ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

# संकर अभिगम आधारित मशीन अनुवाद प्रणाली
# (Hybrid Approach Based on Machine Translation System)

प्रो. हनुमान प्रसाद शुक्ल
अधिष्ठाता
भाषा विद्यापीठ
महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा

***संक्षिप्त (Abstract) :***

*प्रस्तुत शोध-पत्र के माध्यम से संकर आधारित अभिगम का प्रयोग करते हुए मशीन अनुवाद प्रणाली के निर्माण कार्य हेतु एक संगणकीय मॉडल की प्रक्रिया को प्रस्तुत किया गया है। इसके अंतर्गत यह भी बताया गया है कि हिंदी एवं अन्य भारतीय भाषाओं के संदर्भ में मशीन अनुवाद के निर्माण कार्य हेतु व्याकरणिक-कोश एवं कार्पस संकलन आवश्यक प्रक्रिया है। मशीन अनुवाद प्रणाली के निर्माण कार्य में आने वाली सभी प्रकार की समस्याओं का निदान भी निहित है। वस्तुत: मशीन अनुवाद की समस्या में सबसे बड़ी समस्या संदिग्धार्थकता की आती है। जिसके निराकरण हेतु अर्थीय संजाल, अर्थीय डाटाइंडेक्स आदि जैसे महत्वपूर्ण अनुप्रयोग का निर्माण कर किया जा सकता है। अत: व्याकरणिक-कोश निर्माण, कार्पस संकलन एवं सांस्कृतिक डाटासेट के माध्यम से एक मशीन अनुवाद प्रणाली का निर्माण किया जाएगा।*

***मुख्य-शब्द*** *: मशीन अनुवाद, व्याकरणिक-कोश, कार्पस संकलन, संकर आधरित अभिगम आदि।*

## 1. प्रस्तावना (Introduction) :

आज का युग अभियांत्रिकी का युग है। जिससे मानव द्वारा संपन्न किए जा रहे, समस्त कार्य न सिर्फ सुगम हुए; अपितु कम समय में ज्यादा-से-ज्यादा कार्य करने की क्षमता का भी विकास हुआ है। आधुनिक तकनीकी के माध्यम से मानव स्वयं के भावों, विचारों, संदेशों, सूचनाओं और उपलब्धियों आदि को अतिशीघ्र दूसरों तक प्रेषित करने में सक्षम हो पाया है। इस तकनीकी के युग में यदि मशीन में प्राकृतिक भाषा की समझ स्थापित कर दिया जाए, तो मशीन भी मानव की तरह व्यवहार करने में सक्षम हो पाएगी। जिससे मशीन के द्वारा समस्त मानवीय कार्य आसानी के साथ संपन्न किए जा सकते हैं। भाषा अभियांत्रिक लगातार इस विषय में अपना मत जाहिर करते हुए कहते हैं कि मशीन को कृत्रिम-बुद्धि से युक्त करना आवश्यक है। मशीन को इतना बुद्धिमान बना दिया जाए कि मशीन मानव की तरह ही व्यवहार कर सके। इसलिए यहां पर कृत्रिम-बुद्धि पर भी संक्षिप्त परिचय देना आवश्यक

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

हो जाता है। कृत्रिम-बुद्धि एक ऐसा ज्ञान क्षेत्र है, जिसके अंतर्गत मशीन को तर्क करने, सीखने, किसी भी समस्या का समाधान करने, स्वयं निर्णय लेने एवं विचार करने आदि के योग्य बनाया जाता है। इसे इस तरह भी समझा जा सकता है कि एक ऐसी विशेषज्ञ प्रणाली (Expert System) का निर्माण करना है, जिसके द्वारा मानव एवं मशीन के बीच न सिर्फ संवाद स्थापित किया जा सके; बल्कि मानव के आज्ञानुसार सौंपे गए, सभी क्रियाकलापों को मशीन के द्वारा बड़ी ही सजगता व विद्वता के साथ संपन्न कर सके। यह तभी संभव हो सकता है, जब मानव एवं मशीन के बीच संबंध स्थापित हो सके, जहाँ पर प्राकृतिक भाषा संसाधन अपनी केंद्रीय भूमिका निभाती है।

अनुवाद कार्य एक वृहद संकल्पना है, जिसमें स्रोत-भाषा के पाठ को लक्ष्य-भाषा के पाठ में परिवर्तित किया जाता है। यह परिवर्तन सिर्फ शब्द, वाक्य के विश्ले षण/पुर्नगठन तक ही सीमित नहीं होता है, बल्कि इसमें भाषा की अर्थीय संरचना बहुत ही महत्वपूर्ण भूमिका अदा करती है। जब तक स्रोत-भाषा के अर्थतत्व लक्ष्य-भाषा में समाहित नहीं होते हैं, तब तक अनुवाद अधूरा माना जाता है। इस समस्या को केंद्रित करते हुए ही मानवीय अनुवाद की प्रक्रिया को पूरा किया जाता है। जब यही कार्य स्रोत-भाषा के पाठ/वाक को मशीन के द्वारा लक्ष्य-भाषा के पाठ/वाक में परिवर्तित किया जाता है, तो उसे मशीन अनुवाद की संज्ञा दी जाती है। जब बात मशीन अनुवाद की आती है, तो स्रोत-भाषा के अर्थीय संरचना की गंभीरता को समझना ज्यादा महत्वहपूर्ण हो जाता है। यह हम सभी जानते हैं कि मशीन के पास आपनी कोई विचार प्रक्रिया, सोचने या तर्क करने की क्षमता नहीं होती है। मशीन को मानव द्वारा प्रदान किए गए भाषिक-सामग्री का विश्लेषण, संसाधन एवं संश्लेषण कर व्याकरणिक नियम-गुच्छों एवं सांख्यकीय विधियों के प्रयोग द्वारा प्रदान ज्ञान तक ही सीमित होता है। इन्हीं ज्ञान का प्रयोग कर मशीन को दक्ष बनाया जाएगा एवं मशीन अनुवाद प्रणाली को विकसित किया जाएगा, जिसके फलस्वरूप हिंदी से अन्य भारतीय भाषाओं एवं इसके विपरीत क्रम में विभिन्न, साहित्य को उपलब्ध किया जा सके।

## 2. मशीन अनुवाद अभिगम (Approach for Machine Translation):

मशीन अनुवाद प्रणाली के निर्माण कार्य में एकाधिक विधियां कार्यरत हैं, जिन्हें निम्न में समझा जा सकता है :

### i. नियम आधारित (Rule Based):

ऐसा अभिगम है, जिसके अंतर्गत भाषा का भाषावैज्ञानिक आधार पर विश्लेषण कर नियम-गुच्छ को प्रतिपादित किया जाता है, जिसे संगणक में कार्यान्वित कर भाषा का संसाधन सहजता के साथ किया जा सके। वर्तमान समय में भाषा का संसाधन नियम आधारित किया जा रहा है, संभवत: सबसे पहला नियम आधारित भाषा संसाधन प्रणाली का निर्माण क्लेन एवं सिम्पसन (1963) ने किया था। जिसके अंतर्गत हाथों के द्वारा प्रतिपादित नियमों का

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

गुच्छ एवं त्रुटियों को संभालने के लिए एक छोटी सी शब्द-कोश का निर्माण किया गया था। किसी भी भाषा के संपूर्ण घटक भाषा के संदर्भ आधारित नियमों का प्रतिपादन किया जाता है। इस अभिगम सबसे बड़ी कमी है कि भाषावैज्ञानिक आधार पर नियमों को प्रतिपादित एवं कार्यान्वित करने हेतु ज्यादा मेहनत एवं भाषावैज्ञानिक पृष्ठभूमि की आवश्यकता होती है।

### ii. सांख्यकीय आधारित (Statistical Based) :

ऐसा अभिगम है, जिसके अंतर्गत भाषा को भाषावैज्ञानिक आधार पर विश्लेषण की आवश्यकता नहीं होती है। इसके अंतर्गत भाषा के घटकों के मध्य संभाव्यता के आधार पर भाषा का विश्लेषण किया जाता है। चूँकि किसी भी भाषा को एक निश्चित नियम में बांधना अत्यंत मुश्किल एवं कठिन कार्य है। इसलिए एक ऐसे अभिगम की आवश्यकता महसूस हुई, जिसमें भाषा के घटकों की संभाव्यता के आधार पर विश्लेषण एवं संश्लेषण किया जाए। वर्तमान समय में सांख्यकीय आधारित अभिगम को नियम आधारित अभिगम से ज्यादा अच्छा माना जा रहा है । जिसके मुख्य कारक बड़े संग्रहित डाटा स्रोत है, जिसमें डाटा की आवृत्ति एवं अन्य सांख्यकीय डाटा, वृहद कार्पोरा (मुख्यत: एनोटेटेड), द्विभाषी सामानांतर कार्पोरा आदि का प्रयोग किया जाता है। उक्त साधन का प्रयोग करके बड़े-से-बड़े कार्पोरा के शब्द, पदबंध एवं वाक्य के आधार पर एक नमूना (Pattern) तैयार किया जाता है, जिससे समान अनुक्रम वाले समस्त शब्द, पदबंध एवं वाक्य) का संसाधन आसानी के साथ किया जा सके।

### iii. संकर आधारित (Hybrid Based):

ऐसा अभिगम, जिसके अंतर्गत नियम आधारित एवं सांख्यकीय आधारित दोनों अभिगमों का प्रयोग किया गया हो, संकर आधारित अभिगम कहलाता है। यह सबसे अच्छा अभिगम माना गया है क्योंकि भाषा को न पूर्णरूपेण नियमों में ही बांधा जा सकता है एवं न ही पूर्णरूपेण संभाव्यता निकाली जा सकती है। वर्तमान समय में समस्त भाषा संसाधन प्रणालियों में संकर आधारित अभिगमों का प्रयोग आपेक्षित माना जा रहा हैं अथवा किया जा रहा है। अत: संकर आधारित अभिगम में आवश्याकतानुसार भाषावैज्ञानिक नियमों का एवं विस्तृत कार्पोरा के डाटा की संभाव्यता दोनों का प्रयोग किया जाना आवश्यक हो जाता है। जिससे एक स्वस्थ्य भाषिक प्रणाली का निर्माण किया जा सके, जो औसत दर्जे का परिणाम देने में सफल हो सके।

प्रस्तुत मशीन अनुवाद प्रणाली के अंतर्गत कार्य कर रहे सभी भाषिक अनुप्रयोगों का निर्माण कार्य संकर अभिगम का प्रयोग करते हुए किया जाएगा। चूँकि हिंदी एवं अन्य भारतीय भाषा के भाषिक घटकों के विश्लेषण हेतु पूर्णत: न ही भाषावैज्ञानिक नियमों के निष्पादन एवं सांख्यकीय विधि के द्वारा किया जा सकता है। इसलिए सभी मशीन अनुवाद प्रणाली सहित

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

अन्य भाषिक-अनुप्रयोगों के निर्माण कार्य में भाषावैज्ञानिक एवं सांख्यकीय आधारित दोनों अभिगमों का प्रयोग किया जाएगा।

## 3. सांख्यकीय प्रविधि :

प्रस्तुत मशीन अनुवाद प्रणाली के निर्माण कार्य में सुपरवाइज्ड मशीन लर्निंग के आवश्यक विभिन्न एल्गोरिदम का प्रयोग किया जाएगा। डीप लर्निंग के रूप में आर एन एन एल्गोरिदम का प्रयोग किया जाएगा। इन दोनों ही सांख्यकीय प्रविधि के लिए आवश्यकतानुसार स्वत: की पाइथन लाइब्रेरी का भी निर्माण किया जाएगा। मशीन लर्निंग एवं डीप लर्निंग एल्गोरिदम हेतु पूर्व में ही कुछ महत्वपूर्ण पाइथन लाइब्रेरी का निर्माण किया जा चुका है। अतएव आवश्यकतानुसार इन लाइब्रेरी का उपयोग किया जाएगा। मशीन अनुवाद में प्रयुक्त होने वाले भाषिक प्रणाली के निर्माण कार्य हेतु पाइथन लाइब्रेरी का निर्माण किया जाएगा, जिसकी सहायता से भाषावैज्ञानिक नियमों एवं सांख्यकीय प्रविधि दोनों को एक साथ प्रयुक्त किया जा सके।

## 4. संगणकीय व्याकरणिक रूपवाद :

प्राकृतिक भाषा संसाधन में भाषा संरचना को विश्लेषित एवं संश्लेषित करने हेतु पूर्व निर्धारित व्याकरणिक रूपवाद (Grammatical formalism) के रूप में चॉमस्की का पदबंध संरचना व्याकरण (Phrase Structure Grammar) के सिद्धांत पर कार्य किया जाएगा। विशेष परिस्थिति में आवश्यकतानुसार अन्य व्याकरणिक रूपवाद का प्रयोग किया जाएगा। भाषा वैज्ञानिक नियमों के निष्पादन भारतीय भाषाओं का गहराई से अर्थीय घटकों के साथ विश्लेषण करने के पश्चात् कंप्यूटेशनल एल्गोरिदम का निर्माण किया जाएगा।

5. **उद्देश्य (Objective) :** मशीन अनुवाद प्रणाली (हिंदी एवं अन्य भारतीय भाषा के संदर्भ में) का निर्माण करने हेतु निम्न) भाषिक अनुप्रयोगों का निर्माण किया जाएगा :

I. स्वनिम विश्लेषक/संश्लेषक (Phonological Analyzer/Synthesizer)
II. रूपिम विश्लेषक/संश्लेषक (Morphological Analyzer/Synthesizer)
III. पदबंधीय विश्लेषक/संश्लेषक (Phrasal Analyzer/Synthesizer)
IV. वाक्य विश्लेलषक/संश्लेषक (Sentence Analyzer/Synthesizer)
V. प्रोक्ति विश्लेषक/संश्लेषक (Discourse Analyzer/Synthesizer)
VI. अर्थीय विश्लेषक/संश्लेषक (Semantic Analyzer/Synthesizer)
VII. अर्थीय संजाल (Semantic-net)
VIII. शब्द-भेद टैगर (POS Tagger)
IX. नामपद अभिज्ञानक (Name Entity Recognizer)
X. चंकर (Chunker)

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

XI. पद-विच्छेदक (Parser)
XII. मानक वर्तनी जाँचक (Standard Spell Checker)
XIII. मानक व्याकरण जाँचक (Standard Grammar Checker)

6. **लक्ष्य (Aim) :**

1. प्राकृतिक भाषा संसाधन के सभी अनुप्रयोगों का निर्माण कार्य हेतु। जिनकी सूची निम्न में संलग्न है :

i. ई-भाषा शिक्षण (E-Language Learning)

ii. पाठ-से-वाक प्रणाली (Text-to-Speech System)

iii. वाक-से-पाठ प्रणाली (Speech-to-Text System)

iv. वाक-से-वाक प्रणाली (Speech-to-Speech System)

v. पाठ सारांशक (Text Summarization)

vi. प्रश्नउत्तरीय प्रणाली (Question answering System)

vii. अन्य भाषिक अनुप्रयोग (Other Language Application)

2. एक प्रविष्टि की प्रयोग स्तर पर सभी अर्थ को प्रदर्शित करने हेतु। जिसमें ऐसे अर्थ भी मिलेंगे जो अन्यत्र कहीं और अथवा अन्य शब्द-कोश में उपलब्ध नहीं हैं।

3. हिंदी एवं अन्य भारतीय भाषा की वर्तनीगत शुद्धता/ मानकीकरण (शब्दों, वाक्यों एवं व्याकरणिक दृष्टिकोण में) हेतु। जिसे सर्वमान्य रूप में घोषित किया जा सके।

4. हिंदी एवं अन्यं भारतीय भाषा के वाक्य-प्रयोग को सुरक्षित करने हेतु। वाक्य-प्रयोग के सुरक्षित होने से जब कभी भी वाक्य–प्रयोग का इतिहास लिखा जाएगा, तब यह व्याकरणिक-कोश अपनी बहुत ही महत्वपूर्ण भूमिका निभाएगा।

5. हिंदी एवं अन्य भारतीय भाषा के भाषिक विश्लेषण हेतु। हिंदी भाषा का वाक्य-प्रयोग सहित पूर्ण भाषिक विश्लेषण अभी तक कहीं भी नहीं मिलेगा, जिसका आधार प्रारूप यह व्याकरणिक-कोश प्रदान करेगा।

6. हिंदी एवं अन्य भारतीय भाषा के ई-भाषा-शिक्षण के निर्माण कार्य में।

7. हिंदी एवं अन्य भारतीय भाषा के शब्द–वर्गीय रूप में चिह्नित कार्पस (Tagged Corpus) निर्माण हेतु।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

8. हिंदी एवं अन्य भारतीय भाषा के संगणकीय व्याकरण के निर्माण हेतु।

9. हिंदी एवं अन्य भारतीय भाषा के मानक ध्वानि को सुनने एवं समझने हेतु।

10. हिंदी भाषा के शब्दों को अर्थ से कभी-कभी नहीं समझा जा सकता है, तब यह चित्र के द्वारा आसानी से समझा जा सकता है। हिंदी क्रियाओं को चलचित्र (एनीमेशन) के माध्यम से बहुत ही सरलता से समझा जा सकता है। इन सभी चित्रों एवं चलचित्रों के माध्यमम से चित्र कार्पस (Image Corpus) एवं चलचित्र कार्पस (Animation Corpus) कार्य हेतु।

