

# अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन

ANGREJI-HINDI MASHINI ANUVAD TANTRO KA MULYANKAN

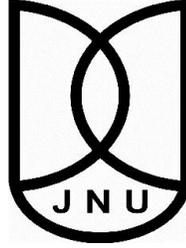
(An Evaluation of English-Hindi Machine Translation Systems)

पीएच.डी. उपाधि हेतु प्रस्तुत शोध-प्रबंध

शोध-निर्देशक  
प्रो.देवेन्द्र कुमार चौबे

सह शोध-निर्देशक  
प्रो.गिरीश नाथ झा

शोधार्थी  
सुमेध खुशालराव हाडके



भारतीय भाषा केंद्र  
भाषा, साहित्य एवं संस्कृति अध्ययन संस्थान  
जवाहरलाल नेहरू विश्वविद्यालय  
नई दिल्ली-110067

2017



जवाहरलाल नेहरू विश्वविद्यालय  
JAWAHARLAL NEHRU UNIVERSITY

भारतीय भाषा केन्द्र

Centre of Indian Languages

भाषा, साहित्य एवं संस्कृति अध्ययन संस्थान  
School of Language, Literature & Culture Studies  
नई दिल्ली-110067, भारत NEW DELHI-110067, INDIA

Dated: 21/07/2017

**DECLARATION**

I hereby declare that the research work done in this Ph.D. Thesis entitled "ANGREJI-HINDI MASHINI ANUVAD TANTRO KA MULYANKAN" (An Evaluation of English-Hindi Machine Translation Systems) submitted by me is an original research work and it has not been previously submitted for any other degree in this or any other university or institution.

Sumedh Khushalrao Hadke  
(Research Scholar)

Prof. Devendra Kumar Choubey  
(SUPERVISOR)

CIL / SLL&CS

JNU New Delhi - 110067



डॉ. देवेन्द्र कुमार चौबे  
Dr. Devendra Kumar Choubey  
प्रोफेसर / Professor  
Centre of Indian Languages  
School of Languages  
Jawaharlal Nehru University  
नई दिल्ली / New Delhi-110067

Prof. Girish Nath Jha  
(CO-SUPERVISOR)

CHAIRPERSON

Special Centre for Sanskrit Studies  
JNU New Delhi - 110067

Prof. Gobind Prasad  
(CHAIRPERSON)

CIL / SLL&CS

JNU New Delhi - 110067

CHAIRPERSON  
CIL/SLL & CS, JNU

## समर्पण

*जीवन में सतत प्रयासरत  
रहने वाले लोगों के लिए..*

## आभार

---

पीएच.डी. शोध एक विशद कार्य होता है, कोई भी शोध कार्य किसी एक व्यक्ति के प्रयास से पूर्ण नहीं होता है। इस शोध कार्य के लिए भी अनेक लोगों से प्रत्यक्ष और अप्रत्यक्ष रूप से सहायता ली गई है।

सर्वप्रथम मैं भारतीय भाषा केंद्र के शिक्षकों का आभार व्यक्त करता हूँ जिन्होंने मुझे शोध-प्रबंध का विषय 'अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन' पर कार्य करने की अनुमति दी।

मैं अपने शोध-निर्देशक प्रो. देवेन्द्र कुमार चौबे, भारतीय भाषा केन्द्र, जे.एन.यू. के प्रति आभार व्यक्त करता हूँ, जिन्होंने शोध विषय के चयन से लेकर इस शोध कार्य को पूरा करने के दौरान समय-समय पर उचित सुझाव एवं मार्गदर्शन किया और इस शोध को पूरा करने के लिए स्वतंत्रता प्रदान की जिससे यह शोध कार्य पूर्ण हुआ है।

अपने सह शोध-निर्देशक प्रो. गिरीश नाथ झा सर के प्रति भी आभार व्यक्त करता हूँ, जिन्होंने इस शोध विषय के चयन के लिए सुझाव दिया और मशीनी अनुवाद मूल्यांकन और त्रुटि विश्लेषण के विभिन्न पहलुओं पर चर्चा की। साथ ही संगणकीय और तकनीकी पक्षों में आनेवाली समस्याओं से अवगत कराया और इस विषय पर मेरी सोच विकसित की। इसके साथ ही इस शोध कार्य के लिए आवश्यक भाषाई सामग्री (कॉर्पोरा) संकलन के लिए विभिन्न वेबसाइट्स तथा संदर्भों के बारे में जानकारी प्रदान की और इस शोध कार्य में आनेवाली समस्याओं का समाधान किया, जिससे यह शोध कार्य पूर्ण हुआ है।

भारतीय भाषा केंद्र, जे.एन.यू. के हिंदी अनुवाद विषय के प्रोफेसर देवशंकर नवीन और डॉ. गंगा सहाय मीणा सर का भी विशेष आभारी हूँ जिन्होंने इस शोध कार्य में प्रत्यक्ष-अप्रत्यक्ष रूप से मार्गदर्शन किया।

इस शोध-प्रबंध को पूरा करने के लिए विभिन्न लोगों की सहायता मिली उनमें मेरे वरिष्ठ मित्र डॉ. अरुण कुमार जिन्होंने इस शोध कार्य के वर्तनी की जांच की और अपने

बहुमूल्य सुझावों से अवगत कराया। साथ ही मुझे शोध के दौरान हमेशा प्रेरित करने में आनंद, डॉ. दिवाकर मिश्रा, पिकी नयनवानी, देवीदीन, संतोष, वासिक, साकेत, हाफिज, मोहन, सद्दे आलम, अतुल ओझा, डॉ. प्रफुल मुन, डॉ. दशरथ कुमार, डॉ. किरन कुंभरे, धिरज थुल आदि लोगों ने सहायता की, इसलिए इनका भी आभार व्यक्त करना चाहूँगा। अपने सहपाठी मित्रों में रमेश कुमारी, अजय, मुन्नी, लखिमा, सेरिंग, दर्शनी, सीमा - सभी का मैं आभार व्यक्त करना चाहता हूँ, जिनका सहयोग इस शोध के दौरान हमेशा मुझे मिलता रहा है।

अपने माता-पिताजी, भाई-बहन और परिजनों के प्रति भी आभार व्यक्त करता हूँ जिनका स्नेह और सहयोग हमेशा मुझे मिलता रहा है, जिससे यह शोध कार्य पूर्ण हुआ है।

भारतीय भाषा केंद्र के रावतजी, रमेश भाई, आशीष का भी आभार व्यक्त करता हूँ। जिन्होंने अपने कार्यालयी कामकाज की व्यस्तताओं के बावजूद मेरे कामकाज में सहायता की तथा इस शोध के दौरान महत्त्वपूर्ण जानकारी देते रहे।

जे.एन.यू. के ग्रंथालय से शोध कार्य संबंधी आवश्यक पुस्तकें कर्मचारियों ने उपलब्ध कराईं इसलिए उन्हें भी धन्यवाद देना चाहूँगा। महात्मा गांधी हिंदी विश्वविद्यालय, वर्धा के ग्रंथालय से भी महत्त्वपूर्ण पुस्तकें उनके कर्मचारियों द्वारा उपलब्ध करायी गयीं इसलिए उनका भी आभार प्रकट करता हूँ।

इस शोध कार्य में कितनी सफलता प्राप्त हुई यह मार्गदर्शक, परीक्षक, अध्येता तथा शोधार्थी ही जान पाएँगे। अगर इस शोध से 'मशीनी अनुवाद मूल्यांकन एवं त्रुटि विश्लेषण' विषय पर कार्य करनेवाले शोधार्थियों, पाठकों, या विद्वज्जनों को कुछ सहायता मिली तो मुझे इस कार्य की सार्थकता और सफलता पर प्रसन्नता होगी।

- सुमेध खुशालराव हाडके

# अनुक्रमणिका

आभार	(iv-v)
अनुक्रमणिका	(vi-xi)
संक्षिप्ताक्षर	(xii-xiv)
टेबल, आकृति और स्क्रीनशॉट की सूची	(xv-xvi)
प्रस्तावना	(xvii-xxii)
<b>प्रथम अध्याय : मशीनी अनुवाद का परिचय और तंत्रों का सर्वेक्षण</b>	<b>(1-50)</b>
1.1 मशीनी अनुवाद का परिचय	
1.2 मशीन अनुवाद का इतिहास	
1.3 मशीनी अनुवाद तंत्र की संरचना	
1.3.1 कॉर्पोरा	
1.3.2 पूर्व संपादन	
1.3.3 शब्द संसाधन	
1.3.4 शब्दतंत्र	
1.3.5 अभिकलनात्मक शब्दकोश	
1.3.6 कृत्रिम बुद्धि	
1.3.7 विश्लेषक	
1.3.8 टैगिंग और पार्सर	
1.3.9 अंग्रेजी प्रजनक	
1.3.10 पाठ संश्लेषण	
1.3.11 अंतरण व्याकरण	
1.3.12 शाब्दिक व्याकरण	
1.3.13 हिंदी प्रजनक	
1.3.14 पश्च संपादन	
1.4 मशीनी अनुवाद की आवश्यकता	

- 1.5 मशीनी अनुवाद के प्रकार
  - 1.5.1 मानव अनुवाद
  - 1.5.2 मशीन साधित मानव अनुवाद
  - 1.5.3 मानव साधित मशीन अनुवाद
  - 1.5.4 मशीनी अनुवाद
- 1.6 मशीनी अनुवाद की प्रक्रिया
- 1.7 मशीनी अनुवाद की पद्धतियाँ
  - 1.7.1 नियम आधारित मशीनी अनुवाद
    - 1.7.1.1 प्रत्यक्ष पद्धति
    - 1.7.1.2 अंतरण पद्धति
    - 1.7.1.3 अंतर्भाषिक पद्धति
  - 1.7.2 उदाहरण आधारित मशीनी अनुवाद
  - 1.7.3 ज्ञान आधारित मशीनी अनुवाद
  - 1.7.4 सांख्यिकीय आधारित मशीनी अनुवाद
  - 1.7.5 मिश्रित मशीनी अनुवाद
- 1.8 मशीनी अनुवाद के तंत्र: एक सर्वेक्षण
  - 1.8.1 गूगल अनुवादक
  - 1.8.2 माइक्रोसॉफ्ट अनुवादक
  - 1.8.3 याहू बेबेल फिश
  - 1.8.4 Apertium मशीनी अनुवाद तंत्र
  - 1.8.5 भारत के प्रमुख मशीनी अनुवाद तंत्र
    - 1.8.5.1 आंग्लभारती
    - 1.8.5.2 अनुसारक
    - 1.8.5.3 अणुभारती
    - 1.8.5.4 मंत्र-राजभाषा
    - 1.8.5.5 आंग्लभारती-II
    - 1.8.5.6 अनुभारती-II

- 1.8.5.7 आंग्लहिंदी
- 1.8.5.8 मात्रा
- 1.8.5.9 शक्ति और शिवा
- 1.8.5.10 अनुवादक
- 1.8.5.11 VAASAANUBAADA द्विभाषी बंगाली मशीनी अनुवाद तंत्र
- 1.8.5.12 अंग्रेजी-कन्नड़ मशीनी अनुवाद तंत्र
- 1.8.5.13 UNL -आधारित अंग्रेजी-हिंदी मशीन अनुवाद तंत्र
- 1.8.5.14 अनुवाद मशीनी अनुवाद तंत्र
- 1.8.5.15 हिंग्लिश मशीनी अनुवाद तंत्र
- 1.8.5.16 आई.बी.एम.(IBM) अंग्रेजी-हिंदी मशीनी अनुवाद तंत्र
- 1.8.5.17 तमिल-हिंदी अनुसारक एवं अंग्रेजी-तमिल मशीनी अनुवाद तंत्र
- 1.8.5.18 असमिया मशीनी अनुवाद तंत्र
- 1.8.5.19 ओडिसी मशीनी अनुवाद तंत्र
- 1.8.5.20 पंजाबी से हिंदी मशीनी अनुवाद तंत्र
- 1.8.5.21 संपर्क मशीनी अनुवाद तंत्र
- 1.8.5.22 संस्कृत सुबंत विश्लेषक
- 1.8.5.23 लीला तंत्र
- 1.8.5.24 अन्य मशीनी अनुवाद तंत्र

## द्वितीय अध्याय : अंग्रेजी-हिंदी भाषा की संरचना और मशीनी अनुवाद की समस्याएँ (51-96)

### 2.1 अंग्रेजी-हिंदी भाषा की संरचना

#### 2.1.1 अंग्रेजी और हिंदी भाषाओं का विकास

#### 2.1.2 लिपि (Script)

##### 2.1.2.1 रोमन लिपि

##### 2.1.2.2 देवनागरी लिपि

#### 2.1.3 वर्णमाला

##### 2.1.3.1 अंग्रेजी और हिंदी भाषाओं की ध्वनियों में अंतर

### 2.1.3.2 संख्यात्मक अंक

## 2.2 अंग्रेजी और हिंदी भाषाओं का तुलनात्मक व्याकरण

### 2.2.1 संज्ञा

#### 2.2.1.1 लिंग

#### 2.2.1.2 वचन

#### 2.2.1.3 कारक

### 2.2.2 सर्वनाम

### 2.2.3 पूर्वसर्ग और परसर्ग

### 2.2.4 विशेषण

### 2.2.5 क्रिया

#### 2.2.5.1 काल

#### 2.2.5.2 पक्ष

#### 2.2.5.3 वृत्ति

#### 2.2.5.4 वाच्य

### 2.2.6 क्रिया-विशेषण

## 2.3 अंग्रेजी और हिंदी की वाक्य संरचना

### 2.3.1 वाक्य संरचना के आधार

#### 2.3.1.1 शब्दक्रम

#### 2.3.1.2 अन्विति

#### 2.3.1.3 कारक संबंध

#### 2.3.1.4 अर्थ-संगति

#### 2.3.1.5 अध्याहार या लोप

### 2.3.2 वाक्य के प्रकार

#### 2.3.2.1 सरल वाक्य

#### 2.3.2.2 मिश्र वाक्य

#### 2.3.2.3 संयुक्त वाक्य

2.4 अंग्रेजी-हिंदी मशीनी अनुवाद की समस्याएँ

**तृतीय अध्याय : शोध का उद्देश्य और प्रविधि**

**(97-112)**

- 3.1 शोध से संबंधित पूर्ववर्ती शोध कार्य
- 3.2 प्रस्तुत शोध का उद्देश्य
- 3.3 शोध के लिए स्रोत सामग्री का प्रयोग
- 3.4 शोध प्रविधि

**चतुर्थ अध्याय : अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन**

**(113-152)**

4.1 मशीनी अनुवाद मूल्यांकन का परिचय

- 4.1.1 मानव मूल्यांकन
- 4.1.2 स्वचालित मूल्यांकन
  - 4.1.2.1 ब्लू मूल्यांकन पद्धति
  - 4.1.2.2 निस्ट मूल्यांकन पद्धति
  - 4.1.2.3 मेटियोर मूल्यांकन पद्धति
  - 4.1.2.4 एफ-परिमाण पद्धति
  - 4.1.2.5 मेटियोर-हिंदी परिमाण पद्धति

4.2 मशीनी अनुवाद तंत्रों का मूल्यांकन

- 4.2.1 कॉर्पोरा निर्माण
- 4.2.2 निर्गत अनुवाद
- 4.2.3 मानव मूल्यांकन
- 4.2.4 कॉर्पोरा का क्षेत्रों के अनुसार मूल्यांकन

**पंचम अध्याय : मशीनी अनुवाद तंत्रों से निर्गत वाक्यों का त्रुटि विश्लेषण**

**(153-194)**

- 5.1 त्रुटि विश्लेषण परिचय
- 5.2 त्रुटियों का वर्गीकरण
- 5.3 मशीन अनूदित वाक्यों का त्रुटि विश्लेषण
  - 5.3.1 वर्ण विचार संबंधी त्रुटियाँ

- 5.3.2 शब्द संरचना संबंधी त्रुटियाँ  
5.3.3 शाब्दिक त्रुटियाँ  
5.3.4 अर्थगत त्रुटियाँ  
5.3.5 वाक्यात्मक त्रुटियाँ  
5.4 मशीनी अनुवाद की त्रुटियों के लिए सुधारात्मक उपाय

उपसंहार	(195-203)
संदर्भ ग्रंथ-सूची	(204-229)
परिशिष्ट	(230-276)
• गूगल अनुवादक आउटपुट स्कोर	(231-242)
• माइक्रोसॉफ्ट अनुवादक आउटपुट स्कोर	(243-253)
• मंत्र-राजभाषा आउटपुट स्कोर	(254-265)
• आंग्ल-हिंदी आउटपुट स्कोर	(266-276)
प्रकाशित शोध-पत्र	(277)

## संक्षिप्ताक्षर (Abbreviations)

A.A.I.	Applied Artificial Intelligence
ALPAC	Automatic Language Processing Advisory Committee
ARPA	Advanced Research Projects Agency
BIS	Bureau of Indian Standards
BLEU	Bi Lingual Evaluation Understudy
C.L.	Computational Linguistics
C-DAC	Centre for Development of Advance Computing
CPG	Computational Paninian Grammar
DeitY	Department of Electronics & Information Technology
DET	Determiners
EHMT	English Hindi Machine Translation
GNMT	Google Neural Machine Translation
HAMT	Human Aided Machine Translation
HEBMT	Hybrid Example Based Machine Translation
IIIT	International Institute of Information Technology
IISc	Indian Institute of Science
ILCI	Indian Language Corpora Initiative
ILMT	Indian Language Machine Translation
M.T.	Machine Translation
MAHT	Machine Aided Human Translation
ManTra	Machine Assisted Translation Technology

METEOR	Metric for Evaluation of Translation with Explicit Ordering
MTE	Machine Translation Evaluation
NCFST	National Centre for Software Technology
NER	Named Entity Recognition
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NP	Noun Phrase
PENN.	University of Pennsylvania (Tagset)
POS	Part of Speech
QUANT	Quantifiers
RBMT	Rule Based Machine Translation
S.L.	Source Language
SMTS	Sampark Machine Translation System
SMTS	Statistical Machine Translation System
SOV	Subject-Object-Verb
SVO	Subject-Verb-Object
T.L	Target Language
TAG	Tree Adjoining Grammar
TDIL	Technical Development for Indian Languages
TDIL	Technology Development for Indian Languages
UNL	Universal Networking Language
VP	Verb Phrase
IL-1	एक भारतीय भाषा
IL-2	अन्य भारतीय भाषा

अं.हि.म.अ.तंत्र	अंग्रेजी-हिंदी मशीनी अनुवाद तंत्र
अ.भा.	अभिकलनात्मक भाषाविज्ञान
अभि.	अभिकर्ता
कृ.बु.	कृत्रिम बुद्धि
क्रि.प.	क्रिया पदबंध
नि.	निर्धारक
प.विशे.	पश्च विशेषक
पू.विशे.	पूर्व विशेषक
प्रा.भा.स.	प्राकृतिक भाषा संसाधन
भा.भा.कॉ.उ.	भारतीय भाषा कॉर्पोरा उपक्रम
भा.भा.प्रौ.वि.	भारतीय भाषाओं के लिए प्रौद्योगिकी विकास
म.अनु.	मशीनी अनुवाद
ल.भा.	लक्ष्य भाषा
विशे.पद.	विशेषण पदबंध
श.मा.प्रा.	शक्ति मानक प्रारूप
सं.प.	संज्ञा पदबंध
संबंध. उप.	संबंधसूचक उपवाक्य
स्था.व्या.नि.	स्थानांतरण व्याकरण नियम
स्रो.भा.	स्रोत भाषा

## टेबल, आकृतियाँ और स्क्रीनशॉट की सूची

टेबल सूची	पृष्ठ संख्या
टेबल 1.1 : भारत के प्रमुख मशीनी अनुवाद तंत्र	36
टेबल 2.1 : अंग्रेजी वर्णमाला	59
टेबल 2.2 : हिंदी वर्णमाला स्वर	60
टेबल 2.3 : हिंदी वर्णमाला व्यंजन	61
टेबल 2.4 : रोमन-देवनागरी अंक	63
टेबल 2.5 : अंग्रेजी-हिंदी भाषा की व्याकरणिक संरचना	65
टेबल 2.6 : एकवचन और बहुवचन के प्रत्यय	67
टेबल 2.7 : कारक में प्रत्यय प्रयोग	68
टेबल 2.8 : अंग्रेजी पूर्वसर्ग और हिंदी परसर्ग का प्रयोग	70
टेबल 3.1 : मूल्यांकन की स्केलिंग पद्धति	111
टेबल 4.1 : मूल्यांकन के लिए 5 पॉइन्ट स्केलिंग पद्धति	129
टेबल 4.2 : गूगल अनुवादक के मूल्यांकन का औसत स्कोर	131
टेबल 4.3 : विषय-क्षेत्रों के अनुसार म. अनु. तंत्रों के मूल्यांकन का औसत स्कोर	133
टेबल 4.4 : मशीनी अनुवाद तंत्रों की बोधगम्यता का कुल औसत स्कोर	148
टेबल 4.5 : मशीनी अनुवाद तंत्रों की बोधगम्यता और अबोधगम्यता का प्रतिशत	149

## आकृतियों की सूची

आकृति 1.1 : मशीनी अनुवाद की प्रक्रिया	21
आकृति 4.1 : गूगल अनुवादक का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर	144
आकृति 4.2 : माइक्रोसॉफ्ट अनुवादक का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर	145
आकृति 4.3 : मंत्र-राजभाषा तंत्र का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर	146
आकृति 4.4 : आंग्ल-हिंदी तंत्र का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर	147
आकृति 4.5: मशीनी अनुवाद तंत्रों की बोधगम्यता का प्रतिशत	150
आकृति 4.6 : मशीनी अनुवाद तंत्रों की अबोधगम्यता का प्रतिशत	151

## स्क्रीनशॉट की सूची

स्क्रीनशॉट 1.1 : गूगल अनुवादक तंत्र की अनुवाद प्रक्रिया	31
स्क्रीनशॉट 1.2 : माइक्रोसॉफ्ट अनुवादक की अनुवाद प्रक्रिया	32
स्क्रीनशॉट 4.1 : गूगल अनुवादक से निर्गत अनुवाद	125
स्क्रीनशॉट 4.2 : माइक्रोसाफ्ट अनुवादक से निर्गत अनुवाद	126
स्क्रीनशॉट 4.3 : मंत्र-राजभाषा से निर्गत अनुवाद	127
स्क्रीनशॉट 4.4 : आंग्ल-हिंदी तंत्र से निर्गत अनुवाद	128

## प्रस्तावना

---

आज वैश्वीकरण के दौर में व्यापारिक प्रतिस्पर्धा के चलते अनुवाद की अधिक मांग बढ़ गयी है। प्रत्येक राष्ट्र अपनी अर्थव्यवस्था को मजबूती प्रदान करना चाहता है। इसके लिए वह विश्व के किसी भी कोने से सर्वोत्तम प्रौद्योगिकी को अपने देश में लाना चाहता है। ऐसे में विभिन्न विषय क्षेत्रों की सामग्री के अनुवाद के लिए अधिक मांग बढ़ गयी है। इसलिए अनुवादक की भूमिका भी अधिक महत्वपूर्ण हो जाती है। लेकिन अनुवादक या किसी व्यक्ति से यह संभव नहीं कि वो विश्व की सभी भाषाओं में पारंगत हो और असंख्य दस्तावेजों तथा पाठ्य सामग्री का अनुवाद कर सके। इसलिए अनुवाद के लिए मशीनी अनुवाद का प्रयोग किया जाता है जिससे मानव के समय, श्रम और धन की बचत हो।

मशीनी अनुवाद के द्वारा एक प्राकृतिक भाषा के पाठ का दूसरी प्राकृतिक भाषा के पाठ में कंप्यूटर के माध्यम से अनुवाद किया जाता है। यह अनुवाद कंप्यूटर में समावेशित भाषाई सामग्री (Corpora), शब्दाकोश (Dictionary), व्याकरणिक संरचना (Grammatical Structure) और अल्गोरिदम (Algorithm) के माध्यम से पूर्ण होता है। इस अनुवाद कार्य में कृत्रिम बुद्धि का योगदान महत्वपूर्ण है। जिसमें कंप्यूटर द्वारा भाषा और शब्द-संसाधन के लिए उसके मस्तिष्क में संचित सूक्ष्म और अव्यक्त रूप में विद्यमान नियमों को एल्गोरिदम या सूत्रों के माध्यम से कंप्यूटर द्वारा परिचालित किया जाता है और अनुवाद प्राप्त किया जाता है।

पिछले सात दशकों से विश्व के कई देशों के संस्थानों और विश्वविद्यालयों में मशीनी अनुवाद पर शोध कार्य चल रहे हैं। जिसमें अनेक मशीनी अनुवाद तंत्रों का विकास किया गया है तथा मशीनी अनुवाद तंत्रों के लिए उपकरणों का निर्माण किया जा रहा है। इस दिशा में भारत में किए गए कार्यों में विभिन्न संस्थानों और विश्वविद्यालयों के कार्य कारगर साबित हुए हैं, जिसमें भारतीय भाषाओं के लिए विविध मशीनी अनुवाद तंत्रों का विकास किया गया है। आज इन मशीनी अनुवाद तंत्रों का प्रयोग संपूर्ण विश्व में किया जाता है, लेकिन इसके अनुवाद की गुणवत्ता में अभाव पाया जाता है। अक्सर ये तंत्र एक भाषा के दूसरी भाषा के अनुवाद में अनेक

गलतियाँ करते हैं, जिससे अनुवाद बोधगम्य (comprehensible) नहीं होता। इन तंत्रों के अनुवाद की गुणवत्ता में सुधार करने के लिए उनका मूल्यांकन करना आवश्यक होता है। मूल्यांकन से अनुवाद में होनेवाली त्रुटियों के बारे में जानकारी प्राप्त होती है और उससे मशीनी अनुवाद तंत्रों के अनुवाद संपादन के लिए बेहतर बनाया जा सकता है।

प्रस्तुत शोध-प्रबंध 'अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन' पर आधारित है। जिसमें अंग्रेजी से हिंदी में मशीनी अनुवाद करने वाले तंत्रों का मूल्यांकन किया गया है। इस मूल्यांकन का उद्देश्य यह है कि ये तंत्र किन विषय क्षेत्रों के लिए सही अनुवाद करते हैं। इन अनुवाद तंत्रों में गूगल, माइक्रोसॉफ्ट, मंत्र-राजभाषा और आंग्ल-हिंदी का समावेश है। इन मशीनी अनुवाद तंत्रों से अनुवाद प्राप्त कर मूल्यांकन और त्रुटि विश्लेषण किया गया है। प्रस्तुत शोध का उद्देश्य यह भी है कि मशीनी अनुवाद मूल्यांकन पद्धतियों के विकास के लिए सुझाव देना। भारत में विभिन्न मशीन अनुवाद तंत्रों का विकास किया गया है, लेकिन अनुवाद की गुणवत्ता जांचने के लिए कोई मानक पद्धति नहीं है जिससे मूल्यांकन के परिणाम विश्वसनीय और सटीक प्राप्त हों। इसके लिए मानव मूल्यांकन ही स्वचालित मूल्यांकन पद्धति से अधिक सटीक परिणाम प्राप्त करने के लिए सक्षम है। इस शोध में भी उपरोक्त मशीनी अनुवाद तंत्रों का मूल्यांकन किया गया है। प्रस्तुत शोध-प्रबंध निम्न बिंदुओं पर आधारित है।

