नई तकनीक का प्रयोग किया गया हैं,जिसके के द्वारा Translatable words को विस्तार करने के लिए विशेष तरह की शब्दकोशों में किया गया हैं विभक्ति शब्द को संभालने के लिए स्टेमिंग और रूपात्मक विश्लेषण किया गया है। फ्रेज अनुवाद के लिए POS टैंगिंग किया गया है। और गलत किए गए अनुवाद को मिटाने के लिए इस पेपर के शोध पीआर कार्य किया गया है।

Synonym sets के माध्यम से query structuring पर आधारित संरचना – शब्दकोश एक सरल और आवश्यक उपकरण है।

सामान्य शब्दकोशों और विशेष शब्दकोशों में समानांतर उपयोग प्रभावशीलता में सुधार करना है।

Proper noun के वर्तनी या विभक्ति में भिन्नता होने के कारण इसका उचित अनुवाद नहीं हो सकता है जिसके कारण समस्या होती है ।

यौगिक शब्द का विभक्ति के रूप में होने या विशेष रुपिम के द्वारा आपस में जुड़ना हो तो यौगिक को सही आधार के रूप में समझना चाहिए।

9. **Reddy and Hanumanthappa (2012**) "Dictionary Based Word Translation in CLIR Using Cohesion Method",

इंटरनेट पर टेक्स्ट डेटा की मात्रा दिन प्रतिदिन बढ़ती जा रही है क्योंकि की लोगो की उनकी प्रकृतिक भाषा में सूचना प्राप्त करने की आवश्यकताए बढ़ती ही जा रही है यही से CLRI मेथड सूचना प्राप्त करने की तकनीक का विकास होते रहा है इस पेपर में कुछ सामान्य अप्रोच का क्रॉस लिंगुअल इन्फॉर्मेशन रिट्रीवल मेथड में विकास किया जा रहा है डिक्शनरी पर आधारित अनुवाद इसका प्रारम्भिक अप्रोच है इस अप्रोच को बहुत आसान तरीके से इनके बीच में समन्वय स्थापित किया जा सकता है और वे एक अच्छा रिट्रीवल अनुवाद के आधार पर एक सर्च के तौर पर कार्य कर सकते है। इस डिक्शनरी में translation ambiguity पर कार्य किया गया है इस में query translation को अधिक महत्व दिया गया है।

10 . **Nofriansyah Dicky et.al** (2014) "Oxford Dictionary Application Use Eclipse's Editor of Android Open Source with Depth First Search Method",

मोबाइल फोन वर्तमान के जीवन शैली में एक अभिन्न भाग होता जा रहा है और प्रारम्भ में उपयोग में आने वाली फ़िक्स्ड – लाइन फोन से अलग भी है और मोबाइल फोन के लिए किये जाने वाले प्रोग्राम बहुत जल्दी विस्तार हुआ क्योंकि समय के अनुसार इसमें कई तरह के विविधताओं का प्रोग्राम बनाए जाने लगा। और इसी के विकास के साथ मोबाइल ऑपरेटर कंपनियाँ आने ग्राहकों को आकर्षित करने के लिए मार्केट में कई तरह के प्रोग्राम को लॉन्च करने लगे जैसे की मुल्टीमीडिया सर्विस ,गेम सर्विस इत्यादि। इस रिसर्च पेपर में depth –first search method का स्ट्रिंग matching algorithm का प्रयोग किया गया है जो की एंडरोइड अप्लीकेशन बेस्ड मोबाइल फोन के लिए एक नयी तकनीकी का प्रयोग हुआ है। एंडरोइड ऑपरेटिंग सिस्टम में विकसित एप्लीकशन में जावा प्रोग्रामिंग भाषा का प्रयोग हुआ है।

11. **Lew (2012)** "How can we make electronic dictionaries more effective? ने इस रिसर्च पेपर में इलेक्ट्रानिक डिक्शनरी को प्रभावी बनाने के लिये बताया गया है –एक डिक्शनरी में ऐसे गुण होने चाहिए की वो यूजर केआवश्यकता को पूर्ण कर सके और उसके द्वारा पूछे गए शब्दों का अर्थ कम से कम समय में दे सके। डिक्शनरी की डेटा अधिक से अधिक मात्रा में कम्प्राहेंसिव होनी चाहिए और साथ ही साथ लेक्सीकोग्राफी के डेटा होने से ही डिक्शनरी का अर्थ पूरा नहीं होता है। डिक्शनरी का मुख्य अर्थ होता है की वह किस तरह के यूजर को ध्यान में रख कर विकसित की जा रही है जैसे की स्कूल के बच्चों के लिए ,भाषा सिखने वालों के लिए ,कंटेंट लर्नर के लिए ,अध्यापक को शिक्षण कार्य में प्रयोग के लिए ,अनुवाद ,टुरिस्ट ,शोधार्थी इत्यादि के लिए किया जा रहा है।