## 7. **मशीन अनुवाद की समस्या एवं चुनौतियाँ (The Problems of Machine Translation) :**

1. संगणकीय भाषावैज्ञानिक नियमों का निष्पादन

2. संदिग्धार्थकता की समस्या

    i. स्वनिमिक संदिग्धार्थकता

    ii. रूपमिक संदिग्धार्थकता

    iii. पदबंधीय संदिग्धार्थकता

    iv. वाक्यीय संदिग्धार्थकता

    v. अन्वादेशिक संदिग्धार्थकता

    vi. अर्थीय संदिग्धार्थकता

3. अर्थीय विश्लेषण/संश्लेषण की समस्या

## 8. **मशीन अनुवाद का महत्व (Importance of Machine Translation) :**

एक समय ऐसा भी था, जब मानव किसी कविता, कहानी, उपन्यास आदि साहित्य का अनुवाद कर अन्य भाषाओं में साहित्य उपलब्ध कराता था एवं 'स्वांत: सुखाय' की अनुभूति करता था। लेकिन आज का समय पूर्णत: बाजारीकरण का समय है। बाजार विभिन्न प्राकृतिक भाषाओं पर आश्रित हो चुका है। आज सामाजिक, सांस्कृातिक, राजनैतिक, व्यापारिक एवं वाणिज्यिक आदि के उद्देश्यों की पूर्ति हेतु अनुवाद कार्य आवश्यक एवं अनिवार्य प्रक्रिया है। बिना अनुवाद किए भाषिक विभिन्नता की दूरी समाप्त नहीं की जा सकती है। अत: इसे संगठित व्यापार कहना ज्यादा उचित होगा। संगठित व्यापार की पूर्ति हेतु अनुवाद की माँग बढ़ गई है। यह माँग इतनी ज्यादा हो चुकी है कि मानवीय अनुवाद के द्वारा कार्य पूर्ण नहीं किया जा सकता। इस परिस्थिति में मशीन अनुवाद ही एक ऐसा

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

विकल्प है, जिसकी सहायता से कम समय में ज्यादा-से-ज्यादा अनुवाद कार्य किया जा सकता है। मानवीय अनुवाद एक धीमी व ज्यादा परिश्रम करने वाली प्रक्रिया है। आज के समय में नई-नई सूचनाएँ एवं ज्ञान-विज्ञान का विकास किया जा रहा है, जिसे विभिन्न भाषाओं में माध्यम से विकसित ज्ञान को वैश्विक-पटल पर सर्वत्र पहुँचाया जा रहा है। इस ज्ञान को वैश्विक-पटल पर पहुँचाने का एकमात्र जरिया अनुवाद ही है। अब मानवीय अनुवाद इतनी जल्दी तो संभव है नहीं। अतएव मशीन अनुवाद का सहारा लेना उचित जान पड़ता है।

अत: यहां पर बड़ी सहजता एवं स्पष्टता के साथ कहना आवश्यक हो जाता है कि अनुवाद कार्य की आवश्यकता वैश्विक-पटल पर अनिवार्य हो चुकी है। जिसका एक प्रमुख विकल्प मशीन अनुवाद ही है। मशीन अनुवाद की जरूरत वैश्विक-जगत के सभी क्षेत्र (समाज, राजनीति, व्यापार, संस्कृति, वाणिज्यिक आदि) में व्याप्त है। इस मशीन अनुवाद का निर्माण कर बड़ी त्वरित गति में अनूदित पाठ की कमी को पूरा किया जा सकता है।

## 9. भाषिक-सामग्री (Language Resources) :

आज का युग मशीन का युग है। हम सभी जानते हैं कि मशीन किसी भी कार्य को बड़ी सरलता एवं त्वरित गति से परिणाम प्रदान करती है। मानवीय जन-जीवन की सभी दैनिक कार्यों को मशीन के द्वारा संपन्न किया जा रहा है, जिसमें अनुवाद जैसे जटिल कार्य भी मशीन के द्वारा सफलतापूर्वक संपन्न किया जा रहा है। मशीन के द्वारा मानव की तरह अनुवाद कार्य पूर्ण करने हेतु मशीन को भी मानव की तरह ही बुद्धिमान बनाना आवश्यक हो जाता है। मशीन के बुद्धिमत्ता का प्रतीक मानव की भांति कार्य को संपन्न करना है। यह तभी संभव है, जब मशीन में मानवीय भाषा की समझ स्थापित कर दी जाए। मशीन में मानवीय भाषा की समझ स्थापित करने हेतु प्राकृतिक भाषा का संसाधन सबसे अहम हो जाता है। प्राकृतिक भाषा संसाधन के सभी अनुप्रयोगों द्वारा अपेक्षित परिणाम प्राप्त करने हेतु भाषिक-सामग्री का संकलन एवं नियमबद्ध प्रारूप तैयार करने पर बल दिया जाता है। मशीन अनुवाद प्रणाली हेतु समानांतर रूप से एकाधिक भाषा के भाषिक-सामग्री का विश्लेषण एवं संश्लेषण करना आवश्यक प्रक्रिया है। इन्हीं भाषिक-सामग्री के आधार पर संगणकीय-कोश, समानांतर संगणकीय व्याकरण एवं अन्य रूपवाद का निर्माण किया जाता है, जिनके माध्यम में मशीन में भाषा की समझ को भी स्थापित किया जा रहा है।

मशीन अनुवाद के कार्य क्षेत्र में सरकारी संस्थाओं के अलावा बड़ी-बड़ी व्यावसायिक कंपनियां कार्य रही हैं, जिसमें भाषिक-सामग्री का अहम योगदान है। मशीन अनुवाद के क्षेत्र में कार्यरत सभी कंपनियों पर यदि एक नजर डाली जाए, तो सर्वप्रथम हमें यही परिणाम प्राप्त होता है कि सभी कंपनियों के पास भाषिक-सामग्री बहुतायत मात्रा में उपलब्ध है। इन्हीं भाषिक-सामग्री के आधार पर ही कंपनियों के भाषिक अनुप्रयोग कार्य रहे हैं और एक

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

औसत दर्जे का परिणाम भी प्रदान करने में सक्षम हैं। ये सभी व्यावसायिक कंपनियां उपयोगकर्ता से ही भाषिक सामग्री को प्राप्त/संग्रह करती हैं और इन्हीं सामग्री के आधार पर परिणाम भी प्रदान करती है। ये बड़ी-बड़ी व्यावसायिक कंपनियां गणितीय फंक्शयन का प्रयोग कर औसत दर्जे का परिणाम प्रदान करने में सक्षम तो होती जा रही हैं, लेकिन भाषाविज्ञान की दृष्टिकोण से सैद्धांतिक एवं कंप्यूटेशनल भाषावैज्ञानिक विश्लेषण नहीं के बराबर कर पा रहीं हैं अथवा अपने ज्ञान को स्वयं तक सीमित करके रखी हुईं हैं। अब इस परिस्थिति में शैक्षणिक संस्थानों का यह उत्तरदायित्व बनता है कि भाषाविज्ञान के नए दृष्टिकोण एवं गणितीय (सांख्यकीय) विधि दोनों को एक साथ एक मंच में अनुप्रयुक्त करते हुए व्याकरणिक रूपवाद का निर्माण करें, जिसे पूर्णत: कंप्यूटर में स्थापित किया जा सके। यह कार्य आधुनिक कंप्यूटेशनल भाषावैज्ञानिकों एवं भाषा-अभियांत्रिकों के ऊपर निर्भर करता है कि इस कार्य को कितनी तीव्र गति से आगे की ओर बढ़ा सकते हैं। चूँकि हिंदी एवं अन्य भारतीय भाषाएँ मुक्त शब्द-क्रम वाली भाषाएँ अतएव इन्हें सिर्फ यह कह देना कि भाषावैज्ञानिक नियमों में बाँधकर किसी एक निश्चित परिणाम तक पहुँचा जा सकता है; पर्याप्त नहीं है। किसी भी भाषा को पूर्णत: नियमों में बांधना एक जटिल और मुश्किल भरा कार्य है। अब यहां एक ऐसी विधा का भी विकास किया जाना आवश्यक महसूस हो रहा है, जिसमें भाषावैज्ञानिक नियमों के साथ-साथ गणतीय (सांख्यकीय) विधि का भी प्रयोग किया जाए। इस दोनों ही विधियों को एकसाथ प्रयोग कर निश्चित ही उच्च-गुणवत्तापूर्ण मशीन अनुवाद एवं अन्य भाषिक अनुप्रयोग का निर्माण किया जा सकता है।

प्रस्तुत मशीन अनुवाद प्रणाली (हिंदी एवं अन्य भारतीय भाषाओं के संदर्भ में) एवं इसमें प्रयुक्त होने वाले सभी भाषिक अनुप्रयोगों के निर्माण कार्य हेतु भाषिक-सामग्री अपनी महत्वपूर्ण भूमिका निभाती है। चूँकि मशीन अनुवाद प्रणाली किसी एक भाषा के विश्लेषण एवं संश्लेषण से तो संभव नहीं है, अतएव समानांतर रूप से एकाधिक भाषाओं के समानांतर भाषिक-सामग्री की आवश्यकता होती है। अब यदि यहां पर यह कहा जाए कि बिना भाषिक-सामग्री के मशीन अनुवाद का निर्माण किया जाए, तो अतिसंयोक्ति ही होगी। मशीन अनुवाद हेतु स्रोत-भाषा एवं लक्ष्य-भाषा के भाषिक-सामग्री का निर्माण/संग्रह समानांतर रूप से किया जाना महत्वपूर्ण कार्य होता है, जिसके माध्यम से ही मशीन अनुवाद (पाठ-से-पाठ एवं वाक्-से-वाक्) का निर्माण कार्य किया जाना संभव होगा। अत: हिंदी एवं अन्य भारतीय भाषा के समानांतर भाषिक-सामग्री का निर्माण एवं संकलन निम्न रूपो में किया जाएगा :

### i. **कार्पस चयन/निर्माण (Corpus Selection/Creation):**

संगणकीय व्याकरणिक-कोश निर्माण एवं भाषिक अवयवों के मध्य संभाव्यता निकालने हेतु समानांतर कार्पस का चयन एवं कार्पस की अनुपलब्धता की स्थिति में समानांतर कार्पस का

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

निर्माण भी किया जाएगा। यह कार्पस हिंदी एवं अन्य भारतीय भाषा में उपलब्ध विभिन्न विषय साहित्य से किया जाएगा। जिसका विवरण निम्न प्रकार है :

i. हिंदी एवं अन्य भारतीय साहित्य
ii. सिनेमा साहित्य
iii. पत्रकारिता साहित्य
iv. सामाजिक विज्ञान साहित्य
v. विधि साहित्य
vi. शिक्षा साहित्य
vii. भारतीय संस्कृति साहित्य
viii. प्रबंधन साहित्य
ix. विज्ञान साहित्य
x. कंप्यूटर विज्ञान साहित्य
xi. भाषाविज्ञान साहित्य
xii. अनुवाद साहित्य
xiii. अन्य साहित्य

चूँकि हिंदी एव अन्य भारतीय भाषा का कार्पस काफी विस्तृत एवं विशाल रूप में प्राप्त होगा, इसलिए संगणकीय व्यारकरणिक-कोश निर्माण हेतु अर्थीय दृष्टिकोण के क्रम से भी साहित्य/कार्पस का चयन महत्वपूर्ण होगा, इसलिए यहाँ पर निम्न रूप में साहित्य /कार्पस का चयन किया जाएगा :

I. अभिधार्थ प्रकट करने वाले साहित्य/कार्पस

a. विभिन्न विषयक शोध-ग्रंथ (प्रयोजन-मूलक साहित्य)

b. निबंध, कथा, कहानी इत्यादि।

II. लाक्षणार्थ प्रकट करने वाले साहित्य/कार्पस

a. उपन्यास, आलोचना इत्यादि।

III. व्यंग्यार्थ प्रकट करने वाले साहित्य/कार्पस

a. व्यंग्य।

i. उपर्युक्त वर्गीकरण का क्रम कार्पस/साहित्य चयन हेतु मात्र इसलिए किया गया है, क्योंकि ज्यादातर अभिधार्थ उपर्युक्त क्रम में ही प्राप्त होने की संभावना अधिक होती है। ऐसा बिल्कुल भी नहीं है कि अभिधार्थ वाली साहित्यिक रचनाओं में लाक्षणार्थ एवं व्यंग्यार्थ अथवा

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

इनके विपरीत क्रम में अर्थ प्राप्त नहीं होंगे। अगर प्राप्त होते हैं, तो उन अर्थो के प्रयोग को भी शामिल किया जाएगा।

ii. वाक्य-प्रयोग की मानकता की दृष्टि से प्रमाणिक एवं ज्यादातर पुराने प्रकाशन की पुस्तकों/कार्पस को चयन/निर्माण का आधार माना जाएगा।

**ii. संगणकीय व्याकरणिक-कोश का प्रारूप (The Format of Computational Grammatical-Dictionary) :**

मशीन अनुवाद प्रणाली के निर्माण कार्य को पूर्ण करने हेतु एक ऐसे संगणकीय व्याकरणिक-कोश का निर्माण किया जाना भी अपेक्षित है, जिसके माध्यम से मशीन अनुवाद में प्रयुक्त होने वाले सभी अनुप्रयोगों का निर्माण कार्य किया जा सके। भाषा के संरक्षण, संवर्धन, अभिवृद्धि एवं वैश्विक स्तर पर स्थापित करने के उद्देश्य से भी यह व्याकरणिक-कोश अपनी महत्वपूर्ण भूमिका निभाएगा। अभी तक जितने भी कोश-ग्रंथों का निर्माण किया गया है, उनमें व्याकरणिक सूचना, व्युत्पत्ति, शाब्दिक अर्थ एवं लाक्षणिक अर्थों तक ही सीमित हैं। ऐसे कोश बहुत ही कम या फिर न के बराबर हैं, जिनमें एक प्रविष्टि संबंधी सभी सूचना विद्यमान हों। अब इस परिस्थिति में हिंदी एवं अन्य भारतीय भाषाओं हेतु समानांतर एक ऐसे संगणकीय व्याकरणिक-कोश की परिकल्पना अपेक्षित मानी जाती है, जिसके द्वारा हिंदी भाषा के सभी भाषिक घटक (पद, पदबंध, वाक्य , मुहावरा/लोकोक्ति, अर्थ, वाक्, चित्र एवं चलचित्र आदि) समाहित हो सकें। ये कार्य तभी संभव है; जब मूल साहित्यिक कार्पस-प्रयोग की दृष्टि से संगणकीय व्याकरणिक-कोश का निर्माण किया जाए।

iii. **मानकता (standardization):**

संगणकीय व्याकरणिक-कोश निर्माण का कार्य करने हेतु सर्वप्रथम भाषिक-विश्लेषण की दृष्टिकोण से भाषिक अवयवों के मानकीकरण की समस्या उत्पन्न होती है। मानकीकरण की समस्या के निवारण हेतु सर्वप्रथम यह जानकारी जुटा लेना आवश्यक हो जाता है कि मानकीकरण की प्रक्रिया क्या है? मानकीकरण एक ऐसी प्रक्रिया है, जिसमें किसी भी भाषा के शब्दों को एकरूपता/शुद्धता/मानकता पर विशेष बल दिया जाता है। एक ही शब्द की वर्तनीगत एकाधिक विकल्पों का प्रयोग किया जाना इस बात का द्योतक है कि भाषा के शब्द या पद का मशीन की दृष्टिकोण से एक निश्चित प्रारूप तय नहीं किया जा सका है। प्राकृतिक भाषा संसाधन के अंतर्गत न सिर्फ शब्द/पद बल्कि वाक्यों का गठन भी व्याकरणिक रूप से शुद्ध एवं मानक होना चाहिए। हालाँकि मानकीकरण की प्रक्रिया भाषावैज्ञानिकों के द्वारा हमेशा से ही विवादग्रस्त रही है और अंतत: यह मान लिया जाता रहा है कि मानव द्वारा प्रयोग किए जा रहे भाषिक इकाईयों को ही यथाउचित प्रयोग किया जाए। यह एक हद तक सत्य है कि मानव द्वारा व्यवहरित शब्दों को ज्यादा तवज्जों दिया जाए, लेकिन यह मशीन के लिए जरूरी नहीं है।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