- अंग्रेजी से हिंदी में 18,000 कॉर्पोरा के वाक्यों का संकलन किया गया है।
- गूगल, माइक्रोसॉफ्ट, मंत्र-राजभाषा और आंग्ल-हिंदी मशीनी अनुवाद तंत्रों से निर्गत अनुवाद किया गया है।
- 5 मानव मूल्यांकनकर्ताओं द्वारा 24 विषय-क्षेत्रों के अनुसार वाक्यों का मूल्यांकन किया गया है।
- मूल्यांकन में पाए गए मशीन अनुवादित वाक्यों का त्रुटि विश्लेषण किया गया है।
- मूल्यांकन और त्रुटि विश्लेषण के लिए सुधारात्मक उपाय दिए गए हैं।

जब दो भाषाओं की संरचना समान होती है तो मशीनी अनुवाद तंत्रों का विकास करना आसान होता है और उनसे अनुवाद भी सरल होता है। लेकिन जब किसी भाषा की संरचना में लिपि, ध्वनि, शब्द, रूप, वाक्य, अर्थ संरचना तथा व्याकरणिक संरचना के स्तर पर भिन्नता पाई जाती है तो मशीनी अनुवाद में भी समस्या उत्पन्न होती है। इस शोध के माध्यम से यह अध्ययन किया गया है कि अंग्रेजी और हिंदी की भाषिक संरचना में पूर्णतः भिन्नता है। इस भाषिक संरचना के कारण मशीनी अनुवाद में कौनसी समस्याएँ उत्पन्न होती हैं। उन समस्याओं का समाधान करने का प्रयास इस शोध-प्रबंध के माध्यम से किया गया है।

प्रस्तुत शोध-प्रबंध के प्रथम अध्याय में मशीनी अनुवाद का परिचय दिया गया है। इसमें मशीनी अनुवाद का इतिहास, मशीनी अनुवाद तंत्र की संरचना, मशीनी अनुवाद की आवश्यकता, मशीनी अनुवाद के प्रकारों का विस्तृत विवेचन किया गया है। मशीनी अनुवाद की प्रक्रिया में एक भाषा से दूसरी भाषा में अनुवाद करने के लिए विश्लेषण, अंतरण, प्रजनन की प्रक्रिया को स्पष्ट किया गया है। साथ ही मशीनी अनुवाद की अनेक पद्धतियों की चर्चा की गयी है। मशीनी अनुवाद तंत्रों के सर्वेक्षण में विश्व स्तर के अनुवाद तंत्र गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, याहू बेबेल फिश तथा भारत के प्रमुख मशीनी अनुवाद तंत्रों में आंग्लभारती, अनुसारक, अणुभारती, मंत्र-राजभाषा, आंग्ल-हिंदी, मात्रा, शक्ति और शिवा, अनुवादक, संपर्क तथा अन्य मशीनी अनुवाद तंत्रों के निर्माण एवं विकास कार्यों के बारे में जानकारी दी गई है।

द्वितीय अध्याय में अंग्रेजी और हिंदी भाषा की संरचना का विवेचन किया गया है। इसमें दोनों भाषाओं की उत्पत्ति, भाषा परिवार, लिपि, वर्णमाला, अंक, ध्वनि, में पाई जानेवाली भिन्नता को स्पष्ट किया गया है। इन दो भाषाओं में शब्दक्रम संरचना में भी भिन्नता पाई जाती है। जैसे- अंग्रेजी भाषा रूप-रचना की दृष्टि से अधिक स्थिर है और इसका शब्दक्रम कर्ता-क्रिया-कर्म संरचना पर तथा हिंदी भाषा का शब्दक्रम कर्ता-कर्म-क्रिया की संरचना पर आधारित है। इन दोनों भाषाओं में व्याकरणिक संरचना में भी भिन्नता पाई जाती है। जिसमें लिंग, वचन, काल, वाच्य, पक्ष, वृत्ति, संज्ञा, सर्वनाम, विशेषण, क्रिया-विशेषण, क्रिया, संयोजक शब्द, पूर्वसर्ग और परसर्ग तथा वाक्य के प्रकार में सरल वाक्य, संयुक्त वाक्य और मिश्र वाक्य आदि घटकों का अध्ययन-विश्लेषण किया गया है। साथ ही इसमें दोनों भाषाओं में

आनेवाली मशीनी अनुवाद की समस्याओं में शब्दक्रम, संदिग्धार्थकता, लिप्यंतरण की समस्या, द्विभाषी कोश का निर्माण, मुहावरों और कहावतों की समस्याएँ, सार्वभौमिक व्याकरण का अभाव, भाषा वैज्ञानिकों का अभाव, कृत्रिम बुद्धि (Artificial Intelligence) के विकास में देरी आदि समस्याओं का विवेचन किया गया है।

तृतीय अध्याय में शोध का उद्देश्य और प्रविधि के बारे में चर्चा की गयी है। इस शोध के उद्देश्य में अंग्रेजी से हिंदी में अनुवाद करने के लिए उपलब्ध मशीनी अनुवाद तंत्रों का गुणवत्ता की दृष्टि से मूल्यांकन और त्रुटि विश्लेषण किया गया है। उसके लिए प्रयुक्त भाषाई सामग्री का संकलन, स्रोत सामग्री का प्रयोग और पूर्ववर्ती शोध कार्य में अंग्रेजी से हिंदी और अंग्रेजी से भारतीय भाषाओं के लिए प्रयुक्त स्वचालित मूल्यांकन पद्धतियों के बारे में चर्चा की गयी है। इस प्रबंध में प्रयुक्त शोध प्रविधि प्रायोगिक तथा सैध्दांतिक स्तर पर आधारित है जिसमें प्रायोगिक रूप में मशीनी अनुवाद तंत्रों का परीक्षण (Testing) किया गया और सैध्दांतिक रूप में मानव द्वारा 4 मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए 5 पॉइन्ट स्केल (0-4) पद्धति का प्रयोग किया गया है। इस मानक पद्धति के द्वारा इस शोध में प्रयुक्त भाषाई सामग्री (कॉर्पोरा) के वाक्यों का मूल्यांकन किया गया है।

चतुर्थ अध्याय में अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन किया गया है। मशीनी अनुवाद का मूल्यांकन मानव द्वारा और स्वचालित मूल्यांकन पद्धति द्वारा किया जाता है। उसमें मानव मूल्यांकन से अधिक स्वचालित मूल्यांकन का प्रयोग समय और श्रम की बचत के लिए तथा निष्पक्षता के लिए अधिक किया जाता है। स्वचालित मूल्यांकन के लिए प्रयुक्त ब्लू, निस्ट, मेटियोर, एफ़ मेट्रिक पद्धति, मेटियोर-हिंदी आदि मूल्यांकन पद्धतियों का विवेचन किया गया है।

इस शोध के मूल्यांकन के लिए प्रयोग की गयी भाषायी सामग्री के वाक्यों का गूगल अनुवादक, माइकोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी मशीनी अनुवाद तंत्रों से अनुवाद किया गया है और अनूदित वाक्यों का मानव द्वारा मूल्यांकन कॉर्पोरा के क्षेत्रों के अनुसार किया गया है। इसमें 24 क्षेत्रों के अनुसार अनुवाद करने वाले मशीनी अनुवाद तंत्रों की बोधगम्यता के आधार पर कुल औसत स्कोर निकाले गए हैं और इन औसत स्कोर से उनकी बोधगम्यता और अबोधगम्यता का प्रतिशत निकाला गया है। जिसमें गूगल अनुवादक सर्वाधिक अर्थात् 71.33

प्रतिशत बोधगम्य अनुवाद करता है। इसके द्वारा स्रोत भाषा के सरल एवं छोटे वाक्यों का अनुवाद अधिक सही होता है। माइक्रोसॉफ्ट अनुवादक 69.67 प्रतिशत बोधगम्य अनुवाद करता है जो गूगल के बाद द्वितीय क्रम में अनुवाद करता है। इसी के साथ मंत्र-राजभाषा केवल 33.33 प्रतिशत बोधगम्य अनुवाद करता है, इसका अर्थ इससे अनुवाद में सर्वाधिक गलतियाँ होती हैं। इसके बाद आंग्ल-हिंदी से अनुवाद की बोधगम्यता का प्रतिशत 53 है, जो तृतीय क्रम में अनुवाद करता है। इन सभी मशीनी अनुवाद तंत्रों के अनुवाद की बोधगम्यता से स्पष्ट हुआ है कि गूगल और माइक्रोसॉफ्ट बोधगम्य अनुवाद प्रदान करने में मंत्र-राजभाषा और आंग्ल-हिंदी से अधिक सक्षम हैं। मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद में अनेक त्रुटियाँ पाई जाती हैं। जिसके लिए पंचम अध्याय में उन त्रुटियों का विश्लेषण किया गया है।

पंचम अध्याय में मशीनी अनुवाद मूल्यांकन में निर्गत वाक्यों के त्रुटियों का विश्लेषण किया गया है। इसके लिए त्रुटियों का भाषाई आधार पर वर्गीकरण किया गया है। इस वर्गीकरण में- वर्ण विचार संबंधी (Orthographic), रूपात्मक संबंधी (morphological), शब्द के स्तर पर (lexical), अर्थ के स्तर पर (semantic) और वाक्यात्मक (syntactic) संबंधी वर्गीकरण कर वाक्यों का विश्लेषण किया गया है। इस विश्लेषण में पाया गया है कि मशीन अक्सर एक भाषा से दूसरे भाषा के अनुवाद में वर्तनी, संक्षिप्तीकरण, लिंग का क्रिया पर प्रभाव, अनअनूदित शब्द, लिप्यंतरण में त्रुटियाँ, गलत शब्द चयन, संदिग्धार्थकता अथवा अस्पष्टता, अर्थगत त्रुटियाँ, वाक्यांश, कहावतें एवं मुहावरों, शब्दक्रम, पूर्वसर्ग तथा परसर्ग, सर्वनाम, विशेषण, आर्टिकल संबंधी त्रुटियाँ, सरल, संयुक्त एवं मिश्रित वाक्यों में पाई गई त्रुटियों का विस्तृत विश्लेषण किया गया है। इस त्रुटि विश्लेषण के बाद मशीनी अनुवाद में इसके लिए सुधारात्मक उपाय भी दिए गए हैं।

प्रस्तुत शोध अंग्रेजी से हिंदी में मशीनी अनुवाद करने वालों तंत्रों में गूगल, माइक्रोसॉफ्ट, मंत्र-राजभाषा, आंग्ल-हिंदी से अनुवाद प्राप्त कर उनका मानव द्वारा विषय क्षेत्रों के अनुसार मूल्यांकन करने के बाद वाक्यों के त्रुटियों का विश्लेषण किया गया है। इस मूल्यांकन से इन मशीनी अनुवाद तंत्रों की अनुवाद करने की गुणवत्ता के बारे में जानकारी मिली तथा ये मशीनी अनुवाद तंत्र किन विषय क्षेत्रों में बोधगम्य अनुवाद करने के लिए अधिक सक्षम हैं, यह भी ज्ञात हुआ है। साथ ही इन मशीनी

अनुवाद तंत्रों में सुधार के लिए सुझाव एवं सुधारात्मक उपाय भी दिए गए हैं। जिससे भविष्य में इन तंत्रों के विकासकर्ताओं के लिए प्रस्तुत शोध में किए गए मूल्यांकन के अध्ययन का उपयोग होगा।

अंत में इस शोध का प्रमुख योगदान यह है कि निम्न 4 मशीनी अनुवाद के तंत्र अनेक विषय क्षेत्रों में बोधगम्य अनुवाद प्रदान करते हैं जिसमें गूगल अनुवादक, कला और संस्कृति, प्रशासनिक सेवा, अर्थव्यवस्था, भूगोल, स्वास्थ्य, विधि, योजनाएँ और पर्यटन आदि विषय क्षेत्रों में अधिक सटीक अनुवाद करता है। माइकोसॉफ्ट भी विशेषतः कला और संस्कृति, शिक्षा, भूगोल, स्वास्थ्य, इतिहास, योजनाएँ, राजनीति और पर्यटन के क्षेत्र में अधिक सहज अनुवाद करता है। इसी के साथ मंत्र-राजभाषा प्रशासन संबंधी दस्तावेजों के अलावा कृषि, बैंक, शिक्षा, वित्त, स्वास्थ्य सेवा, सूचना प्रौद्योगिकी और लघु उद्योग के क्षेत्रों में अनुवाद करता है। आंग्ल-हिंदी भी कुछ क्षेत्रों जैसे- व्यापार, रक्षा, पर्यावरण, अर्थव्यवस्था, भूगोल, स्वास्थ्य, इतिहास, प्रौद्योगिकी और पर्यटन के क्षेत्र में अनुवाद करता है। इससे यह स्पष्ट होता है कि ये सभी मशीनी अनुवाद तंत्र अलग-अलग विषय क्षेत्रों के अनुसार अनुवाद करने के लिए सक्षम हैं, तथापि इनसे अनुवाद में अनेक गलतियाँ पाई जाती हैं। इसलिए इनका मूल्यांकन कर तंत्रों को अद्यतन करने की आवश्यकता है जिससे अनुवाद बोधगम्य होगा और अनुवाद की गुणवत्ता भी बढ़ेगी।

- सुमेध खुशालराव हाडके

## प्रथम अध्याय

# मशीनी अनुवाद का परिचय और तंत्रों का सर्वेक्षण

- 1.1 मशीनी अनुवाद का परिचय
- 1.2 मशीन अनुवाद का इतिहास
- 1.3 मशीनी अनुवाद तंत्र की संरचना
- 1.4 मशीनी अनुवाद की आवश्यकता
- 1.5 मशीनी अनुवाद के प्रकार
- 1.6 मशीनी अनुवाद की प्रक्रिया
- 1.7 मशीनी अनुवाद की पद्धतियाँ
- 1.8 मशीनी अनुवाद के तंत्र : एक सर्वेक्षण

## प्रथम अध्याय

### मशीनी अनुवाद का परिचय और तंत्रों का सर्वेक्षण

(INTRODUCTION TO MACHINE TRANSLATION AND SURVEY OF SYSTEMS)

---

#### 1.1 मशीनी अनुवाद का परिचय (Introduction to Machine Translation)

वैश्वीकरण के दौर में ज्ञान का आदान-प्रदान बहुत तीव्र गति से हो रहा है। आज विश्व में सूचना का आदान-प्रदान विविध भाषा-भाषी समुदायों के द्वारा किए जा रहे नवीनतम प्रयोगों का परिणाम है। विश्व के विभिन्न देशों में अनेक भाषाओं का प्रयोग बोलचाल तथा कार्यालयी कामकाज के लिए किया जाता है। लेकिन वैश्विक स्तर पर देखा जाए तो विज्ञान, तकनीकी, शिक्षा, वाणिज्य, अर्थव्यवस्था और जनसंचार आदि विषय-क्षेत्रों में उपलब्ध जानकारी अधिकतर अंग्रेजी भाषा में ही होती है। अब इन सूचनाओं को विश्व की सभी भाषाओं में उपलब्ध कराने हेतु अनुवाद की आवश्यकता बढ़ रही है। फलस्वरूप विभिन्न भाषाई सामग्री का निर्माण किया जा रहा है। अनुवाद आज के युग की अनिवार्य आवश्यकता है। अनुवाद एक सेतु की तरह काम करता है, जिसके माध्यम से एक भाषा (स्रोत भाषा) का पाठक वर्ग दूसरी भाषा (लक्ष्य भाषा) में संपर्क स्थापित करता है। इस संपर्क में स्रोत भाषा में व्यक्त संदेश या विषय-वस्तु का रूपांतरण, समान अर्थ वाली लक्ष्य भाषा के कथ्य या विचार में अभिव्यक्त होता है। प्रत्येक भाषा की अपनी संरचना, शैली और प्रकृति होती है। उसकी रक्षा करते हुए व्यक्त भावों को सहज रूप में प्रस्तुत करना ही अनुवाद है।

भारत एक बहुभाषिक देश है। यहाँ विभिन्न भाषाओं का परस्पर संप्रेषण होता है। इसलिए यहाँ भिन्न-भिन्न भाषाओं की कला, उनके साहित्य और संस्कृति को जानने के लिए पहले से अनुवाद होता रहा है। लेकिन यह अनुवाद मानव द्वारा काफ़ी श्रमसाध्य और समय लेने वाला रहा है। आज भारत में संविधान स्वीकृत 22 भाषाएँ हैं, जिनका तकनीकी रूप से विकास करना बहुत महत्वपूर्ण कार्य है, ताकि हम विभिन्न भाषाओं के सूचना का आदान-प्रदान तुरंत कर सकें। इस सूचना का संचार करने एवं अनुवाद कार्य

को अधिक सफल बनाने के लिए और मानव की सहायता स्वरूप 'मशीनी अनुवाद' का विकास हुआ है। 'मशीनी अनुवाद, भाषा का अध्ययन और विश्लेषण करने के लिए भाषा का वैज्ञानिक अध्ययन प्रस्तुत करने वाला शास्त्र भाषाविज्ञान का अनुप्रयुक्त क्षेत्र 'अभिकलनात्मक भाषाविज्ञान' (Computational Linguistics) की विधाओं में विकसित हुआ है। 'अभिकलनात्मक भाषाविज्ञान' का मुख्य उद्देश्य प्राकृतिक भाषा (Natural Language) को समझ कर, ऐसे व्यापक मॉडल और सिद्धांतों का विकास करना है, जिनकी सहायता से मानव और मशीन के बीच संवाद स्थापित हो। इसमें कृत्रिम बुद्धि (Artificial Intelligence) का भी समावेश किया गया है ताकि कंप्यूटर भी मानव की तरह सोचे-समझे और अनुवाद करे। इस कृत्रिम बुद्धि के माध्यम से शब्दों और अर्थों का विश्लेषण करके मशीनी अनुवाद ने भाषाई पक्ष को अधिक मजबूत किया है। जिससे वैश्विक स्तर पर विभिन्न देशों ने इस प्रणाली का प्रयोग कर कई मशीनी अनुवाद तंत्रों का निर्माण किया है।

एक स्रोत भाषा (SL) की पाठ्यसामग्री को दूसरी लक्ष्य भाषा (TL) में प्रस्तुत करने की प्रक्रिया अनुवाद है। अनुवाद की इस प्रक्रिया में मशीन (कंप्यूटर) का उपयोग करना मशीनी अनुवाद है, जिससे स्रोत भाषा (SL) की पाठ्य सामग्री का अनुवाद लक्ष्य भाषा (TL) में स्वतः हो जाता है। एक भाषा की सामग्री का दूसरी भाषा में अनुवाद करना, यानी केवल शब्दों का रूपांतरण नहीं है, बल्कि अनुवाद एक तरह की विचार प्रक्रिया है। जब अनुवादक किसी भाषा का अनुवाद करता है, तो वह दोनों भाषाओं के साहित्य, भाव, उनकी संस्कृति और संवेदना को मूल भाषा की तरह बनाए रखने का प्रयास करता है, लेकिन कंप्यूटर द्वारा किए गए अनुवाद में अक्सर गलतियाँ पायी जाती हैं। लेकिन मानव अपनी सहज बुद्धि से उससे जुड़े हुए संदर्भ की जानकारी के आधार पर भाषा की अभिव्यक्तियों की सही निर्वचन करता है। उसके पास भाषा की समझ और सांसारिक ज्ञान (World Knowledge) पहले से होता है, लेकिन कंप्यूटर केवल शब्द का वही अर्थ ग्रहण करता है, जो उसकी स्मृति (Memory) में संचित होता है और उसे प्रोग्रामिंग के द्वारा संचालित किया जाता है। इस संचालन के लिए वाक्यात्मक नियमों का एल्गोरिदम तथा शब्दकोश का डेटाबेस के रूप में कंप्यूटर में संग्रहित करते हुए

मशीनी अनुवाद प्रणालियों का विकास किया जाता है। इस मशीनी अनुवाद की परिकल्पना को समझने के लिए इसकी परिभाषा को जानना महत्वपूर्ण है।

सूरजभान सिंह के अनुसार “मशीनी अनुवाद, अनुवाद की ऐसी प्रक्रिया है, जिसमें कंप्यूटर प्रणाली (System) के जरिए एक भाषा से दूसरी भाषा में अपने आप अनुवाद की जाने वाली सामग्री (Text) को इनपुट के रूप में कंप्यूटर में देते हैं। कंप्यूटर की भीतरी प्रणाली (System), जिसमें दोनों भाषाओं के शब्दों, मुहावरों और व्याकरणिक नियमों का ज्ञान संचित रहता है, जो अपने आप उस सामग्री का दूसरी भाषा में अनुवाद करती है और कुछ ही क्षणों में निर्गत पाठ (Output) के रूप में अनूदित सामग्री प्राप्त हो जाती है।”<sup>1</sup> इनके अनुसार अगर कंप्यूटर में अनुवाद के लिए आवश्यक भाषाई सामग्री का समावेश किया जाए तो कंप्यूटर के माध्यम से स्रोत भाषा से लक्ष्य भाषा में शीघ्र ही अनुवाद प्राप्त किया जा सकता है।

दीपा गुप्ता के अनुसार, “मशीनी अनुवाद एक ऐसी प्रक्रिया है जो पाठ की इकाइयों को एक भाषा (स्रोत भाषा) से दूसरी भाषा (लक्ष्य भाषा) में कंप्यूटर की कृत्रिम बुद्धि के माध्यम से अनूदित करती है। (Machine Translation is the process of translating text units of one language (Source language) into a second language (Target language) by using the artificial intelligence of computers.)”<sup>2</sup> इनके अनुसार कृत्रिम बुद्धि के माध्यम से एक प्राकृतिक भाषा को दूसरी भाषा में अनुवाद सहजता से किया जा सकता है लेकिन इसके लिए कंप्यूटर में कृत्रिम बुद्धि का समावेश किया जाना चाहिए।

**Machine Translation:** “Machine Translation (MT) involves the use of computer programs to translate texts from one natural language into another automatically. It is usually subsumed under the category of computer (based translation, together with Computer-aided translation.)”<sup>3</sup>

---

<sup>1</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 252

<sup>2</sup> शर्मा, राकेश, *मशीनी अनुवाद: स्थिति एवं संभावनाएँ*, अनुवाद पत्रिका, 2014, जनवरी-मार्च, अंक: 158, पृ.66

<sup>3</sup> Baker, Mona, Gabriela, Saldanha (ed.), *Rutledge Encyclopaedia of Translation studies*, 2011, p.162

**Machine translation**, commonly known as MT, can be defined as “translation from one natural language (source language (SL)) to another language (target language (TL)) using computerized systems and, with or without human assistance”<sup>4</sup>

इन परिभाषाओं से यह स्पष्ट होता है कि एक प्राकृतिक भाषा का कंप्यूटर के द्वारा दूसरी भाषाओं में अनुवाद करना ही मशीनी अनुवाद है। यह प्रक्रिया विभिन्न संसाधनों, सामग्री के माध्यम से पूर्ण होती है। जिसमें कॉर्पोरा, व्याकरण जाँचक, रूपवैज्ञानिक विश्लेषक, पार्सर, द्विभाषिक शब्दावली का समावेश मशीनी अनुवाद के लिए किया जाता है।

## 1.2 मशीन अनुवाद का इतिहास (History of Machine Translation)

विश्व में मशीन अनुवाद का इतिहास अगर देखा जाए तो यह ज्ञात होता है कि मशीनी अनुवाद के क्षेत्र में प्रयास 1949 के बाद वारेन विवर के ‘भाषा विश्लेषण और अनुवाद’ संबंधी शोध प्रस्ताव से शुरू हुआ है। लेकिन यह प्रस्ताव स्रोत भाषा से लक्ष्य भाषा के शब्द प्रतिस्थापन पर आधारित था। इसी दिशा में मशीनी अनुवाद तंत्र निर्माण करने का पहला प्रयास 1956 में जार्ज टाउन विश्वविद्यालय में शुरू हुआ था, जिस पर 1964 में अमेरिकी प्रशासन द्वारा गठित एल्पेक (ऑटोमेटिक लैंग्वेज प्रोसेसिंग एडवाइज़री कमेटी) ने 1966 में रिपोर्ट प्रस्तुत की और उसमें अनुवाद तंत्र को विकसित करने के लिए मूलभूत शोध करने की आवश्यकता को बताया गया। इस रिपोर्ट की मान्यताओं को अस्वीकार कर अमेरिकी सरकार ने अनुवाद संबंधी शोध के लिए वित्तीय सहायता बंद कर दी। इससे अमेरिका में अनुवाद संबंधी शोध कार्य काफी समय तक बंद रहा, लेकिन यूरोप के अन्य देशों में तथा सोवियत संघ में अनुवाद कार्य जारी रहा। उसके बाद जापान में भी इस दिशा में प्रयास शुरू हुए। तो कुछ निजी कंपनियों ने भी भारी मात्रा में वित्तीय सहायता प्राप्त कर मशीनी अनुवाद के लिए प्रयास शुरू किए।<sup>5</sup>

---

<sup>4</sup> Hutchins, W.J., & Somers, H. L., *An introduction to machine translation*, 1992, p.3

<sup>5</sup> शर्मा, प्रमोद, *मशीनी अनुवाद : स्थिति, सीमाएँ और संभावनाएँ*, अनुवाद शतक विशेषांक, 1999, पृ.229

सन 1970 से 1980 तक विश्वभर में मशीनी अनुवाद में काफी रुचि बढ़ी और मशीनी अनुवाद एवं मशीन साधित अनुवाद प्रणालियों को विकसित कर उसका प्रयोग सरकारी प्रशासन, व्यापार और उद्योग के दस्तावेजों के अनुवाद के लिए किया जाने लगा। उसके बाद से 60 के दशक के पूर्वार्ध तक मशीनी अनुवाद संबंधी प्रयासों का तीन चरणों में अनुवाद की विधियों के रूप में विकास हुआ। इस प्रत्यक्ष विधि का प्रयोग स्रोत भाषा की सामग्री का विश्लेषण करने और उसे लक्ष्य भाषा के शब्दों के स्थान पर रूपांतरण करने तथा वाक्य-विन्यास के अनुरूप उसे संयोजित करने के लिए किया जाता था। इस विधि के फलस्वरूप 1964 में जॉर्ज टाउन विश्वविद्यालय में रूसी से अंग्रेजी अनुवाद तंत्र गैट (GAT) को विकसित किया और 1970 में इसे सिस्ट्रान (SYSTRAN) के रूप में परिवर्तित किया गया। इसकी सहायता से NASA ने अपोलो सोयूज के सहयोग से संबंधित सामग्री का अनुवाद किया है। यह तंत्र विश्व का प्रथम व्यावसायिक मशीनी अनुवाद तंत्र था। इस विधि के आधार पर हांगकांग के चीनी विश्वविद्यालय ने भी CULT (1968) नामक मशीनी अनुवाद तंत्र विकसित कर गणित और भौतिकी विषयों से संबंधित शोध पत्रों का अंग्रेजी में अनुवाद किया था। अन्य सफल प्रणालियों में TITUS तथा LOGOS आदि उल्लेखनीय मशीनी अनुवाद तंत्र विकसित किए गए हैं।<sup>6</sup>

कनाडा में अंग्रेजी के साथ फ्रेंच के राजभाषा बनने के बाद मॉट्रियल विश्वविद्यालय में मौसम संबंधी जानकारी का अंग्रेजी से फ्रेंच में अनुवाद करने के लिए एक पूर्ण स्वचालित मशीनी अनुवाद तंत्र TAUM METEO विकसित किया गया। यह अनुवाद तंत्र आज भी कार्यरत है। अमेरिका में टेक्सास यूनिवर्सिटी में व्यावहारिक मशीनी अनुवाद तंत्र के विकास के लिए METAL प्रोजेक्ट चल रहा है जिसे जर्मनी की सीमेन्स कंपनी वित्तीय सहायता देती है। 1984 में विकसित SUSY तंत्र का विकास रूसी-जर्मन प्रोटोटाइप सिस्टम द्वारा अंतरण विधि (Transfer Method) के अंतर्गत किया गया और फ्रांस के ग्रनेल विश्वविद्यालय ने इस विधि के अंतर्गत GETA तंत्र को सफलतापूर्वक विकसित किया। यूरोप में यूरोपियन आर्थिक समुदाय (European Economic Community) के व्यावहार हेतु नौ भाषाएँ स्वीकृत हैं।