डिक्शनरी का प्रयोग इनमें से किस लिए किया जा रहा हैं – लेखन के लिए ,टेक्स्ट के रिविजन,अनुवाद ,भाषा सीखने ,स्क्रेब्बल को खेलने के लिए ,इसका अधिक ध्यान दिया जाये।

लेक्सीकोग्राफी डेटा के प्रभावी रूप को एक्सेस करने –

डिक्शनरी में डेटा को लोकेट करने के लिए कई तरह के स्टेप्स और अलग अलग तरह की स्किल्स होती है और यह कार्य मेटल लेक्सीकोग्राफार्स करता है।

शीर्ष शब्द Headwordidentification : -

डिक्शनरी की सबसे बड़ी समस्या head word identification की होती क्योंकि की डिक्शनरी में नेटिव लैंगवेज़ में हेड वर्ड के नहीं होने पर समस्या उत्पन्न होती है इसलिए इसका निर्धारण करना बहुत ही आवश्यक होता है और इसके होने से ही टार्गेट भाषा के शब्दों को सर्च किया जा सकता है

मल्टीलींगुअल एक्सेस :- डिक्शनरी में बाकी शब्दों को जोड़ने के लिए मल्टीलिंगुअल एक्सेस करने की फासिलिटी होती है।

**12. Adda Martin et.al (2005) "Multilingual Dictionary".**

ने इस पेपर में स्पीच टेक्नोलोजी से संबन्धित बात की गयी है जिसमें बताया गया है की आज के समय में स्पीच सिस्टम बहुत ही विकसित रूप में फैल गया है स्पीच recognizer एक सेंसेटिव पर्फार्मेंस देता है। ये दो तरह का होता है एक acoustic और दूसरा उच्चारण के तौर पर होता है। यदि डिक्शनरी में साउंड का प्रयोग किया गया है तो नेटिव स्पीकर कभी भी दूसरी भाषा को नहीं समझ सकता है ,तो ऐसी स्थिति में स्पीच recognizer के द्वारा वो उसे आसानी से समझ सकता है और साथ ही साथ इसमें बहुभाषिक शब्दकोश के बारे में बताया गया है। जिसमें कैसे दो से अधिक भाषाओं के डिक्ट को बनाया जा सकता है और उसके लिए किस तरह की डेटा की आवश्यकता होती है।

## निष्कर्ष :

साहित्य पुनरवलोकन की सर्वेक्षण एवं समीक्षा करने के बाद यह ज्ञात होता कि 'भारतीय भाषाओं कि समानांतर शब्दावली' के क्षेत्र में भारतीय और विदेशी भाषाओं में शोधकर्ताओं के हर दृष्टिकोण से बहुत सारे अनुसंधान किए जा रहे हैं। प्रस्तुत शोध भारतीय भाषाओं कि समानांतर शब्दावली के , पत्र में –निर्माण से संबन्धित पूर्व में किए गए शोधकार्य और उनमें उपयोग कि जाने वाली विभिन्न ऑनलाइन शब्दावली तकनीकों और दृष्टिकोणों का संछिप्त विवरण प्रस्तुत किया गया है। यह विभिन्न भाषाओं में ऑनलाइन कोश होने के कारण बहुभाषिक ऑनलाइन शब्दावली और प्राकृत भाषा संसाधन के लिए उपयुक्त और तकनीकी के निर्माण में प्रभावी तरीकों से मदद करेगा। इसको संकर तथा रूल आधारित तरीकों से आज के दौर के आवश्यकतानुसार तैयार किया जा सकता हैं। भारतीय भाषाओं के समानांतर शब्दों को लेकर कोश का निर्माण किया जा सकता है जिससे भारत सरकार द्वारा चलायी गई योजना त्रि –स्तरीय भाषा का निरूपण प्रभावी तरीके से किया जा सकता हैं।

## संदर्भ-सूची:


- Granell. Ximo( ) "Multilingual Information Management: Information, Technology and Translations"
- Barneva.P. et.al (2002) के द्वारा "Multilingual Multimedia Electronic Dictionary For Children"
- Kamakoti.B(2015) "multilingual electronic dictionary for tamil,hindi,english languages" F. no : 5-175/2014 (HRP)
- Burke G Abbot (21 December 2018) "A Brief Sanskrit Glossary: A Spiritual Student's Guide to Essential Sanskrit Terms"
- Pirkola A. et.al (2001) University of Tampre, Finland "Dictionary- Based Cross-Language Information Retrieval: Problems, Method, and Research Findings"
- . Reddy and Hanumanthappa (2012) "Dictionary Based Word Translation in CLIR Using Cohesion Method"
- **Lew (2012) "How can we make electronic dictionaries more effective?**
- **Qin Jialun et.al (2002)** "Supporting Multilingual Information Retrieval in Web Application: An English-Chinese Web Portal Experiment"-. The University Of Arizona
- **Gearailt (2005) Dictionary Characteristics in Cross-language Information Retrieval,**
- Angelov krasimir (2020) "A Parallel Word Net for English, Swedish and Bulgarian
- Vijipriya. j, et.al (2017) . "A study on development of multilingual dictionary software", IJCERT 4(9): PP: 367- 372.
- Lee, K.-J., Kim, J.-E., & Yun, B.-H. ( .(2013Extracting Multiword Sentiment Expressions by Using a Domain-Specific Corpus and a Seed Lexicon. *ETRI Journal, (pp..(848-838*
- Moreno-Ortiz, A., Perez-Hernandez, C., & Del-Olmo, M. A. ( .(2013Managing Multiword Expressions in a Lexicon-Based Sentiment Analysis System for Spanish. Proceedings of the 9the Workshop on Multiword Expressions (MWE (2013, (pp. 1– .(10Atlanta, Georgia.