उक्त कार्य को पूर्ण करने हेतु कंप्यूटेशनल व्याकरण की आवश्यकता होती है, न कि सामान्य व्याकरण की। सामान्य व्याकरण मानव द्वारा व्यवहरित भाषा को नियमबद्ध एवं विश्लेषण करके सिद्धांत प्रतिपादित करता है जबकि कंप्यूटेशनल व्याकरण भाषा का विश्लेषण एवं नियमबद्ध करके मशीन में स्थापित करता है। जब बात मशीन में भाषा को स्थापित करने की आती है, तो भाषा को नियमों में बांधना और नियमों के आधार पर भाषा के सभी घटकों का प्रारूप तय करना होता है। इस परिस्थिति में मानकीकरण की प्रक्रिया अपनी अहम भूमिका निभाती है कि मानवीय व्यवहार में प्रयोग होने वाली भाषा के सभी घटकों का निश्चित प्रारूप तय किया जा सके। भाषा के शब्द/पद को मानक बनाना मुश्किल भरा कार्य भी है, अत: इस कार्य को पूर्ण करने हेतु विद्वत-जनों के साथ विचार-विमर्श आवश्यक होता है। हिंदी एवं अन्य भारतीय भाषाओं के वर्तनीगत मानकीकरण संबंधी एकरूपता/शुद्धता/मानकता इत्यादि का संस्कृत, हिंदी एवं उर्दू के विषय-विशेषज्ञों से गहन विचार-विमर्श के पश्चात् निर्धारण का निर्णय लिया जाएगा। हिंदी भाषा में प्रयुक्त निम्न यूनीकोड आधारित देवनागरी लिपि के वर्णमाला को मानक माना जाएगा।

**I. स्वर (Vowel) :**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| अ | आ | इ | ई | उ | ऊ | ऋ | ए | ऐ | ओ | औ |

तालिका क्र. 01

**II. मूल व्यंवजन(Consonant) :**

| | | | | |
|---|---|---|---|---|
| क | ख | ग | घ | ङ |
| च | छ | ज | झ | ञ |
| ट | ठ | ड | ढ | ण |
| त | थ | द | ध | न |
| प | फ | ब | भ | म |
| य | र | ल | व | |
| श | ष | स | ह | |
| ड़ | ढ़ | | | |

तालिका क्र. 02

III. संयुक्त व्यंजन (Combine-consonant) :

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

| | | | |
|---|---|---|---|
| क्ष (क् + ष) | त्र (त् + र) | ज्ञ (ज् +ञ) | श्र (श् + र) |

तालिका क्र. 03

IV. संविधान के अनुच्छेद 343 (1) के अनुसार संघ के राजकीय प्रयोजनों के लिए प्रयुक्त होने वाले अंकों का रूप भारतीय अंकों का अंतरराष्ट्रीय रूप को मान्यता दे दिया गया है, अत: इस कोश में भी अंको के अंतरराष्ट्रीय रूप का ही प्रयोग किया जाएगा।

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |

तालिका क्र. 03

V. संयुक्त व्यंजन के रूप में पंचमाक्षर के बाद सवर्गीय चारों वर्णों के एकरूपता एवं खोज की सुविधा के अनुसार अनुस्वार का ही प्रयोग कर प्रविष्टि बनाई जाएगी। जैसे- गङ्गाजल को 'गंगाजल' के रूप में लिखा जाएगा।

VI. यदि पंचमाक्षर के बाद किसी अन्य वर्ग का कोई वर्ण आए तो पंचमाक्षर अनुस्वार के रूप परिवर्तित नहीं किया जाएगा, यथा- अन्य, उन्मुख इत्यादि।

VII. हिंदी के शब्दों में उचित यथास्थान चंद्रबिंदु का प्रयोग अनिवार्यत: किया जाएगा।

VIII. अरबी-फारसी शब्दों के मानक वर्तनी उर्दू-कोश विषय-विशेषज्ञों के द्वारा तय किया जाएगा। जैसे : गजल (ग़जल, गज़ल, ग़ज़ल)।

IX. क, त और फ/फ़ का संयुक्ताक्षर बनाते समय दोनों वर्णों के हुक को हटाकर संयुक्ताक्षर बनाया जाएगा। जैसे संयुक्त (संयुक्त नहीं), पक्का (□□□□□ नहीं) इत्यादि।

### iv. **व्याकरणिक-कोश की विशेषताएँ (Characteristic of Grammatical-Dictionary):**

हिंदी एवं अन्य भारतीय भाषाओं के समानांतर इस व्याकरणिक-कोश की अपनी कुछ भिन्न विशेषताएँ हैं, जो निम्नवत् हैं :

1. **मानकता :** हिंदी एवं अन्य भारतीय भाषाओं का प्रयोग विभिन्न माध्यमों द्वारा शब्द/पद एवं व्याकरणिक दृष्टि से भी अनेक रूपों में लिखा जाता है। इस कोश के माध्यम से विश्व-भर में व्याप्त अमानकता को एक मानकता के सूत्र में पिरोया जाएगा।

**2. प्रविष्टि का चयन :** हिंदी एवं अन्य भारतीय भाषाओं के मूल साहित्यिक कार्पस-प्रयोग की दृष्टि से प्रविष्टि का चयन किया जाएगा, जो अन्य शब्द-कोश से बिल्कुल ही अलग होगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**3. उच्चारण रोमन संकेत :** हिंदी एवं अन्य सभी भारतीय भाषाओं के प्रविष्टि का समानांतर उच्चारण रोमन संकेत उच्चारण-ईकाई (Syllabic unit) के प्रारूप में ही लिखा जाएगा।

**4. आई.पी.ए. :** हिंदी एवं अन्य सभी भारतीय भाषाओं के प्रविष्टि का आई.पी.ए. प्रारूप में लिखा जाएगा।

**5. प्रविष्टि की प्राप्त वर्तनीगत अन्य रूप :** हिंदी एवं अन्य भारतीय भाषाओं के शब्दों को एकाधिक वर्तनीगत रूप में भी लिखा जाता है। इसप्रकार के वर्तनीगत रूपों को अलग से लिखा जाएगा।

**6. शब्द-वर्ग :** मूल साहित्यिक कार्पस-प्रयोग के आधार पर प्रविष्टि के शब्द-वर्ग को दिया जाएगा। ऐसी भारतीय भाषा जिसका कार्पस/साहित्य उपलब्ध नहीं हैं; हिंदी भाषा के समानांतर पाठ का अनूदित कर प्रयोग किया जाएगा।

**7. व्याकरणिक कोटि :** हिंदी एवं अन्य भारतीय भाषाओं के सभी प्रविष्टि के व्याकरणिक-कोटियों का भी निर्धारण किया जाएगा। इन व्याकरणिक कोटियों की सहायता से ही भाषा-संसाधन के सभी नियम-गुच्छ का निर्माण कार्य किया जाएगा।

**8. व्युत्पत्ति :** अर्थीय विश्लेषण हेतु हिंदी एवं अन्य भारतीय भाषाओं के शब्द/पद को तत्सम, तद्भव, देशज एवं आगत के आधार पर व्युत्पत्ति दी जाएगी।

**9. व्युत्पत्तिगत विश्लेषण :** संबंधित प्रविष्टि की व्युत्पत्तिगत स्रोत-भाषा के आधार पर विश्लेषित किया जाएगा। अन्य सभी भारतीय भाषाओं के शब्द/पद के व्युत्ति का भी विश्लेषण किया जाएगा।

**10. रूपसाधक रूपिम :** हिंदी एवं अन्य भारतीय भाषाओं के सभी प्रविष्टियों का रूपसाधक प्रत्यय लगने से बनने वाले सभी रूपों को लिखा जाएगा।

**11. व्युत्पादिक रूपिम :** हिंदी एवं अन्य भारतीय भाषाओं के मूल शब्द/पद में लगने वाले सभी उपसर्ग/प्रत्ययों को अलग-अलग विश्लेषित करते हुए लिखा जाएगा।

**12. विभिन्न अर्थ का निर्धारण :** हिंदी एवं अन्य भारतीय भाषाओं के प्रविष्टियों का अर्थ सबसे महत्वपूर्ण भाग होगा, जो सभी प्रविष्टि में समाहित सभी सूचनाओं को प्रकट करेगा।

**13. वाक:** हिंदी एवं अन्य भारतीय भाषाओं के सभी प्रविष्टियों के मानक स्वरूप के आधार पर वाक को अभिलेखबद्ध करके संग्रहित किया जाएगा।

**14. चित्र :** कार्पस वाक्य–प्रयोग में प्राप्त मूर्त संज्ञा शब्दों का प्रयोग आधारित चित्र का निर्माण किया जाएगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**15. चलचित्र (एनीमेशन) :** हिंदी एवं अन्य भारतीय भाषाओं की ऐसी क्रियाएँ जिससे किसी कार्य की अनुभूति होगी, उन क्रियाओं के चलचित्र (एनीमेशन) का निर्माण किया जाएगा।

### v. प्रविष्टि-चयन (Entry Selection) :

मशीन अनुवाद प्रणाली में स्रोत एवं लक्ष्य-भाषा के पाठ के अनुरूप भाषा का संसाधन किया जाता है। भाषा संसाधन की प्रक्रिया में संरचनागत एवं अर्थगत स्तर अपनी महत्वपूर्ण भूमिका निभाते है। इसलिए यहां पर प्रविष्टि का चयन वाक्य के अर्थीय आधार पर किया जाएगा, जिसके माध्यम से पूरे भाषा का विश्लेषण किया जा सके। प्रविष्टि चयन हेतु निम्न बातों का ध्यान रखना आवश्यक होगा:

I. प्रविष्टि चयन का मुख्य आधार प्रयुक्त घटक का अर्थीय आधार होगा; अर्थात् वाक्य संरचना एवं अर्थ संरचना के आधार पर प्रविष्टि का चयन किया जाएगा।
II. वाक्य संरचना के प्रयुक्त सभी शब्द-वर्ग के अवयवों की प्रविष्टि बनाई जाएगी।
III. वर्तनीगत शुद्धता/मानकता का पूर्ण पालन किया जाएगा।
IV. प्रविष्टि का निर्माण पुल्लिंग एवं स्त्रीलिंग दोनों रूपों में होगा, लेकिन बहुवचन के रूप में नहीं होगा।
V. पारिभाषिक शब्दों के अलावा प्राप्त शब्दों की पूरी कोशिश होनी चाहिए कि एक वाक्य–प्रयोग से केवल एक ही प्रविष्टि बनाई जाए, परंतु यह अनिवार्य नहीं है, क्योंकि ऐसी प्रविष्टि कहीं अन्यत्र प्राप्त नहीं हुई, तो प्रविष्टि के छूट जाने की संभावना बनी रहेगी।
VI. प्रविष्टि की एकाधिक शब्द-वर्ग में प्रयुक्त होने पर उसकी अलग-अलग प्रविष्टि निर्माण की जाएगी, जैसे : गधा (संज्ञा पुल्लिंग), गधा (विशेषण)।
VII. ऐसी प्रविष्टियाँ जो वर्तनीगत समान (समनामी) हो; किंतु अर्थीय एवं व्युत्पत्ति आधारित अलग-अलग हों, उनकी प्रविष्टि भी अलग-अलग की जाएगी। जैसे : 1. आम (हिंदी)- एक फल 2. आम (अरबी) – साधारण, सामान्य।
VIII. वर्तनीगत एक ही शब्द/पद के एक से अधिक रूपों की अलग प्रविष्टि नहीं बनाई जाएगी। बल्कि प्रविष्टि की 'प्राप्त वर्तनीगत अन्य रूप' के कॉलम में इसे लिखा जाएगा। जैसे: दर्शन ('दरसन/दरशन/दर्सन' आदि)।
IX. हिंदी वाक्य-संरचना में एक से अधिक अर्थ को प्रकट करने वाले अँग्रेजी शब्दों की प्रविष्टि बनाई जाएगी।
X. संख्यावाचक विशेषण की प्रविष्टि बनाई जाएगी।
XI. यौगिक/सामासिक शब्दों के बीच में योजक चिह्न लगाकर प्रविष्टि बनाई जाएगी। जैसे : तूल-कलाम, लक्ष्मण-रेखा, उच्च-न्यायालय, शेख-चिल्ली, खुदा-हाफिज आदि।
XII. जब 'वाला/वान' शब्द संज्ञा के साथ जुड़कर एक इकाई का प्रकार्य करे, तब संज्ञा के साथ 'वाला/वान' प्रत्यय को जोड़कर प्रविष्टि बनाई जाएगी। संज्ञा के अलावा

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

अन्य शब्द-वर्ग के साथ 'वाला/वान' को जोड़कर प्रविष्टि नहीं बनाई जाएगी। जैसे : गाड़ीवाला, गाड़ीवान इत्यादि।

XIII. हिंदी वर्तनी व्यवस्था के अनुरूप 'हल' चिह्न लगाकर प्रविष्टि बनाने की आवश्यकता नहीं है। जैसे- 'पश्चात्-पश्चात' आदि।

XIV. प्रविष्टि-चयन हेतु आवश्यकतानुसार अन्य नियमों को शामिल भी किया जाएगा।

v. **रोमन उच्चानरण संकेत (Roman Pronunciation Symbol):** उच्चारण संकेत उच्चारण-ईकाई (Syllabic unit) के प्रारूप में ही लिखा जाएगा। जैसे : कलम (ka-la-m)। इसे समझने हेतु निम्न तालिका का अवलोकन किया जाएगा :

**स्वर :**

| अ | आ | इ | ई | उ | ऊ | ए | ऐ | ओ | औ |
|---|---|---|---|---|---|---|---|---|---|
| A | ā | I | Ī | U | Ū | E | ai | O | Au |

| ऋ | अं | अँ | अः |
|---|---|---|---|
| Ṛi | ṃ | m̐ | ḥ |

**व्यंजन :**

| | | | | |
|---|---|---|---|---|
| क | ख | ग | घ | ङ |
| K | Kh | G | Gh | ṅ |
| च | छ | ज | झ | ञ |
| C | Ch | J | Jh | ṉ |
| ट | ठ | ड | ढ | ण |
| ṭ | ṭh | ḍ | ḍh | ṇ |
| | | ड़ | ढ़ | |
| | | ṛ | ṛh | |
| त | थ | द | ध | न |
| T | Th | D | Dh | N |
| प | फ | ब | भ | म |
| P | Ph | B | Bh | M |

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

| | | | | |
|---|---|---|---|---|
| य | र | ल | व | श |
| Y | R | L | V | ṡ |
| ष | स | ह | | |
| ṣ | S | H | | |
| क्ष | त्र | ज्ञ | | |
| kṣa | Tr | jn̲a | | |

आगत वर्ण :

| | | | | | | |
|---|---|---|---|---|---|---|
| क़ | ख़ | ग़ | ज़ | फ़ | | ऑ |
| Q | Qh | g̶ | Z | F | | Ŏ |

तालिका क्र. 04

vi. **आई.पी.ए. (IPA):** प्रविष्टि के सभी वर्णों के निर्धारित हिंदी आई.पी.ए. (International Phonetic Alphabet) के प्रारूप में ही लिखा जाएगा। जैसे : कलम (kələmə)। इसे समझने हेतु निम्न तालिका का अवलोकन किया जाएगा :

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**स्वर :**

| अ | आ | इ | ई | उ | ऊ | ए | ऐ | ओ | औ |
|---|---|---|---|---|---|---|---|---|---|
| ə | aː | I | iː | U | uː | E | ɛː | oː | ɔː |
| ऋ | अं | अँ | अः | | | | | | |
| r̥ | ⁿ | ˜ | H | | | | | | |

**व्यंजन :**

| क | ख | ग | घ | ङ |
|---|---|---|---|---|
| K | kʰ | g | gʰ | Ŋ |
| च | छ | ज | झ | ञ |
| tʃ | tʃʰ | dʒ | dʒʰ | ɲ |
| ट | ठ | ड | ढ | ण |
| ʈ | ʈʰ | ɖ | ɖʰ | ɳ |
| त | थ | द | ध | न |
| t̪ | t̪ʰ | d̪ | d̪ʰ | N |
| प | फ | ब | भ | म |
| P | pʰ | b | bʰ | M |
| य | र | ल | व | श |
| J | R | l | ʋ | ʃ |
| ष | स | ह | | |
| ʂ | S | ɦ | | |
| क्ष | त्र | ज्ञ | | |
| kʂə | t̪rə | dʒnə | | |

आगत वर्ण :

| क़ | ख़ | ग़ | ज़ | फ़ |
|---|---|---|---|---|
| Q | X | ɣ | Z | F |

तालिका क्र. 05

vii. **प्रविष्टि की प्राप्त वर्तनीगत अन्य रूप :** हिंदी भाषा में एक शब्द को एकाधिक रूपों में प्रयोग किया जाता है, कभी-कभी विकृत रूप भी प्राप्त होते हैं। इसलिए जिन प्रविष्टियों को मानक नहीं माना जाएगा, उन्हें वर्तनीगत अन्य रूप के तौर पर सुरक्षित किया जाएगा। जैसे : दर्शन की जगह दरसन/दर्सन/दरशन आदि।

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

viii. **पर्यायवाची :** हिंदी एवं अन्य सभी भारतीय भाषाओं की प्रविष्टि (ऐसे शब्द जिनकी पर्यायवाची लिखने योग्य हो) के लगभग सभी पर्यायवाची शब्दों/पदों/पदबंधों/मुहावरा को लिखा जाएगा।

ix. **विलोम :** प्रविष्टि के (ऐसे शब्द जिनकी विलोम लिखने योग्य हो) के लगभग सभी विलोम शब्दों/पदों/पदबंधों/मुहावरा को लिखा जाएगा।

x. **व्याकरणिक कोटि का चयन :** व्याकरणिक कोटि का चयन निम्नानुसार किया जाएगा :