---

<sup>6</sup> शर्मा, राजमणि, *अनुवाद विज्ञान*, 2004, पृ.184

1986 में यूरोपीय संसद ने मशीनी अनुवाद के विकास पर काफी पैसा खर्च किया था जिसके परिणामस्वरूप आज मशीनी अनुवाद पर यूरोप में कई परियोजनाओं में शोध-अध्ययन हो रहा है। EUROTRA और GETA वहाँ सरकारी वित्तीय सहायता प्राप्त प्रमुख परियोजनाएँ हैं। मशीनी अनुवाद के क्षेत्र में जापान ने भी विकसित टेक्नोलॉजी का प्रयोग कर 1980 में अनुवाद की दिशा में जापानी से अंग्रेजी में अनुवाद के लिए परियोजना शुरू की। 1988 से CICC (Centre for International Cooperation for Computerization) जापान ने बहुभाषिक मशीनी अनुवाद तंत्र के विकास पर काम शुरू कर इसमें अंग्रेजी, जापानी के अलावा थाई, मलेशियाई, इंडोनेशियाई और चीनी भाषाओं को मशीनी अनुवाद तंत्र में शामिल किया गया है। इसके साथ ही ATLAS-I, ATLAS-II, TAURUS, MU, PIVOT ये विश्व के कुछ प्रमुख व्यावहारिक मशीनी अनुवाद तंत्र हैं।<sup>7</sup> जिनका विकास विभिन्न देशों द्वारा विशिष्ट विषय-क्षेत्रों में अनुवाद के लिए किया गया है।

2000 के बाद मशीनी अनुवाद के क्षेत्र में काफी बदलाव आए हैं। अभी अधिकतर शोध सांख्यिकी मशीनी अनुवाद और उदाहरण आधारित मशीनी अनुवाद पर किए जा रहे हैं। स्पीच प्रौद्योगिकी से संबंधित शोध पर भी ध्यान दिया जा रहा है कि सीमित क्षेत्र में किस तरह इस प्रणाली का विकास किया जा सकता है। यूरोप और संयुक्त राज्य अमरीका में संसदीय भाषणों और समाचारों का प्रसारण करने के लिए स्वचालित मशीनी अनुवाद तंत्रों का विकास किया गया है। इसी तरह फ्रेंच-जर्मन 'Quaero' परियोजना मशीनी अनुवाद तंत्र का प्रयोग कर बहुभाषी इंटरनेट में उपलब्ध सामग्री जैसे- वेबपेज, ऑडियो-वीडियो सामग्री, स्पीच आदि के अनुवाद के लिए कार्य कर रही है।<sup>8</sup> इन मशीनी अनुवाद तंत्रों की तरह विश्व में अनेक परियोजनाएँ हैं जो इस तरह के तंत्रों को विकसित करने का प्रयास कर रही हैं। उसमें से कुछ महत्वपूर्ण प्रणाली या तंत्र आज भी काम कर रहे हैं। आज विश्व में कुछ मशीनी अनुवाद तंत्रों का प्रयोग सार्वभौमिक रूप से हो रहा है। जिसमें गूगल अनुवादक (Google Translate), माइक्रोसॉफ्ट बिंग ट्रान्सलेटर (Microsoft

---

<sup>7</sup> विकास, ओम, *मशीनी अनुवाद की समस्याएँ*, (संपा. डॉ.नगेन्द्र), अनुवाद विज्ञान सिद्धांत और अनुप्रयोग, 1993, पृ. 341-342

<sup>8</sup> [https://en.wikipedia.org/wiki/History\\_of\\_machine\\_translation](https://en.wikipedia.org/wiki/History_of_machine_translation) (retrieved on 13.02.2017)

Bing Translator), बेबेल फिश (Babel Fish), एशिया ऑनलाइन (Asia Online) का योगदान महत्वपूर्ण है।

### 1.3 मशीनी अनुवाद तंत्र की संरचना

स्रोत भाषा से लक्ष्य भाषा में स्वचालित अनुवाद करने के लिए मशीन में विभिन्न घटकों का समावेश किया जाता है जिससे किसी भी वाक (Speech) तथा पाठ (Text) का दूसरी भाषा में अनुवाद होता है। इस मशीनी अनुवाद तंत्र की संरचना में विभिन्न घटक, सामग्री तथा उपकरणों के समावेश होता है। जिससे मशीनी अनुवाद की प्रक्रिया पूर्ण होती है।

#### 1.3.1 कॉर्पोरा (Corpora)

मशीनी अनुवाद तंत्र विकसित करने के लिए कॉर्पोरा की आवश्यकता होती है। कॉर्पोरा किसी भी भाषा की सामग्री का संकलन होता है। जिसमें भाषा विश्लेषण के लिए लिखित या वाचिक पाठ के रूप में वाक्यों का संकलन होता है। वास्तविक भाषायी सामग्री के आधार पर भाषा का विश्लेषण और विवरण कार्पस भाषा विज्ञान (Corpus Linguistics) के अंतर्गत किया जाता है। कविता रस्तोगी के अनुसार मशीनी अनुवाद की प्रणाली विकसित करने के लिए प्रथम कॉर्पोरा की आवश्यकता होती है। कॉर्पोरा से अभिप्राय क्षेत्र विशेष से संबंधित अध्ययन सामग्री से होता है। इसकी सहायता से किसी विषय की सम्पूर्ण शब्दावली प्राप्त की जा सकती है। शब्दावली एकत्रित करने के उपरान्त वाक्य संरचना का विश्लेषण किया जाता है। तत्पश्चात विश्लेषक की आवश्यकता के अनुसार संरचना की दृष्टि से समान दस्तावेजों को अलग स्तरों में व्यवस्थित किया जाता है। जब स्रोत भाषा के दो या अधिक समान शब्दों के लिए लक्ष्य भाषा में किसी एक ऐसे समानार्थी शब्द चयन करने की आवश्यकता पड़ती है, जो लक्ष्य भाषा के अधिक से अधिक शब्दों का समानार्थी बनने की क्षमता रखता हो, उनके लिए अनेक शब्दों के विकल्प दिए जाते हैं जिससे संदर्भ के अनुसार उपयुक्त शब्द का चयन किया जाता है।<sup>9</sup>

---

<sup>9</sup> रस्तोगी, कविता, *समसामयिक अनुप्रयुक्त भाषाविज्ञान*, 2010, पृ.102

### 1.3.2 पूर्व संपादन (Pre-editing)

इसमें अनुवाद से पहले स्रोत भाषा पाठ का संपादन किया जाता है। अनुवाद करने से पहले ही पाठ के अनुवाद में आनेवाली सभी कठिनाइयों को दूर किया जाता है। जटिल व्याकरणिक संरचना, द्विअर्थक शब्द और वाक्यगत विशिष्टताओं को दूर किया जाता है। व्यावहारिक रूप से मशीनी अनुवाद दक्षता पूर्व संपादन कार्य है, जिसमें मानव के अनुवाद संबंधी ज्ञान की भूमिका महत्वपूर्ण होती है। उसी आधार पर पाठ का संपादन होता है।<sup>10</sup> इस प्रक्रिया में अनुवादक संपादन करते समय स्रोत पाठ के संदर्भ और अर्थ के आशय को बड़ी सूक्ष्मता से परिवर्तित करता है, जिससे मशीनी अनुवाद से आउटपुट में अधिक समस्या नहीं आती।

### 1.3.3 शब्द संसाधन (Word processing)

शब्द संसाधक (Word Processor) वस्तुतः टंकण कार्य का आधुनिक स्वरूप है। शब्द संसाधन करना भाषा संसाधन का आरंभिक सोपान है। कंप्यूटर द्वारा शब्द संसाधन की सहायता से संशोधन, परिवर्तन के अलावा हजारों पृष्ठों की सामग्री स्मृतिकोश में संग्रहीत की जा सकती है। आवश्यकता पड़ने पर उसे स्क्रीन पर फिर से संशोधित और परिवर्तित किया जा सकता है। इसमें संपूर्ण पाठ की एक फाइल से दूसरी फाइल में अंतरण किया जा सकता है। भारत और विदेशों में शब्द संसाधन के अनेक पैकेज विकसित किए गए हैं, जिनमें द्विभाषिक या फिर बहुभाषिक रूप में विभिन्न भारतीय भाषाओं के लिए शब्द संसाधन के कार्य किए जा सकते हैं।<sup>11</sup> इसलिए भारत सरकार के राजभाषा विभाग ने केन्द्र सरकार के कार्यालयों में कंप्यूटर संबंधी भाषा नीति की घोषणा करते हुए स्पष्ट किया था कि किसी भी कंप्यूटर को तभी द्विभाषिक माना जाएगा, जब उसमें शब्द संसाधन के साथ-साथ डेटा संसाधन की सुविधा भी हिंदी-अंग्रेजी में अर्थात् द्विभाषिक रूप में भी उपलब्ध होगी।

---

<sup>10</sup> मशीनी अनुवाद पर पाठ्य सामग्री, इग्नू, पृ. 34-35

<sup>11</sup> गोस्वामी, कृष्ण कुमार, *मशीनीकरण से कंप्यूटरीकरण*, Language Technology online e-journal (Retrieved on 05.06.2016)

### 1.3.4 शब्दतंत्र (Wordnet)

किसी भी भाषा का एक बहुत बड़ा शब्द संचय (Database) होता है, जिसमें संज्ञा, विशेषण, क्रिया और क्रिया-विशेषण संज्ञानात्मक समानार्थक शब्दों या पर्यायों के समूह में एकत्रित किया गया होता है और जिनमें से प्रत्येक शब्द एक पृथक संकल्पना को स्पष्ट करते हैं। शब्दों के ये पर्याय-समूह संकल्पनीय अर्थों तथा शाब्दिक संबंधों के माध्यम से एक-दूसरे से जुड़े हुए होते हैं। एक वाक्य में कहें तो शब्दतंत्र अर्थों के माध्यम से सुव्यवस्थित किया गया होता है और मशीन के पढ़ने योग्य एक शाब्दिक संचय होता है। शब्दतंत्र की संरचना प्राकृतिक भाषा संसाधन के क्षेत्र में हो रहे शोध के लिए भी उपयोगी साबित हो रही है। शब्दतंत्र शब्दों के मध्य विभिन्न शाब्दिक तथा अर्थपरक संबंधों को एक साथ लाने के लिए निर्मित तंत्र है। यह शाब्दिक जानकारियों को शब्दों के अर्थों के रूप में सुव्यवस्थित किया जाता है।<sup>12</sup>

हिंदी शब्दतंत्र शब्दों के मध्य विभिन्न प्रकार के सम्बन्धों को दर्शाने का एक संगणकीय शब्दतंत्र है। इसकी संरचना अंग्रेजी शब्दतंत्र पर आधारित है। हिंदी शब्दतंत्र में अर्थ की समानता के आधार पर पर्याय-समूह (Syn-set) का निर्माण एक स्पष्ट व्याख्या और वाक्य-प्रयोग के साथ किया जाता है, ताकि शब्दों की अनेकार्थता के कारण अर्थ की स्पष्टता बाधित न हो। वास्तव में, ये पर्याय समूह (Syn-set) ही हिंदी शब्दतंत्र के आधार हैं। आज इन शब्दतंत्रों का विकास विभिन्न भाषाओं में करने के लिए देश में अनेक शोध कार्य चल रहे हैं।

### 1.3.5 अभिकलनात्मक शब्दकोश

इस शब्दकोश का निर्माण मशीनी अनुवाद के लिए महत्वपूर्ण है जिसमें विभिन्न भाषाओं की शब्दावली का समावेश होता है। शब्दकोश संसाधनों के कार्यान्वयन से संबंधिक कार्य के रूप में अभिकलनात्मक शब्दकोश रचना की व्याख्या प्राकृतिक भाषा

---

<sup>12</sup> पाण्डेय, प्रभाकर, *हिंदी शब्दतंत्र की संरचना*, Language Technology online e-journal (Retrieved on 05.06.2016)

संसाधन में की जाती है। देखा जाए तो शब्दकोश रचना का कार्य आधार सामग्री पर अधिक जोर देने वाला है। इसलिए मशीन के द्वारा मिलने वाला सहयोग इस कार्य में अधिक महत्वपूर्ण हो जाता है। इस कार्य में कंप्यूटर, कोशकार और उपयोगकर्ता के लिए दोहरी भूमिका निभाता है। पहले तो कोशकार को संपादन करने के लिए शब्द प्रतिनिधित्व और शब्दकोशीय आधार सामग्री संसाधन करने में सहायता करता है तो दूसरी ओर कोश के उपयोगकर्ता को इलेक्ट्रॉनिक रूप में शब्द को खोजने में सहायता करता है।<sup>13</sup> इस संपूर्ण प्रक्रिया को पूर्ण रूप प्रदान करने के लिए एन.एल.पी. में वाचिक और लिखित दोनों रूपों में शब्दकोशीय संसाधन में सूचना आधार सामग्री की आवश्यकता होती है।

इस अभिकलनात्मक शब्दकोश में अनेक भाषाओं की शब्दावली का समावेश किया जाता है जिससे अनुवाद सरल शब्दों में हो। शब्दकोश में एक शब्द के लिए विभिन्न अर्थ का समावेश होता है तथा इसमें शब्दावली को विभिन्न रूपों में वर्गीकृत किया जाता है। इस संबंध में हिंदी में डॉ. कैलाशचन्द्र भाटिया तथा डॉ. जगन्नाथन के कार्य उल्लेखनीय हैं।<sup>14</sup> साथ ही भारत सरकार की ओर से हिंदी से अन्य भारतीय भाषाओं में शब्दकोश निर्माण के लिए 'भारतीय भाषाओं के लिए प्रौद्योगिकी विकास' संस्थान कार्यरत है।<sup>15</sup>

### 1.3.6 कृत्रिम बुद्धि (Artificial intelligence)

कृत्रिम बुद्धि का प्रयोग मशीनी अनुवाद तंत्र को विकसित करने के लिए किया जा रहा है। इसमें टेक्नोलॉजी के कई नए आयाम भी जुड़ गए हैं, जैसे इंटरनेट के जरिए तुरंत ऑनलाइन अनुवाद की सुविधा का उपलब्ध होना और कृत्रिम बुद्धि का उपयोग कर भाषा की संरचना के नियमों को यथासंभव अनुवाद तंत्र में समाकलित (integrate)

---

<sup>13</sup> कांबले, प्रकाश, *हिंदी-मराठी मशीनी अनुवाद में शाब्दिक अस्पष्टता*, एम.फिल.शोध प्रबंध 2009, पृ. 142

<sup>14</sup> शर्मा, प्रमोद, *मशीनी अनुवाद : स्थिति, सीमाएँ और संभावनाएँ*, 'अनुवाद' शतक विशेषांक, 1999, पृ. 244

<sup>15</sup> <http://tdil.mit.gov.in/> (retrieved on 14.07.2017)

करना।<sup>16</sup> यदि हम किसी तरह इन नियमों और तर्कों को व्यक्त करने में समर्थ हो सकें तो कंप्यूटर द्वारा समझे जा सकने वाले अल्गोरिदम या सूत्रों में परिवर्तित कर कंप्यूटर भी इस कृत्रिम बुद्धि के आधार पर वही निर्णय और परिणाम दे सकता है जो मनुष्य देता है। इसलिए कृत्रिम बुद्धि का समावेश मशीनी अनुवाद तंत्रों में किया जा रहा है जिससे अनुवाद स्वचालित रूप में प्राप्त हो।

### 1.3.7 विश्लेषक (Analyser)

विश्लेषक के अंतर्गत वाक्य रचना (Syntactic) विश्लेषक, अर्थगत (Semantic), विश्लेषक और संदर्भगत (Contextual) विश्लेषक आते हैं। वाक्य रचना विश्लेषक वाक्य से विभक्ति को हटाकर (Parsing) शब्दरूप विश्लेषण (Morphological Analysis) करता है। जैसे goes से go; 'जाता है', तथा 'जा रहा है' से 'जा' मूल शब्द निकलता है। इन्हें शब्द संहिता में देखा जा सकता है। अर्थगत विश्लेषण में शब्द स्तर पर संबंधों की जानकारी दी जाती है। संदर्भगत विश्लेषक पूर्व कोडित जानकारी से संदर्भ सूचना निकालता है। यह जानकारी संकल्पनात्मक शब्द संहिता से विशाल संरचना के रूप में संग्रहित की जाती है। वाक्य रचना विश्लेषण के बाद व्याख्यायित पद-निरूपण किया जाता है। अर्थगत और संदर्भगत विश्लेषण के बाद दिए गए वाक्य का अंतःनिरूपण (Intermediate representation) फ्रेम संरचना (Frames structure) के रूप में किया जाता है। साथ ही पाठ में आनेवाली अनेकार्थता को वाक्यगत विश्लेषण से दूर किया जा सकता है।<sup>17</sup>

### 1.3.8 टैगिंग और पार्सर (Tagging and Parser)

टैग एक ऐसा व्याकरण स्वरूप है जो वृक्ष संरचना पर आधारित है। इसके माध्यम से वाक्य के आंतरिक लक्षणों को शब्द वर्ग के माध्यम से उन्हें संज्ञा, क्रिया,

---

<sup>16</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 257

<sup>17</sup> विकास, ओम, *मशीनी अनुवाद की समस्याएँ*, अनुवाद विज्ञान सिद्धांत और अनुप्रयोग-

डॉ.नगेन्द्र, 1993, पृ. 346

विशेषण आदि में चिन्हित (Marking up) किया जाता है और पार्सर वाक्य की विभिन्न इकाइयों को पद-विच्छेदन हेतु प्रयुक्त व्याकरण के आधार पर परिभाषित या नामित करता है। किसी भी भाषा के वाक्य या पाठ के लिए पार्सर का उपयोग वाक्य के विभिन्न घटकों से संबंधित सूचना और संज्ञा व क्रिया के कारकीय संबंधों की सूचना के लिए किया जाता है। इस सूचना को विभिन्न अनुवाद तंत्र अपनी-अपनी प्रक्रिया के अनुरूप प्रयोग करते हैं। अनुवाद तंत्र विकसित करने में एक समस्या उचित पार्सर के लिए उचित व्याकरणिक प्रविधि के चयन को लेकर भी आती है। इसके लिए कारक व्याकरण, ट्री एडजाएनिंग व्याकरण, लेक्सिकल फंक्शनल ग्रामर, फ्रेज स्ट्रक्चर ग्रामर आदि में से किसी एक का चयन कर उसके आधार पर अलग-अलग तरह के पार्सर विकसित किए गए हैं।<sup>18</sup> इसके समाधान के लिए किसी ऐसी विधि का चयन करना आवश्यक है जो भारतीय भाषाओं की अपनी हो और भारतीय भाषा से कंप्यूटर वैज्ञानिक परिचित हों। इस आधार पर पाणिनि के नियमबद्ध संस्कृत व्याकरण के आधार पर पार्सर विकसित किए जा सकते हैं।

### 1.3.9 अंग्रेजी प्रजनक (English generator)

स्रोत भाषा से लक्ष्य भाषा में अनुवाद करने की प्रक्रिया विश्लेषण-अंतरण-प्रजनक पर आधारित होती है। स्रोत भाषा के किसी भी वाक्य या पाठ के अनुवाद के लिए उसका विश्लेषण होता है। विश्लेषण के बाद लक्ष्य भाषा संरचना के लिए शब्दावली का चयन किया जाता है और प्रजनक की प्रक्रिया में लक्ष्य भाषा की संरचना के अनुरूप व्याकरणिक नियमों का प्रयोग कर मशीन द्वारा अनुवाद किया जाता है।<sup>19</sup> इसी तरह अंग्रेजी से हिंदी की संरचना में अनुवाद करने के लिए प्रजनक में कुछ व्याकरणिक नियमों का निर्माण करना होता है। अंग्रेजी प्रजनक पाठ में निहित पूर्वसर्ग, संज्ञा,

---

<sup>18</sup> शर्मा, प्रमोद, *मशीनी अनुवाद: स्थिति, सीमाएँ और संभावनाएँ*, 'अनुवाद' शतक विशेषांक - 100-101, 1999, पृ.234

<sup>19</sup> Dave S., Parikh J., Bhattacharyya, P. *Interlingua-Based English-Hindi Machine Translation and Language Divergence*, 2001, p. 251-304

सर्वनाम, विशेषण, क्रिया विशेषण, क्रिया, अव्यय और व्याकरणिक कोटियों में लिंग, वचन, काल, पुरुष, वृत्ति आदि की संरचना को लक्ष्य भाषा में प्रजनित करने का कार्य करता है।

### 1.3.10 पाठ संश्लेषण (Text generation)

अंतः निरूपित अभिव्यक्ति से वाक्य या संपूर्ण पाठ को तैयार करने की क्रिया जटिल है। यह प्रक्रिया 'विश्लेषक' की प्रक्रिया का प्रतिबिंब है। संसाधन की जटिलता को कम करने के उद्देश्य से शब्द संहिता (Lexicon) का आकार भी सीमित रखा जाता है। कुछ शब्द लक्ष्य भाषा में संस्कृति के कारण स्वीकार्य नहीं होते हैं, ऐसे शब्दों को समझाने के लिए जानकारी ज्ञान कोश से ली जाती है। यह निर्णय प्रायः मानव प्रयोक्ता की सहायता से लिया जाता है।

### 1.3.11 अंतरण व्याकरण (Transfer grammar)

मशीनी अनुवाद में अंतरण व्याकरण की भूमिका महत्वपूर्ण है। इस व्याकरण से नियम निर्माण के लिए दोनों भाषाओं की संपूर्ण संरचना की जानकारी होना आवश्यक है। इस नियम के माध्यम से मशीन द्वारा प्राप्त अनुवाद में गुणवत्ता को अधिक सुधारा जा सकता है। मशीन में नियमों का समावेश (Installation) करने के साथ उसे प्रयुक्त करने के लिए अल्गोरिदम (Algorithm) बनाने की आवश्यकता होती है, जो नियमों को व्यवस्थित रूप में समावेशित और संचालित करते हैं। इस अल्गोरिदम के माध्यम से मशीन द्वारा संपूर्ण क्रियाकलाप (Process) किया जाता है। मशीन अनुवाद में स्थानांतरण व्याकरण (Transfer Grammar) के दो प्रयोजन हैं। पहला, यह अनुवाद प्रक्रिया के दौरान स्रोत भाषा (Source Language) के शब्दों और अभिव्यक्तियों के लिए लक्ष्य भाषा (Target Language) के समानार्थी शब्दों और अभिव्यक्तियों को उपलब्ध कराने का कार्य किया जाता है। दूसरा, इन समानार्थी शब्दों और अभिव्यक्तियों को लक्ष्य भाषा की वाक्य संरचना में सही स्थान पर और उचित क्रम में स्थापित होने का निर्देश दिया जाता है। भारत में आई.आई.टी. हैदराबाद ने संपर्क तंत्र के लिए स्थानांतरण व्याकरण नियमों का निर्माण किया है जिसमें एक भाषा से अनेक भाषाओं के लिए

नियमों का निर्माण किया गया है। यह नियम एक भारतीय भाषा (IL-1) से दूसरी भारतीय भाषा (IL-2) में अनुवाद के लिए महत्वपूर्ण है। इस पर अविनेश (IIT HYDERABAD) के शोध-प्रबंध कार्य में संपर्क तंत्र (अंग्रेजी-हिंदी मशीनी अनुवाद तंत्र) निर्माण में स्थानांतरण व्याकरणिक नियमों के प्रयोग को स्पष्ट किया गया है।<sup>20</sup>

### 1.3.12 शाब्दिक व्याकरण (Lexical grammar)

शब्दगत व्याकरण में शब्द जब अपने सभी संगत व्याकरणिक लक्षणों से चिन्हित होकर व्यक्त होते हैं तो उसे लेक्सिकन कहा जाता है। इन व्याकरणिक लक्षणों में शब्द-वर्ग की पहचान (संज्ञा, विशेषण, क्रिया आदि); वचन-लिंग की पहचान; सहप्रयोगों (collocation) की पहचान, यानी ऐसे परसर्गों या शब्दों की पहचान जो अनिवार्य रूप से इन शब्दों के साथ-साथ प्रयुक्त होते हैं। लेक्सिकल व्याकरण के पीछे यह मान्यता है कि हर शब्द का अपना एक अलग व्याकरण होता है जिसकी वाक्य रचना में अहम भूमिका होती है। उदाहरण के लिए of शब्द का यह व्याकरण है कि यह हमेशा दो संज्ञाओं के बीच में आता है और पहली संज्ञा से पहले आर्टिकल the/a का प्रयोग होता है।<sup>21</sup> लेक्सिकल व्याकरण का भी मशीनी अनुवाद तंत्र के निर्माण के लिए प्रयोग किया जाता है। कंप्यूटर में इसका समावेश किया जाता है जिससे मशीनी अनुवाद में व्याकरणिक गलतियाँ कुछ हद तक कम होती हैं।

### 1.3.13 हिंदी प्रजनक (Hindi generator)

मशीनी अनुवाद में जब अंग्रेजी से हिंदी भाषा में अनुवाद किया जाता है तो हिंदी प्रजनक, लक्ष्य भाषा की संरचना के अनुरूप कार्य करता है। जिसमें दोनों भाषाओं की भाषाई सामग्री, द्विभाषी शब्दकोश, रूपवैज्ञानिक विश्लेषक, पार्सर, चंकिंग और व्याकरण जाँचक का समावेश मशीनी अनुवाद तंत्रों में किया जाता है। इन सभी घटकों की प्रक्रिया

---

<sup>20</sup> Avinesh, P.V.S., *Transfer Grammar Engine and Automatic Learning of Reorder Rules in Machine Translation* (M.Sc- Thesis), 2010, p.18

<sup>21</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003 पृ.266

के बाद हिंदी प्रजनक लक्ष्य भाषा में अनुवाद करता है। अगर पाए जानेवाले अनुवाद में कठिनाईयाँ आती हैं तो कंप्यूटर में समावेशित शब्दकोश से इसका समाधान किया जाता है और अनुवाद किया जाता है।

#### 1.3.14 पश्च संपादन (Post editing)

यह संपादन अनुवाद के निर्गम पाठ पाने के बाद किया जाता है। इस संपादन में मशीनी अनुवाद से निर्गत वाक्य प्राप्ति के बाद लक्ष्य भाषा की संरचना के अनुसार संपादन किया जाता है और इसमें शब्दों के क्रम को बदलने की सुविधा भी होती है। दो छोटे वाक्यों को एक बड़े वाक्य में रखने और इसके विपरित एक बड़े वाक्य को दो छोटे वाक्यों में करने की आवश्यकता पड़ती है। पश्च संपादन यह अनुवाद की अंतिम प्रक्रिया होती है। इसमें अनुवाद पूर्णतया सहज और सटीक रूप में प्राप्त किया जा सकता है।<sup>22</sup> इसमें निर्गत अनुवाद में सुधार करने की सुविधा मशीनी अनुवाद तंत्रों में होती है।

इन सभी संरचनात्मक इकाईयों से गुजरने के बाद मशीनी अनुवाद से स्रोत भाषा का पाठ लक्ष्य भाषा में अनूदित होता है। अनूदित होने वाला पाठ कितना सही अनूदित हुआ है यह मशीनी अनुवाद तंत्र पर निर्भर करता है कि उसका निर्माण किन विषय-क्षेत्रों के अनुवाद के लिए किया गया है।