- Neha R.et. al ( 2012). IndoWordNet Application Programming Interface
- Tsvetkov, Y., & Wintner, S. ( .(2010Extraction of Multi-word Expressions from Small Parallel Corpora. *Coling :2010Poster Volume, (pp.1256–.(1264*
- पोल, प्रशांत . जुलाई2013 .(7कम्प्यूटर पर आईटी ने खोले भारतीय भाषाओं के रास्ते .*पांचजन्य*, (pp. .(31-30
- Jurafsky, D., & Martin, J. H. ( .(2009*Speech and Language Processing*. Pearson Education.
- Aggarwal, K.K., & Yogesh Singh. ( .(2008*Software Engineering*. New Delhi: New Age International (P) Limited.
- पंडित रामचंद पाठक () भार्गव हिंदी डिक्शनरी.प्रकाशक भार्गव बुक डिपो, चौक वाराणसी
- हशनवारिश (2013) भोजपुरी से मानक भाषा. राज्य शिक्षा शोध और एवं प्रशिक्षणपरिषद् द्वारा, विहार।
- राजनाथ भट्ट (2019) "संस्कृत –हिन्दी –कश्मीरी त्रि-भाषीय शब्द सूची"
- नीरन अरुणेश et.al (2018) "भोजपुरी-हिंदी-अंग्रेजी शब्दकोश"
- श्री अरविन्द et.al (2009) 'भोजपुरी-हिंदी-अंग्रेजी लोक शब्दकोश', केंद्रीय हिंदी संस्थान आगरा
- गुप्त, ओमप्रकाश .(1960) .*मुहावरा मीमांसा*.परिषद-राष्ट्रभाषा-बिहार :पटना .
- प्रसाद, धनजी .(2011) .*भाषाविज्ञान का सैद्धांतिक, अनुप्रयुक्त एवं तकनीकी पक्ष*.प्रिय साहित्य सदन :दिल्ली .
- प्रसाद, धनजी .(2012) .*सी*.प्रकाशन संस्थान :नयी दिल्ली .*शार्प प्रोग्रामिंग एवं हिंदी के भाषिक टूल्स* .
- पाण्डेय, पृथ्वीनाथ .(2015) .*मानक सामान्य हिंदी*..लि .प्रा (इ) अरिहंत पब्लिकेशंस :मेरठ .
- तिवारी, भोलानाथ .(2009) .*भाषा विज्ञान प्रवेश एवं हिंदी भाषा* नयी .दिल्ली.किताबघर प्रकाशन :
- गुप्ता, एस .(2015) ..पी .*अनुसंधान संदर्भिकासम्प्रत्यय -, कार्यविधि एवं प्रविधि*.शारदा पुस्तक भवन :इलाहाबाद .
- सिंह, सूरजभान .(2000) .*हिंदी का वाक्यात्मक व्याकरण*.साहित्य सहकार :दिल्ली .
- ओझा, त्रिभुवन .(1994) .*हिन्दी में अनेकार्थता का अनुशीलन*.विश्वविद्यालय प्रकाशन :वाराणसी .
- तिवारी, भोलानाथ एवं किरण बाला .(1992) .*हिंदी भाषा की आर्थी संरचना*.साहित्य सहकार :दिल्ली .
- कपूर, बदरीनाथ.लोकभारती प्रकाशन :इलाहाबाद .*हिंदी मुहावरे और लोकोक्ति कोश* .*(*2005) .
- तरुण, हरिवंश .(2009) .*मानक हिन्दी मुहावरा लोकोक्ति कोश*.प्रकाशन संस्थान :नयी दिल्ली .
- सिंह, राजवीर .(2012) .*सरस्वती मुहावरा कोश*.सरस्वती विहार :नोएडा .
- पाण्डेय, भवदेव. ( .(2011*हिंदी मुहावरा कोश*. दिल्ली.वाणी प्रकाशन :

- तिवारी, भोलानाथ, शशिप्रभा एवं किरणबाला. ( .(2003*वृहत हिंदी मुहावरा कोश* :दिल्ली .साहित्य सहकार.
- शर्मा, सुनील .(2010) .*मेधावी मुहावरा कोश*.मेधा बुक्स :दिल्ली .
- सिंह, मृत्युंजय कुमार. ( .(2011*छात्रोपयोगी मुहावरा कोश*.कुनाल प्रकाशन :दिल्ली .