1. **संज्ञा :** जिससे कल्पित सृष्टि की किसी वस्तु का नाम सूचित हो। संज्ञा के प्रकार को निम्न रूप में चयन किया जाएगा :

a. **व्यक्तिवाचक संज्ञा :** जिस संज्ञा से एक ही पदार्थ व पदार्थों के एक ही शमूह का बोध हो। जैसे : प्रकृति, परमात्मा, राम इत्यांदि।
b. **जातिवाचक संज्ञा :** जिस संज्ञा से संपूर्ण पदार्थों व उनके समूहों का बोध हो। जैसे : मनुष्य, घर, नदी इत्यादि।
c. **भाववाचक संज्ञा :** जिस संज्ञा से पदार्थ में पाए जाने वाले किसी धर्म का बोध हो। जैसे : बुढ़ापा, नम्रता, चतुराई इत्यादि।

i. जातिवाचक संज्ञा के रूप में प्रयुक्त होने वाली संज्ञाएँ जब भाव का बोध करवाएँगी, तब उन्हें भाववाचक संज्ञा के अंतर्गत रखा जाएगा। जैसे : बुढ़ापा, लड़कपन, मित्रता
ii. विशेषण शब्दों से बनने वाली संज्ञाएँ भाववाचक संज्ञा के अंतर्गत आएँगी। जैसे : गरम-गरमी, कठोर-कठोरता, मीठा-मिठास इत्यादि।
iii. क्रिया-विशेषण से बनने वाली संज्ञाएँ भी भाववाचक संज्ञा के अंतर्गत आएँगी। यथा- जिसका भीतर-बाहर एक सा हो, यहाँ की भूमि अच्छी हो।

**2. क्रियार्थक संज्ञा :** क्रियार्थक संज्ञाएँ क्रिया के धातु रूप से ही निर्मित होती हैं, अत: इसे अलग शब्द-वर्ग में रखा गया है। किन्तु इनके व्याकरणिक कोटि में इन्हें भाववाचक संज्ञा के रूप में ही प्रविष्टि करनी है। जैसे : घबराहट, सजावट, चढ़ाई इत्यादि। धातुओं का प्रयोग भी भाववाचक संज्ञा के समान होता है। जैसे : हम नाच नहीं देखते।

3. **सर्वनाम :** जो पूर्वापर संबंध से किसी भी संज्ञा के बदले उपयोग में आता हो। यहां पर निम्न रूप में सर्वनाम का चयन किया जाएगा :

i. **पुरुषवाचक :** मैं, तू, आप (आदरसूचक)।

**ii. निजवाचक :** आप (वह अपने को सुधार रहा है)।

**iii. निश्चयवाचक :** यह, वह, सो आदि।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**iv. संबंधवाचक :** जो (जो हरिश्चंद्र ने किया वह तो अब कोई भी भारतवासी न करेगा।)।

**v. प्रश्नवाचक :** कौन, क्या आदि।

**vi. अनिश्चयवाचक :** कोई, कुछ आदि।

i. उपर्युक्त में केवल मूल सर्वनामों का ही वर्गीकरण किया गया है, इनके रूपांतरित रूपों को भी इन्हीं वर्गीकरण के अंतर्गत रखा जाएगा।
ii. उत्तम पुरुष 'मैं' और मध्यम पुरुष 'तू' को छोड़कर शेष सर्वनाम और सभी संज्ञाएँ अन्य पुरुष में आती हैं।
iii. निजवाचक 'आप' अलग-अलग स्थानों में अलग-अलग पुरुषों के बदले आ सकता है।
iv. सभी सर्वनामों की प्रविष्टि बनाई जाएगी।

**4. विशेषण :** जिस शब्द से संज्ञा की व्याप्ति मर्यादित होती हो।

i. **सार्वनामिक :** पुरुषवाचक एवं निजवाचक सर्वनामों को छोड़कर शेष सर्वनामों का प्रयोग विशेषण के समान होता है। जैसे : वह, क्या, यह इत्यादि (वह एक नौकर आ गया। क्या आदमी हो। ऐसा कब हो सकता है कि मुझे भी दोष लगे।)

ii. **कालवाचक :** जब समय का बोध हो। जैसे : नया, पुराना, भविष्य इत्यादि।

iii. **स्थानवाचक :** जब स्थान का बोध हो। जैसे : लंबा, चौड़ा, ऊँचा, नीचा इत्यादि।

iv. **आकारवाचक :** जब आकार का बोध हो। जैसे : गोल, चौकोर, सुडौल इत्यादि।

v. **रंगवाचक :** जब रंग का बोध हो। जैसे : लाल, पीला, नीला इत्यादि।

vi. **दशावाचक :** जब दशा का बोध हो। जैसे : दुबला, पतला, मोटा इत्यादि।

vii. **गुणवाचक :** जब गुण का बोध हो। जैसे : भला, बुरा, उचित इत्यादि।

viii. **संख्यावाचक :** जिसमें संख्या का बोध हो। संख्या का बोध निम्न तीन रूपों में व्याप्त होती है :

a. **निश्चित:** जब निश्चित संख्या का बोध हो। जैसे : एक, पाँच, आध, पाव आदि।

b. **अनिश्चित :** जब किसी निश्चित संख्या का बोध नही हो। जैसे : कम, कुछ, बहुत आदि।

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

c. **परिणामबोधक :** जब किसी वस्तु के नाप या तौल का बोध हो। जैसे : सब, सारा, समूचा इत्यादि।

**ix. क्रियार्थक :** निम्न रूप में प्राप्त सभी प्रकार के विशेषण को क्रियार्थक संज्ञा के रूप में ही प्रविष्टि करनी है :

a. **क्रियार्थक संज्ञा से :** जैसे : तुमको परीक्षा करनी हो तो लो।

b. **कर्तृवाचक संज्ञा से:** जैसे : जाने वाला नौकर ।

**c. वर्तमानकालिक कृदंत से:** जैसे : बहता पानी।

**d. भूतकालक कृदंत से:** मरा घोड़ा, गिरा घर, किया हुआ काम।

- बड़ा-सा, खोटी-सी, किया हुआ, जाने वाला आदि सभी की प्रविष्टि बनाई जाएगी।
- वर्तमानकालिक एवं भूतकालिक कृदंत से बनने वाले विशेषणों के भी प्राप्त रूप में प्रविष्टि बनाई जाएगी, यथा- बहता, मरा, किया हुआ इत्यादि।

**5. क्रिया :** जब किसी वस्तु के विषय में कुछ विधान करते हैं। क्रिया का चयन निम्न दो रूपों में किया जाएगा।

**i. सर्कमक क्रिया :** जिस क्रिया से सूचित होने वाले व्यापार का फल कर्ता से निकलकर किसी दूसरी वस्तु पर पड़ता हो। जैसे : पकड़ना, लाना इत्यादि।

ii. **अर्कमक क्रिया :** जिस क्रिया से सूचित होने वाला व्यापार और उसका फल कर्ता ही पर पड़े। जैसे : चलना, सोना इत्यादि।

iii. **मूल धातु :** जो किसी दूसरे शब्दों की सहायता से न बने हो। जैसे : करना, बैठना, चलना इत्यादि।

iv. **यौगिक धातु :** जो किसी दूसरे शब्दों की सहायता से बने हों। जैसे : चलाना, रंगना, चिकनाना इत्यादि।

- धातु के अंत में 'ना' प्रत्यय जोड़कर क्रिया का साधारण रूप बनाया जाएगा।
- कुछ क्रियाएँ दोनों रूपों में प्रयुक्त होती हैं, अत: इनके रूप को वाक्य-प्रयोग से स्पाष्ट किया जाएगा। जैसे : खुजलाना, भरना, लजाना, भूलना आदि।

6. **प्रेरणार्थक क्रिया :**

i. **प्रथम प्रेरणार्थक :** जैसे : चलाना, गिराना इत्यादि।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

ii. **द्वितीय प्रेरणार्थक :** जैसे : चलवाना, गिरवाना इत्यादि।

- सभी प्रेरणार्थक क्रियाएँ सकर्मक होती हैं।

7. **मिश्र क्रिया :** संज्ञा/विशेषण/क्रियांगी + क्रियाकर से मिलकर बनने वाली क्रिया। जैसे : इंतजार करना, मना करना इत्यादि।

- अर्थीय स्तर पर मिश्र क्रिया के दोनों घटक मिलकर एक आर्थी इकाई का निर्माण करते हैं।

8. **संयुक्त क्रिया :** क्रिया धातु + रंजक क्रिया से मिलकर बनने वाली क्रिया। जैसे : भाग जाना, रो पड़ना इत्यादि।

**9. यौगिक क्रिया :** मुख्य क्रिया + मुख्य क्रिया से मिलकर बनने वाली क्रिया। जैसे : ले जाना, कर दिखाना इत्यादि।

**10. सहायक क्रिया :** जब मुख्य क्रिया के साथ सहायक रूप में प्रयुक्त हो। जैसे : है, था इत्यादि।

**11. क्रिया-विशेषण :** जिस अव्यय से क्रिया की कोई विशेषता जानी जाती हो। क्रिया-विशेषण का चयन निम्न रूपों में किया जाएगा।

a. **मूल:** जो किसी दूसरे शब्दों के योग से नहीं बनते। जैसे : ठीक, दूर, अचानक, आज, फिर इत्यादि।
b. **यौगिक:** जो किसी दूसरे शब्दों में प्रत्यय व शब्द जोड़ने से बनते हैं। इस प्रकार के क्रिया-विशेषण का चयन प्राप्त निम्न रूप में चयन करेंगे :
   - संज्ञा से बनने वाली : जैसे : प्रेम-पूर्वक, मन से इत्यादि।
   - सर्वनाम से बनने वाली : जैसे : यहाँ, इसलिए, तिसपर इत्यादि।
   - विशेषण से बनने वाली : जैसे : चुपके, भूले से, सहज में इत्यादि।
   - धातु से बनने वाली : देखते हुए, चाहे इत्यादि।
   - अव्यय से बनने वाली : यहाँ तक, कब का।
   - क्रिया-विशेषण के साथ निश्चयवाचक शब्दों से बनने वाली : अब-अभी, यहाँ-यहीं इत्यादि।

**C. संयुक्त:** संयुक्त क्रिया-विशेषण निम्न रूप में प्राप्त होती है। इन सभी क्रिया-विशेषण को निम्न प्राप्त रूप में चयन किया

जाएगा:

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

- संज्ञाओं की द्विरुक्ति से : जैसे : घर-घर, घड़ी-घड़ी।
- भिन्न-भिन्न संज्ञाओं से : जैसे : रात-दिन, सांझ-सवेरे।
- विशेषण की द्विरुक्ति से : जैसे : एका-एक, ठीक-ठाक।
- क्रिया-विशेषण की द्विरुक्ति से : जैसे : धीरे-धीरे, जहाँ-जहाँ।
- भिन्न क्रिया-विशेषण से :जैसे : जहाँ-तहाँ, जब-तब।
- क्रिया-विशेषण के बीच 'न' लगने से : जैसे : कभी-न-कभी, कुछ-न-कुछ।
- अनुकरणवाचक की द्विरुक्ति से : जैसे : गटगट, सटासट, तड़तड़।
- संज्ञा एवं विशेषण से : जैसे : एक-साथ, एक-बार।
- अव्यय के साथ अन्य शब्दों के मेल से : जैसे : यथा-क्रम, बे-फायदा, प्रतिदिन।
- विशेषण एवं पूर्वकालिक कृदंत के मेल से : जैसे : विशेष-करके, एक-एक-करके।

D. **स्थानीय:** दूसरे शब्द-वर्ग जो बिना किसी रूपांतर के क्रिया-विशेषण के समान उपयोग में आते हों। इनके बनने की प्रक्रिया के निम्न आधार पर चयन किया जाएगा :

- संज्ञा से बनने वाले :जैसे : तुम मेरी मदद पत्थर करोगे।
- सर्वनाम से बनने वाले :जैसे : लीजिए महाराज, मैं यह चला।
- विशेषण से बनने वाले : जैसे : मनुष्य उदास बैठा है।
- पूर्वकालिक कृदंत से बनने वाले : जैसे : तुम दौड़कर चलते हो।
- उपर्युक्त चारों प्रकार के विशेषणों की प्रविष्टि प्राप्त रूप में ही बनाई जाएगी। यथा-भूल से, कभी-न-कभी, दौड़कर, सहज में इत्यादि।

**12. संबंध-सूचक :** जो अव्यय संज्ञा (अथवा संज्ञा के समान उपयोग में आने वाले शब्द) के बहुधा आगे आकर उसका संबंध वाक्य के किसी दूसरे शब्द के साथ मिलाता है। संबंध-सूचक शब्द का चयन निम्न रूप में किया जाएगा :

**i. संबद्ध :** ऐसे संबंध-सूचक शब्द संज्ञाओं के विभक्तियों के आगे आते हों। जैसे : धन के बिना, पूजा से पहले इत्यादि।

**ii. अनुबद्ध :** ऐसे संबंध-सूचक शब्द संज्ञा के विकृत रूप (बिना विभक्ति के) के साथ आते हों। जैसे : किनारे तक, लड़के सरीखा, कटोर भर इत्यादि।

**13. समुच्चय-बोधक :** जो अव्यय एक वाक्य का संबंध दूसरे वाक्य से मिलाता हो। समुच्य-बोधक का चयन निम्न रूप में किया जाएगा :

**i. संयोजक :** जिनके द्वारा दो या अधिक मुख्य वाक्य का संग्रह होता है। जैसे : और, व, तथा, एवं, भी इत्यादि।

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

**ii. विभाजक :** जिन अव्ययों से दो या अधिक वाक्यों व शब्दों में से किसी एक का ग्रहण अथवा दोनों का त्याग हो। जैसे : या, वा, अथवा, कि, न-न, या-या, चाहे-चाहे, क्या–क्या, न कि, नहीं तो इत्यादि।

**iii. विरोधदर्शक :** जो अव्यय दो वाक्यों में से पहले का निषेध व परिमिति सूचित करते हैं। जैसे : पर, परंतु, किंतु, लेकिन, मगर, बल्कि इत्यादि।

**iv. परिणामदर्शक :** जिन अव्यय से आगे के वाक्य का अर्थ पिछले वाक्य के अर्थ का फल है। जैसे : इसलिए, सो, अत, अतएव इत्यादि।

**v. कारणवाचक:** जिन अव्ययों से आरंभ होने वाले वाक्य पूर्व वाक्य का समर्थन करते हों। जैसे : क्योंकि, जोकि, इसलिए इत्यादि।

**vi. उद्देश्यवाचक:** 'न' अव्यय के पश्चात आने वाला वाक्य दूसरे वाक्य का उद्देश्य व हेतु सूचित करता है। जैसे : कि, जो, ताकि इत्यादि।

**vii. संकेतवाचक :** इनमें 'कि' को छोड़कर शेष शब्द, संबंधवाचक और नित्य-संबंधी सर्वनामों के समान, जोड़े में आते हों। जैसे : जो-तो, यदि-तो, यद्यपि-तथापि, चाहे-परंतु इत्यादि।

**viii. स्वरूपवाचक:** जिन अव्ययों के द्वारा जुड़े हुए शब्दों या वाक्यों में से पहले शब्द व वाक्य का स्वरूप (स्पष्टीकरण) पिछले शब्द व वाक्य से जाना जाता हो। जैसे : अर्थात्, मानो इत्यादि।

**14. विस्मयादि-बोधक :** जिन अव्ययों का संबंध वाक्य से नहीं रहता और जो वक्ता के केवल हर्ष-शोकादि भाव सूचित करते हो। विस्मयादि-बोधक का चयन निम्न रूप में किया जाएगा :

**i. हर्षबोधक :** जैसे : आहा! वाह वाह! इत्यादि।

**ii. शोकबोधक :** जैसे : आह! ऊह! हाय! इत्यादि।

**iii. आश्चर्यबोधक :** जैसे : ऐं! ओहो क्या! इत्यादि।

**iv. अनुमोदनबोधक:** जैसे : ठीक! वाह! अच्छा! इत्यादि।

**v. तिरस्कारबोधक:** जैसे : छि:! हट! अरे! इत्यादि।

**vi. स्वीकारबोधक:** जैसे : हाँ! जी! हाँ! बहुत अच्छा! इत्यादि।

**vii. संबोधबोधक :** जैसे : अरे! रे! अजी! इत्यादि।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**15. परसर्ग :**

i. कर्तृवाचक: जैसे : ने।

ii. कर्मवाचक: जैसे : को।

iii. करणवाचक: जैसे : से।

iv. संप्रदानवाचक: जैसे : को।

v. अपादानवाचक: जैसे : से।

vi. संबंधवाचक: जैसे : का, के, की इत्यादि।

vii. अधिकरणवाचक: जैसे : में, पर इत्यादि।

viii. संबोधनवाचक: जैसे : हे, अजी, अहो इत्यादि।

16. **निपात :** जैसे : ही, भी, तक इत्यादि।

**xii. लिंग :** लिंग का चयन निम्न रूप में चयन किया जाएगा :

1. **पुल्लिंग :** जिससे यथार्थ व कल्पित पुरूष का बोध हो। जैसे- लड़का, बैल, पेड़ इत्यादि।