#### 1.4 मशीनी अनुवाद की आवश्यकता (Requirement of Machine Translation)

मानव सभ्यता के विकास के बाद उत्तरोत्तर ज्ञान में वृद्धि हो रही है। विभिन्न संस्कृतियों को जानने के लिए भाषा एक सेतु का कार्य कर रही है। पहले धर्म-प्रसार करने के लिए भाषाई माध्यम तथा दुभाषियों के द्वारा अनुवाद को एक नई दिशा प्राप्त हुई जिसके चलते धर्म का प्रचार और प्रसार करना आसान हुआ। आज विभिन्न समुदायों और देशों के विकास के चलते सूचना का आदान-प्रदान शीघ्र हो रहा है जिसके फलस्वरूप अनुवाद की आवश्यकता भी बढ़ती गई। आज सूचना क्रांति के युग में विभिन्न

---

<sup>22</sup> मशीनी अनुवाद पर पाठ्य सामग्री, इग्नू, पृ. 34

अंतरराष्ट्रीय कंपनियों को अपने उत्पादों को बेचने के लिए यह सहज माध्यम के रूप में उपलब्ध हुआ है। आज किसी भी उत्पाद की सूचनाओं को हम अंग्रेजी के अलावा स्थानीय भाषा में भी पढ़ सकते हैं। यह सूचना आज कंप्यूटर के माध्यम से लोगों तक तत्काल पहुंच रही है जिसमें मशीनी अनुवाद की बहुत बड़ी भूमिका है। जहाँ एक ओर मशीनों से शारीरिक श्रम से अधिक राहत मिली, वहीं दूसरी ओर कंप्यूटरों से अनुवाद जैसे बौद्धिक कार्यों में मदद भी मिलने लगी है।

मशीनी अनुवाद में कंप्यूटर का इस्तेमाल होता है। शब्द संसाधन से दिन-प्रतिदिन मनुष्य का कार्य करना आसान हो रहा है। कार्यालयी कामकाज में रिपोर्ट लेखन, सरकारी दस्तावेजों का आधिकारिक भाषा में अनुवाद करना अनिवार्य हो गया है। इस प्रकार के दस्तावेजों को पठनीय और उपयोगी बनाने के लिए अन्य क्षेत्रों में भी इन तंत्रों के विकास करने की आवश्यकता है। भारत में 'संपर्क' तथा 'मंत्र-राजभाषा' मशीनी अनुवाद तंत्र का विकास इस क्षेत्र में उपयोगी साबित हुआ है। आज भारत के केवल 12 प्रतिशत लोग ही अंग्रेजी भाषा में बात करते हैं।<sup>23</sup> अधिकांश सूचना सामग्री अंग्रेजी भाषा में उपलब्ध होने के कारण सामान्य व्यक्ति महत्वपूर्ण जानकारी से अनभिज्ञ रहता है। इस दूरी को मशीनी अनुवाद कम करने का काम करेगा। इसलिए विश्व में करोड़ों लोगों के बीच सीधे संवाद स्थापित करना तथा विश्व की सभी स्थानीय भाषाओं में तकनीकी एवं भाषाई संसाधन के माध्यम से विकास करना ही इसका उद्देश्य है। इस तंत्र की महत्ता निम्नलिखित बिंदुओं से ज्ञात होती है-

- मशीन द्वारा अनुवाद अत्यंत अल्प समय में होने से पाठ का तुरंत आशय सामने आता है। अधिकांश स्थितियों में पाठ का सार समझने में इतना ही पर्याप्त होता है।
- बैंकिंग, तकनीकी आदि विषयों में मशीनी अनुवाद विशेष रूप से सहायक सिद्ध हो रहा है।
- कंप्यूटर एवं इंटरनेट के अतिरिक्त खर्च शून्य होंगे।

---

<sup>23</sup>[https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_English-speaking\\_population](https://en.wikipedia.org/wiki/List_of_countries_by_English-speaking_population)  
[Retrieved on 13.02.2017]

- गोपनीयता एवं निजत्व की रक्षा की जाती है।
- भारत की सभी भाषाएँ सजातीय होने के चलते एक ही प्रोग्राम के माध्यम से अनेक भाषाओं में अनुवाद किया जा सकता है।
- अनुवादक अधिकांशतः दो ही भाषाओं में प्रवीण होता है।
- किसी पाठ को मशीन द्वारा अनुवाद करने के बाद उसे किसी प्रशिक्षित व्यक्ति द्वारा सुधार किया जा सकता है या पश्च संपादन किया जा सकता है।<sup>24</sup>

इन बिंदुओं से यह ज्ञात होता है कि भारत जैसे बहुभाषी देश के लिए मशीनी अनुवाद तंत्रों का विकास करना अधिक आवश्यक है। साथ ही मशीनी अनुवाद की यह बढ़ती उपयोगिता इस बात को भी प्रमाणित करती है कि मशीनी अनुवाद तंत्रों को विकसित करनेवाली परियोजनाओं को सरकार के अलावा औद्योगिक फर्मों से भी वित्तीय सहायता मिल रही है। जिसके चलते इस क्षेत्र में अधिक शोध एवं अध्ययन हो रहे हैं।

## 1.5 मशीनी अनुवाद के प्रकार (Types of Machine Translation)

अनेक भाषावैज्ञानिकों ने अनुवाद के प्रकारों की तरह मशीनी अनुवाद के प्रकारों को भी विवेचित किया है। प्रो. राजमणि शर्मा<sup>25</sup> ने भी मानव की सहभागिता या सहायता के आधार पर मशीनी अनुवाद के प्रकार दिए हैं। मशीनी अनुवाद एक स्वचालित प्रक्रिया है लेकिन भाषावैज्ञानिकों द्वारा कहा जाता है कि अनुवाद प्राप्ति के बाद उसका पश्च संपादन मानव द्वारा किया जाना चाहिए जिससे अनुवाद सहज और सरल दूसरे भाषा में होता है।

### 1.5.1 मानव अनुवाद (Human Translation)

---

<sup>24</sup> शर्मा, राकेश, *मशीनी अनुवाद: स्थिति एवं संभावनाएँ*, अनुवाद पत्रिका, 2014, जनवरी-मार्च, अंक: 158, पृ.68

<sup>25</sup> शर्मा, राजमणि, *अनुवाद विज्ञान*, 2004, पृ.183

मानव अनुवाद की प्रक्रिया में मानव स्वतः अनुवाद करता है। वह किसी भी विषय क्षेत्र के पाठ का अनुवाद करता है। साहित्यिक अनुवाद के पाठ में संस्कृति, भाव, संवेदना, लक्षणा एवं व्यंजना रहती है, जिसे मानव अपनी सहज ग्रहण की हुई मस्तिष्क क्षमता के आधार पर वह अनुवाद प्रदान करता है, जिसमें मशीन की कोई सहायता नहीं ली जाती।

### 1.5.2. मशीन साधित मानव अनुवाद (Machine Aided Human Translation)

मशीन साधित मानव अनुवाद की प्रक्रिया में कंप्यूटर के स्मृति कोश (Memory) में संग्रहीत उपयुक्त शब्दावली का प्रयोग कर अनुवादक अनुवाद करता है। यह तकनीकी और वैज्ञानिक अनुवाद के लिए ज्यादा उपयुक्त है, क्योंकि इस अनुवाद में शब्द-प्रति-शब्द अनुवाद किया जाता है।

### 1.5.3. मानव साधित मशीन अनुवाद (Human Aided Machine Translation)

मानव साधित मशीनी अनुवाद में अनुवादक ही पूर्व संपादन करता है और वह पाठ कंप्यूटर को अनुवाद के लिए देता है। इसमें अनुवादक की आवश्यकता पड़ने पर वाक्यों की संदिग्धता को स्पष्ट करके वाक्य को सरल बनाता है और अनुवाद पूर्ण होने के बाद पश्च-संपादन किया जाता है।<sup>26</sup>

### 1.5.4. मशीनी अनुवाद (Machine Translation)

मशीनी अनुवाद में एक भाषा से दूसरी भाषा में अनुवाद के लिए कंप्यूटर में संचालित कृत्रिम बुद्धि के माध्यम से शब्दगत विश्लेषण और अर्थगत संबंधों को जोड़कर संकल्पनाओं का स्वीकार्य पाठ में निरूपण किया जाता है। इसमें मशीनी अनुवाद द्वारा की जानेवाली भाषाओं के संदर्भ में कॉर्पोरा का स्थान बहुत महत्वपूर्ण है, जिसमें किसी भाषाई सामग्री के विभिन्न घटकों का समावेश होता है। इसमें शब्द की परिभाषा, प्रयोग

---

<sup>26</sup> Karunkaran, k., Jaykumar, Translation as Synthesis, 'Human Aided Machine Translation: Problems and Perspective', 1988, p.183

और व्याकरणिक अभिलक्षणों के साथ शब्दगत तथा अर्थगत संबंधों का उल्लेख भी होता है। जिससे अनुवाद बिना मानव की सहायता लिए स्वतः हो जाता है।

मशीनी अनुवाद के प्रकारों से यह स्पष्ट होता है कि अभी तक के मशीन अनुवाद में मानव की सहायता महत्वपूर्ण रही है लेकिन अब मशीन के माध्यम से पूर्णतः अनुवाद प्राप्त करने का प्रयास हो रहा है।

## 1.6 मशीनी अनुवाद की प्रक्रिया (Process of Machine Translation)

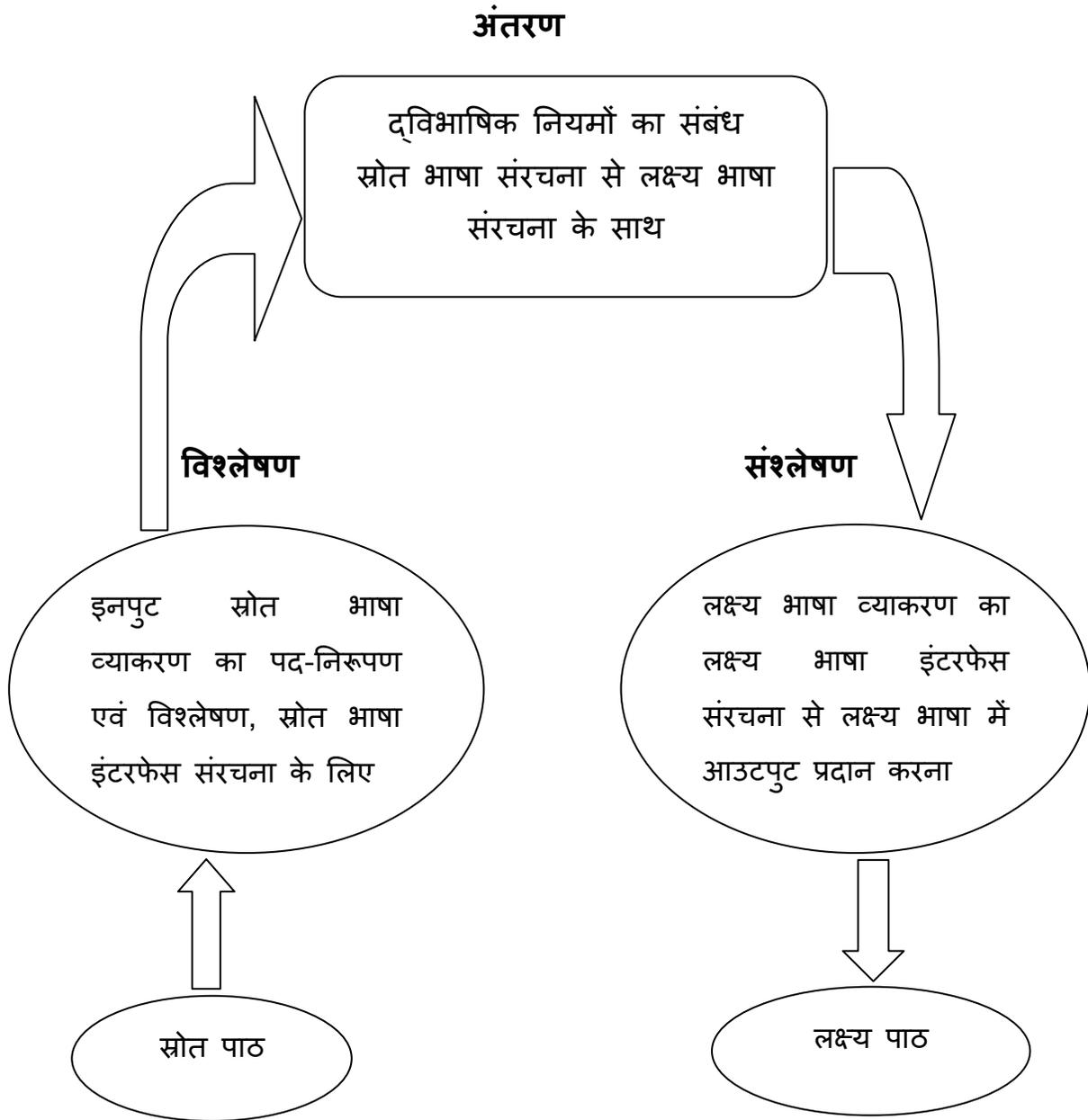
एक भाषा से दूसरी भाषा में अनुवाद करना यह एक तरह से बौद्धिक प्रक्रिया है। इसके लिए युजीन ए. नाएडा ने अनुवाद प्रक्रिया के तीन चरण बताए हैं- विश्लेषण, अंतरण और पुनर्रचना। इन तीन चरणों का बड़ा महत्व है जिसके बिना अनुवादक की सैद्धांतिक प्रक्रिया पूर्ण नहीं होती। उसी तरह मशीनी अनुवाद में भी भाषा संबंधी विश्लेषण, अंतरण और संश्लेषण या प्रजनन के पक्ष समाहित होते हैं। जिस प्रक्रिया से मशीन द्वारा अनुवाद संपादन होता है।<sup>27</sup> मशीनी अनुवाद में प्रथम कार्य को पद-अन्वय (Parsing) कहा जाता है, जिसमें पाठ के अंतर्गत वाक्य के घटकों (Components) के मध्य कार्यरत प्रकार्यात्मक संबंधों की पहचान कर वाक्य का पद-विच्छेदन किया जाता है और अंतरण (Transfer) के चरण में यह कार्य 'लेक्सिकन ट्रान्सफर' (LT) के द्वारा किया जाता है। यह स्रोत भाषा पाठ के लिए आगत (Input) शब्दों एवं अभिव्यक्तियों के लिए समानार्थी शब्द प्रदान करता है। अनुवाद में पुनर्गठन की तरह प्रजनक मशीनी अनुवाद में लक्ष्य भाषा के अनुवाद को लिपि, शब्द, वाक्य, शैली तथा रूप में प्रस्तुत करता है अर्थात् पार्सर से आए शब्दों द्वारा लक्ष्य भाषा की संरचनानुसार वाक्य गठन का कार्य प्रजनक (Generator) करता है।<sup>28</sup> इस प्रक्रिया में मशीनी अनुवाद तंत्र निर्माण के लिए समावेशित घटकों का बड़ा योगदान होता है, जिसका प्रयोग कर अनुवाद स्रोत भाषा

---

<sup>27</sup> गोस्वामी, कृष्ण कुमार, *अनुवाद विज्ञान की भूमिका*, 2012, पृ.168-170

<sup>28</sup> रस्तोगी, कविता, *समसामयिक अनुप्रयुक्त भाषाविज्ञान*, 2010, पृ.101

से लक्ष्य भाषा में किया जाता है। मशीनी अनुवाद की इस प्रक्रिया को निम्नलिखित आकृति में दिया गया है।



आकृति 1.1 : मशीनी अनुवाद की प्रक्रिया <sup>29</sup>

<sup>29</sup> Arnold D., Balkan L., et.al. (ed.), *Machine Translation: an Introductory Guide*, 1994, p.68

मशीनी अनुवाद प्रक्रिया के लिए आवश्यक संसाधन के घटक पूर्ण रूप में विकसित न होने की वजह से भारत में मशीनी अनुवाद तंत्र (MT System) से प्राप्त अनुवाद तथा तंत्र निर्माण में विभिन्न समस्याएँ पायी जाती हैं। ये समस्याएँ भाषिक संरचना, वाक्यगत भाषिक अस्पष्टता, संदर्भ परक अस्पष्टता, मुहावरों एवं कहावतों के अनुवाद की समस्या, शब्दगत संबंध की समस्या तथा विकसित नए शब्दों के अनुवाद की समस्या होती है, जिससे अनुवाद में दुरुहता उत्पन्न होती है और मशीनी अनुवाद द्वारा अनुवाद में अक्सर गलतियाँ पायी जाती हैं। इन समस्याओं का मूल कारण भाषिक संसाधनों का पूर्ण विकास न होना है। इसलिए मशीनी अनुवाद तंत्रों को विकसित करने की आवश्यकता है। इसके लिए मशीनी अनुवाद के तंत्र विभिन्न पद्धतियों पर तथा दृष्टिकोण पर आधारित है।

### 1.7 मशीनी अनुवाद की पद्धतियाँ

कंप्यूटर के विकास के बाद इसकी कार्यक्षमता और कार्य करने की गति का लाभ उठाने के उद्देश्य से अनुवाद के क्षेत्र में काम करने वाले प्रारंभिक अनुसंधानकर्ताओं (1964) ने मशीन अनुवाद को मूल रूप से एक शुद्ध इंजीनियरींग प्रक्रिया के रूप में देखा और इसके व्याकरणिक पक्ष को नजर अंदाज करते हुए मशीन अनुवाद प्रणाली विकसित करनी शुरू की और उसके बाद इसके जो परिणाम सामने आए वे बहुत आशावर्धक नहीं थे। अंग्रेजी-रूसी मशीनी अनुवाद तंत्र के अनुवाद में गलतियों के बाद अनुसंधानकर्ताओं को यह स्पष्ट होने लगा कि मशीनी अनुवाद वास्तव में कंप्यूटर विज्ञान और भाषाविज्ञान दोनों के माध्यम से ही मशीनी अनुवाद तंत्र का विकास किया जा सकता है। कोई भी मशीनी अनुवाद तंत्र तब तक सफल नहीं हो सकता जब तक अर्थ, वाक्य-विन्यास, शब्द रचना और यहाँ तक कि पाठ में आए पदबंधों और मुहावरों की व्याकरणिक संरचना का समावेश इस तंत्र में नहीं होता।<sup>30</sup> उसके बाद मनुष्य की बुद्धि की तरह उन्होंने कृत्रिम बुद्धि (Artificial Intelligence) को बनाने का प्रयास किया। जिसे 1980-90 के दौरान उसके संतोषजनक परिणाम मिले और रोबोट टेक्नोलॉजी की संकल्पना को साकार रूप

---

<sup>30</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 256

मिलना शुरू हुआ। तब से यह आशा बढ़ने लगी कि कृत्रिम बुद्धि की सहायता से मशीनी अनुवाद प्रणाली को भी एक सीमा तक बुद्धियुक्त किया जा सकता है। इस कृत्रिम बुद्धि में भाषा के मूलभूत व्याकरणिक नियम या सूत्र संचित किए जाते हैं। जिससे मशीन द्वारा अनुवाद करना आसान हो रहा है। आज मशीनी अनुवाद के विभिन्न तंत्र अनेक दृष्टिकोण तथा पद्धतियों पर आधारित हैं। जैसे- नियम आधारित, उदाहरण आधारित, ज्ञान आधारित, सांख्यिकीय आधारित, मिश्रित संरचना पर तथा कॉर्पस आधारित संरचना आदि पद्धतियों पर मशीनी अनुवाद तंत्र आधारित हैं। जिनका नीचे विवरण दिया गया है।

### 1.7.1 नियम आधारित मशीनी अनुवाद (Rule Based Machine Translation)

यह पद्धति पहली व्यावसायिक मशीनी अनुवाद प्रणाली में से एक है। जिसमें शब्द के अलग-अलग स्थानों पर संदर्भ के अनुसार शब्दों के अर्थ का प्रयोग भाषाई नियमों से किया जाता है। नियम आधारित मशीनी अनुवाद में भाषाई नियमों का बृहद संग्रहण कर तीन अलग चरण- विश्लेषण, अंतरण और प्रजनन में प्रयोग किया जाता है। इन भाषाई नियमों का निर्माण मानव भाषा विशेषज्ञानिकों और प्रोग्रामर से किया जाता है। यह तंत्र मानव द्वारा निर्मित 'अनुवाद शब्दकोश' पर निर्भर रहता है, जिसमें अनुवाद को संपादित और सुधारित करने की सुविधा उपयोगकर्ताओं को होती है। इस पद्धति में स्रोत भाषा की संरचना के आधार पर लक्ष्य भाषा में नियमों का प्रतिपादन किया जाता है, उस आधार पर अनुवाद होता है। प्रमुखतः नियम आधारित मशीनी अनुवाद तंत्र भाषाई नियम और द्विभाषिक शब्दकोश के माध्यम से वाक्यगत और अर्थगत संरचना के अनुसार स्रोत भाषा पाठ से लक्ष्य भाषा में अनुवाद करते हैं।<sup>31</sup>

**1.7.1.1 प्रत्यक्ष पद्धति (Direct method):** यह पद्धति सबसे पहले जॉर्जटाउन विश्वविद्यालय द्वारा 1964 में रूसी-अंग्रेजी अनुवाद के संदर्भ में विकसित की गई थी। यह प्रणाली मूल रूप से दो भाषाओं के शब्दकोशों में उपलब्ध जानकारी, शब्द-रचना के बृहद नियमों और सतही वाक्य-संरचना पर आधारित थी। इसमें व्याकरणिक नियम तथा

---

<sup>31</sup> [https://en.wikipedia.org/wiki/Rule-based\\_machine\\_translation](https://en.wikipedia.org/wiki/Rule-based_machine_translation) [Retrieved on 08.06.2016]

पद निरूपण (Parsing) को विशेष महत्व नहीं दिया गया। इसलिए इस विधि से प्राप्त ज्यादातर अनुवाद शाब्दिक या कृत्रिम होता था, जिसमें कई बार अर्थ का अनर्थ होने की संभावना रहती थी। बाद में 1976 में इस पद्धति का परिष्कृत रूप सिस्ट्रान के नाम से विकसित हुआ। पहले इसमें लगभग 20 से अधिक भाषाई युग्मों के अनुवाद की सुविधा थी। आज यह मशीनी अनुवाद तंत्र निर्माण के लिए विकासकर्ता को प्लेटफार्म प्रदान करती है। इसके माध्यम से अनुवाद की शुद्धता 60-70 प्रतिशत के आस-पास मानी जाती है। आज इंटरनेट के कुछ वेबसाइट्स में अंग्रेजी, स्पेनिश, फ्रांसीसी, जर्मन आदि भाषाओं के ऑनलाइन अनुवाद के लिए इसी के अद्यतन रूप का प्रयोग किया जाता है।<sup>32</sup>

**1.7.1.2 अंतरण पद्धति (Transfer method):** इस पद्धति के अंतर्गत अनुवाद की प्रक्रिया तीन चरणों से होकर गुजरती है। पहले चरण में स्रोत भाषा के वाक्य का व्याकरणिक विश्लेषण कर उसकी एक सूक्ष्म संरचना निर्मित की जाती है। दूसरे चरण में स्रोत भाषा की इस संरचना को लक्ष्य भाषा के समक्ष (Corresponding) संरचना में अंतरित (Transfer) किया जाता है। तीसरे चरण में इस आउटपुट को लक्ष्यभाषा में पुनर्गठित (Reorganize/ reconstruct) किया जाता है जो इस प्रणाली के अनुवाद का निर्गम (Output) रूप होता है।<sup>33</sup>

**1.7.1.3 अंतर्भाषिक पद्धति (Interlingua method) :** इस पद्धति में स्रोत भाषा (Source Language) और लक्ष्य भाषा (Target language) के बीच एक अन्य मध्यस्थ (Intermediate) भाषा जिसे 'सार्वभौम भाषा' कहा जाता है जिसका सृजन (Creation) किया जाता है। यह मध्यस्थ भाषा वास्तविक भाषा न होकर एक ऐसी भाषानिरपेक्ष भाषा होती है जिसमें सांसारिक ज्ञान (World knowledge) को संकल्पनाओं के प्रतीकों के रूप में निरूपित किया जाता है। सांसारिक ज्ञान या विश्वज्ञान के अंतर्गत विश्व की सभी वस्तुएँ, प्राणी, विचार और संबंध आदि संकल्पनाएँ शामिल हैं। व्याकरण के रूप में विश्वज्ञान का निरूपण (representation) और वर्गीकरण के अलावा इसमें प्रत्येक कोटि

---

<sup>32</sup> गोस्वामी, कृष्ण कुमार, *अनुवाद विज्ञान की भूमिका*, 2012, पृ.168-170

<sup>33</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003 पृ. 257

को अर्थपरक लक्षणों (Semantic features) के रूप में प्रस्तुत किया जाता है। इस पद्धति का भविष्य में यह लाभ होगा कि यदि इस आधार पर विश्वज्ञान निरूपण की कोई सफल सार्वभौम (Universal) भाषा विकसित होती है तो इस एक युक्ति से एक से अधिक भाषाओं के लिए मशीनी अनुवाद तंत्र का विकास करना संभव हो सकेगा। लेकिन इसमें शब्दों के आर्थिक विश्लेषण और कृत्रिम बुद्धि के भाषिक पक्ष पर अधिक शोध करने की आवश्यकता है।<sup>34</sup>

### 1.7.2 उदाहरण आधारित मशीनी अनुवाद (Example Based Machine Translation)

इस पद्धति में कॉर्पोरा (बृहद शब्दसंग्रह) के आधार पर मशीन के लिए कृत्रिम बुद्धि से समानांतर कॉर्पोरा तैयार किया जाता है और उसी के आधार पर नियम दिए जाते हैं। समानांतर कॉर्पोरा, मशीनी अनुवाद की कृत्रिम बुद्धि के सृजन के लिए प्रयोग किया जाता है। वास्तव में कृत्रिम बुद्धि के माध्यम से मानव मस्तिष्क की तरह अनुवाद करने के लिए इसका इस्तेमाल मशीनी अनुवाद तंत्रों में किया जाता है। यह कृत्रिम बुद्धि मशीन में दिए गए नियमों से परिचालित होती है। कॉर्पोरा मानव अनूदित कार्य है, जिससे कंप्यूटर में समानांतर कॉर्पोरा द्वारा दो भाषाओं के बीच विभिन्न वाक्यों का समावेश किया जाता है। मशीनी अनुवाद तंत्र में अनुवाद के लिए दिए गए वाक्य की पहचान कॉर्पोरा से उसी प्रकार के वाक्य से कराई जाती है। जिसमें स्रोत भाषा के वाक्य को लक्ष्य भाषा के समानांतर संरचना में अनुवाद किया जाता है। इस प्रक्रिया के लिए समानांतर कॉर्पोरा का समावेश कंप्यूटर में किया जाता है तथा उसे अनुवाद के लिए प्रोग्राम के माध्यम से संचालित किया जाता है।<sup>35</sup> उदाहरण आधारित मशीनी अनुवाद तंत्रों को (Example Based MT Systems) सीमित उदाहरण आधार पर विकास किया जा सकता है। इसके प्रशिक्षण (Training) और परीक्षण (Testing) में पाया गया कि यह भाषा की प्रकृति के अनुसार अच्छी तरह से अनुवाद कार्य करती है।

---

<sup>34</sup> वही, पृ. 257

<sup>35</sup> गोस्वामी, कृष्ण कुमार, *अनुवाद विज्ञान की भूमिका*, 2012, पृ.172

### 1.7.3 ज्ञान आधारित मशीनी अनुवाद (Knowledge Based Machine Translation)

ज्ञान आधारित प्रणाली में प्रथमतः स्रोत भाषा पाठ की आंतरिक संरचना को समझकर उसका लक्ष्य भाषा में अनुवाद किया जाता है। यह पद्धति अंतरभाषा प्रणाली पर आधारित है। ज्ञान आधारित मशीनी अनुवाद में शब्दकोश से वाक्य विन्यास और अर्थ विन्यास की जानकारी रूपात्मक सहज ज्ञान के आधार पर प्राप्त करती है। इसमें ज्ञान आधारित समझ विकसित करने के कारण इससे कई भाषाओं में अनुवाद किया जा सकता है। यह अन्य अंतरभाषा प्रणाली से भिन्न है क्योंकि यह स्रोत भाषा का विश्लेषण विश्व ज्ञान के आधार पर करती है। इस पद्धति में विश्व ज्ञान के साथ इसमें विभिन्न भाषाओं के शब्दों के अर्थ का समावेशन किया जाता है। इस प्रकार, भाषाओं के अर्थ का प्रतिनिधित्व करने के लिए एक विशिष्ट भाषा की आवश्यकता होती है। इस पद्धति में स्रोत भाषा का विश्लेषण करने के बाद उसमें संचित ज्ञान के आधार पर लक्ष्य भाषा के वाक्य में रूपांतरित किया जाता है। यह प्रणाली स्वचालित रूप से सटीक अनुवाद प्रदान करती हैं। विभिन्न भाषाओं में पाठ का सटीक अनुवाद करने के लिए इसमें भाषाई ज्ञान का बृहद मात्रा में संकलन तथा संचालन करना होता है। हालांकि, इसमें ज्ञान की बड़ी मात्रा में आवश्यकता के कारण यह उत्पादन करने में काफी महंगी होती हैं। अंग्रेजी से वियतनामी मशीनी अनुवाद तंत्र इस पद्धति पर आधारित है।<sup>36</sup>

### 1.7.4 सांख्यिकीय आधारित मशीनी अनुवाद (Statistical Based Machine Translation)

यह मशीन अनुवाद का प्रतिमान (Paradigm) है जिसमें अनुवाद के लिए द्विभाषिक पाठ कॉर्पोरा का विश्लेषण कर सांख्यिकीय मॉडल से निर्गत (output) अनुवाद होता है। इस मशीनी अनुवाद का दृष्टिकोण आम तौर पर शब्द आधारित होता है लेकिन नए विकसित तंत्रों में अनुवाद, पदबंध संरचना के आधार पर होता है। द्विभाषिक पाठ कॉर्पोरा का विश्लेषण (स्रोत से लक्ष्य भाषाओं में) और एकभाषिक कॉर्पोरा का प्रयोग सांख्यिकीय मॉडल के लिए किया जाता है, जिससे एक भाषा से अन्य भाषाओं में अनुवाद

---

<sup>36</sup> Goyal, Vishal, *Development of a Hindi to Punjabi Machine Translation System*, PhD Thesis, 2010, p.630-631

होता है। इस मशीनी अनुवाद में मानव और डेटा संसाधनों का उपयोग अधिक कार्यक्षम रूप में किया जाता है। जिसमें अधिक मात्रा में समानांतर कॉर्पोरा का मशीन पठनीय प्रारूप (Machine readable format) में और उससे भी अधिक एकभाषिक डेटा का समावेश अनुवाद के लिए किया जाता है।<sup>37</sup> जिससे सांख्यिकीय आंकड़ों के आधार पर मशीनी अनुवाद प्राप्त होता है। इस मशीनी अनुवाद की पद्धति में इन प्रकारों का समावेश किया गया है- शब्द आधारित अनुवाद (Word-based translation), पदबंध आधारित अनुवाद (Phrase-based Translation), वाक्य विन्यास आधारित अनुवाद (Syntax-based Translation) पदानुक्रमित वाक्यांश आधारित दृष्टिकोण (Hierarchical Phrase based approaches) आदि। विभिन्न दस्तावेजों का अनुवाद स्रोत भाषा से लक्ष्य भाषा में संभाव्यता के साथ गणितीय सूत्रों के आधार पर किया जाता है।

गूगल अनुवादक तंत्र भी इसी तरह अनुवाद के लिए बृहद कॉर्पोरा का प्रयोग करता है। इसमें सिमैन्टिक वेब प्रौद्योगिकी को विकसित करने के प्रयास चल रहे हैं जिससे प्राकृतिक भाषा का संसाधन करने में सहायता होगी और अनुवाद भी अर्थ की दृष्टि से पूर्ण रूप में प्राप्त होगा। गूगल अनुवादक में अनुवाद के लिए भाषाओं के नियम सिखाने (Machine learning) के बजाय उसे कॉर्पोरा के बृहद वाक्य के संकलन से अनुवाद किया जाता है। यह अनुवाद तंत्र पर छोड़ दिया जाता है कि वह अनुवाद की जानेवाली भाषा के अनुसार उसके नियम का निर्माण करें। इसके लिए वह मानव द्वारा अनुवाद किए हुए लाखों दस्तावेजों के कॉर्पोरा का समावेश इसमें किया जाता है। इस कॉर्पोरा के संकलन से ही इन तंत्रों द्वारा अनुवाद किया जाता है। इसके अनुवाद में कुछ गलतियाँ भी होती हैं जिसमें अर्थगत अनुवाद की समस्या प्रमुख होती है जो मशीनी अनुवाद के लिए चुनौती निर्माण करती है।<sup>38</sup> यदि कॉर्पोरा का संकलन अधिक मात्रा में किया जाए तो सांख्यिकीय और नियम आधारित पद्धति के प्रयोग से अधिक सटीक अनुवाद किया जा सकता है।

---

<sup>37</sup> [https://en.wikipedia.org/wiki/Statistical\\_machine\\_translation](https://en.wikipedia.org/wiki/Statistical_machine_translation)  
[Retrieved on 17.05.2017]

<sup>38</sup> Santosh Kumar T.S., *Word Sense Disambiguation Using Semantic Web for Tamil to English Statistical Machine Translation*. IRA-International Journal of Technology & Engineering (ISSN 2455-4480), 5(2), 22-31

सांख्यिकीय मशीन अनुवाद प्रणाली में वाक्य विन्यास आधारित पद्धति महत्वपूर्ण है। जिसमें एक शब्द या पदबंध संरचना के बजाय वाक्यात्मक इकाइयों पर आधारित अनुवाद किया जाता है। अभी इसमें संदर्भ मुक्त व्याकरण का प्रयोग अनुवाद की गुणवत्ता में सुधार करने के लिए किया जा रहा है लेकिन इसके अनुवाद करने की गति अन्य मशीनी अनुवाद तंत्रों से काफी कम है। इसमें भाषा युग्म (Language pairs) के वाक्यगत भिन्नता को क्रमानुसार (reordering) करने और व्याकरण को अनुवाद के अनुकूल करने के लिए इसमें प्रयास हो रहे हैं।<sup>39</sup>

### 1.7.5 मिश्रित मशीनी अनुवाद (Hybrid Machine Translation)

यह मशीनी अनुवाद की ऐसी पद्धति है जिसमें विविध मशीनी अनुवाद दृष्टिकोण का प्रयोग इसे विकसित करने के लिए किया जाता है। इसे विकसित करने की प्रेरणा अनुवाद में आनेवाली संदिग्धता और उसकी शुद्धता से है। इसमें अनुवाद के लिए नियम आधारित पद्धति (Rule based) तथा सांख्यिकीय आधारित पद्धति का मिश्र रूप में समावेश किया जाता है। सांख्यिकीय का प्रयोग निर्गत अनुवाद को सुधार करने के लिए किया जाता है। स्रोत से लक्ष्य भाषा में वाक्यों के अनुवाद में पश्च संपादन की प्रक्रिया सांख्यिकीय दृष्टिकोण से स्वचालित रूप से होती है। इस मशीनी अनुवाद में भाषा युग्म और विषय-क्षेत्र की जटिलता के लिए विशिष्ट दृष्टिकोण का प्रयोग अलग-अलग रूप में किया जाता है।<sup>40</sup> इन मशीनी अनुवाद प्रणालियों के अतिरिक्त भी प्रणालियाँ आज कार्य कर रही हैं, जिसके विकास एवं निर्माण के लिए अनुसंधान कार्य हो चल रहे हैं।

इसमें स्रोत से लक्ष्य भाषा वाक्यों के अनुवाद में पश्च संपादन की प्रक्रिया सांख्यिकीय दृष्टिकोण से स्वचालित रूप से होती है। इस मशीनी अनुवाद में भाषा युग्म और विषय-क्षेत्र की जटिलता के लिए विशिष्ट दृष्टिकोण का प्रयोग अलग-अलग रूप में किया जाता है। मिश्रित मशीनी अनुवाद पद्धति का प्रयोग मुख्य रूप से अनुवाद तंत्र से

---

<sup>39</sup> <http://www.asiaonline.net/EN/MachineTranslation/default.aspx?QID=17>  
[Retrieved on 08.06.2016]

<sup>40</sup> [https://en.wikipedia.org/wiki/Hybrid\\_machine\\_translation](https://en.wikipedia.org/wiki/Hybrid_machine_translation) [Retrieved on 08.06.2016]

अधिक सही अनुवाद करने के लिए किया जाता है जिसमें सभी तंत्रों में स्वतंत्र रूप में प्रयोग की गई पद्धतियों का समावेश इसमें किया जाता है। सिस्ट्रान भी यह मशीनी अनुवाद की कंपनी 1968 में स्थापित हुई है जो इस दृष्टिकोण का अभी प्रयोग कर रही है। इसके पहले यह नियम आधारित और सांख्यिकीय आधारित दृष्टिकोण का प्रयोग कर रही थी।<sup>41</sup>

उपर्युक्त मशीनी अनुवाद की विभिन्न पद्धतियों का प्रयोग विकासकर्ता मशीनी अनुवाद तंत्र के निर्माण के लिए करते हैं। जिसमें प्रमुखतः स्रोत और लक्ष्य भाषा के समानांतर भाषाई कॉर्पोरा, व्याकरणिक संरचना (Grammatical structure), द्विभाषिक शब्दावली (Bilingual dictionary), पूर्व संसाधक (Pre-processor), पद निरूपक (Parser), रूपवैज्ञानिक विश्लेषक (Morphological analyzer), प्रजनक (Generator) शब्द संचयन आदि का होना आवश्यक है। आज इन पद्धतियों का प्रयोग अनेक मशीनी अनुवाद तंत्र करते हैं। इन मशीनी अनुवाद तंत्र के विकास की दिशा में विश्व स्तर पर तथा भारत में चल रहे प्रयासों को नीचे विश्लेषित किया गया है।

## 1.8 मशीनी अनुवाद के तंत्र : एक सर्वेक्षण

आज विश्व में मशीनी अनुवाद पर काफी शोध कार्य हो रहे हैं। इनका महत्वपूर्ण उद्देश्य है कि एक भाषा के माध्यम से दूसरी भाषा में मानव के लिए सहज अनुवाद प्रदान करना। अगर यह अनुवाद सही होता है, तो इसका लाभ विभिन्न कार्यों के लिए कर सकते हैं फिर चाहे वह क्षेत्र, उद्योग-व्यापार के साथ सूचना प्रौद्योगिकी का हो या विज्ञान एवं तकनीकी का हो। इनके अतिरिक्त भी अनुवाद के क्षेत्र हो सकते हैं। वैश्विक स्तर पर कुछ मशीनी अनुवाद तंत्र कई भाषाओं के लिए अनेक विषय-क्षेत्रों में अनुवाद कर रहे हैं, जिनका नीचे विस्तार से विवेचन किया गया है। □ □ □

### 1.8.1 गूगल अनुवादक (Google Translate 2006)

---

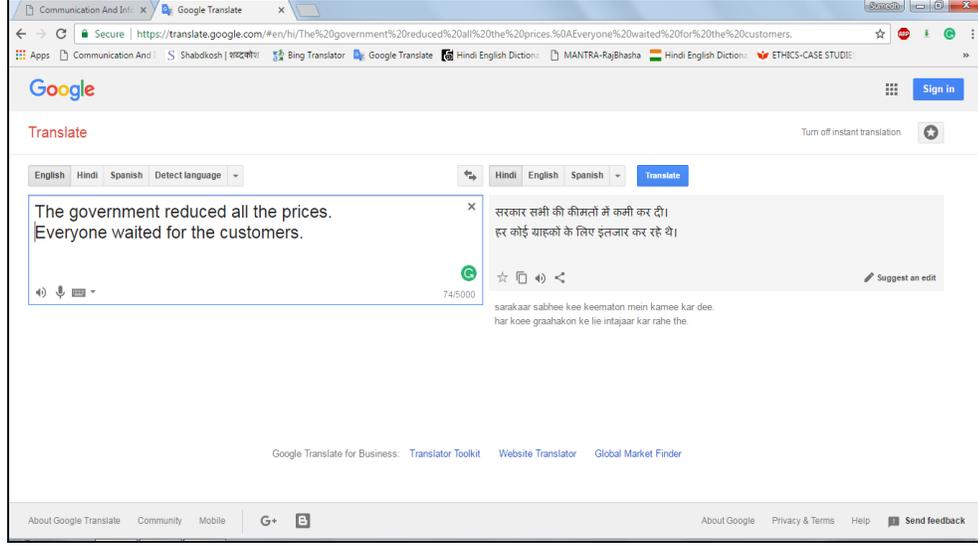
<sup>41</sup> Sachdeva, Kunal, *Hindi to English Machine Translation*, 2016, (M.Sc. Thesis) p.9

गूगल अनुवादक बहुभाषिक मशीनी अनुवाद तंत्र है जो विश्व की 103 भाषाओं में पाठ, स्पीच, तथा वेब साइट का अनुवाद करता है। उसमें भारत की प्रमुख 10 भाषाएँ (हिंदी, बांग्ला, गुजराती, कन्नड, तमिल, तेलुगू, मराठी, पंजाबी, सिंधी, उर्दू) और अंग्रेजी के साथ अनुवाद करता है। साथ ही 15 भाषाओं के लिए वाक् इनपुट प्रणाली (Speech input System) का समावेश भी इसमें किया गया है। यह तंत्र एक भाषा के पाठ (Text) का या वेबपेज (Webpage) का दूसरी भाषा में अनुवाद करता है। गूगल अनुवादक (Google Translate), सांख्यिकीय मशीनी अनुवाद (Statistical Machine Translation Based) पर आधारित अनुवाद साफ्टवेयर (Software) है। जिसे गूगल इन्कॉर्पोरेशन (Google Incorporation) ने सिस्ट्रान (SYSTRAN) आधारित अनुवाद पद्धति का प्रयोग कर विकसित किया है। सिस्ट्रान मशीनी अनुवाद की एक पुरानी कंपनी है, जिसे डॉ. पीटर टोमा (Peter Toma) ने 1968 में स्थापित किया था। सिस्ट्रान ने संयुक्त राष्ट्र अमेरिका के रक्षा विभाग और यूरोपीय आयोग के लिए व्यापक अनुवाद कार्य किए हैं। इसके साथ ही याहू बेबेल फिश (Yahoo Babel Fish), अमेरिकन ऑनलाइन (AOL) तथा अन्य अनुवाद तंत्रों की सेवाएँ भी आज सिस्ट्रान का प्रयोग कर रही हैं।

प्रारंभ में गूगल ने अंग्रेजी से अरबी में अनुवाद की शुरुआत की थी और आज विश्व की अनेक भाषाओं में अनुवाद कर रहा है। नवंबर, 2016 में गूगल ने घोषणा की है कि गूगल अनुवादक में अनुवाद की शुद्धता और सटीकता बढ़ाने के लिए 'Google Neural Machine Translation' में इसका प्रतिस्थापन किया जा रहा है। इस नई तकनीक से पूरे वाक्य का अनुवाद तत्काल रूप में प्राप्त करने के लिए व्यापक संदर्भ का प्रयोग किया जाता है। साथ ही इसमें भाषा के अनुरूप वाक्य संरचना और उचित व्याकरण का समावेश भी किया जाएगा। GNMT तंत्र उदाहरण आधारित मशीनी अनुवाद पद्धति पर आधारित है जिसमें पहले अंग्रेजी से चीनी, फ्रेंच, जर्मन, जापानी, कोरियाई, पुर्तगाली, स्पेनिश और तुर्की आदि आठ भाषाओं के लिए विकसित किया गया है। इस वर्ष में हिंदी, रसियन और वियतनामी भाषाओं के अनुवाद के लिए भी इस पद्धति का प्रयोग किया जाएगा।<sup>42</sup> गूगल अनुवादक की अनुवाद प्रक्रिया को निम्न स्क्रीनशॉट में दिखाया गया है।

---

<sup>42</sup> [https://en.wikipedia.org/wiki/Google\\_Translate](https://en.wikipedia.org/wiki/Google_Translate) [Retrieved on 14.02.2017]



### स्क्रीनशॉट 1.1 : गूगल अनुवादक तंत्र की अनुवाद प्रक्रिया

Source taken from <https://translate.google.com/> [Retrieved on 14.02.2017]

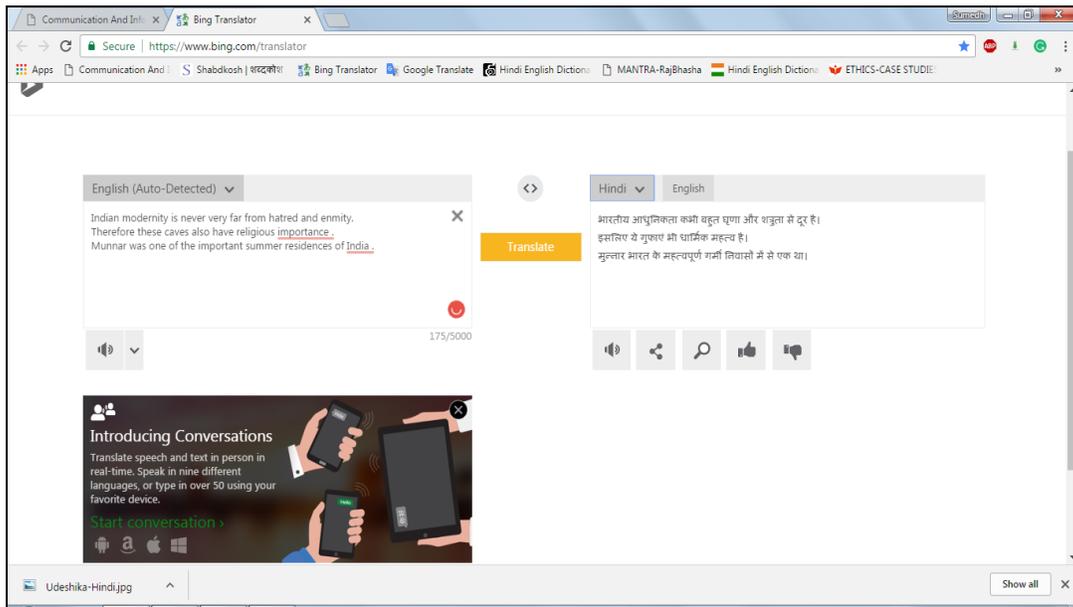
यह तंत्र प्रशासनिक दस्तावेज, मनोरंजन, कृषि, राजनीति, विज्ञान और तकनीकी, कला एवं संस्कृति, पर्यटन, स्वास्थ्य, खेल-कूद, अर्थव्यवस्था, बैंकिंग आदि विषय-क्षेत्रों का अनुवाद करता है। गूगल अनुवादक तंत्र बेहतर अनुवाद कार्य करता है, फिर भी इसके अनुवाद की कुछ सीमाएँ हैं। अगर अनुवाद के लिए सीमित पैराग्राफ की संख्या और तकनीकी शब्दावली (Technical Terms) दी जाए तो उसका अनुवाद गूगल द्वारा सही होता है, लेकिन पूर्णतः सही अनुवाद प्राप्त नहीं कर सकते। फिर भी गूगल के अनुवाद में कुछ गलतियाँ और विषमताएँ पायी जाती हैं क्योंकि गूगल अनुवादक तंत्र शब्दकोश या व्याकरणिक नियमों के बजाय सांख्यिकीय मॉडल पर आधारित है, लेकिन इसका अनुवाद सामान्य तौर पर अन्य अनुवाद तंत्रों से बेहतर होता है।

### 1.8.2 माइक्रोसॉफ्ट अनुवादक (Microsoft Translator 2009)

विभिन्न भाषाओं के अनुवाद के लिए स्वचालित अनुवाद (Automatic Translation) तंत्र का निर्माण 'माइक्रोसॉफ्ट रिसर्च' ने माइक्रोसॉफ्ट अनुवादक नाम से किया है। यह विश्व के प्रमुख मशीनी अनुवाद तंत्रों में से एक है। यह फरवरी 2017 से विश्व की 60 भाषाओं में अनुवाद प्रदान करता है। भारत की हिंदी और उर्दू भाषा का भी समावेश इस तंत्र में किया गया है। माइक्रोसॉफ्ट अनुवादक के द्वारा एक भाषा के पाठ

(Text) का या वेबपेज (webpage) का दूसरी भाषा में अनुवाद किया जाता है। साथ ही पाठ (Text) से वाक् (Speech) प्रणाली के अंतर्गत, निर्गत (Output) अनुवाद को वाक् (Speech) के रूप में भी सुन सकते हैं। अद्यतन रूप में यह 9 वाक् अनुवाद तंत्र (Speech Translation System) को सेवाएँ दे रहा है। मशीन से संबंधित पाठ का अनुवाद माइक्रोसॉफ्ट वाक्य-विन्यास आधारित सांख्यिकीय मशीनी अनुवाद (Syntax-Based Statistical Machine Translation) प्रणाली से करता है। इस तंत्र से प्रत्येक दिन 10 लाख पृष्ठों का अनुवाद किया जाता है। यह तंत्र परिवहन, कृषि, स्वास्थ्य और प्रौद्योगिकी तथा अन्य विषय क्षेत्रों का भी अनुवाद करता है।<sup>43</sup>

माइक्रोसॉफ्ट अनुवादक से अनुवाद करने में माइक्रोसॉफ्ट अनुवाद हब (Microsoft Translator Hub) की भूमिका महत्वपूर्ण है। यह हब इस तंत्र को बेहतर सेवाएँ प्रदान करता है और अन्य अनुवाद तंत्रों के निर्माण में सहायता भी करता है। इस मशीनी अनुवाद तंत्र को इस निम्न स्क्रीनशॉट में देखा जा सकता है।



स्क्रीनशॉट 1.2 : माइक्रोसॉफ्ट अनुवादक की अनुवाद प्रक्रिया

Source taken from <https://www.bing.com/translator> [Retrieved on 14.02.2017]

<sup>43</sup> [https://en.wikipedia.org/wiki/Microsoft\\_Translator](https://en.wikipedia.org/wiki/Microsoft_Translator) [Retrieved on 14.02.2017]

माइक्रोसॉफ्ट अनुवाद तंत्र से प्राप्त अनुवाद पूर्ण रूप से सही है, यह तो नहीं कह सकते लेकिन यह 70 प्रतिशत अनुवाद सही करता है तो कुछ अनअनुदित (Untranslated) शब्द को लिप्यंतरित करता है। इस अनुवाद तंत्र से प्राप्त अनुवाद को देखते हुए इसमें आनेवाले समय में बदलाव के लिए माइक्रोसॉफ्ट प्रयास कर रहा है।

### 1.8.3 याहू बेबेल फिश (Yahoo Babel Fish 2008)

याहू बेबेल फिश (Babel fish) यह अल्टा विस्टा (AltaVista) द्वारा विकसित याहू वेबसाइट पर एक वेब आधारित अनुप्रयोग (Application) है। यह तंत्र एक भाषा के पाठ (Text) का या वेबपेज (Webpage) का किसी दूसरी भाषा में अनुवाद करता है। यह अनुवाद प्रौद्योगिकी (Translation Technology) बेबेल फिश को सिस्ट्रान (SYSTRAN) द्वारा दी गयी है जिसने अन्य अनुवाद की सेवाओं के लिए यह प्रौद्योगिकी प्रदान की है। यह विश्व की प्रमुख 38 भाषाओं के साथ अंग्रेजी से हिंदी में अनुवाद करता है। इस तंत्र को ऑस्कर ए. जोफ़ (Oscar A Jofre) और उनकी टीम ने जनवरी, 1995 में संचालित किया था। जिसका प्रयोग डगलस एडम्स (Douglas Adams's) की श्रृंखला 'The Hitchhikers Guide to the Galaxy' के तत्कालिक भाषिक अनुवाद के लिए किया जाता था। जून 2013 से यह तंत्र माइक्रोसॉफ्ट बिंग अनुवादक के लिए सेवाएँ दे रहा है। यह तंत्र अनेक भाषाओं के विषय-क्षेत्रों के साथ युग्म अनुवाद (Pair Translation) करता है।<sup>44</sup>

### 1.8.4 Apertium मशीनी अनुवाद तंत्र

यह नियम आधारित मशीनी अनुवाद तंत्र है। शुरुआत में इस तंत्र का प्रयोग संबंधित भाषाओं के युग्म अनुवाद के लिए किया जाता था। बाद में इसका प्रयोग विभिन्न भाषाओं के अनुवाद के लिए किया जाने लगा। इसे उथल-अंतरण (Shallow-Transfer) मशीनी अनुवाद प्रणाली पर विकसित किया गया है। जिसमें शाब्दिक रूपांतरण, शब्द-भेद टैगिंग (Part of speech tagging) और शब्द श्रेणी अस्पष्टता (word category disambiguation) और सीमाबद्ध स्वचालन (Finite state transducers) का प्रयोग किया है।

---

<sup>44</sup> [https://en.wikipedia.org/wiki/Babel\\_Fish\\_\(website\)](https://en.wikipedia.org/wiki/Babel_Fish_(website)) [Retrieved on 14.02.2017]

इसमें अभिनव व्याकरण टैगर (Constraint Grammar Taggers) का प्रयोग कुछ भाषा युग्मों (Language Pairs) के अनुवाद के लिए किया जाता है। वर्तमान में अपर्तिअम 35 भाषाओं में युग्म अनुवाद (Pair translation) करता है। इसमें हिंदी और उर्दू के साथ अंग्रेजी भाषा का भी समावेश किया है। यह काफी हद तक सही अनुवाद करता है और इसे अधिक कार्यक्षम करने के लिए इसमें विकास किया जा रहा है।<sup>45</sup>

विश्व स्तर पर इन मशीनी अनुवाद तंत्रों के अलावा भी अनुवाद तंत्रों का विकास किया गया है लेकिन ये प्रमुख मशीनी अनुवाद तंत्र हैं, जिनका प्रयोग विश्व जगत में अधिक मात्रा में अनुवाद के लिए किया जाता है। इन तंत्रों के विकास को देखते हुए भारत में भी कई मशीनी अनुवाद तंत्रों तथा इसके लिए आवश्यक भाषाई संसाधनों का विकास किया जा रहा है।

#### 1.8.5 भारत के प्रमुख मशीनी अनुवाद तंत्र (Major Machine Translation Systems in India)

भारत एक बहुभाषिक देश होने के साथ इसमें एकाधिक भाषा परिवारों की भाषाएँ हैं। भारत की अन्य भाषाएँ भी शब्द संपदा तथा व्याकरणात्मक और संरचनात्मक दृष्टि से हिंदी और संस्कृत के अत्यधिक निकट हैं। इसलिए अन्य भाषाओं के लिए भाषाई संसाधन तथा उपकरणों का निर्माण करना आसान हो रहा है। अगर देखा जाए तो मशीनी अनुवाद तंत्रों का विकास किसी एक भाषा के सीमित क्षेत्र या विषय में सही और सटीक किया जा सकता है। लेकिन इसके लिए प्राकृतिक भाषा की सामग्री को विकसित करने की आवश्यकता है। इसमें भाषा के कॉर्पस (Corpus), द्विभाषिक कोश (Bilingual Dictionary), शब्द संसाधक (Word processor), विच्छेदक (Parser), रूपवैज्ञानिक विश्लेषक (Morphological analyzer), प्रजनक (Generator), वर्तनी जाँचक (Spell checker), वाक्य विश्लेषक (Syntactic analyzer) और वाक् प्रौद्योगिकी (Speech Technology) आदि उपकरणों (Tools) का निर्माण करने की आवश्यकता है।