**2. स्त्रीलिंग :** जिससे यथार्थ व कल्पित स्त्री का बोध हो। जैसे- लड़की, गाय, लता इत्यादि।

**xiii. वचन** : वचन का चयन निम्न रूप में किया जाएगा :

i. **एकवचन :** संज्ञा के जिस रूप से एक ही वस्तु का बोध हो। जैसे- लड़का, कपड़ा, टोपी इत्यादि।
ii. **बहुवचन :** संज्ञा के जिस रूप से एकाधिक वस्तुओं का बोध हो। जैसे- लड़के, कपड़े, टोपियाँ इत्यादि।

**xiv. पुरुष :** पुरूष का चयन निम्न रूप में किया जाएगा :

**1. उत्तम पुरुष :** जिसमें स्वयं वक्ता या लेखक का बोध हो। जैसे- मैं

**2. मध्यम पुरुष :** जिसमें श्रोता या पाठक का बोध हो। जैसे – तू, तुम, आप (आदरसूचक)

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

**3. अन्य पुरुष :** जिसमें वक्ता या श्रोता को छोड़कर किसी अन्य का बोध हो। जैसे- यह, वह, वो आदि। सभी संज्ञाएं अन्य

पुरुष हैं।

**xv. काल:** काल का चयन निम्न आधार पर किया जाएगा :

**1. वर्तमान :** जिसमें वर्तमान काल का बोध हो।

**2. भूत :** जिसमें भूतकाल का बोध हो।

**3. भविष्य :** जिसमें भविष्य काल का बोध हो।

**xvi. वृत्ति :** वृत्ति का चयन निम्न आधार पर किया जाएगा :

**1. आज्ञार्थक :** वाक्य में जब आज्ञा का भाव हो। जैसे- पढ़, पढ़ो, पढ़िए।

**2. संभावनार्थक :** वाक्य में जब संभावना प्रकट हो रही हो। जैसे- शायद वह भी आए। मैं चाहता हूँ कि वह भी आए।

**3. अनिश्चित संभाव्य :** वाक्य में जब अनिश्चित संभावना प्रकट हो रही हो। जैसे- शायद वह आता हो।

**4. निश्चित संभाव्य :** वाक्य में जब निश्चित संभावना प्रकट हो रही हो। जैसे- वह यही काम करता होगा। उसने यही काम किया होगा।

**5. संकेतार्थ :** वाक्य में जब संकेत प्रकट हो रहा हो। जैसे- अगर वह कहता तो मैं जरूर आता।

**6. भविष्यत :** ऐसी वाक्य-संरचना जिसमें भविष्य में कार्य व्यापार पूर्ण हो। जैसे- वह जाएगा।

**xvii. पक्ष :** पक्ष का चयन निम्न आधार पर किया जाएगा।

**1. पूर्ण पक्ष :** जब कार्य व्यापार की पूर्ण अवस्था का बोध हो। जैसे- रमेश लौटा।

**2. अपूर्ण पक्ष :** जब कार्य व्यापार की किसी अपूर्ण अवस्था का बोध हो। जैसे- सूरज पूरब से निकलता है।

**xviii. वाच्य :** वाच्य का चयन निम्न रूप में किया जाएगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**1. कर्तृवाच्य :** कर्ता द्वारा किए गए कार्य का फल जब क्रिया पर पड़े। जैसे – नौकर दरवाजा खोल रहा है। बच्चों ने गीत गाए।

**2. कर्मवाच्य** : कर्म का फल जब क्रिया पर पड़े। जैसे- दरवाजा खोला जा रहा है। नौकर से दरवाजा नहीं खोला जाता।

**3. भाववाच्य :** जब क्रिया स्वतंत्र हो। जैसे- जमीन पर नहीं बैठा जाता।

**xix. अर्थीय घटक (Semantic Component) :**

भाषा-संसाधन में भाषा के अवयवों का सटीक परिणाम प्राप्त करने हेतु अर्थीय घटकों का संग्रहण भी आवश्यक होता है। यह सर्वविदित है कि भाषा-विश्लेषण में अर्थीय घटक अपनी महत्वपूर्ण योगदान देते हैं, जिनकी सहायता से एकाधिक अर्थ में प्राप्त शब्द/पद, पदबंध, वाक्य की पहचान की जा सके। भाषा-संसाधन में भाषा का विश्लेषण जितनी गहराई के साथ किया जाएगा, परिणाम उतने ही सटीक प्राप्त होंगे। सभी भाषिक अवयवों के अर्थीय घटक अलग-अलग होते हैं, जिन्हें निम्न रूप में दर्शाया गया है :

| **क्र.** | **संज्ञा हेतु** | |
|---|---|---|
| 1. | सजीव | निर्जीव |
| 2. | मानव | अमानव |
| 3. | मूर्त | अमूर्त |
| 4. | गणनीय | अगणनीय |
| 5. | निश्चित | अनिश्चित |
| 6. | चेतन | अचेतन |
| 7. | स्वेच्छायुक्त | स्वेच्छारहित |
| **क्रिया हेतु** | | |
| 8. | घटना | प्रक्रिया |
| 9. | स्थितिसूचक | स्थिति-परिवर्तक |
| 10. | स्थित्यात्मक | अस्थित्यात्मक |
| 11. | मानसिक | शारीरिक |
| 12. | परिणामी | अपरिणामी |
| 13. | अभिप्रेत | अनभिप्रेत |
| 14. | तथ्यात्मक | धारणात्मक |
| **विशेषण हेतु** | | |

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

| | | |
|---|---|---|
| 15. | मूल्य-निर्णयपरक | वस्तुपरक |
| 16. | निरपेक्ष अवस्थासूचक | सापेक्षिक अवस्थासूचक |
| 17. | रूपपरक | गुणपरक |
| **क्रियाविशेषण- हेतु** | | |
| 18. | अधिकरणात्मक | प्रयोजनसूचक |
| 19. | गंतव्यसूचक | गतिसूचक |
| 20. | गणनावाचक | |
| 21. | क्रमवाचक | |
| 22. | आवृत्तिवाचक | |
| 23. | समुदायवाचक | |
| 24. | प्रत्येक बोधक | |
| 25. | पूर्णांक-बोधक | अपूर्णांक-बोधक |
| **संबंध-वाचक** | | |
| 26. | कालवाचक | |
| 27. | स्थानवाचक | |
| 28. | दिशावाचक | |
| 29. | साधनवाचक | |
| 30. | हेतुवाचक | |
| 31. | विषयवाचक | |
| 32. | व्यतिरेकवाचक | |
| 33. | विनिमयवाचक | |
| 34. | सादृश्यवाचक | |
| 35. | विरोधवाचक | |
| 36. | सहचारवाचक | |
| 37. | संग्रहवाचक | |
| 38. | तुलनावाचक | |
| प्रस्तुत कोश निर्माण में यदि संज्ञा, सर्वनाम, विशेषण एवं क्रिया विशेषण के अन्य अर्थीय घटक प्राप्त होते हैं, तो उन्हें भी शामिल किया जाएगा। | | |

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

तालिका क्र. 06

## xx. **मुहावरा एवं लोकोक्ति:**

हिंदी एवं अन्य भारतीय भाषाओं के पाठ में प्रयुक्त मुहावरा एवं लोकोक्ति बहुतायत मात्रा में प्राप्त होते हैं, जिनका यथाउचित संसाधन भी किया जाना चाहिए। मुहावरा एवं लोकोक्ति के माध्यम से ही भाषा अर्थीय स्तर पर समृद्ध होती है। जिनकी पहचान कर, विश्लेषण एवं संश्लेषण आवश्यक होता है। मुहवरा की संरचना वाक्यो में ही निहित होती है, अतएव मुहावरा में प्रयुक्त सभी घटकों का व्याकरणिक विश्लेषण महत्वपूर्ण हो जाता है। लोकोक्तियों की बहुधा अपनी पूरी अलग वाक्य संरचना होती है, इसलिए इनकी व्याकरणिक विश्लेषण की आवश्यकता तो नहीं बल्कि एकाधिक अर्थों का संचय आवश्यक होता है। इससे संबंधित निम्न नियमों का पालन किया जाएगा :

- किसी मुहावरे में परसर्ग, अव्यय के प्रयोग से बनने वाले यथा- बाल बाँका होना, बाल बाँका न होना, बाल बाँका भी न होना, में से सभी रूपों की प्रविष्टि बनाई जाएगी।
- मुहावरा की संरचना यदि दो क्रियाओं से मिलकर निर्मित होती है, तब यह देखा जाएगा कि दोनों क्रियाओं में कौन-सी मुख्य क्रिया है। यदि प्रथम मुख्य-क्रिया है, तब एक ही क्रिया से समाप्त कर प्रविष्टि बनाई जाएगी अन्यथा दोनों क्रियाओं को मिलाकर प्रविष्टि बनाई जाएगी। जैसे : 'धूल में मिला देना' की 'धूल में मिलाना' जबकि 'पैसे खड़े कर लेना' का 'पैसे खड़ा करना' नहीं।
- मुहावरा की संरचना वाक्य में ही निहित होती है, इसलिए इसमें प्रयुक्त, सभी घटकों की व्याकरणिक विश्लेषण किया जाएगा, जैसे : धूल में मिलाना (संज्ञा + परसर्ग + क्रिया)।
- लोकोक्तियों की बहुधा अपनी पूरी अलग वाक्य-संरचना होती है, अत: जिस रूप में यह वाक्य-प्रयोग में प्रयुक्त होगी, उसी रूप में इसकी प्रविष्टि की जाएगी।
- लोकोक्तियों की अपनी पूरी वाक्य-संरचना होती है, एतएव इसके सभी अवयवों के व्याकरणिक विश्लेषण की आवश्यकता नहीं होगी।

## xxi. **व्युत्पत्ति एवं व्युत्पत्तिगत विश्लेषण :**

भाषा का विश्लेषण वाक्य संरचना में प्रयुक्त घटकों के व्युत्पत्ति भी अपनी महत्वपूर्ण भूमिका अदा करता है। व्युत्पत्ति के आधार पर घटक के अर्थ पर भी प्रभाव पड़ता है। इसलिए भाषिक घटकों के व्युत्पत्ति के आधार पर भी विश्लेषण किया जाना ज्यादा उपयोगी होगा। इसमें तत्सम, तद्भव, देशज एवं आगत एवं शब्द/पद, पदबंध के निर्माण प्रक्रिया का

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

इतिहास भी शामिल किया जाना उचित होगा। इसे निम्न तालिका का अनुसरण कर विश्लेषण करने में सटीक परिणाम प्राप्त किए जा सकेंगे। व्युत्पत्तिगत विश्लेषण को निम्न रूप में समझ सकते हैं :

**1. तत्सम :** वे संस्कृत शब्द जो अपने असली स्वरूप में हिंदी भाषा में प्रचलित हैं। जैसे : राजा, पिता, कवि इत्यादि।

**2. तद्भव :** वे शब्द जो या तो सीधे प्राकृत से या प्राकृत के द्वारा संस्कृत से निकले हैं। जैसे : राय, खेत, दाहिना, किसान इत्यादि।

- हिंदी के क्रिया, सर्वनाम एवं परसर्ग शब्द प्राय: सब के सब तद्भव हैं।

**3. देशज :** वे शब्द, जो किसी संस्कृत या प्राकृत मूल से निकले हुए नहीं जान पड़ते एवं जिनकी व्युत्पत्ति का पता नहीं लगता है। जैसे : खिड़की, घूआ, ठेस इत्यादि।

- जिनकी व्युत्पत्ति का पता नहीं लगता, उन्हें देशज माना जाएगा।
- अनुकरणवाचक शब्दों की व्युत्पत्ति हिंदी ही मानी जाएगी।

**4. आगत :** विदेशी भाषा से जो शब्द हिंदी में आए हैं। जैसे : आदमी, अदालत, कोतल, कमरा, अपील इत्यादि।

- ऐसी प्रविष्टियाँ जिनकी व्युत्पत्ति का संयोजन संस्कृत+संस्कृत, तद्भव+तद्भव, देशज+देशज एवं आगत+आगत से हैं, उन्हें क्रमश: तत्सम, तद्भव, देशज एवं आगत ही रखा जाएगा एवं व्युत्पत्ति विश्लेषण में एक ही रूप में लिखा जाएगा, जैसे : [सं.] रामराज्य।
- ऐसी प्रविष्टियाँ जिनकी व्युत्पत्ति का भिन्न संयोजन (अरबी + फारसी आदि) हो, उन्हें व्युत्पत्तिगत विश्लेषण में खासदान- [अ.] खास + [फा.] दान के रूप में लिखा जाएगा।
- मुहावरे एवं लोकोक्तियों की व्युत्पत्ति हिंदी ही मानी जाए।

**xxii. रूपसाधक रूपिम :** रूपसाधक रूपिम विश्लेषण के उद्देश्य से मुख्य डाटाबेस में रूपसाधक प्रत्यय लगने से बनने वाले सभी रूपों को लिया जाएगा, ताकि सभी रूपिम का डाटा संग्रह किया जा सके। यथा– लड़का, लड़के, लड़को, लड़कों इत्यादि।

**xxiii. व्युत्पादक रूपिम :** व्युत्पादक रूपिम विश्लेषण तब ज्यादा महत्वपूर्ण हो जाता है, जब एक ही मूल शब्द से विभिन्न प्रत्ययों के संयोजन से नए रूपिम का निर्माण किया जा सके।

यथा –

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

i. ईमानदारी– ईमान (मूल) + दार (प्रत्यय1) + ई (प्रत्यय2)

ii. राष्ट्रीयतावादी– राष्ट्र (मूल) + ईय(प्रत्यय1) + ता(प्रत्यय2) +वाद(प्रत्यय3)+ ई(प्रत्यय4)

iii. असफलता- अ(उपसर्ग1) + स(उपसर्ग2) + फल (मूल) + ता (उपसर्ग) आदि

xxiv. **चित्र एवं चलचित्र :** मुख्य डाटाबेस में चित्र एवं चलचित्र (एनीमेशन) का भी संग्रहण एवं सुरक्षित किया जाना अपेक्षित है। जिसके माध्यम से संगणकीय-कोश के अर्थीय आधार पर सभी आयामों को पूरा किया जा सके। इसका प्रारूप निम्न रूप में होगा :

**1. संज्ञा :** ऐसी प्रविष्टियाँ जिससे किसी भी प्रकार की संज्ञा (मूर्त रूप) की अनुभूति हो, उनके चित्रों को भी सुरक्षित किया जाएगा।

**2. क्रिया :** ऐसी प्रविष्टियाँ जिससे किसी भी कार्य (Action) की अनूभूति हो, उन प्रविष्टियों के चलचित्र (एनीमेशन) को भी सुरक्षित किया जाएगा।

**xxv. ध्वनि :** शुद्ध उच्चरित/मानक प्रविष्टियों के उच्चरित रूप अर्थात् ध्वनि को भी सुरक्षित किया जाएगा। यह ध्वनि वाक्य संरचनागत प्रयोग के अनुरूप होनी चाहिए।

**xxvi. अर्थ निर्धारण :** वाक्य संरचना में प्रयुक्त अर्थ ही सबसे महत्वपूर्ण भाग है, जिसके बिना भाषा-संसाधन का कार्य अपूर्ण माना जाता है। इन्हीं अर्थों की सहायता से भाषिक तत्वों के अनेकार्थकता/संदिग्धार्थकता आदि की समस्या का निवारण संभव है। प्रविष्टि संबंधित सभी अर्थों को सूचीबद्ध तरीके से संग्रहित एवं सुरक्षित किया जाएगा। अर्थ-निर्धारण हेतु निम्न नियमों का पालन किया जाएगा:

1. प्रविष्टि का अर्थ कार्पस में प्रयुक्त वाक्य-प्रयोग में प्रयुक्त अर्थ के रूप में ही दिया जाएगा।

2. प्रविष्टि के अर्थ को समझाने के लिए व्याख्यायित अर्थ को पहली वरीयता दी जाएगी।

3. जिन प्रविष्टियों के अर्थ को व्याख्यायित नहीं किया जा सकता, उनके प्रचलित पर्याय (अधिकतम दो) का प्रयोग किया जाएगा।

4. दो पर्यायवाची शब्दों के बीच अर्धविराम (;) का प्रयोग किया जाएगा।

5. प्रविष्टि के रूप में चयनित शब्द/पद का प्रयोग उसके अर्थ अथवा व्याख्या में नहीं दिया जाएगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

6. अर्थ पूर्ण होने के पश्चात पूर्ण विराम (।) का प्रयोग किया जाएगा।

7. अभिधार्थ, लाक्षणार्थ एवं व्यंजनार्थ के क्रम में अर्थों को रखा जाएगा, परंतु इन्हें अर्थ में कहीं भी नहीं लिखा जाएगा।