---

<sup>45</sup><https://en.wikipedia.org/wiki/Apertium> [Retrieved on 14.02.2017]

भारत में इस विकास कार्य में भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (TDIL), आई.आई.टी., कानपुर; आई.आई.आई.टी. हैदराबाद; हैदराबाद विश्वविद्यालय; सी-डैक बंगलोर; सी-डैक, पुणे; राष्ट्रीय सॉफ्टवेयर प्रौद्योगिकी केंद्र, मुंबई; अन्ना विश्वविद्यालय, चेन्नई; जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली; आई.बी.एम. रिसर्च इंडिया आदि संस्थानों तथा विश्वविद्यालयों के कार्य महत्वपूर्ण है। भारत में विकसित मशीनी अनुवाद के तंत्र नीचे टेबल में दिए गए हैं।

मशीनी अनुवाद तंत्र (MT Systems)	भाषाएँ (Languages)	पद्धति (Approach)	क्षेत्र (Domain)
आंग्लभारती	अंग्रेजी से भारतीय भाषाएँ (हिंदी)	अन्तर्भाषिक/ Pseudo – interlingua	सार्वजनिक स्वास्थ्य
आंग्लभारती-II	अंग्रेजी से भारतीय भाषाएँ	उदाहरण आधारित	सामान्य
आंग्ल हिंदी	अंग्रेजी-हिंदी	अन्तर्भाषिक	सामान्य क्षेत्र
अनुसारक	तेलुगू, कन्नड़, बंगाली, पंजाबी, मराठी से हिंदी	प्रत्यक्ष विधि	सामान्य (बच्चों की कहानीयां)
अणुभारती	हिंदी-अंग्रेजी- भारतीय भाषाओं में	उदाहरण/हाइब्रिड आधारित	सामान्य
मंत्र-राजभाषा	अंग्रेजी से भारतीय भाषाएँ	अंतरण आधारित	राजभाषा एवं प्रशासनिक दस्तावेज
मात्रा	अंग्रेजी से हिंदी	अंतरण आधारित	समाचारों, वार्षिक प्रतिवेदनों और तकनीकी पदबंधों
शक्ति	अंग्रेजी से हिंदी, मराठी और तेलुगु	अंतरण आधारित	सामान्य
शिवा	अंग्रेजी से हिंदी	उदाहरण आधारित	सामान्य
अनुवादक	अंग्रेजी से भारतीय भाषाएँ (हिंदी)	--	अंग्रेजी के सरल वाक्यों का हिंदी

			अनुवाद
UCSG-मैट	अंग्रेजी से कन्नड़	अंतरण आधारित	सरकारी परिपत्रों के अनुवाद
UNL आधारित	अंग्रेजी, हिंदी, मराठी और भारतीय भाषाएँ	अन्तर्भाषिक	सामान्य
तमिल-हिंदी, अंग्रेजी-तमिल	तमिल से हिंदी, अंग्रेजी से तमिल	अंतरण आधारित	सामान्य
जाधवपूर-अनुवाद	अंग्रेजी से बंगाली	उदाहरण आधारित	समाचार के शीर्षकों का अनुवाद
अंग्रेजी-असमिया	अंग्रेजी से असमिया	--	अंग्रेजी से असमिया में सरल वाक्यों का अनुवाद
अंग्रेजी-उड़िया	अंग्रेजी से उड़िया	--	सामान्य
आई.बी.एम.	अंग्रेजी से भारतीय भाषाएँ	सांख्यिकीय आधारित	सामान्य
संपर्क	भारतीय भाषाएँ (पंजाबी, हिंदी, तमिल, तेलुगू, उर्दू, मलयालम)	नियम और सांख्यिकीय आधारित	पर्यटन और स्वास्थ्य
हिंग्लिश	हिंदी से अंग्रेजी	उदाहरण आधारित	सामान्य
पंजाबी-हिंदी	पंजाबी से हिंदी	प्रत्यक्ष विधि	सामान्य

टेबल 1.1 : भारत के प्रमुख मशीनी अनुवाद तंत्र

#### 1.8.5.1 आंग्लभारती (Anglabharati 1991)

आंग्लभारती योजना का प्रारंभ 1991 में प्रो.आर.एम.के. सिन्हा के निर्देशन में भारतीय प्रौद्योगिकी संस्थान, कानपुर (IITK) में हुआ था। इस तंत्र का निर्माण अंग्रेजी (English) से भारतीय भाषाओं में अनुवाद के लिए किया गया था। प्राथमिक रूप से

इसका प्रयोग अंग्रेजी से हिंदी अनुवाद के लिए किया गया। इस तंत्र में नियम आधारित प्रणाली (Rule Base System) का प्रयोग करते हुए इससे स्वास्थ्य सेवाओं तथा सामान्य विषयों का अनुवाद किया जाता है। अंग्रेजी भाषा की वाक्य संरचना कर्ता-क्रिया-कर्म (SVO) पर आधारित है, जबकि भारतीय भाषाओं की वाक्यसंरचना मुक्त शब्दक्रम (Free Word Order) अर्थात् कर्ता-कर्म-क्रिया (SOV) पर आधारित है। अंग्रेजी से अन्य भारतीय भाषा में अनुवाद के लिए आंग्ल-भारती ने Pseudo - interlingua (Pseudo Lingua for Indian Languages (PLIL)) प्रणाली का प्रयोग किया है। इस प्रणाली के अंतर्गत यह अंग्रेजी वाक्य का पहले विश्लेषण करती है और फिर भारतीय भाषाओं के लिए मध्यवर्ती संरचना (intermediate structure) की अस्पष्टता को दूर करते हुए लक्ष्य भाषा पाठ का प्रजनन (text generation) करती है।<sup>46</sup> इसके द्वारा 90 प्रतिशत अनुवाद प्राप्त किया जा सकता है और 10 प्रतिशत मानव के द्वारा पश्च संपादन (Post-editing) के द्वारा पूर्ण किया जा सकता है। आंग्लभारती तंत्र में स्रोत भाषा से लक्ष्य भाषा में अनुवाद के लिए सेन्स डिसेम्बिग्युएटर्स (sense disambiguators), लक्ष्य पाठ्य प्रजनक (Target text generator), बहुभाषिक शब्दकोश (multilingual dictionary), नियम आधारित अधिग्राहक (Rule-base acquirer) आदि घटकों का समावेश किया है।<sup>47</sup> इस परियोजना का दूसरा चरण आंग्लभारती द्वितीय है।

### 1.8.5.2 अनुसारक (Anusaaraka 1995)

प्रो. राजीव संगल के निर्देशन में आई.आई.टी. कानपुर ने 'अनुसारक' मशीनी अनुवाद तंत्र परियोजना की शुरुआत की। बाद में इस परियोजना को आई.आई.आई.टी.(IIT), हैदराबाद में स्थानांतरित किया गया है। इस तंत्र में पाणिनीय (Paninian) व्याकरण के सिद्धान्तों का प्रयोग कर भारतीय भाषाओं में अनुवाद को उन्नत

---

<sup>46</sup> Sinha, R. M. K., R. Jain, A.Jain, *Translation from English to Indian languages: ANGLABHARTI Approach*, in proceedings of Symposium on Translation Support System STRANS 2001, IIT Kanpur, India, 2001, p.167-172

<sup>47</sup> Ainura, Asamidinova; *Knowledge Base for Russian-English Machine Translation Divergences*, PhD. Thesis, 2007, p.56

बनाने का कार्य किया गया है।<sup>48</sup> यह स्रोत भारतीय भाषा के पाठ को लक्ष्य भारतीय भाषा के निकटतम और ग्राह्य रूप में अनूदित करता है। इसकी मूल अवधारणा भारतीय भाषाओं की व्याकरणिकता को ज्यादा महत्व न देते हुए वाक्य से अर्थ ग्रहण करने पर उसकी बोधगम्यता को आधार बनाती है और उसमें मानव की सहायता लेने के लिए भी सुझाव देती है।

एक तरह से यह तंत्र सूचना स्थानान्तरण का कार्य करता है। यह तंत्र तेलुगू, कन्नड़, बांग्ला, मराठी, और पंजाबी से हिंदी आदि भारतीय भाषाओं में अनुवाद करता है। इस तंत्र का प्रयोग सामान्य अनुवाद और बच्चों की कहानियों का अनुवाद करने में किया गया है। आई.आई.आई.टी. (IIIT), हैदराबाद और केंद्रीय विश्वविद्यालय, हैदराबाद ने तेलुगू से हिंदी तंत्र तथा संस्कृत अध्ययन विभाग, केंद्रीय विश्वविद्यालय, हैदराबाद ने अंग्रेजी-हिंदी तंत्र विकसित किया है। अनुसारक तंत्र में, रूपवैज्ञानिक विश्लेषक, स्थानीय शब्द समूहक (Local Word Group), बहु-भाषिक शब्दकोश (Bi-lingual Dictionaries), स्रोत भाषा से लक्ष्य भाषा के लिए सम्मिलक (Mapper from Source Language to Target Language), शाब्दिक संसाधक (Word Synthesizer), पश्च संपादन (Post-editing interface)<sup>49</sup> आदि घटक इस तंत्र में महत्वपूर्ण रूप में कार्य करते हैं।

### 1.8.5.3 अणुभारती (Anubharati 1995)

आई.आई.टी., कानपुर की दूसरी परियोजना 'अणुभारती' है। इस तंत्र का निर्माण 1995 में प्रो.आर.एम.के. सिन्हा के निर्देशन में हिंदी से अंग्रेजी एवं अन्य भारतीय भाषाओं में अनुवाद के लिए किया था। यह तंत्र उदाहरण आधारित प्रणाली (Example Base System) पर आधारित है, जिसमें स्रोत भाषा से लक्ष्य भाषा में उदाहरण आधारित कॉर्पोरा के माध्यम से अनुवाद किया जाता है। साथ ही ये तंत्र दोनों भाषा के वाक्यों का अनुवाद

---

<sup>48</sup> <http://anusaaraka.iiit.ac.in/> [Retrieved on 18.04.2016]

<sup>49</sup> झा, गिरीश नाथ, दि.मणि, दि. मिश्र, *भारत में भाषा प्रौद्योगिकी: एक सर्वेक्षण*, गवेषणा, 2008, अप्रैल-जून, अंक 90, पृ. 41

शब्द स्तर पर करता है।<sup>50</sup> उदाहरण आधारित दृष्टिकोण से हिंदी वाक्य के इनपुट (Input) देने के बाद उसे पदबंध स्तर पर विश्लेषित कर शब्दक्रम के अनुसार वाक्य की संरचना की जाती है। साथ ही हिंदी के मानकीकृत वाक्य को अनूदित वाक्य के साथ संयोजित किया जाता है। इसमें कुछ संयोजन नहीं हुआ तो इनपुट वाक्य को Shallow Chunker के द्वारा विच्छेद कर फिर से पदानुक्रम के अनुसार संयोजित किया जाता है।

#### 1.8.5.4 मंत्र-राजभाषा (Mantra-Rajbhasha 1999)

मंत्र-राजभाषा एक मशीन साधित अनुवाद सिस्टम है, जो राजभाषा के प्रशासनिक, वित्तीय, कृषि, लघु उद्योग, सूचना प्रौद्योगिकी, स्वास्थ्य रक्षा, शिक्षा एवं बैंकिंग क्षेत्रों के दस्तावेजों का अंग्रेजी से हिंदी में अनुवाद करता है। मंत्र टैक्नॉलाजी पर आधारित यह तंत्र सी-डैक, पुणे के एप्लाइड आर्टिफिशियल इंटेलीजेंस ग्रुप ने विकसित किया गया है। इसमें अंग्रेजी से हिंदी व्याकरण, पद-विच्छेदन तथा प्रजनन के लिए प्रो. अरविंद जोशी द्वारा विकसित कोशीयकृत वृक्ष संलग्न व्याकरण (Lexicalized Tree Adjoining Grammar) का प्रयोग किया गया है।<sup>51</sup> यह भारत सरकार के गृह मंत्रालय, राजभाषा विभाग के अधीन राजभाषा और प्रशासनिक दस्तावेजों का अनुवाद तथा कृषि, बैंकिंग, शिक्षा, वित्त, स्वास्थ्य सुरक्षा और सूचना प्रौद्योगिकी आदि विषय-क्षेत्रों का अंग्रेजी से हिंदी में अनुवाद करने के लिए किया जाता है। वर्तमान में अन्य भाषाएँ अंग्रेजी-बांग्ला, अंग्रेजी-तेलुगू, अंग्रेजी-गुजराती, हिंदी-बांगाली, हिंदी-मराठी और हिंदी-अंग्रेजी के विकास के लिए कार्य चल रहा है। यह अनुवाद तंत्र प्रशासनिक क्षेत्र में 90 प्रतिशत और इतर क्षेत्र में 60-65 प्रतिशत अनुवाद करने में सक्षम है।<sup>52</sup> यह अब ऑनलाइन के लिए इस लिंक <https://mantra-rajbhasha.rb-aai.in/> पर उपलब्ध है।

---

<sup>50</sup> Vishwabharat, A Ministry of Information Technology publication on Language Technology in India, Issues: 2001(May), p.6

<sup>51</sup> <https://mantra-rajbhasha.rb-aai.in/> [Retrieved on 15.02.2017]

<sup>52</sup> [http://www.prayaslt.blogspot.in/2009\\_08\\_01\\_archive.html](http://www.prayaslt.blogspot.in/2009_08_01_archive.html) [Retrieved on 15.02.2017]

#### 1.8.5.5 आंग्लभारती-II (Anglabharati -II, 2004)

आंग्ल-भारती को विकसित करने के लिए (IIT) आई.आई.टी. कानपुर में प्रो. सिन्हा के निर्देशन में इसका निर्माण किया गया और इसे भारतीय इलेक्ट्रॉनिक्स अनुसंधान एवं विकास केन्द्र, नोएडा (वर्तमान में सी-डैक नोएडा) ने अद्यतन किया है, जिसे आंग्ल-भारती द्वितीय कहते हैं। यह तंत्र सामान्यीकृत उदाहरण आधारित (Generalized Example Base) और कच्चे उदाहरण आधारित (Raw Example-Base) प्रणाली का मिश्रित रूप है। यह अंग्रेजी से भारतीय भाषाओं के लिए सामान्य विषय-क्षेत्रों के सामग्री का अनुवाद करता है। इसमें अनुवाद के लिए पूर्व-संपादित (Pre-editing), रूप प्रक्रियापरक विश्लेषण (Morphological Analysis), हिंदी पाठ प्रजनन (Text Generator) और पश्च संपादन (Post editing) जैसे घटकों का प्रयोग किया गया है।<sup>53</sup> यह तंत्र आंग्लभारती प्रथम की तुलना में अनुवाद करने के लिए अब अधिक सक्षम है।

#### 1.8.5.6 अनुभारती-II (Anubharti-II, 2004)

यह तंत्र अनुभारती प्रथम का सुधारित रूप है। जिसमें हिंदी भाषा से भारतीय भाषा में अनुवाद के लिए मशीन साधित अनुवाद प्रणाली का प्रयोग किया गया है। जिसकी संरचना उदाहरण आधारित और कार्पस आधारित दृष्टिकोण तथा प्राथमिक व्याकरण विश्लेषण के संयोजन से बनी है। इसमें पारंपरिक उदाहरण आधारित प्रणाली की तरह बृहद उदाहरणों की आवश्यकता को कम किया गया है। इसमें अनुवाद मुख्य रूप से वाक्यात्मक समूहों की पहचान के द्वारा अन्य भाषा में अनुवाद किया जाता है। यह तंत्र सामान्य विषय क्षेत्र का अनुवाद करता है।<sup>54</sup>

---

<sup>53</sup> Dwivedi, S.K., Sukhadeve, P.P., *Machine Translation System in Indian Prospectives*, Journal of Computer Science, 2010, 6 (10), p. 1113

<sup>54</sup> Sitender, Bawa Seema, *Survey of Indian Machine Translation Systems*, International Journal of Computer Science and Technology, 2012, 3(1), p.286

### 1.8.5.7 आंग्लहिंदी (Anghlahindi, 2003)

आंग्लहिंदी यह अंग्रेजी से हिंदी मशीन साधित मशीनी अनुवाद तंत्र है। जिसके निर्माण में आंग्लभारती के प्रतिरूपक (Modules) का प्रयोग किया गया है। यह संज्ञा पदबंध और क्रिया पदबंध के सही और सटीक अनुवाद के लिए इसमें नियम आधारित, उदाहरण आधारित और सांख्यिकीय विधियों का प्रयोग किया गया है। यह तंत्र 20 शब्दों तक के सरल, जटिल और मिश्रित वाक्य का लगभग 90 प्रतिशत सही अनुवाद कर सकता है। अनुवाद प्राप्ति के बाद इसमें पश्च संपादन की आवश्यकता होती है।<sup>55</sup> अगर कोई अज्ञात शब्द शब्दकोश में नहीं मिलता है और जहाँ वर्तनी कि गलतियाँ होती है तो वहाँ यह हिंदी (देवनागरी) में लिप्यंतरित करता है। इसके अलावा, इसमें अंग्रेजी ओसीआर (ocr) के माध्यम से किसी भी अंग्रेजी में मुद्रित पृष्ठ को अनुवाद के लिए सम्मिलित कर सकते हैं, जिसमें हिंदी में अनुवाद किया जाता है। इसमें पाठ (Text) से वाक (Speech) प्रणाली का भी समावेश किया गया है।

### 1.8.5.8 मात्रा (MaTra 2004)

मात्रा यह पूर्णतः स्वचालित अंग्रेजी से हिंदी मशीन अनुवाद तंत्र है। इसे राष्ट्रीय सॉफ्टवेयर प्रौद्योगिकी केंद्र, मुंबई (National Centre for Software Technology) ने विकसित किया है। वर्तमान में सी-डैक मुंबई में इसे स्थानांतरित किया गया है। मात्रा ने प्राकृतिक भाषा विश्लेषण की पारंपरिक समस्याओं पर अभिनव दृष्टिकोण को अपना कर नई फ्रेम संरचित प्रणाली का प्रयोग किया है।<sup>56</sup> यह तंत्र वाक्यों का अर्थपरक विश्लेषण करता है और इसके अंतर्गत अंतःक्रियात्मक संरचना शोधक (Interactive Structural Editor) उपलब्ध है, जिसके माध्यम से निवेशित अंग्रेजी वाक्यों का संरचनात्मक प्रतिरूपण (Structural Representation) प्राप्त हो जाता है। जहाँ अंग्रेजी के मिश्र वाक्य में

---

<sup>55</sup> Sinha, R.M.K., A. Jain, *Anghlahindi: An English to Hindi machine-aided translation system*. 2003, pp: 494-497

<sup>56</sup> Vishwabharat, A Ministry of Information Technology publication on Language Technology in India, Issues: 2003(Oct), p.106

कठिनाई आती है हिंदी में अनुवाद के लिए, उन वाक्यों का सामान्य वाक्य के रूप में विलग कर विश्लेषण किया जाता है। वर्तमान समय में मात्रा तंत्र का प्रयोग सामान्य क्षेत्र, समाचारों, वार्षिक प्रतिवेदनों (Annual reports) और तकनीकी पद (Technical Phrase) के अनुवाद के लिए किया जाता है।<sup>57</sup>

#### 1.8.5.9 शक्ति और शिवा (Shakti and Shiva 2004)

शक्ति मशीनी अनुवाद तंत्र का विकास अंग्रेजी से भारतीय भाषाओं में अनुवाद के लिए किया गया है। वर्तमान में यह तंत्र तीन भाषाओं अंग्रेजी से हिंदी, मराठी और तेलुगू का अनुवाद करता है। इस तंत्र के लिए जो संयुक्त नियम बनाए गए हैं, वे भाषाई प्रकृति (Linguistic Nature), उदाहरण आधारित और सांख्यिकीय दृष्टिकोण (Statistical approach) पर आधारित हैं। सांख्यिकीय दृष्टिकोण से परिणाम निकालने का या भाषाई सूचना प्राप्त करने का कार्य किया जाता है। तो कुछ मॉड्यूलस का प्रयोग अर्थ सूचना (Semantic Information) प्राप्त के लिए होता है। यह तंत्र एकाधिक दृष्टिकोणों से संयोजित है, जिसका उद्देश्य भाषाई विश्लेषण करना है।<sup>58</sup> इस तंत्र की बनावट (Architecture) अत्यंत सरल है। सभी मॉड्यूलस का आधार सामग्री के आधार पर परिचालन किया जाता है, जिसका प्रारूप 'शक्ति मानक' (Shakti Standard Format) है। दूसरी ओर शिवा तंत्र अंग्रेजी से हिंदी में अनुवाद करता है, जो उदाहरण आधारित मशीनी अनुवाद (EBMT) तंत्र है। इन दोनों तंत्रों का विकास अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान (IIIT) हैदराबाद, कार्नेगी मेलन विश्वविद्यालय, अमेरिका और भारतीय विज्ञान संस्थान (IIS) बंगलूर के संयुक्त तत्त्वावधान में हुआ है।<sup>59</sup> यह तंत्र ऑनलाइन रूप में इस <http://shakti.iiit.ac.in/> लिंक पर उपलब्ध है।

---

<sup>57</sup> Ananthkrishnan, R., M.Kavitha, et.al.; *MaTra: A Practical Approach to Fully-Automatic Indicative English-Hindi Machine Translation*, 2006

<sup>58</sup> Dwivedi, S.K., P.P.Sukhadeve, *Machine Translation System in Indian Prospectives*, Journal of Computer Science 6 (10), 2010, p. 1111-1116

<sup>59</sup> Naskar, S., S.Bandyopadhyay, *Use of Machine Translation in India: current Status*, 2002, p. 23-32

#### 1.8.5.10 अनुवादक (Anuvaadak)

नई दिल्ली की एक निजी संस्था सुपर इन्फोसाफ्ट ने श्रीमती अंजली राव चौधरी के निर्देशन में अनुवादक 5.0 तंत्र का विकास किया है। जिसका प्रयोग सामान्य प्रयोजनार्थ अंग्रेजी के सरल वाक्यों का हिंदी अनुवाद के लिए किया जाता है। इसमें कार्यालयी, प्रशासन, कृषि, भाषा विज्ञान, तकनीकी आदि विषय-क्षेत्रों के अंग्रेजी-हिंदी शब्दावली और पदबंधीय कोश का समावेश किया है। इससे अंग्रेजी दस्तावेजों का वाक्य-विन्यास कर हिंदी व्याकरण के नियमों के अनुसार हिंदी में अनुवाद किया जा सकता है। इस तंत्र में दिए गए शब्दकोश में नए शब्दों को जोड़ने तथा बदलने एवं शब्दों में संशोधन (पश्च संपादन) करने की सुविधा उपलब्ध है। साथ ही अंग्रेजी और हिंदी दोनों भाषाओं के लिए वर्तनी जाँचक (Spell Checker) की सुविधा भी इस तंत्र में है। इस तंत्र को विंडोज की किसी भी ऑपरेटिंग प्रणाली पर चलाया जा सकता है।<sup>60</sup>

#### 1.8.5.11 VAASAANUBAADA द्विभाषी बंगाली मशीनी अनुवाद तंत्र

द्विभाषिक बंगाली-आसामी स्वचालित मशीनी अनुवाद तंत्र का निर्माण समाचार पाठ के अनुवाद के लिए इस तंत्र को विकसित किया गया है। इसमें 2002 में उदाहरण आधारित मशीन अनुवाद (EBMT) तकनीक का प्रयोग किया गया है। इसमें अनुवाद वाक्य स्तर पर किया जाता है। जिसमें कुछ वाक्यों के अनुवाद के लिए पूर्व-संसाधन तथा पश्च संपादन किया जाता है। सटीक अनुवाद के लिए वाक्य अगर बृहद है तो विराम चिन्ह द्वारा उसे खंडित किया जाता है और सटीक अनुवाद प्राप्त किया जाता है।

#### 1.8.5.12 अंग्रेजी-कन्नड़ मशीनी अनुवाद तंत्र (2002)

यह मशीन साधित अनुवाद तंत्र (Machine aided Translation system) है। जिसे हैदराबाद विश्वविद्यालय के डॉ. के.नारायण मूर्ति ने अंग्रेजी से कन्नड़ में अनुवाद करने के

---

<sup>60</sup> नैटियाल, ज.प्र., *भाषा विज्ञान और सूचना प्रौद्योगिकी*, गवेषणा (अंक 88), 2007 (अक्टूबर-दिसंबर), पृ. 41

लिए विकसित किया है। इसे कर्नाटक सरकार द्वारा वित्तीय सहायता भी प्राप्त है। इसका प्रयोग इसमें अंतरण आधारित दृष्टिकोण का प्रयोग सरकारी परिपत्रों के अनुवाद के लिए किया जाता था। साथ ही इसमें सार्वभौमिक उपवाक्य संरचना व्याकरण (Universal Clause Structure Grammar) का प्रयोग हुआ जिससे अंग्रेजी के इनपुट वाक्य (Input Sentences) को UCSG पार्सर के द्वारा विश्लेषित किया जाता है और अनुवाद कन्नड में द्विभाषी शब्दकोश (Bilingual Dictionary) के द्वारा किया जाता है। इसमें कन्नड के लिए रूपवैज्ञानिक विश्लेषक (Morphological analyzer) का भी समावेश किया गया है। अंत में अनुवाद पदबंध और शब्द समूहों के द्वारा लक्ष्य भाषा में किया जाता है।<sup>61</sup>

#### 1.8.5.13 UNL -आधारित अंग्रेजी-हिंदी मशीन अनुवाद तंत्र (2003)

‘यूनिवर्सल नेटवर्किंग लैंग्वेज’ यह संयुक्त राष्ट्र विश्वविद्यालय, टोकियो की अंतरराष्ट्रीय परियोजना है, जिसका उद्देश्य अंतरराष्ट्रीय स्तर की सभी मानव भाषाओं को अंतरभाषा (Interlingua) के माध्यम से विकसित करना है। UNL के माध्यम से अंग्रेजी से हिंदी को संयुक्त राष्ट्र संघ की भाषाओं से जोड़ने के लिए आई.आई.टी. (IIT), मुंबई ने इस परियोजना में महत्वपूर्ण भूमिका निभाई है। प्रो.पुष्पक भट्टाचार्या अंग्रेजी से मराठी और बांग्ला भाषाओं में मशीनी अनुवाद तंत्र निर्माण के लिए प्रयास कर रहे हैं। इस तंत्र में हिंदी से यू.एन.एल. में रूपांतरण के लिए संपरिवर्तक (enconverter) और यू.एन.एल. से हिंदी में रूपांतरण के लिए विपरिवर्तक (deconverter) का प्रयोग किया गया है।<sup>62</sup> यह तंत्र सामान्य विषय-क्षेत्र का अनुवाद करता है। ये तंत्र ऑनलाइन <http://www.cfilt.iitb.ac.in/machine-translation/eng-hindi-mt/> इस लिंक पर है।

#### 1.8.5.14 ‘अनुवाद’ मशीनी अनुवाद तंत्र (2004)

जादवपुर विश्वविद्यालय ने समाचार पत्रों के शीर्षकों के अंग्रेजी से बंगाली में अनुवाद के लिए संकरित दृष्टिकोण से मशीन अनुवाद तंत्र का निर्माण किया है। इस तंत्र

---

<sup>61</sup> Goyal, Vishal; *Development of a Hindi to Punjabi Machine Translation System*, PhD Thesis, 2010, p.675

<sup>62</sup> Rao Durgesh, *Machine Translation in India: Brief survey*, 2001, p.1-5

में उदाहरण आधारित दृष्टिकोण के साथ इसमें अंतरण पद्धति (Transfer approach) का भी समावेश किया गया है। यह तंत्र समाचार के शीर्षकों का अनुवाद ज्ञान आधार (Knowledge bases) और उदाहरण संरचना के आधार पर करता है। उसी तरह एस.जी. कुमार ने 2005 में अंग्रेजी से बंगाली भाषा में मशीनी अनुवाद तंत्र को विकसित किया जो 98 प्रतिशत सही अनुवाद करता है। जिसमें बांग्ला उदाहरण आधारित प्रणाली का समावेश किया गया है। साथ ही इसमें नियम और अंतरण प्रणाली का भी समावेश किया है।<sup>63</sup> वर्तमान में इस तंत्र का नया संस्करण वाक्य स्तर का अनुवाद कर रहा है।

#### 1.8.5.15 हिंग्लिश (Hinglish 2004)

यह तंत्र मानक हिंदी से अंग्रेजी में अनुवाद करता है, इसे प्रो. सिन्हा और ठाकुर ने विकसित किया है। इसे विकसित करने के लिए अंग्रेजी से हिंदी आंग्लभारती (II) और हिंदी से अंग्रेजी अनुभारती (II) की प्रणाली का प्रयोग किया गया है। हिंग्लिश से प्राप्त किया जाने वाला अनुवाद 90 प्रतिशत स्वीकार्य होता है। यह तंत्र केवल बहुअर्थक क्रियापदों के अर्थ को हल करने में केवल पूर्ण रूप से सक्षम नहीं है।<sup>64</sup>

#### 1.8.5.16 आई.बी.एम (IBM) अंग्रेजी-हिंदी मशीनी अनुवाद तंत्र (2006)

आई.बी.एम (International Business Machines) रिसर्च इंडिया, नई दिल्ली ने पहले उदाहरण आधारित दृष्टिकोण का प्रयोग किया। बाद में इस मशीनी अनुवाद तंत्र में सांख्यिकीय दृष्टिकोण (Statistical) को अपनाया। इसे अंग्रेजी से अन्य भारतीय भाषाओं में मशीनी अनुवाद के विकास के लिए आई.बी.एम कंपनी महत्वपूर्ण भूमिका निभा रही है।<sup>65</sup>

---

<sup>63</sup> Asamidinova, Ainura; *Knowledge Base for Russian-English Machine Translation Divergences*, PhD.Thesis, 2007, p.66

<sup>64</sup> Dwivedi, S.K., P.P.Sukhadeve, *Machine Translation System in Indian Perspectives*, Journal of Computer Science 6 (10), 2010, p. 1111-1116

<sup>65</sup> Sitender, Seema Bawa, *Survey of Indian Machine Translation Systems*, JCST, 2012, p.288

### 1.8.5.17 तमिल-हिंदी अनुसारक एवं अंग्रेजी-तमिल मशीनी अनुवाद तंत्र

इस तंत्र का विकास के.बी. चन्द्रशेखर शोध केन्द्र, अन्ना विश्वविद्यालय, चेन्नई के प्रो. सी.एन. कृष्णन के निर्देशन में हुआ है। यह तंत्र अनुसारक के आधार पर कार्य करता है, जिसमें तमिल से हिंदी में अनुवाद किया जाता है। इस तंत्र में रूपवैज्ञानिक विश्लेषक (Morphological Analyzer), सम्मिलक इकाई (Mapping unit), प्रजनक (Generator)<sup>66</sup> आदि घटकों का समावेश है। यह तंत्र शाब्दिक स्तर का अनुवाद करता है, लेकिन यह अर्थीय विश्लेषण (Semantic analysis) पर आधारित नहीं है। इस तंत्र के लिए तमिल रूपवैज्ञानिक विश्लेषक और तमिल-हिंदी द्विभाषी शब्दकोश का निर्माण किया गया है। इसी शोध केंद्र ने अंग्रेजी-तमिल मशीन साधित अनुवाद तंत्र को विकसित किया है। जिसमें सीमित शब्दावली (100-150) के अंतरण नियमों (Transfer rules) का लघु समुच्चय (Small Sets) है और संपूर्ण वाक्यात्मक विश्लेषण (Syntactical analysis) का समावेश भी किया गया है।<sup>67</sup>

### 1.8.5.18 असमिया मशीनी अनुवाद तंत्र

आई.आई.टी. गुवाहटी ने अंग्रेजी से असमिया मशीनी अनुवाद तंत्र को विकसित किया है। यह तंत्र मूलतः नियम आधारित है जो अंग्रेजी से असमिया द्विभाषिक शब्दकोश के आधार पर अनुवाद करता है। इस तंत्र के शब्दकोश में 5000 मूल शब्दों (Root word) का समावेश किया गया है। यह अंग्रेजी से अनुवाद करते समय इनपुट वाक्य का (रूपवैज्ञानिक, वाक्यगत और अर्थ के स्तर पर) विश्लेषण करता है। फिर अनुवाद की प्रक्रिया स्रोत भाषा से लक्ष्य भाषा में शब्द प्रति शब्द और पद से पद

---

<sup>66</sup>[http://www.au-kbc.org/research\\_areas/nlp/projects/anusaaraka.html](http://www.au-kbc.org/research_areas/nlp/projects/anusaaraka.html) [Retrieved on 15.02.2017]

<sup>67</sup> Dwivedi, S.K., P.P.Sukhadeve, *Machine Translation System in Indian Prospectives*, Journal of Computer Science 6 (10), 2010, p. 1111-1116

द्विभाषी शब्दकोश के माध्यम से होती है।<sup>68</sup> यह तंत्र वर्तमान में अंग्रेजी से असमिया में सरल वाक्यों का अनुवाद करता है।

#### 1.8.5.19 ओडिसी मशीनी अनुवाद तंत्र

इस तंत्र को उत्कल विश्वविद्यालय की डॉ. संघमित्रा मोहंती ने विकसित किया है। इसमें स्रोत भाषा अंग्रेजी है और लक्ष्य भाषा ओड़िया है। अनुवाद के लिए विभिन्न उपकरण (Tools) ओड़िया से अंग्रेजी ई-शब्दकोश (e-Dictionary Oriya to English), ओड़िया वर्डनेट (Oriya WorldNet), ओड़िया वर्तनी जाँचक (Oriya Spell checker), ओड़िया रूपवैज्ञानिक विश्लेषक (Morphological Analyzer), ओड़िया शब्द-संसाधक (Oriya Word Processor) आदि घटकों का विकास किया जा रहा है।<sup>69</sup>

#### 1.8.5.20 पंजाबी से हिंदी मशीनी अनुवाद तंत्र (2008, 2009)

इस तंत्र को वर्ष 2008 में जोसन और लेहल (Josan and Lehal) के निर्देशन में पंजाब विश्वविद्यालय, पटियाला में विकसित किया गया। यह 92.8 प्रतिशत शुद्धता के साथ शब्द-प्रति-शब्द अनुवाद पर आधारित तंत्र में विभिन्न स्रोतों का प्रयोग किया गया है। जैसे- धातु कोश, रूपप्रक्रियात्मक कोश, संदिग्धार्थक शब्दकोश आदि। अनुवाद मॉड्यूल्स वाक्य की विभिन्न वाक्यगत संरचनाओं में नामीय पद अभिज्ञान, शब्द आशय विसंदिग्धक (Word Sense Disambiguator), पुनरावृत्ति, लिप्यंतरण आदि का प्रयोग करते हैं। इसी संस्था में वर्ष 2009 में विशाल गोयल और प्रो. लेहल (Goyal and Lehal) के निर्देशन में हिंदी से पंजाबी मशीनी अनुवाद तंत्र का विकास किया गया। यह भी 95 प्रतिशत शुद्धता के साथ शब्द प्रति शब्द अनुवाद अभिगम पर आधारित है। इस तंत्र के लिए हिंदी-पंजाबी

---

<sup>68</sup> Vishwabharat, A Ministry of Information Technology publication on Language Technology in India, Issues, 2003 (October), p.108

<sup>69</sup> Vishwabharat, A Ministry of Information Technology publication on Language Technology in India, Issues, 2003 (October), p.71

कोश, रूपप्रक्रियात्मक विश्लेषक, शब्द आशय विसंदिग्धक (Word Sense Disambiguator) जैसे मॉड्यूल बनाए गए हैं।<sup>70</sup>

#### 1.8.5.21 संपर्क (Sampark 2009)

इस तंत्र का विकास सभी भारतीय भाषाओं में अनुवाद के लिए किया जा रहा है। इस परियोजना के विकास कार्य में 11 विश्वविद्यालयों, संस्थानों का समावेश है। भारतीय भाषा से भारतीय भाषाओं में मशीनी अनुवाद तंत्र के निर्माण में 9 भाषाओं का 18 भाषाओं के युग्म अनुवाद के लिए प्रयोग किया जाता है। वर्तमान में इस परियोजना के अंतर्गत पंजाबी से हिंदी, हिंदी से पंजाबी, तेलुगू से तमिल, उर्दू से हिंदी, हिंदी से उर्दू, तमिल से हिंदी, तेलुगू से हिंदी, मराठी से हिंदी अनुवाद तंत्र का विकास किया है।<sup>71</sup> सबसे पहले, संपर्क भाषा के विश्लेषण के लिए कम्प्यूटेशनल पाणिनी व्याकरण के दृष्टिकोण का प्रयोग किया गया है और मशीन लर्निंग को भी इसमें जोड़ा गया है। इसमें पारंपरिक नियम, शब्दकोश आधारित एल्गोरिदम के साथ सांख्यिकी का प्रयोग मशीन को सीखने की प्रक्रिया के लिए किया गया है।

#### 1.8.5.22 संस्कृत सुबंत विश्लेषक (Sanskrit Karka Analyzer)

जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली के संस्कृत अध्ययन केन्द्र के एम. फिल. शोधार्थी सुभाष चंद्रा द्वारा यह ऑनलाइन उपकरण तैयार किया गया है। वह इस वेबसाइट <http://sanskrit.jnu.ac.in/subanta/rsubanta.jsp> पर उपलब्ध है।

---

<sup>70</sup> निधि, करुणा; *मशीनी अनुवाद : भूत, वर्तमान एवं भविष्य*, भाषा प्रौद्योगिकी (E-journal) ISSN : 2231-4989

<http://hinditech.in/journal/E0%A4%BF%E0%A4%A7%E0%A4%BF.csp?cid=27>  
[Retrieved on 15.02.2017]

<sup>71</sup><http://sampark.iiit.ac.in/sampark/web/index.php/registration/validuser>  
[Retrieved on 15.02.2017]

### 1.8.5.23 लीला तंत्र (Leela software)

सी-डैक पुणे के AAI (एप्लाइड आर्टिफिसियल इंटेलिजेन्स) ग्रुप द्वारा लीला तंत्र विकसित किया गया है। Leela (Learning Indian Languages Through Artificial Intelligence) का अर्थ है कृत्रिम बुद्धि की सहायता से भारतीय भाषाओं का अधिगम कार्यक्रम। भारत सरकार के गृह मंत्रालय के अंतर्गत कार्यरत राजभाषा विभाग द्वारा प्रस्तुत इस तंत्र ने हिंदी पाठ्यक्रम को तीन स्तर 'प्रबोध', 'प्रवीण' और 'प्राज्ञ' पर संयोजित किया है। यह तंत्र विभिन्न सरकारी कर्मचारियों को कक्षा शिक्षण तथा दूरस्थ शिक्षण प्रणाली द्वारा हिंदी शिक्षण प्रदान करता है। यह कार्यक्रम विशेषतः उन कर्मचारियों के लिए उपयोगी है जो नियमित रूप से कक्षा में उपस्थित रहकर हिंदी सीख पाने में असमर्थ हैं।<sup>72</sup> आज यह सॉफ्टवेयर कार्यक्रम सभी कर्मचारियों को ऑनलाइन रूप में शिक्षा दे रहा है तथा इस विषय से संबंधित पाठ्यक्रम भी विभिन्न संस्थानों में पढाए जा रहे हैं।

### 1.8.5.24 अन्य मशीनी अनुवाद तंत्र (Other Machine Translation System)

भारत में मशीनी अनुवाद पर किए जा रहे प्रयासों से लगता है कि यह कार्य कठिन होने के बावजूद भी बहुत से क्षेत्रों में सफल हो रहा है। वास्तव में मशीनी अनुवाद का उद्देश्य शीघ्र गति से अनुवाद प्राप्त कर समय और श्रम की बचत करना है और उसका प्रयोग सामान्य प्रयोक्ता के लिए भी साध्य बनाना है। लेकिन भाषा वैज्ञानिकों के लिए आज भी यह चुनौती है कि वह मानव सहायता रहित मशीनी अनुवाद के तंत्रों को विकसित करें। इसलिए मशीनी अनुवाद तंत्रों में प्राकृतिक भाषा (NL) को विकसित कर कृत्रिम बुद्धि (Artificial Intelligence) का समावेश किया गया है। इस कृत्रिम बुद्धि के निर्माण द्वारा मशीन को मानव बुद्धि की तरह ही सृजनशील बनाने का प्रयास किया जा रहा है, जिससे स्वचालित मशीनी अनुवाद का उद्देश्य सफल हो सके। हमें मानना पड़ेगा कि कंप्यूटर से अनुवाद 100 प्रतिशत सही नहीं हो रहा है लेकिन 70 से 80 प्रतिशत का

---

<sup>72</sup> डबीरू, शिरीषा, *हिंदी एवं दक्षिण भारतीय भाषाओं में ई-लर्निंग का कार्यान्वयन*, गवेषणा पत्रिका 2007, पृ. 92

अनुवाद सही मात्रा में पाया जा सकता है। लेकिन आज भी इसके लिए मानव की सहायता आवश्यक है।

भारतीय भाषाओं में मशीनी अनुवाद की दिशा में अन्य भी तंत्र कार्य कर रहे हैं और कुछ तंत्रों के लिए प्रयास चल रहा है। बंगाली से आसामी में समाचार अनुवाद के लिए मशीनी अनुवाद तंत्र एवं तेलुगू से तमिल मशीनी अनुवाद तंत्र को हैदराबाद विश्वविद्यालय ने विकसित किया है। बंगाली से हिंदी में मशीनी अनुवाद तंत्र आई.आई.टी. खड़गपुर ने विकसित किया है। अंग्रेजी से मणिपुरी मशीनी अनुवाद तंत्र के लिए शिवाजी बंद्योपाध्याय के निर्देशन में जादवपुर विश्वविद्यालय, कोलकाता में कार्य चल रहा है। इसी तरह और भी मशीनी अनुवाद के तंत्रों का विकास विश्वविद्यालयों, संस्थानों, निजी कंपनियों के द्वारा अधिक तेज़ गति से हो रहा है। जिससे मानव को अनुवाद में सहायता मिल रही है।

निष्कर्ष में यह कहा जा सकता है कि विश्व के साथ भारत में मशीनी अनुवाद के अनेक तंत्रों का निर्माण किया गया है, लेकिन अनुवाद में गुणवत्ता का अभाव रहा है। वास्तव में मशीनी अनुवाद सीमित क्षेत्र के विषयों में अनुवाद प्रदान करने के लिए सक्षम है। जैसे की सी-डैक पुणे ने मंत्र-राजभाषा का निर्माण किया है। वह कार्यालयीन और प्रशासनिक दस्तावेजों का अनुवाद सक्षम रूप में करता है। लेकिन जब साहित्य के अनुवाद की बात आती है तो अनुवाद में गलतियाँ होती हैं और प्राप्त अनुवाद में संदिग्धता भी उत्पन्न होती हैं। इसके लिए भाषावैज्ञानिक को मशीन अनुवाद के क्षेत्र में अधिक कार्य करने होंगे ताकि मशीन द्वारा सही अनुवाद हो। वर्तमान में मशीनी अनुवाद तंत्र विकसित करने के लिए बृहत् स्तर पर प्रयास चल रहा है। इसके लिए भारत सरकार की तरफ से भी विविध परियोजनाएँ चलायी जा रही हैं। इससे यह संभावना दर्शायी जा सकती है कि भविष्य में प्रत्येक भाषा में मशीनी अनुवाद तंत्र विकसित होंगे, जो अनुवाद में कार्यक्षमता के साथ गुणवत्ता भी प्रदान करेंगे।

# द्वितीय अध्याय

## अंग्रेजी-हिंदी भाषा की संरचना और मशीनी अनुवाद की समस्याएँ

2.1 अंग्रेजी-हिंदी भाषा की संरचना

2.2 अंग्रेजी और हिंदी भाषाओं का तुलनात्मक व्याकरण

2.3 अंग्रेजी और हिंदी की वाक्य संरचना

2.4 अंग्रेजी-हिंदी मशीनी अनुवाद की समस्याएँ

## द्वितीय अध्याय अंग्रेजी-हिंदी भाषा की संरचना और मशीनी अनुवाद की समस्याएँ

### (LANGUAGE STRUCTURE OF ENGLISH-HINDI AND PROBLEMS OF MACHINE TRANSLATION)

---

भाषा संप्रेषण का एक माध्यम है, जो मानव समूह में परस्पर संवाद का संप्रेषण कराती है। संप्रेषण का संबंध मानव और मानवेत्तर प्राणियों में भी पाया जाता है, लेकिन मानव अपनी अभिव्यक्ति को मुखोच्चरित ध्वनियों द्वारा व्यक्त करता है। इसलिए कहा जाता है कि भाषा और समाज का संबंध एक दूसरे के साथ अभिन्न है। प्रत्येक भाषा की संरचना अलग होती है और उस भाषा की संरचना में ध्वनि, शब्द, व्याकरण और अर्थ का स्थान महत्वपूर्ण होता है।

ध्वनि भाषा का भौतिक एवं मूर्त पक्ष है जिसे हम ज्ञानेंद्रियों के द्वारा ग्रहण करते हैं। व्याकरण भाषा की संरचना का व्यवस्थापरक पक्ष है और अर्थ भाषा का मानसिक पक्ष है जो मूलतः वक्ता या श्रोता के मस्तिष्क में निहित होता है। मूल रूप से ध्वनियाँ स्वयं में अर्थहीन होती हैं, लेकिन व्याकरण व्यवस्था के माध्यम से शब्द तथा वाक्यों का आकार प्राप्त कर, वे अर्थयुक्त हो जाती हैं। इसलिए कहा जाता है कि ध्वनि को अर्थ तक ले जाने वाली कड़ी व्याकरण है। इन घटकों के साथ भाषा में वाक्य का भी स्थान महत्वपूर्ण होता है, क्योंकि वाक्य ही भाषा की संरचना को सार्थक बनाने का कार्य करता है।<sup>1</sup> वाक्य एक या एक से अधिक शब्दों से बनता है, जो संदर्भ विशेष में व्याकरणिक दृष्टि से पूर्ण होता है और उसमें अर्थ संरचना भी निहित होती है। इसलिए भाषा की संपूर्ण संरचना में वाक्य की भूमिका प्रमुख होती है।

---

<sup>1</sup> सिंह, सुरजभान, *हिंदी का वाक्यात्मक व्याकरण*, 1985, पृ. 20

प्रस्तुत अध्याय में, अंग्रेजी-हिंदी भाषा की संरचना को स्पष्ट किया गया है। दोनों भाषाओं में रूप, शब्द, संज्ञापद, विशेषण, क्रिया-विशेषण, कारक, वाक्य, तथा व्याकरण के स्तर पर काफी भिन्नताएँ पाई जाती हैं, जिसे अनुवाद के लिए समझना आवश्यक है। इसलिए अनुवाद के स्वरूप दोनों भाषाओं की संरचना में मशीनी अनुवाद में आनेवाली समस्याओं का विवेचन किया गया है।

## 2.1 अंग्रेजी-हिंदी भाषा की संरचना

संसार में लगभग तीन हजार भाषाएँ बोली जाती हैं। इनमें बहुत-सी भाषाओं का पारिवारिक रूप से आपस में संबंध है, अर्थात् वे मूलतः किसी एक भाषा से ही निकली हैं। ध्वनि, व्याकरण तथा शब्द-समूह का तुलनात्मक अध्ययन करके तथा भौगोलिक निकटता का विचार करके विद्वानों ने भाषाई पारिवारिक संबंधों का पता लगाया है।

भारत भी बहुभाषिक देश है जहाँ 22 संवैधानिक भाषाएँ हैं, जिनमें दस भाषाओं की अपनी-अपनी लिपि है। भारतीय भाषाओं को उनके मूल और समानता के अनुसार पांच समूहों में वर्गीकृत किया गया है- इसमें इंडो-आर्यन परिवार (हिंदी, बांग्ला, आसामी, पंजाबी, मराठी, उड़िया और गुजराती) तथा द्रविड़ परिवार (तमिल, तेलुगु, कन्नड़ और मलयालम), ऑस्ट्रो-एशियाई परिवार, तिब्बती-बर्मोज़ परिवार और अंदमानी परिवार आदि का समावेश किया गया है। इनमें से अधिकांश भाषाएँ संरचनात्मक रूप से समान हैं। भारतीय भाषाएँ शब्द संरचना की दृष्टि से अधिक विभक्तिपरक है। अपेक्षाकृत मुक्त शब्द क्रम पर आधारित है। भारतीय भाषाओं की वाक्य संरचना देश के भीतर की सभी भाषाओं के समान है। अंग्रेजी भाषा की वाक्य संरचना कर्ता-क्रिया-कर्म (SVO) पर और हिंदी भाषा की संरचना कर्ता+कर्म+क्रिया (SOV) पर आधारित है।<sup>2</sup> इन दोनों भाषाओं में संरचनात्मक स्तर पर भिन्नता है। यदि दो भाषाओं में भाषाई संरचनात्मक स्तर पर

---

<sup>2</sup> Goyal, Vishal, *Development of a Hindi to Punjabi Machine Translation System*, PhD Thesis, 2010, p. 687

समानता है तो यह माना जाता है कि मशीन अनुवाद तंत्रों को सहज रूप से विकसित किया जा सकता है और उस दृष्टिकोण का प्रयोग अन्य भाषाओं के विकास के लिए किया जा सकता है।

### 2.1.1 अंग्रेजी और हिंदी भाषाओं का विकास

अंग्रेजी भाषा का मूल स्रोत पूर्व ऐतिहासिक काल का आदिभारोपीय (Proto Indo-European) भाषा परिवार माना जाता है। इस परिवार में जर्मनिक, पश्चिम जर्मनिक, एंग्लो फ्रिसियाई, एंग्लिक और अंग्रेजी का समावेश होता है। अंग्रेजी भाषा का समावेश जर्मन भाषा के पश्चिमी जर्मन समूह के अंतर्गत किया जाता है। कहा जाता है कि अंग्रेजी भाषा के बोलने वाले 5000 वर्ष पूर्व बाल्टिक सागर के समीपवर्ती क्षेत्र में रहते थे। कालांतर में जर्मनिक जनजातियों में से एक एंगल्स जनजाति केंद्रीय ब्रिटेन में जाकर बस गई। जो बाद में एंगल लैंड कहलाया और इसका रूपांतरण बाद में इंग्लैंड में हो गया। बाद में एंगल्स, जूट्स और सेक्सन जनजातियों की भाषा अंग्रेजी कहलाई गयी। प्रारंभिक मध्ययुगीन काल में अन्य जर्मनिक भाषाओं का अंग्रेजी की शब्दावली पर काफी प्रभाव पड़ा और बाद में रोमन भाषाओं में से विशेषतः फ्रेंच भाषा का इस पर काफी प्रभाव पड़ा।<sup>3</sup>

आज अंग्रेजी भाषा का प्रयोग लगभग 60 देशों में आधिकारिक भाषा तथा राजकाज की भाषा के रूप में किया जाता है। यह यूनाइटेड किंगडम, संयुक्त राज्य अमेरिका, कनाडा, ऑस्ट्रेलिया, आयरलैंड और न्यूजीलैंड में सबसे अधिक बोली जाने वाली भाषा है। कैरेबियन, अफ्रीका और दक्षिण एशिया के कुछ क्षेत्रों में भी व्यापक रूप में बोली जाती है। 2016 के अनुसार, 400 मिलियन लोग मातृभाषा के रूप में अंग्रेजी बोलते हैं और करीब 1.1 अरब लोग द्वितीय भाषा के रूप में प्रयोग करते हैं। यह मंदारिन और स्पेनिश के बाद दुनिया में तीसरी सबसे बोली जानेवाली भाषा है। इसे संयुक्त राष्ट्र, यूरोपीय संघ, तथा

---

<sup>3</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 44

वैश्विक और क्षेत्रीय स्तर के अंतर्राष्ट्रीय संगठन ने अधिकारिक भाषा के रूप में अपनाया है। 17 वीं सदी से लेकर 20 वीं सदी तक ब्रिटिश साम्राज्य के वैश्विक प्रभाव के चलते आधुनिक अंग्रेजी का दुनिया भर के सभी राष्ट्रों में प्रचार-प्रसार हुआ। आज मुद्रण और इलेक्ट्रॉनिक माध्यमों के चलते तथा अमेरिका के वैश्विक महाशक्ति के रूप में उभरने के बाद अंग्रेजी अंतर्राष्ट्रीय संवाद की प्रमुख भाषा बन गई है। इसलिए आधुनिक अंग्रेजी को वैश्विक रूप में सामान्य भाषा या लोगभाषा (lingua franca) कहते हैं।<sup>4</sup>

अंग्रेजी यह आज आम भाषा है जो भारत के बहुभाषी समाज को एक रूप में बांधने का काम करती है और यह सभी सरकारी, व्यापार और शैक्षणिक कार्याकलाप में प्रयोग की जाने वाली भाषाओं में से एक है। लगभग भारत के सभी विश्वविद्यालय तथा संस्थानों में अंग्रेजी में शिक्षा दी जाती है और इसे लगभग सभी प्राथमिक और माध्यमिक विद्यालयों में दूसरी या तीसरी भाषा के रूप में भी इसे पढ़ाया भी जाता है। हालांकि, शिक्षित वर्गों के बीच शहरी भारत में अंग्रेजी के प्रति लोकप्रियता और सफलता के बावजूद, ग्रामीण क्षेत्रों में रहने वाले लोगों के साथ-साथ तुलना की जाए तो अंग्रेजी न बोलने वालों की तादाद का प्रतिशत ज्यादा है अर्थात् ग्रामीण क्षेत्र में अंग्रेजी बोलनेवालों का प्रतिशत न के बराबर है।<sup>5</sup>

हिंदी का संबंध भारोपीय परिवार के इंडो-आर्यन से है। यह परिवार मुख्यतः भारत तथा यूरोप के आसपास फैला हुआ है, अंतः इसे भारत-यूरोपीय या संक्षेप में भारोपीय कहते हैं। भारत में आर्यों के आने के बाद से यहाँ अनेक भाषाओं का विकास हुआ है। जिसके स्वरूप वैदिक संस्कृत, पालि, प्राकृत, अपभ्रंश, हिंदी (खड़ीबोली), बांग्ला, ओड़िया, असमिया, मराठी, सिन्धी और पंजाबी आदि भाषाओं का समावेश आर्यभाषा परिवार में किया जाता है।<sup>6</sup> इसके अतिरिक्त अन्य भाषाएँ भी भारत के अन्य राज्यों में बोली जाती हैं, जिनका भाषा परिवार अलग-अलग है।

---

<sup>4</sup> [https://en.wikipedia.org/wiki/English\\_language](https://en.wikipedia.org/wiki/English_language) (retrived on 10.04.2017)