**xxvii. वाक्य-प्रयोग :**

विभिन्न भाषा के साहित्य में प्रयुक्त शब्दों की ही प्रविष्टि बनाई जाएगी। भाषा की मानकता मानव समाज द्वारा प्रयोग किए गए शब्द/पद में ही निहित है। वाक्य-प्रयोग ही यह बताता है कि भाषा के अंतर्गत प्रयुक्त भाषिक घटक कितने प्रकार से प्रयोग किए जा सकेंगे। वाक्य–प्रयोग का निर्धारण, अर्थ-निर्धारण की सीमा तक दिया जाना आवश्यक होगा। यदि संबंधित प्रविष्टि का अर्थ एक वाक्य संरचना में निहित नहीं है, तो वाक्य अधूरा माना जाएगा एवं अर्थ भी स्पष्ट नहीं होगा। वाक्य-प्रयोग हेतु निम्न नियमों का पालन किया जाएगा :

1. प्रविष्टि के कार्पस वाक्य-प्रयोग का अर्थ निर्धारण की सीमा तक दिया जाएगा, जिसे प्रविष्टिकर्ता स्वयं तय करेगा।

2. वाक्य-प्रयोग एक अथवा एकाधिक वाक्य का भी हो सकता है।

3. वाक्य-प्रयोग का निर्धारण करते समय यह भी सुनिश्चित किया जाएगा कि वाक्य-प्रयोग भी मानक (शब्द एवं व्याकरणिक दृष्टिकोण से) हो।

**xxviii. वाक्य–संदर्भ :**

मुख्य डाटाबेस में संग्रहित किए जा रहे वाक्य-प्रयोग की प्रमाणिकता हेतु पूर्ण संदर्भ देना आवश्यक होता है, जिससे यह स्पष्ट किया जा सके कि संबंधित वाक्य-प्रयोग किसने और कहां प्रयोग किया था। वाक्य-प्रयोग की प्रमाणिकता हेतु लेखक द्वारा लिखे गए साहित्य का पूर्ण संदर्भ दिया जाएगा, जिससे आवश्यकतानुसार वाक्य-प्रयोग को आसानी से प्राप्त किया जा सके। इसी प्रकार प्रयुक्त सभी अन्य वाक्य-प्रयोग के संदर्भ को भी दिया जाएगा। वाक्य-संदर्भ को सुरक्षित करने का प्रारूप कृति कोड/पृ.सं. (जैसे : 2/28) होगा। कृति कोड साहित्य चयन के पश्चात प्रदान किया जाएगा एवं पृ.सं. में चयनित साहित्य के 'पृष्ठ संख्या' को लिखा जाएगा। सुरक्षित करने के पश्चात वाक्य-संदर्भ को प्रदर्शित करने के लिए ए.पी.ए. प्रारूप का प्रयोग किया जाएगा, जो इस प्रकार है :

जैसे : प्रेमचंद. (1947). मानसरोवर भाग-एक. बनारस: सरस्वती प्रेस, पृष्ठ सं. 14.

**xxix. संदर्भ-पृष्ठ :** वाक्य-प्रयोग द्वारा पूर्ण संदर्भ की अनुभूति नहीं होने स्थिति में संबंधित साहित्य के पृष्ठ-चित्र को भी सुरक्षित किया जाएगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

## 10. **ऑनलाइन डाटा इंट्री पोर्टल (Online Data Entry Portal) :**

उक्त में वर्णित व्याकरणिक-कोश एवं कार्पस संकलन/संग्रहण का कार्य पूर्णत: ऑनलाइन होगी। ऑनलाइन के माध्यम से ही डाटाबेस में डाटा का संग्रहण एवं भंडारण किया जाएगा। संग्रहण एवं भंडारण के पश्चात ऑनलाइन ही सार्वजनिक रूप से व्याकरणिक-कोश एवं हिंदी एवं अन्य भारतीय भाषा के अनूदित कार्पस को उपलब्ध कराया जाएगा। सार्वजनिक ऑनलाइन प्रारूप का निर्धारण मुख्य डाटाबेस में 25% डाटा संग्रहित करने के पश्चात् किया जाएगा। ऑनलाइन प्रारूप में एक मुख-पृष्ठ का निर्माण किया जाएगा, जिसमें सभी भाषिक अनुप्रयोग को सूचीबद्ध तरीके से लिंक प्रदान किया जाएगा। सभी भाषिक-सामग्री की पृथक-पृथक पृष्ठ का निर्माण किया जाएगा। मुख्य डाटाबेस से हिंदी एवं अन्य भारतीय भाषाओं के व्याकरणिक-कोश को भी अलग पृष्ठ प्रदान किया जाएगा। इस पृष्ठ पर प्रयोक्ता को किसी भी प्रविष्टि की खोज करने हेतु खोज-पृष्ठ पर संबंधित प्रविष्टि की एक सुनिश्चित प्रारूप में सभी सूचनाएँ प्रदर्शित होंगी। प्रयोक्ता द्वारा किसी भी प्रविष्टि की खोज हेतु पहली प्राथमिकता सटीक परिणाम प्रदान करनी की होगी, यदि वृहद डाटाबेस में खोजकर्ता/प्रयोक्ता द्वारा खोजी गई प्रविष्टि का मिलान नहीं हुआ, तब इस परिस्थिति में भी प्रयोक्ता के पास खोजी गई प्रविष्टि से संबंधित अन्य विकल्प की सुविधा सुलभ होगी।

इस प्रबंधन प्रणाली की सहायता से न सिर्फ मुख्य डाटाबेस में व्याकरणिक-कोश का निर्माण किया जाएगा बल्कि व्याकरणिक-कोश में प्रयुक्त कार्पस का समानांतर अनूदित पाठ को संग्रहित करने हेतु ऑनलाइन कार्पस का प्रदान किया जाएगा। ऑनलाइन अनुवाद कराने के पश्चात् अनूदित पाठ को भी डाटाबेस में संग्रहित व सुरक्षित किया जाएगा। कार्पस का अनुवाद कार्य करने से संबंधित सभी सुविधाओं को साफ्टवेयर के माध्यम से उपलब्ध कराया जाएगा।

ऑनलाइन व्याकरणिक-कोश एवं कार्पस का अनुवाद कार्य करने हेतु सभी कार्यरत लोगों को प्रविष्टिकर्ता, संशोधनकर्ता, निरीक्षणकर्ता एवं संपादनकर्ता के अनुरूप ही सुविधा उपलब्ध होगी। इस प्रबंधन प्रणाली के द्वारा व्याकरणिक-कोश एवं कार्पस अनुवाद का कार्य न सिर्फ प्रयोगशाला से अपितु घर या अन्य जगह जहां संबंधित सुविधा उपलब्ध हो; इस कार्य को सुगमता के साथ पूर्ण किया जा सकेगा।

यदि प्रयोक्ता किसी भी प्रविष्टि/अनूदित कार्पस से संबंधित किसी भी बिंदु पर अपना कोई सुझाव/संशोधन अथवा टिप्पणी करना चाहे तो उसे खोज-पृष्ठ पर इसकी सुविधा उपलब्ध होगी। यह सुविधा प्रत्येक प्रविष्टि/अनूदित कार्पस की अलग-अलग प्रदान की जाएगी। इस कॉलम में एकाधिक प्रयोक्ता/पुरोधा एक साथ आपस में विचार-विमर्श के पश्चात एक सटीक निर्णय तक पहुँच सकते हैं।

## 11. **व्याकरणिक-कोश निर्माण पोर्टल :**

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**1. प्रविष्टि :** हिंदी एवं अन्य भारतीय भाषाओं के संबंधित प्रविष्टि को लिखा जाएगा।

**2. रोमन उच्चारण :** हिंदी एवं अन्य भारतीय भाषाओं के रोमन उच्चारण को लिखा जाएगा।

**3. आई.पी.ए. :** हिंदी एवं अन्य भारतीय भाषाओं के आई.पी.ए. को लिखा जाएगा।

**4. वर्तनीगत प्राप्त अन्य रूप :** हिंदी एवं अन्य भारतीय भाषाओं के वर्तनीगत प्राप्त अन्य रूप को लिखा जाएगा।

5. **चित्र/चलचित्र :** प्रविष्टि संबंधी चित्र/चलचित्र की प्रविष्टि की जाएगी।

**6. वाक:** हिंदी एवं अन्य भारतीय भाषाओं के वाक की प्रविष्टि की जाएगी।

**7. शब्द-वर्ग :** हिंदी एवं अन्य भारतीय भाषाओं के शब्द-वर्ग को लिखा जाएगा।

**8. व्याकरणिक-कोटि :** संबंधित प्रविष्टि की व्याकरणिक-कोटि को लिखा जाएगा।

**9. अर्थीय घटक** : संबंधित प्रविष्टि के अर्थीय घटक को लिखा जाएगा।

**10. पर्यायवाची :** संबंधित प्रविष्टि की पर्यायवाची को लिखा जाएगा।

**11. विलोम :** संबंधित प्रविष्टि के विलोम शब्दों को लिखा जाएगा।

**12. व्युत्पत्ति :** संबंधित प्रविष्ट की व्युत्पत्ति को लिखा जाएगा।

**13. व्युत्पत्तिगत विश्लेषण** : संबंधित प्रविष्टि के व्युत्पत्तिगत विश्लेषण को लिखा जाएगा।

**14. रूपसाधक रूपिम** : संबंधित प्रविष्टि के सभी रूपसाधक रूपिम को लिखा जाएगा।

**15. व्युत्पादक रूपिम** : संबंधित प्रविष्टि के सभी व्युत्पादक रूपिम को लिखा जाएगा।

**16. अर्थ:** संबंधित प्रविष्टि के प्राप्त: सभी अर्थों को लिखा जाएगा।

**17. कार्पस-प्रयोग :** संबंधित प्रविष्टि की प्राप्त विभिन्न कार्पस-प्रयोग को लिखा जाएगा।

**18. कार्पस-संदर्भ :** संबंधित प्रविष्टि की विभिन्न कार्पस-संदर्भ को लिखा जाएगा।

**19. संदर्भ-पृष्ठ :** संबंधित प्रविष्टि के कार्पस-पृष्ठ की चित्र का चयन किया जाएगा।

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

**20. टिप्पणी :** व्याकरणिक-कोश संबंधी प्राप्त किसी भी प्रकार की समस्या, सुझाव अथवा साफ्टवेयर में अनुपलब्ध सुविधा को टिप्पणी के तौर पर लिखा जाएगा।

**21. प्रविष्टि-कर्ता :** साफ्टवेयर के माध्यम से मुख्य डाटाबेस में प्रविष्टि-कर्ता के तौर पर (पहली बार बनाई गई प्रविष्टि एवं अन्य सभी सूचनाओं के साथ) अधिकृत व्यक्ति का नाम सुरक्षित किया जाएगा।

22. **प्रविष्टि-तिथि :** साफ्टवेयर के द्वारा पहली बार बनाई गई प्रविष्टि की दिनांक को सुरक्षित किया जाएगा।

**23. संशोधन-कर्ता :** साफ्टवेयर के माध्यम से डाटाबेस में संग्रहित प्रविष्टि में किसी भी प्रकार की संशोधन करने के पश्चात् संशोधन-कर्ता के तौर पर व्यक्ति का नाम सुरक्षित किया जाएगा। प्रविष्टि-कर्ता किसी भी प्रविष्टि को संशोधित कर सकता है।

**24. संशोधन-तिथि :** साफ्टवेयर के द्वारा प्रविष्टि में दूसरी बार संशोधित करने पर संशोधन दिनांक को सुरक्षित किया जाएगा।

**25. निरीक्षण-कर्ता** : निरीक्षण-कर्ता के द्वारा प्रविष्टि का निरीक्षण करने के पश्चात् निरीक्षण-कर्ता के नाम को सुरक्षित किया जाएगा।

**26. निरीक्षण-तिथि :** साफ्टवेयर के द्वारा प्रविष्टि की निरीक्षण करने पर निरीक्षण दिनांक को सुरक्षित किया जाएगा।

**27. संपादन-कर्ता :** संपादन-कर्ता के द्वारा साफ्टवेयर के माध्यम से डाटाबेस में प्रविष्टि को संपादित करने के पश्चात् संपादन-कर्ता का नाम सुरक्षित किया जाएगा।

**28. संपादन-तिथि :** साफ्टवेयर के द्वारा प्रविष्टि का संपादन करने पर संपादन दिनांक को सुरक्षित किया जाएगा।

**29. संपादन-स्थिति :** संपादन-कर्ता द्वारा संपादित की गई प्रविष्टि की स्थिति 'संपादित' के रूप में सुरक्षित कर लिया जाएगा।

**30. प्रसारण-स्थिति :** प्रविष्टि के संपादित होने के पश्चात् प्रसारण स्थिति को 'प्रसारित' के रूप में सुरक्षित कर लिया जाएगा। प्रसारण-स्थिति की प्रक्रिया के पूर्ण होने के पश्चात संबंधित प्रविष्टि ऑनलाइन के माध्यम से सार्वजनिक तौर पर प्रदर्शित होनी शुरू हो जाएगी।

**31. लॉगइन-आईडी :** प्रयोग-कोश के साफ्टवेयर में आधिकारिक रूप से नई प्रविष्टि करने, प्रविष्टि में संशोधन करने, निरीक्षण करने अथवा संपादन करने हेतु

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

लॉगइन आईडी प्रदान की जाएगी। लॉगइन आईडी के रूप में ई-मेल आईडी का प्रयोग किया जाएगा।

**32. पॉसवर्ड :** उक्त सभी आधिकारिक कार्य को पूर्ण करने हेतु एक पॉसवर्ड प्रदान किया जाएगा, अधिकृत व्यक्ति अपनी सुविधानुसार अपने पॉसवर्ड में बदलाव कर सकता है।

## 12. अनुवाद प्रबंधन प्रणाली :

एक ऐसे ऑनलाइन मंच का निर्माण किया जाएगा, जिसमें अनुवाद किए जाने वाले साहित्य का चयन करके अनुवादक द्वारा चुने गए संयोजन के आधार पर हिंदी से अन्य भारतीय भाषाओं तथा इसके विपरीत क्रम में मानवीय अनुवाद करने की सुविधा प्रदान की जाएगी। यहाँ पर अनुवाद किए जाने वाले पृष्ठ पर स्रोत अनुवाद सामग्री पहले से मौजूद होगी एवं नीचे लक्ष्य-भाषा में अनुवाद करने हेतु एक टेक्स्ट इडीटर का विकल्प मौजूद होगा। जहाँ पर अनुवादक स्रोत-भाषा के पाठ को लक्ष्या-भाषा के पाठ के रूप में टंकित कर सकेगा। टंकण के तत्पश्चात 'प्रविष्ट करें' नामक विकल्प की सहायता से अपने पाठ को डाटाबेस में संग्रहित कर सकेगा।

## I. स्रोत भाषा का चयन :

सर्वप्रथम स्रोत-भाषा के रूप में अन्य भारतीय भाषाओं में उपलब्ध साहित्य को हिंदी में अनुवाद कार्य हेतु प्रथम वरीयता दी जाएगी एवं तत्पश्चात हिंदी से अन्य भारतीय भाषाओं में अनुवाद पूर्ण किया जाएगा। अनुवाद कार्य पूर्ण करने एवं स्रोत-भाषा का चयन करने हेतु सर्वप्रथम अनुवादक के द्वारा निम्न चरण प्रक्रिया को पूर्ण करना आवश्यक होगा :

### i. स्रोत/लक्ष्य भाषा :

| क्र. | स्रोत/लक्ष्य भाषा | |
|---|---|---|
| 1. | मराठी-हिंदी | हिंदी-मराठी |
| 2. | गुजराती-हिंदी | हिंदी-गुजराती |
| 3. | मैथिली-हिंदी | हिंदी-मैथिली |
| 4. | संस्कृत-हिंदी | हिंदी-संस्कृत |

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

| | | |
|---|---|---|
| 5. | ओडिया-हिंदी | हिंदी-ओडिया |
| 6. | आसामी-हिंदी | हिंदी-आसामी |
| 7. | बंगाली-हिंदी | हिंदी-बंगाली |
| 8. | बोडो-हिंदी | हिंदी-बोडो |
| 9. | डोंगरी-हिंदी | हिंदी-डोंगरी |
| 10. | कन्नड-हिंदी | हिंदी-कन्नड |
| 11. | कश्मीरी-हिंदी | हिंदी-कश्मीरी |
| 12. | कोंकणी-हिंदी | हिंदी-कोंकणी |
| 13. | मलयालम-हिंदी | हिंदी-मलयालम |
| 14. | मणिपुरी-हिंदी | हिंदी-मणिपुरी |
| 15. | नेपाली-हिंदी | हिंदी-नेपाली |
| 16. | पंजाबी-हिंदी | हिंदी-पंजाबी |
| 17. | संताली-हिंदी | हिंदी-संताली |
| 18. | सिंधी-हिंदी | हिंदी-सिंधी |
| 19. | तमिल-हिंदी | हिंदी-तमिल |
| 20. | तेलुगु-हिंदी | हिंदी-तेलुगु |

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023

| | | |
|---|---|---|
| 21. | उर्दू-हिंदी | हिंदी-उर्दू |
| 22. | विदेशी भाषा-हिंदी | हिंदी-विदेशी भाषा |

तालिका क्र. 07

ii. **लेखक/रचनाकार का नाम :** अनुवादक को स्रोत-भाषा का चयन करने के पश्चात् लेखक/रचनाकार के नाम का चयन करना होगा।

**iii. कृति/रचना का प्रकार :**

a. प्रयोजन मूलक हिंदी (शोध-ग्रंथ)

b. कहानी

c. उपन्यास

d. व्यंग्य

e. आलोचना

**iv. रचना का नाम :** रचनाकार/लेखक द्वारा रचित कृति के नाम का उल्लेख करना आवश्यक होगा। यदि अनुवादक कार्पस का चयन अपनी इच्छानुसार करना चाहे तो चयन करने का विकल्प प्रदान किया जाएगा।

v. **प्रकाशक :** अनुवादक को कृति के प्रकाशक के नाम का उल्लेख करना होगा।

vi. **प्रकाशन वर्ष :** कृति के प्रकाशन वर्ष का उल्लेख करना होगा।

उपर्युक्त साहित्य में बतौर स्रोत-भाषा में जब प्रयोजनमूलक साहित्य का चयन किया जाएगा, तब अनुवादकर्ता को अधिकतम 50 पृष्ठ के पाठ को उपलब्ध किया जाएगा। परंतु जब कहानी, उपन्यास एवं व्यंग्य का चयन किया जाएगा, तब पृष्ठ आवश्यक न होकर संपूर्ण संदर्भ आवश्यक हो जाएगा।

**vii. समय-सीमा निर्धारण :**

अनुवाद कार्य को संपन्न करने हेतु एक निश्चित समय सीमा का निर्धारण अत्यंत आवश्यक है, जिससे अनुवाद कार्य को तेजी से अग्रसर किया जा सके। समय सीमा का निर्धारण संपादक महोदय के द्वारा स्रोत-सामग्री के आधार पर किया जाएगा। निर्धारित समय सीमा के अंदर ही अनुवादक अपने द्वारा संपन्न किए गए कार्य को ऑनलाइन ही प्रविष्ट कर सकेगा।

## II. सुविधाएँ :

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

इस परियोजना में आधिकारिक तौर पर कार्य करने वाले अथवा अन्य व्यक्ति जो इस कार्य के प्रति इच्छुक हैं, उन्हें यह प्रबंधन प्रणाली कुछ अधिकार प्रदान करेगा, जो निम्नवत् होंगे :

i. अनुवादक एवं संशोधन-कर्ता को प्रदान की जाने वाली रचना के अनूदित सामग्री को प्रविष्ट करने, संशोधन करने एवं स्वअनूदित पाठ को हटाने की सुविधा प्रदान की जाएगी।
ii. निरीक्षण-कर्ता को अनुवादक द्वारा अनूदित सामग्री को निरीक्षण करने, संशोधन करने, स्वयं द्वारा अनूदित पाठ को प्रविष्ट करने एवं स्वअनूदित पाठ को हटाने की सुविधा प्रदान की जाएगी।
iii. संपादनकर्ता को उक्त संबंधित सभी सुविधाएँ प्रदान की जाएगी।
iv. अनुवादक अपने गत-पृष्ठ पर स्वयं के द्वारा पूर्ण किए गए कार्य को तालिका के माध्यम से देख सकता है।

III. **अनुभव प्रमाण-पत्र :** कार्पस का अनुवाद कार्य करने वाले सभी सदस्यों को निर्धारित समय सीमा में पूर्ण किए गए कार्य का ई-अनुभव प्रमाण-पत्र भी प्रदान किया जाएगा।

**IV. अनुवाद-पृष्ठ में उपलब्ध सुविधाएँ :** अनुवादक के व्यक्तिगत-पृष्ठ पर अनुवाद करने एवं अनूदित सामग्री को डाटाबेस में संग्रहित करने हेतु कुछ आवश्यवक मूल-भूत सुविधाएँ प्रदान की जाएगी, जो निम्नवत् होंगी:

i. **प्रविष्टि करें :** अनुवादक द्वारा अनूदित सामग्री को साफ्टवेयर के माध्यम से टंकित करने हेतु 'प्रविष्ट करें' नामक विकल्प का चयन करना होगा, तत्पश्चात अनूदित सामग्री डाटाबेस में सुरक्षित हो जाएगी।
ii. **संशोधन करें :** अनुवादक/संशोधन-कर्ता को अनूदित सामग्री को संशोधित करने हेतु 'संशोधन करें' नामक विकल्प का चयन करना होगा, तत्पाश्चात् डाटाबेस में संग्रहित अनूदित सामग्री का अद्यतन होकर डाटाबेस में संग्रहित हो जाएगी।
iii. **निरीक्षण करें :** संशोधित अनूदित सामग्री का निरीक्षण-कर्ता के द्वारा निरीक्षित करने हेतु 'निरीक्षण करें' नामक विकल्प का चयन करना होगा, तत्पश्चात डाटाबेस में निरीक्षित सामग्री का अद्यतन हो जाएगा।
iv. **हटाएँ :** यदि अनूदित सामग्री को हटाना चाहें तो 'हटाएँ' नामक विकल्प का चयन करना होगा, तत्पश्चात् उक्त सामग्री मुख्य डाटाबेस से हट जाएगी।
v. **टिप्पणी :** अनुवादक/संशोधन-कर्ता/निरीक्षण-कर्ता के द्वारा अनुवाद संबंधी किसी भी प्रकार समस्या, सुझाव अथवा साफ्टवेयर में अनुपलब्ध सुविधा को टिप्पणी के तौर पर लिखा जाएगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

V. **अनुवादक :** साफ्टवेयर के माध्यम से मुख्य डाटाबेस में अनुवादक के तौर पर (पहली बार अनुवाद किए गए पाठ) अधिकृत किए गए व्यक्ति का नाम सुरक्षित किया जाएगा।

VI. **अनुवाद-तिथि :** साफ्टवेयर के द्वारा पहली बार अनुवाद किए गए पाठ के दिनांक को सुरक्षित किया जाएगा।

**VII. संशोधन-कर्ता :** साफ्टवेयर के माध्यम से डाटाबेस में संग्रहित प्रविष्टि में किसी भी प्रकार का संशोधन करने के पश्चात संशोधन-कर्ता के नाम को सुरक्षित किए जाएगा। अनुवादक किसी भी अनुवाद को संशोधित कर सकता है।

**VIII. संशोधन-स्थिति** : साफ्टवेयर के द्वारा अनुवाद किए गए पाठ में दूसरी बार संशोधित करने पर संशोधन दिनांक को सुरक्षित किया जाएगा।

**IX. निरीक्षण-कर्ता :** निरीक्षण-कर्ता के द्वारा अनुवाद का निरीक्षण करने के पश्चात निरीक्षण-कर्ता के नाम को सुरक्षित किया जाएगा।

**X. निरीक्षण-तिथि :** साफ्टवेयर के द्वारा अनुवाद का निरीक्षण करने पर निरीक्षण दिनांक को सुरक्षित किया जाएगा।

**XI. संपादन-कर्ता :** संपादन-कर्ता के द्वारा साफ्टवेयर के माध्यम से डाटाबेस में अनूदित पाठ को संपादित करने के पश्चात संपादन-कर्ता का नाम सुरक्षित किया जाएगा।

**XII. संपादन-तिथि :** साफ्टवेयर के द्वारा अनुवाद के पाठ का संपादन करने पर संपादन दिनांक को सुरक्षित किया जाएगा।

XIV. **संपादन-स्थिति :** संपादन-कर्ता द्वारा संपादित की गई अनुवाद-सामग्री की स्थिति को 'संपादित' के रूप में सुरक्षित कर लिया जाएगा।

XV. **प्रसारण-स्थिति :** अनुवाद कार्य के संपादित होने के पश्चात् प्रसारण स्थिति को 'प्रसारित' के रूप में सुरक्षित कर लिया जाएगा। प्रसारण संबंधित स्थिति की प्रक्रिया के पूर्ण होने के पश्चात अनूदित पाठ सामग्री ऑनलाइन के माध्यम से सार्वजनिक तौर पर प्रदर्शित होनी शुरू हो जाएगी।

**XV. पंजीकरण :** अनुवाद कार्य करने वाले अधिकृत व्यक्तियों को पंजीकरण करना आवश्यक होगा, पंजीकरण प्रपत्र में अनुवादक की व्यक्तिगत जानकारी निम्नवत् होगी :

i. **नाम:** अनुवादक का नाम
ii. **संस्थान का नाम:** अनुवादकर्ता के संबंधित संस्थान का नाम होगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

iii. **विषय विशेषज्ञता :** अनुवादकर्ता अपने विषय-क्षेत्र (हिंदी साहित्यक, विज्ञान, सामाजिक विज्ञान आदि) संबंधित जानकारी।

iv. **भाषा-संयोजन :** अनुवादक को भाषा-संयोजन का उल्लेख करेगा, जिसका विवरण स्रोत-भाषा के अंतर्गत दिया गया है।

v. **लॉगइन-आईडी :** व्याकरणिक-कोश एवं अनुवाद प्रबंधन प्रणाली दोनों के लिए अलग-अलग लॉगइन-आईडी का निर्धारण किया जाएगा। इस प्रणाली में आधिकारिक रूप से कार्य करने वाले सभी व्यक्तियों को लॉगइन आईडी प्रदान की जाएगी। लॉगइन आईडी के रूप में ई-मेल आईडी का प्रयोग किया जाएगा।

vi. **पासवर्ड :** उक्त सभी आधिकारिक कार्य को पूर्ण करने हेतु एक पॉसवर्ड प्रदान किया जाएगा, अधिकृत व्यक्ति अपनी सुविधानुसार अपने पॉसवर्ड में बदलाव भी कर पाएगा।

## 13. भारतीय संस्कृति कोश (Indian Cultural Dictionary):

किसी भी भाषा में समय, काल, संस्कृति की सापेक्षता हमेशा से जुड़ी रही है, जो भाषा को संदर्भगत बनाकर अर्थ को प्रकट करती है। भाषिक व्यवहार में अभिव्यक्ति के संदर्भ और सापेक्षताएं ही भाषा में अर्थ प्रदान करती है, जहां संदर्भ एवं संस्कृति के बदलने पर व्यक्तियों के अर्थबोध भी बदल जाते हैं। भाषा संस्कृति की संवाहक भी होती है, अतएव भारतीय भाषाओं के सम्यक विश्लेषण हेतु भारतीय सांस्कृतिक पहलुओं का भी संरक्षण, संवर्धन एवं सर्वेक्षण आवश्यक हो जाता है। सभी भारतीय भाषाएं अपनी प्रांतीय संस्कृति से भी ओत-प्रोत होती हैं। इनका संबंध भारतीय सभ्यता एवं परंपरा से है। भारतीय भाषाएं भी भारतीय संस्कार से संस्कारित हैं। भाषा के अर्थ को प्रभावित करने वाले कारकों में सम्मिलित होने वाले कारक धर्म, जाति, भोजन, पहनावा, नृत्य आदि होते हैं। चूँकि भाषा संस्कृति को अपने अंदर समाहित किए रहती है, इसलिए इसे विलगाने का अर्थ है; भाषा की अपूर्णता को दर्शाना। प्रौद्योगिकी में भी भाषा के संपूर्ण आयाम को अनुप्रयुक्त करने हेतु संग्रहित व भंण्डारित करना भी आवश्यक प्रक्रिया है, जिसे केंद्रित करते हुए एक ऐसे डाटाबेस का निर्माण किया जाएगा, जिसमें भाषिक संस्कृति के सभी कारक तत्व समाहित हो सकें। इस सांस्कृतिक डाटाबेस की उपयोगिता न सिर्फ हिंदी एवं अन्य भारतीय भाषाओं के विश्लेषण व संश्लेषण करने में भी होगी अपितु प्राकृतिक भाषा संसाधन के महत्वपूर्ण भाषिक अनुप्रयोगों के निर्माण कार्य में भी सहायक सिद्ध होगा।

**उद्देश्य:** इस कोश के उद्देश्य निम्नवत् होंगे :

1. भारतीय संस्कृति के विभिन्न कारक तत्वों का वृहद डाटाबेस तैयार करना।
2. विभिन्न भारतीय संस्कृति की विविधता का अध्ययन व विश्लेषण करना।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

3. भारतीय जातियों का सर्वेक्षण करना।

4. भारतीय बोली/भाषा का सर्वेक्षण करना।

5. भारतीय लोकगीत/लोकसाहित्य का संकलन करना।

6. भारतीय संस्कृति के सभी कारकों का गहन अध्ययन व विश्लेषण करना।

7. भारतीय जातिगत इतिहास का पुनर्लेखन करना।

8. भारतीय संस्कृति का भारतीय भाषा को प्रभावित करने वाले कारक तत्वों का विश्लेषण करना।

9. भारतीय संस्कृति का ऐतिहासिक व तुलनात्मक वर्णन करना।

10. पारंपरिक भारतीय संस्कृति का ऐतिहासिक वर्गीकरण करना।

11. प्राकृतिक भाषा संसाधन के अनुप्रयोगों में भारतीय भाषा के विश्लेषण एवं संश्लेषण में सांस्कृतिक शब्दों /पदों का प्रयोग करना।

## 14. भारतीय संस्कृति कोश पोर्टल :

भारतीय संस्कृति कोश को निर्माण करने हेतु एक ऑनलाइन सॉफ्टवेयर का निर्माण किया जाएगा, जिसकी सहायता से समस्त सूचनाएं सम्मिलित कर संग्रहित व भण्डारित की जाएगी। प्रस्तुत भारतीय संस्कृति कोश में अन्य सांस्कृतिक कारक तत्व के प्राप्त की परिस्थिति में उसे भी शामिल किया जाएगा। निम्न कारक तत्वों में यदि किसी ऐसी भारतीय संस्कृति जिसमें वह कारक तत्व लागू नहीं होता है, उसे रिक्त छोड़ दिया जाएगा। इस भारतीय सांस्कृातिक कोश को भी ऑनलाइन प्रारूप में सार्वजनिक रूप में प्रस्तुत किया जाएगा, जिसका प्रारूप 25% डाटाबेस पूर्ण होने के पश्चात् निर्माण किया जाएगा। इस कोश में भी भाषा संबंधी मानकता को पूर्णत: लागू किया जाएगा। यहां पर लिपि की समस्या आने पर देवनागरी लिपि को ही मानक माना जाएगा, किंतु आवश्यकतानुसार अन्य लिपि का भी समानांतर रूप से प्रयोग किया जा सकता है। इस कोश के अंतर्गत आवश्यकतानुसार प्रविष्टि आधारित रंग, रूप, ऊचाई, चौड़ाई आदि या अन्य शारीरिक विविधता, आवास, पेय पदार्थ, ग्रामीण/शहरी या तौर-तरीका आदि एवं यदि भारत के इतर प्रवासी मूल साहित्य या प्रवासी सांस्कृतिक सर्वेक्षण प्राप्त होता है, तो इन दोनों के आधार पर भारतीय संस्कृति के कारक तत्वों संबंधी अन्य सुविधाएं सुलभ कराई जाएगी। इस साफ्टवेयर का प्रारूप निम्न प्रकार से होगा :

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**1. धर्म :** भारतीय देश में पाई जाने वाले सभी धर्मो की प्रविष्टि की जाएगी, जैसे: हिंदू, मुस्लिम, सिक्ख, ईसाई आदि।

**2. मुख्य जाति:** भारतीय संस्कृति के आधार पर मुख्य जाति की प्रविष्टि की जाएगी, जैसे: ब्राह्मण, क्षत्रिय आदि।

**3. उपजाति:** भारतीय संस्कृाति के आधार पर उपजाति की प्रविष्टि की जाएगी, जैसे: तिवारी, द्विवेदी आदि।

**4. राष्ट्र:** हिंदी एवं अन्य भारतीय भाषाओं के बोलने वाले व्यक्तियों के राष्ट्र का उल्लेख करते हुए प्रविष्टि की जाएगी, जैसे: भारत, नेपाल, सूरीनाम आदि।

**5. प्रदेश:** तत्पश्चात् देश के आधार पर प्रदेश की प्रविष्टि की जाएगी, जैसे: महाराष्ट्र, मध्यप्रदेश आदि।

**6. भाषा/बोली :** संबंधित प्रविष्टि के आधार पर भाषा अथवा बोली की प्रविष्टि की जाएगी।

**7. भोजन:** भारतीय संस्कृति के आधार पर मुख्य भोजन की प्रविष्ट की जाएगी, जैसे: चावल, दाल, रोटी, सब्जी, मछली, चावल, इडली, सांभर आदि।

**8. वस्त्र** : भारतीय संस्कृति के आधार पर मुख्य वस्त्र की प्रविष्टि की जाएगी, जैसे: कुर्ता, पजामा/खादी वस्त्र आदि।

**9. संस्कार:** भारतीय संस्कृति के आधार पर जन्म से लेकर मृत्यु तक के सभी संस्कारों की प्रविष्टि की जाएगी,जैसे: बरहौ संस्कार, विवाह संस्कार आदि।

**10. लोकनृत्य:** भारतीय संस्कृति के आधार पर लोकनृत्य की प्रविष्टि की जाएगी।

**11. लोक-गीत :** भारतीय संस्कृति के आधार पर सभी लोकगीत की प्रविष्टि की जाएगी, जैसे: सोहर, दादर, सोहाग आदि।

**12. लोक-कथा:** संबंधित लोक-कथा की प्रविष्टि की जाएगी।

**13. लोक-कहानी:** भारतीय संस्कृति के आधार पर संबंधित लोक-कहानी की प्रविष्टि की जाएगी।

**14. लोक-साहित्य:** भारतीय संस्कृति के आधार पर लोक-साहित्य की प्रविष्टि की जाएगी।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**15. अराध्य-देवी/देवता :** भारतीय संस्कृति के आधार पर अराध्य-देवी अथवा अराध्य देवता की प्रविष्टि की जाएगी।

**16. पारंपरिक पेशा:** भारतीय संस्कृति के आधार पर पारंपरिक पेशा की प्रविष्टि की जाएगी।

**17. खेल:** भारतीय संस्कृति के आधार पर पारंपरिक खेल की प्रविष्टि की जाएगी।

**18. धर्मग्रंथ:** भारतीय संस्कृति के आधार पर धर्मग्रंथ की प्रविष्टि की जाएगी।

**19. विवाह:** भारतीय संस्कृति के आधार पर विवाह के प्रकार की प्रविष्टि की जाएगी।

**20. हस्तशिल्प:** भारतीय संस्कृति के आधार पर संबंधित सभी हस्तशिल्प की प्रविष्टि की जाएगी।

**21. हस्तकला:** भारतीय संस्कृति के आधार पर संबंधित हस्तकला की प्रविष्टि की जाएगी।

**22. त्योहार:** भारतीय संस्कृति के आधार पर संबंधित त्योहार की प्रविष्टि की जाएगी।

**23. पारंपरिक मेला:** भारतीय संस्कृति के आधार पर संबंधित पारंपरिक मेला की प्रविष्टि की जाएगी।

**24. लोकनाट्य:** भारतीय संस्कृति के आधार पर संबंधित पारंपरिक लोकनाटक की प्रविष्टि की जाएगी।

**25. टिप्पणी:** उक्त सभी सांस्कृतिक कारक तत्वों से संबंधित प्राप्त किसी भी प्रकार की समस्या, सुझाव अथवा साफ्टवेयर में अनुपलब्ध सुविधा को टिप्पणी के तौर पर लिखा जाएगा।

**26. प्रविष्टि-कर्ता :** साफ्टवेयर के माध्यम से मुख्य डाटाबेस में प्रविष्टि-कर्ता के तौर पर (पहली बार बनाई गई प्रविष्टि) अधिकृत व्यक्ति का नाम सुरक्षित किया जाएगा।

**27. प्रविष्टि-तिथि :** साफ्टवेयर के द्वारा पहली बार बनाई गई प्रविष्टि की दिनांक को सुरक्षित किया जाएगा।

**28. संशोधन-कर्ता :** साफ्टवेयर के माध्यम से डाटाबेस में संग्रहित प्रविष्टि में किसी भी प्रकार की संशोधन करने के पश्चात् संशोधन-कर्ता के तौर पर व्यक्ति का नाम सुरक्षित किया जाएगा। प्रविष्टि-कर्ता किसी भी प्रविष्टि को संशोधित कर सकता है।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

**29. संशोधन-तिथि :** साफ्टवेयर के द्वारा प्रविष्टि में दूसरी बार संशोधित करने पर संशोधन दिनांक को सुरक्षित किया जाएगा।

**30. निरीक्षण-कर्ता** : निरीक्षण-कर्ता के द्वारा प्रविष्टि का निरीक्षण करने के पश्चात् निरीक्षण-कर्ता के नाम को सुरक्षित किया जाएगा।

**31. निरीक्षण-तिथि :** साफ्टवेयर के द्वारा प्रविष्टि की निरीक्षण करने पर निरीक्षण दिनांक को सुरक्षित किया जाएगा।

**32. संपादन-कर्ता :** संपादन-कर्ता के द्वारा साफ्टवेयर के माध्यम से डाटाबेस में प्रविष्टि को संपादित करने के पश्चात् संपादन-कर्ता का नाम सुरक्षित किया जाएगा।

**33. संपादन-तिथि :** साफ्टवेयर के द्वारा प्रविष्टि का संपादन करने पर संपादन दिनांक को सुरक्षित किया जाएगा।

**34. संपादन-स्थिति :** संपादन-कर्ता द्वारा संपादित की गई प्रविष्टि की स्थिति 'संपादित' के रूप में सुरक्षित कर लिया जाएगा।

**35. प्रसारण-स्थिति :** प्रविष्टि के संपादित होने के पश्चात् प्रसारण स्थित को 'प्रसारित' के रूप में सुरक्षित कर लिया जाएगा। प्रसारण-स्थिति की प्रक्रिया के पूर्ण होने के पश्चात संबंधित प्रविष्टि ऑनलाइन के माध्यम से सार्वजनिक तौर पर प्रदर्शित होनी शुरू हो जाएगी।

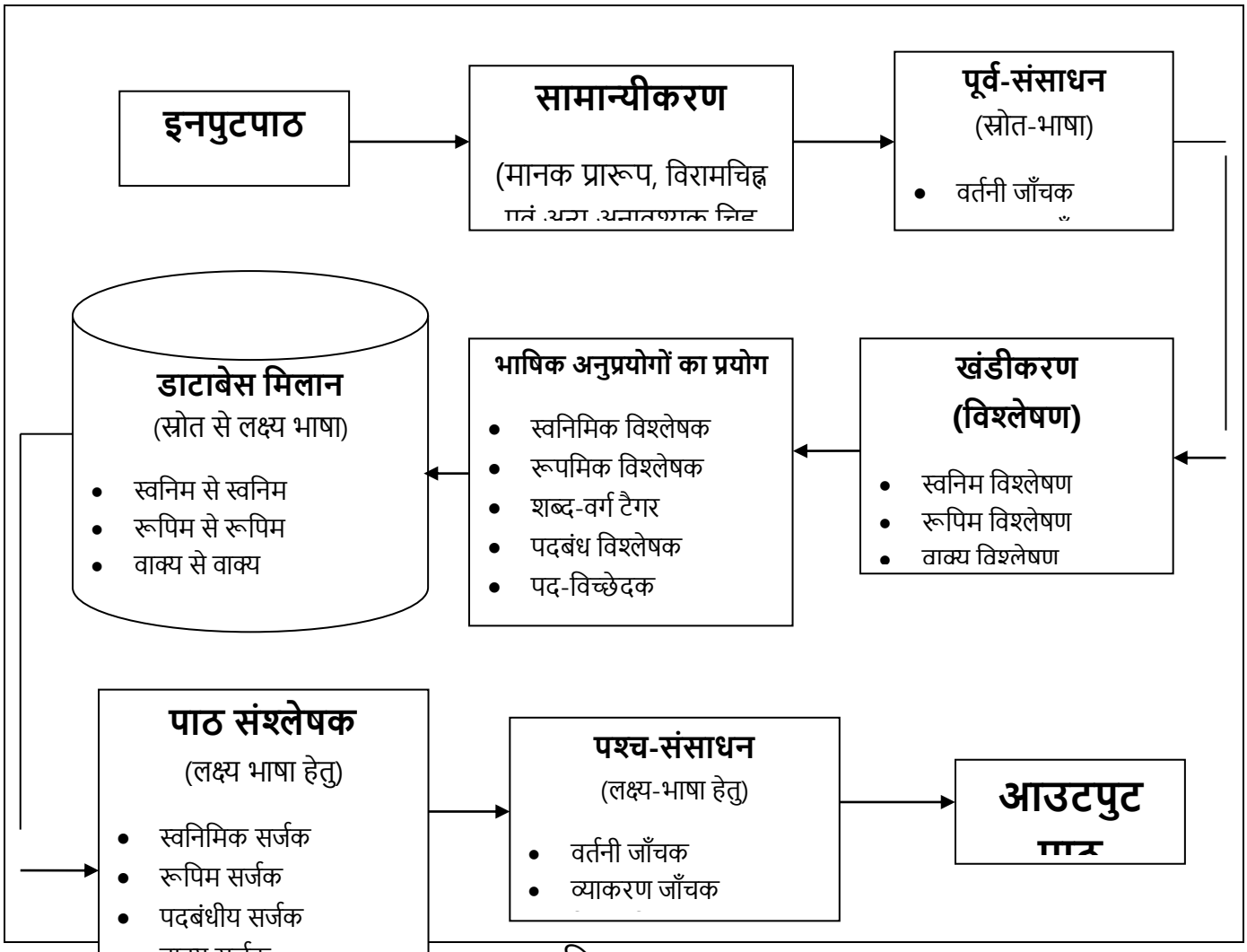
**36. लॉगइन-आईडी :** भारतीय सांस्कृतिक कोश के साफ्टवेयर में आधिकारिक रूप से नई प्रविष्टि करने, प्रविष्टि में संशोधन करने, निरीक्षण करने अथवा संपादन करने हेतु लॉगइन आईडी प्रदान की जाएगी। लॉगइन आईडी के रूप में ई-मेल आईडी का प्रयोग किया जाएगा।

**37. पॉसवर्ड :** उक्त सभी आधिकारिक कार्य को पूर्ण करने हेतु एक पॉसवर्ड प्रदान किया जाएगा, अधिकृत व्यक्ति अपनी सुविधानुसार अपने पॉसवर्ड में बदलाव कर सकता है।

15. **मशीन अनुवाद प्रक्रिया :**

| **मशीन अनुवाद प्रणाली की प्रक्रिया** |
|---|
| |

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023




तालिका क्र. 08

i. **इनपुट पाठ (स्रोत-भाषा) :** यह मशीन अनुवाद प्रणाली का प्रथम चरण होगा, जिसकी सहायता से स्रोत-भाषा (हिंदी अथवा अन्य भारतीय भाषा) के पाठ को इनपुट किया जाएगा।

ii. **पाठ सामान्यीकरण :** इस चरण में पाठ का सामान्यीकरण, वर्तनीगत एकरूपता/मानकीकरण, विराम चिह्न आदि को आधार पर पाठ को संसाधन हेतु त्रुटिविहीन रूप में निर्मित करना।

iii. **पूर्व-संसाधन :** इस चरण में स्रोत-भाषा के पाठ को वर्तनी जाँचक एवं व्याकरण जाँचक द्वारा त्रुटिविहीन करना।

iv. **पाठ विश्लेषण** : यह चरण सबसे महत्वपूर्ण चरण होगा, क्योंकि स्रोत-भाषा के भाषिक संरचना के आधार पर सभी स्तर (स्वनिम, रूपिम, पदबंधीय, वाक्य) स्तर पर पाठ का विश्लेषण किया जाएगा।

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

v. **भाषिक अनुप्रयोगों का प्रयोग :** इस चरण में स्रोत-पाठ के भाषिक अनुप्रयोग जैसे : स्वनिम विश्लेषक/संश्लेषक, रूपमिक विश्लेषक/संश्लेषक, पदबंधीय विश्लेषक/संश्लेषक, शब्द-वर्ग टैगर, नामपद अभिज्ञानक, पद-विच्छेदक आदि का प्रयोग किया जाएगा।

vi. **पाठ मिलान :** इस चरण में विश्लेषित स्रोत-भाषा के पाठ को मुख्य डाटाबेस में संग्रहित लक्ष्य-भाषा के पाठ से मिलान किया जाएगा। यहां पर विश्लेषित पाठ का स्वनिम से स्वनिम, रूपिम से रूपिम, पदबंध से पदबंध, वाक्य से वाक्य एवं अर्थ से अर्थ का मिलान किया जाएगा। अर्थीय स्तर के मिलान हेतु अर्थों इंडेक्स मॉडल तैयार किया जाएगा, जिसकी सहायता से लक्ष्य अर्थ को प्राप्त किया जाएगा।

vii. **पाठ-पुनर्गठन** : इस चरण में प्राप्त विश्लेषित लक्ष्य-भाषा के पाठ का पुनर्गठन किया जाएगा। पुनर्गठन का प्रारूप विश्लेषित वाक्य-संरचना से संश्लेषित वाक्य-संरचना के अनुसार किया जाएगा। अर्थात् स्वनिम से रूपिम, रूपिम से पदबंध, पदबंध से वाक्य के आधार पर पुनर्गठन किया जाएगा।

viii. **पश्च-संसाधन :** इस चरण में लक्ष्य भाषा के पुनर्गठित वाक्य-संरचना के शब्दों का वर्तनी जाँचक एवं व्याकरण जाँचक के प्रयोग से पाठ को शुद्ध एवं मानक बनाया जाएगा, जिससे प्राप्त लक्ष्य–भाषा का पाठ साधारण रूप में प्राप्त हो सके।

ix. **आउटपुट :** इस अंतिम चरण में लक्ष्य–भाषा का पाठ आउटपुट के रूप में प्राप्त होगा, जिसे साधारण रूप में पढ़ा व समझा जा सके।

**16. अग्रसिरा (Front-End) :** ऑनलाइन डाटाबेस प्रबंधन प्रणाली के अग्रसिरा का निर्माण HTML, CSS, ASP.NET, C#.NET, Bootstrap एवं मशीनी आनुवाद प्रणाली के निर्माण हेतु पाइथन (Python) प्रोग्रामिंग भाषा का चयन किया जाएगा, क्यों कि पाइथन प्रोग्रामिंग भाषा स्ट्रिंग मैनिपुलेशन एवं गणितीय प्रविधि दोनों को एक साथ अनुप्रयुक्त करने की सुविधा प्रदान करती है, जिसके माध्ययम से भाषा का संसाधन सहजता के साथ किया जा सकता है।

**17. पश्च-सिरा (Back-End) :** डाटा का संकलन एवं विश्लेषण करने के पश्चात डाटा का अंत:संबंध एवं पुर्नप्राप्ति हेतु डाटाबेस प्रबंधन का कार्य किया जाएगा। चूँकि यह डाटाबेस प्रबंधन प्रणाली पूर्णत: ऑनलाइन प्रणाली है, इसलिए डाटा के प्रबंधन कार्य हेतु माइक्रोसॉफ्ट एसक्यूएलसर्वर (M.S. SQL Server) नाम सॉफ्टवेयर का प्रयोग किया जाएगा। इस सॉफ्टवेयर के माध्यसम से विश्लेषित/संश्लेषित एवं अनूदित सभी प्रकार के डाटा का संग्रहण कार्य आसानी से किया जा सकता है।

**18. निष्कर्ष:**

ISSN PRINT 2319 1775 Online 2320 7876
*Research paper* **© 2012 IJFANS. All Rights Reserved,** **UGC CARE Listed ( Group -I) Journal Volume 12, Iss 1, 2023**

अत: सारांशत: यह कहा जा सकता है कि संकर आधारित अभिगम का प्रयोग करते हुए मशीन अनुवाद प्रणाली का विकास कार्य किया जा सकता है। इस मशीन अनुवाद प्रणाली संगणकीय व्याकरणिक-कोश एवं समानांतर अनूदित कार्पस एक अनिवार्य तत्व होंगे, जिसमें सभी भाषिक घटक मौजूद होंगे। इस कोश एवं कार्पस की सहायता से मशीन अनुवाद में प्रयोग होने वाले सभी अनुप्रयोगों का विकास किया जाएगा, तत्पश्चात मशीन अनुवाद का निर्माण किया जा सकेगा।

## 19. संदर्भ-सूची :

1. गुरु, कामताप्रसाद. (2009). हिंदी व्याकरण. इलाहाबाद: लोकभारती प्रकाशन.

2. सिंह, सूरजभान. (1985). हिंदी का वाक्यात्मक व्याकरण. दिल्ली : साहित्य सहकार प्रकाशन.

3. सिंह, सूरजभान. (2000). अंग्रेजी-हिंदी अनुवाद व्याकरण. दिल्ली : प्रभात प्रकाशन.

4. टेकचंदाणी, रवि प्रकाश. (2016). देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण. नई-दिल्ली: केंद्रीय हिंदी निदेशालय, उच्चातर शिक्षा विभाग. मानव संसाधन विकास मंत्रालय.

5. ओझा, त्रिभुवन. (1986). हिंदी में अनेकार्थकता का अनुशीलन. वाराणसी : विश्वविद्यालय प्रकाशन.

6. मल्होत्रा, विजय कुमार. (2002). कंप्यूटर के भाषिक अनुप्रयोग. नई दिल्ली: वाणी प्रकाशन.

7. द्विवेदी, डॉ. कपिल देव. (2000). अर्थविज्ञान और व्याकरण दर्शन. वाराणसी: विश्वविद्यालय प्रकाशन.

8. King, Margaret. (1983). Parsing Natural Language. London, New York: Academic Press, A Subsidiary of Harcourt Brace Jovanovich.

9. Indurkhya, Nitin; Damerau, Fred J. (2010). Handbook of Natural Language Processing, Second Edition, CRC Press.

10. Hardeniya, Nitin; Perkins, Jacob, Chopra, Deepti; Joshi, Nisheeth; Mathur, Iti. (2016). Natural Language Processing: Python and NLTK. Packt, Mumbai: Birmingham.