<sup>5</sup> Banerjee, Esha, "Intra-phrasal agreement within the Hindi Noun Phrase: A Grammar Checker Approach" 2011,P.6 (M.Phil Dissertation)

<sup>6</sup> शर्मा, रामकिशोर, *भाषाविज्ञान हिंदी भाषा और लिपि*, 2007, पृ. 216

हिंदी की उत्पत्ति संस्कृत भाषा से होते हुए शौरसेनी प्राकृत और शौरसेनी अपभ्रंश के माध्यम से हुई है। हिंदी पर द्रविड़, तुर्किक, फारसी, अरबी, पुर्तगाली और अंग्रेजी भाषा का भी प्रभाव पड़ा है। 7वीं शताब्दी ईसवी में प्राकृत भाषा से 10वीं शताब्दी ईसवी में हिंदी को अपभ्रंश भाषा ने स्थिरता दिलाई। आज ब्रजभाषा, भोजपुरी, अवधी, खड़ी बोली इत्यादि हिंदी की बोलीयाँ हैं। यह हिंदुस्तानी के रूप में मानक हिंदी खड़ी बोली पर आधारित है। हिंदी भाषा की मानक हिंदी मूलतः खड़ी बोली है जिसे मेरठ और आस-पास के इलाके में बोली जाती है। उर्दू - हिन्दुस्तानी का एक और रूप है जिसने बाद में मुगल काल (18 वीं सदी) में भाषाई प्रतिष्ठा अर्जित की थी और 19 वीं सदी के उत्तरार्ध में हिंदी का मानक रूप अपनाने के लिए एक आंदोलन ने उर्दू से हिन्दुस्तानी भाषा को अलग कर दिया जिससे हिंदी भाषा मानक हिंदी के रूप में विकसित हुई। आज हिंदी, विश्व में मंदारिन, स्पेनिश और अंग्रेजी के बाद चौथी सबसे अधिक बोली जानेवाली भाषा है। भारत के बाहर इसका आधिकारिक भाषा के रूप में प्रयोग किया जाता है। जैसे- फिजी-हिंदी के रूप में फिजी में प्रयोग किया जाता है। मॉरीशस, त्रिनिदाद, टोबैगो, गुयाना, सूरीनाम में यह एक मान्यता प्राप्त क्षेत्रीय भाषा है। नेपाल, संयुक्त राज्य अमेरिका, इंग्लैंड, दक्षिण अफ्रीका आदि देशों में भी हिंदी बोली जाती है तथा दुनिया भर के देशों में बड़ी संख्या में हिंदी को विदेशी भाषा के रूप में पढ़ाया जाता है।<sup>7</sup>

आज संवैधानिक रूप से हिंदी भारत की राजभाषा है और लिपि देवनागरी है। साथ ही अंग्रेजी भाषा का प्रयोग सह-राजभाषा (द्वितीय भाषा) के रूप में किया जाता है। हिंदी भाषा के बोलने वालों की संख्या भारत में 42 करोड़ 20 लाख (41.3%) (2001 की जनगणना के अनुसार) है। कुल मिलाकर भारत और अन्य देशों में 60 करोड़ से अधिक लोग हिंदी बोलते, पढ़ते और लिखते हैं। फिजी, मॉरीशस, गुयाना, सूरीनाम की अधिकतर और नेपाल की कुछ जनता हिंदी भाषा बोलती है। हिंदी का प्रयोग प्रमुख रूप से भारत में उत्तर प्रदेश, बिहार, मध्यप्रदेश, हरियाणा, दिल्ली, हिमाचल प्रदेश, झारखंड, राजस्थान, उत्तराखंड और छत्तीसगढ़ आदि राज्यों में किया जाता है।<sup>8</sup> इस भाषा की विभिन्न बोलियों

---

<sup>7</sup> <https://en.wikipedia.org/wiki/Hindi> (retrieved on 13.04.2017)

<sup>8</sup> Kachru, Yamuna, *Hindi*, 2006, page no.1

(Dialects) का प्रयोग अलग-अलग प्रदेशों में किया जाता है तथा इनकी बोलियों में भी काफी भिन्नता पायी जाती है।

### 2.1.2 लिपि (Script)

अंग्रेजी में रोमन लिपि और हिंदी में देवनागरी लिपि का प्रयोग होता है। दोनों लिपियों की अक्षर संरचना एकदम अलग है। जिसे नीचे दिया गया है।

#### 2.1.2.1 रोमन लिपि

रोमन लिपि अंग्रेजी सहित अनेक यूरोपीय भाषाओं की लिपि है, जैसे- फ्रेंच, जर्मन, स्पेनिश, इतालवी, लैटीन, ग्रीक, डच, नॉर्वेजियन, स्वीडिश आदि। रोमन लिपि का उदगम सेमेटिक वर्णमाला (1500 ई.पू.) और प्रारंभिक ग्रीक (100 ई.पू.) से माना जाता है। चूंकि इस बीच उसके कई रूप विकसित हो चुके थे, इसलिए 350 ई.पू. के आस-पास इसका मानकीकरण किया गया। पंद्रहवीं शताब्दी में इसके कैपिटल और छोटे अक्षरों का मानकीकरण हुआ। वर्णों की आकृति की दृष्टि से रोमन रेखीय लिपि है अर्थात् इसमें प्रत्येक वर्ण, चाहे वह व्यंजन हो या स्वर, एक के बाद एक स्वतंत्र रूप से रेखीय क्रम में लिखा जाता है, जैसे- remuneration , documentary आदि। इसमें किसी विशेष चिन्हों, मात्राओं और संयुक्त व्यंजनों का प्रयोग नहीं होता। टंकण (typing), मुद्रण और कंप्यूटर की दृष्टि से यह लिपि आसान है। इस लिपि में लिप्यंतरण (transliteration) और लिप्यंकन (transcription) करना आसान है। भाषाओं की ध्वनियों को लिप्यंकित करने के लिए रोमन लिपि में इंटरनेशनल फ़ोनोटिक अल्फाबेट (IPA) का प्रयोग किया जाता है, जिसका आज विश्व भर में अधिकतर भाषाएँ प्रयोग करती हैं। देवनागरी में अंतरराष्ट्रीय स्तर पर इस तरह की प्रणाली का कोई विकास नहीं हुआ, लेकिन केंद्रीय हिंदी निदेशालय ने भारतीय भाषाओं की विशिष्ट ध्वनियों के लिप्यंकन के लिए 1980 में देवनागरी मॉडल विकसित किया है।<sup>9</sup>

---

<sup>9</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ.76

### 2.1.2.2 देवनागरी लिपि

भारत में प्राचीन लिपियों में ब्राह्मी और खरोष्ठी लिपि मिलती है। देवनागरी लिपि का मूल स्रोत ब्राह्मी लिपि (500 ई.पू.) माना जाता है जो लगभग सभी भारतीय लिपियों की मातृलिपि है। ब्राह्मी लिपि का (जिसके उद्गम स्थान को लेकर विद्वानों में मतभेद है) भारत में दो शैलियों में विभाजन था - उत्तरी शैली और दक्षिणी शैली। उत्तरी शैली से ऐतिहासिक क्रम में गुप्त लिपि (400-500 ई.), कुटील लिपि (600 ई.), प्राचीन नागरी (900-1100 ई.) और देवनागरी लिपि (1200) का विकास हुआ। लेकिन जिस रूप में आज हम देवनागरी को पहचानते हैं उसका विकास पंद्रहवीं शताब्दी में ही हुआ था। इसी प्राचीन नागरी से आर्य परिवार की अन्य लिपियाँ (बांगला, गुजराती, उड़ीया, असमिया, गुरुमुखी आदि) भी विकसित हुई हैं। दूसरी ओर ब्राह्मी लिपि की दक्षिणी शैली में नंदनागरी लिपि और फिर उससे द्रविड़ भाषाओं की चार लिपियाँ (तमिल, तेलुगु, कन्नड़ और मलयालम) का विकास हुआ है।<sup>10</sup>

हिंदी भाषा देवनागरी लिपि में लिखी गई है। इसे बायें से दायें लिखा जाता है। दो हजार से अधिक वर्षों की अवधि के दौरान संस्कृत और अन्य भारतीय भाषाओं में लिखने के लिए देवनागरी लिपि का प्रयोग किया जाता रहा है। औपनिवेशिक काल में भी इस लिपि का विशेष उपयोग (विशेष रूप से यूरोपीय विद्वानों द्वारा) संस्कृत में काम प्रकाशित करने के लिए किया गया था। नागरी नाम कैसे पड़ा इसको लेकर विद्वानों में मतभेद हैं। लेकिन कहा जाता है कि गुजरात के नागर ब्राह्मणों द्वारा विशेष रूप से प्रयुक्त होने के कारण यह नागरी कहलाई, तो कुछ लोगों के अनुसार नगरों में प्रचलित होने के कारण इसका नाम 'नागरी' पड़ा। अन्य लोगों के अनुसार देवनागर अर्थात् काशी में प्रचार के कारण यह 'देवनागरी' कहलाई। विद्वानों के अनुसार नाम की व्युत्पत्ति का प्रश्न अभी तक अनिर्णीत है। इस लिपि पर फ़ारसी के नुक्ता/बिंदु का और मराठी के अ, ळ आदि के स्थान पर अ, ल, या, आ, ई, उ आदि तथा गुजराती-अंग्रेजी के प्रभाव से देवनागरी में अँ का प्रयोग शुरू हुआ है।<sup>11</sup> फिर भी यह कहा जा सकता है कि बीसवीं

<sup>10</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 77

<sup>11</sup> तिवारी, भोलानाथ, *हिंदी भाषा*, 2005, पृ.189-190

सदी में आज भी देवनागरी लिपि का उपयोग तेजी से हो रहा है। लिप्यंतरण और अंतरराष्ट्रीय शब्दावली की दृष्टि से उनके प्रचलित रूप को अंग्रेजी रूप में ही अपनाना चाहिए और हिंदी एवं अन्य भारतीय भाषाओं की प्रकृति के अनुसार ही उनका लिप्यंतरण करना चाहिए।

### 2.1.3 वर्णमाला (Alphabet)

दो भिन्न भाषा परिवार होने के कारण अंग्रेजी और हिंदी की ध्वनि व्यवस्था में काफी अंतर है। दोनों भाषाओं की वर्णमालाओं (alphabets) का विन्यास, वर्णों (characters) के उच्चारण मूल्य (pronunciation values), ध्वनिगत विशिष्टताएँ (phonetic features), बलाघात (stress) और अनुतान (intonation) स्तर पर दिखाई देता है। वर्णों के उच्चारण समूह को 'वर्णमाला' कहते हैं। वर्ण हमारी उच्चारित भाषा या वाणी की सबसे छोटी इकाई है। इन्हीं इकाइयों को मिलाकर शब्द समूह और वाक्यों की रचना होती है।

स्वर उन ध्वनियों को कहते हैं जिनके उच्चारण में कोई अवरोध उत्पन्न नहीं होता। व्यंजन उन ध्वनियों को कहते हैं जिनके उच्चारण में वायु मार्ग में अवरोध उत्पन्न होता है। व्यंजन का उच्चारण स्वरों की सहायता से होता है। अंग्रेजी वर्णमाला प्राचीन रोमन वर्णमाला पर आधारित है। रोमन वर्णमाला में स्वरों और व्यंजनों का अलग-अलग वर्गीकरण नहीं है और न ही उसका क्रम विन्यास (sequence) संस्कृत वर्णमाला की तरह है। अंग्रेजी के स्वर (a,e,i,o,u) वर्णमाला के बीच में स्थित हैं जबकि हिंदी में स्वरों का एक अलग वर्ग है। अंग्रेजी वर्णमाला में 21 व्यंजन (consonants) और 5 स्वर (vowel) हैं।

A	B	C	D	E
F	G	H	I	J
K	L	M	N	O
P	Q	R	S	T
U	V	W	X	Y
Z				

टेबल 2.1 : अंग्रेजी वर्णमाला

अंग्रेजी में मानक उच्चारण के लिए यूनाइटेड किंगडम और संयुक्त राज्य अमेरिका में किए जाने वाले उच्चारण को आधार माना जाता है। जिसे ध्वन्यात्मक प्रतीकों के लिए 'अंतर्राष्ट्रीय ध्वन्यात्मक वर्णमाला' (आईपीए) का प्रयोग किया जाता है।<sup>12</sup>

हिंदी और भारत की लगभग सभी भाषाओं (उर्दू को छोड़कर) की वर्णमाला व्यवस्था मूल रूप से संस्कृत वर्णमाला पर आधारित है। संस्कृत वर्णमाला का संयोजन भाषावैज्ञानिक और ध्वन्यात्मक लक्षणों (phonetics features) पर आधारित है। हिंदी वर्णमाला देवनागरी का क्रम विन्यास, स्वरों और व्यंजनों के उच्चारण स्थान में कंठ्य (Velar), तालव्य (palatal), मूर्धन्य (retroflex), दंत्य (dental), ओष्ठ्य (labial), नासिक्य (nasal); प्रयत्न या उच्चारण रीति (manner of articulation) में अघोष, सघोष और अल्पप्राण, महाप्राण की संकल्पना पर आधारित है। यमुना कचरू ने 'हिंदी' पुस्तक में हिंदी वर्णमाला में 50 वर्ण दिए हैं। उसमें 12 स्वर और 38 व्यंजन दिए हैं जिसे निम्न टेबल में दिया गया है।

क) हिंदी स्वर

अ (a)	आ(ā)	इ (i)	ई (ī)	उ (u)	ऊ (ū)	ए (ē)
ऐ (ai)	ओ(ō)	औ (au)	ऋ (ri)	ं(m)	ँ(m)	

टेबल 2.2: हिंदी वर्णमाला स्वर

ख) हिंदी व्यंजन

क k	ख kh	ग g	घ gh	ङ ṅ
च c	छ ch	ज j	झ jh	ञ ṅ
ट ṭ	ठ ṭh	ड ḍ	ढ ḍh	ण ṇ
त t	थ th	द d	ध dh	न n

<sup>12</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ.69 -70

प p	फ ph	ब b	भ bh	म m
य y	र r	ल l	व v	
श sh	ष sh	स s	ह h	
क्ष ksh	त्र tra	ज्ञ jna	श्र shra	
ड़ (ṛa)	ढ़ (ṛha)			

टेबल 2.3: हिंदी वर्णमाला व्यंजन

दोनों भाषाओं की वर्णमाला में काफी भिन्नता है। यह देखा गया है कि अंग्रेजी में किसी भी तरह के अक्षर के साथ मात्रा का प्रयोग नहीं होता, लेकिन हिंदी में स्वर के साथ मात्रा का प्रयोग हमेशा होता है जिससे व्यंजन का निर्माण होता है जैसे- ा, ि, ी, ु, ू, ृ, े, ै, ो, ौ, ं, ः आदि। इन मात्राओं के अलावा देवनागरी में क, ख, ग, ज, ड, ढ, फ, य इन अक्षरों का भी प्रयोग होता है। हलन्त (्), अनुस्वार (ं) - स्वर के बाद आनेवाला व्यंजन है, जिसकी ध्वनि नाक से निकलती है। जैसे- अंगूर, कंकण। विसर्ग (ः) - इसका भी प्रयोग स्वर के बाद होता है। यह व्यंजन है और इसका उच्चारण 'ह' की तरह होता है। जैसे- अतः, स्वतः, दुःख इत्यादि। संयुक्त व्यंजन में ज (ज+ञ), क्ष (क+श), श्र (श+र), त्र (त+र); र का प्रयोग कर क्र, प्र, द्र, व्र; इस मात्रा का प्रयोग कर्, हर्, शर् म आदि संयुक्त व्यंजन के निर्माण के लिए किया जाता है।<sup>13</sup> युग्मक व्यंजन - पक्की, कच्चा, अच्छा आदि शब्दों का निर्माण दो शब्दों के एक साथ प्रयोग से होता है।

### 2.1.3.1 अंग्रेजी और हिंदी भाषाओं की ध्वनियों में अंतर

अंग्रेजी और हिंदी की ध्वनियों में भी महाप्राणत्व, बलाघात, अनुनासिकता के माध्यम स्वरूप भिन्नता दिखाई देती है।

क) हिंदी में शब्द के अंत में 'अ' का लोप होता है। जैसे- राम (rama)।

<sup>13</sup> प्रसाद, वासुदेवनन्दन, *आधुनिक हिंदी व्याकरण और रचना*, 2005, पृ.18-19

ख) संध्यस्वर (Diphthong) : संध्यस्वर दो स्वरों के योग से बनी एक इकाई है। जैसे- अंग्रेजी में *tail* (टेइल) और हिंदी में *मैया* (मइया) का उच्चारण, इनमें *ai* और *अइ* संध्यस्वर है।

ग) महाप्राणत्व (Aspirate) : अंग्रेजी के स्पर्श व्यंजन (Plosives) शब्द के प्रारंभ में प्रयुक्त होने पर हल्के महाप्राणत्व के साथ उच्चारित होते हैं। जैसे- *cat*- कैट और *खैट* के बीच का उच्चारण।

घ) अनुनासिकता (Nasalization) (ँ) : जिन स्वरों का उच्चारण नाक और मुँह से होता है और उच्चारण में लघुता रहती है। जैसे- *गाँव*, *दाँत*।

ङ) 'र' का लोप: अंग्रेजी उच्चारण प्रणाली में स्वर के बाद आने वाले (*r*) का उच्चारण प्रायः नहीं होता। उसकी जगह *अ* स्वर का आगम हो जाता है। उदाहरण के लिए *father* शब्द का उच्चारण *फ़ादअ* होना।

च) बलाघात (stress) : बलाघात हमेशा शब्द के स्तर पर होता है और अनुतान (intonation) वाक्य के स्तर पर। जैसे- *import* में *im'port*, *frequent* में *fre'quent* आदि। हिंदी में कभी-कभी दो व्यंजन और एक स्वर मिलकर बलाघात बन सकता है। जैसे- *आइए* (*आ-इ-ए*), *पिला* (*प+ई-ला*), *जल्दी* (*ज+ल-द+ई*) आदि।

छ) विराम चिह्न : हिंदी में विराम चिह्न के रूप में (।) पूर्ण विराम का प्रयोग होता है और अंग्रेजी में (.) की तरह पूर्ण विराम का प्रयोग करते हैं। अन्य चिह्नों को हमने अंग्रेजी से लिया है, इसलिए दोनों भाषाओं के बाकी चिह्न लगभग समान हैं।

ज) संक्षिप्तीकरण (abbreviation): संक्षिप्त रूपों का प्रयोग अंग्रेजी और हिंदी दोनों भाषाओं में किया जाता है। जैसे- *प्रो.*(*Pro.*), *डॉ.*(*Dr.*), *ई.*(*i.*), *पू.*(*pū.*)। दोनों भाषाओं की ध्वनियों के अंतर तथा उच्चारण से हिंदी अक्षर प्रधान भाषा है, जिसमें बलाघात का ज्यादा प्रयोग नहीं होता लेकिन शब्द के उच्चारण में समान बल के साथ उच्चारित किया जाता है।<sup>14</sup>

---

<sup>14</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 72-76

### 2.1.3.2 संख्यात्मक अंक (Numerals)

दोनों भाषाओं में अंक अलग स्वरूप में लिखे जाते हैं जिसे टेबल में दिया गया है।

रोमन	0	1	2	3	4	5	6	7	8	9	10
देवनागरी	०	१	२	३	४	५	६	७	८	९	१०

टेबल 2.4 : रोमन-देवनागरी अंक

दोनों भाषाओं की लिपि असमान होने की वजह से उनके अंक लिखने में भिन्नता पाई जाती है। लेकिन आज हिंदी में भी रोमन अंकों का प्रयोग मुद्रण और कंप्यूटर के चलते अधिक हो रहा है।

### 2.2 अंग्रेजी और हिंदी भाषाओं का तुलनात्मक व्याकरण

दोनों भाषाओं की व्याकरणिक संरचना में संज्ञा, सर्वनाम, पूर्वसर्ग, परसर्ग, विशेषण, क्रिया-विशेषण, क्रिया, लिंग, वचन, कारक का विवेचन महत्वपूर्ण है। जिसमें भाषा के व्याकरणिक घटकों का अध्ययन किया गया है। इस अनुरूप अंग्रेजी और हिंदी भाषा की संरचना के घटकों में किस तरह की भिन्नता है, इसे टेबल द्वारा देखा जा सकता है। विभिन्न भाषाओं के संपर्क में आने से और समाज के विभिन्न व्यवहार क्षेत्रों में विभिन्न मातृभाषियों द्वारा प्रयोग किए जाने के कारण आधुनिक काल में अंग्रेजी की उच्चारण व्यवस्था, शब्दावली और व्याकरणिक संरचना में परिवर्तन हुआ है। अगर भारतीय भाषाओं की संरचना की बात करें तो इन भाषाओं की संरचना एक दूसरे के समान है। लेकिन हिंदी भाषा की तुलना अंग्रेजी भाषा से करें तो यह ज्ञात होता है कि दोनों भाषाओं की संरचना में भिन्नता पाई जाती है। जिसे नीचे टेबल द्वारा देखा जा सकता है।

अंग्रेजी	हिंदी
कर्ता, क्रिया, कर्म (शब्दक्रम)	कर्ता, कर्म, क्रिया (शब्दक्रम)
तीन लिंग : पुल्लिंग, स्त्रीलिंग और नपुंसक लिंग (Neuter)	दो लिंग : पुल्लिंग और स्त्रीलिंग
वचन : एकवचन और बहुवचन [ बहुवचन में संज्ञा के बाद -s, -es, -z, (boy-boys) आदि का प्रयोग]	वचन : एकवचन और बहुवचन [बहुवचन में कर्तृकारक (कमरा), तिर्यक कारक (Oblique case) (कमरे में) का प्रयोग]
अंग्रेजी वचन का प्रभाव : संज्ञा, पुरुषवाचक सर्वनाम (he-they), निर्देशवाचक सर्वनाम (this-these) और क्रिया पर भी होता है।	हिंदी वचन का प्रभाव : संज्ञा (वह लड़का), सर्वनाम (वह-वे, यह-ये), विशेषण (छोटा कमरा-छोटे कमरे) और क्रिया पर भी होता है।
संज्ञा के पहले और बाद में पूर्वसर्ग का प्रयोग	संज्ञा के बाद परसर्ग का प्रयोग (सर्वनाम के साथ कारक प्रयोग) (राम ने, उसमें)
कर्मकारक का प्रयोग (accusative)	कर्तृकारक का प्रयोग (ergative)
संबंध कारक के लिए सिर्फ of का प्रयोग	संबंध कारक में (का, की, के) का प्रयोग
अंग्रेजी में and, or, but, then, because, so, either or का प्रयोग अधिक किया जाता है।	समुच्चयबोधक का प्रयोग दो या दो से अधिक शब्द या वाक्यों को जोड़ने का काम करता है। और, कि, लेकिन, पर, या, तथा, एवं, इसलिए आदि का प्रयोग अधिक होता है।
वाच्य : कर्तृवाच्य, कर्मवाच्य,	वाच्य का प्रयोग : कर्तृवाच्य, कर्मवाच्य और भाववाच्य

विशेषण के पहले the आर्टिकल का प्रयोग होता है तो उसके बाद संज्ञा शब्द प्रयुक्त नहीं होता। (The poor)	विशेषण का संज्ञा के रूप में प्रयोग (सुंदर लड़की), क्रिया के रूप में प्रयोग (अच्छा करना)
संज्ञा का क्रिया पर प्रभाव नहीं पड़ता (Geeta cooks food/ Sudhir cooks food)	संज्ञा का (लिंग से) क्रिया पर प्रभाव (गीता खाना <u>पकाती</u> है/ सुधीर खाना <u>पकाता</u> है)

टेबल 2.5 : अंग्रेजी-हिंदी भाषा की व्याकरणिक संरचना

### 2.2.1 संज्ञा (Nouns)

संज्ञा उस विकारी शब्द को कहते हैं, जिससे किसी विशेष वस्तु, भाव और जीव के नाम का बोध हो। संज्ञा शब्द संरचना एवं वाक्य संरचना का महत्वपूर्ण अंग है। संज्ञा के अनेक प्रकार वैयाकरणों ने स्पष्ट किए हैं, लेकिन प्रमुख रूप से संज्ञा के प्रकार निम्न हैं।<sup>15</sup>

अंग्रेजी में संज्ञा के प्रकारों में Proper noun (व्यक्तिवाचक संज्ञा): *श्याम, सुरेश*; Common noun (जातिवाचक संज्ञा): *मकान, कुर्सी, घोड़ा*; Collective noun (समूहवाचक संज्ञा): *सभा, दल, गुच्छा*; Abstract noun (भाववाचक संज्ञा): *मित्रता, चतुराई, लड़कपन*; Countable: *book, table* और Uncountable noun: *water, cheese, advice*; इसे हिंदी में द्रववाचक संज्ञा: *लोहा, सोना, दूध, पानी* इत्यादि के अंतर्गत रखते हैं। यही संज्ञा के प्रकार हिंदी में भी पाए जाते हैं। संज्ञा के रूपांतर तथा विकार में लिंग, वचन और कारक इन तीन घटकों का स्थान महत्वपूर्ण होता है। इसलिए यहाँ लिंग, वचन और कारक को विश्लेषित किया गया है।<sup>16</sup>

<sup>15</sup> प्रसाद, वासुदेवनन्दन, *आधुनिक हिंदी व्याकरण और रचना*, 2005, पृ.69

<sup>16</sup> Koul, Omkar N., *Modern Hindi Grammar*, 2008, p.33-35

### 2.2.1.1. लिंग (Gender)

व्याकरणिक गठन में सबसे पहला अंतर संज्ञा शब्दों की लिंग व्यवस्था में अनुभव होता है। अंग्रेजी में तीन लिंगों की व्यवस्था है, पुल्लिंग (masculine), स्त्रीलिंग (feminine) और नपुंसक लिंग (Neuter Gender)। निर्जीव वस्तुओं को अंग्रेजी में नपुंसक लिंग में रखा जाता है। हिंदी में नपुंसक लिंग न होने के चलते *Pencil, Car, Books, 'समाज', 'रूमाल'* आदि शब्दों को अब पुल्लिंग और स्त्रीलिंग इन दो प्रकारों में ही समाहित किया जाता है। हिंदी में केवल पुल्लिंग और स्त्रीलिंग ये दो लिंग हैं। पुरुष जाति का जहाँ बोध होता है उसे पुल्लिंग कहते हैं। जैसे- *लड़का, नर* आदि। स्त्री जाति का जहाँ बोध होता है उसे स्त्रीलिंग कहते हैं। *लड़की, नारी* ये शब्द स्त्री जाति का बोध कराते हैं। संज्ञा में प्रत्ययों के प्रयोग से लिंग बदल जाता है। सामान्यतः पुल्लिंग में '-आ, -आव, -पा, -पन, -न' प्रत्ययों का प्रयोग शब्द के अंत में होता है। जैसे- *मोटा, चढ़ाव, बुढ़ापा, लड़कपन, लेन-देन* आदि शब्दों में प्रत्ययों का समावेश अतर्निहित है। अन्य कई प्रकार के शब्दों का भी समावेश पुल्लिंग में किया जाता है। उसी तरह पुल्लिंग से स्त्रीलिंग बनाने के लिए स्त्री प्रत्यय '-ई, -इया, -इन, -नी, -आनी, -आइन, -आ, -नी' (*घोड़ी, चिड़िया, धोबिन, मोरनी, नौकरानी, पंडिताइन, बाला* आदि शब्दों में) प्रयोग होता है।

### 2.2.1.2. वचन (Number)

अंग्रेजी और हिंदी में वचन के दो प्रकार हैं- एकवचन और बहुवचन। शब्दों के जिस रूप से एक ही वस्तु का बोध हो, उसे एकवचन कहते हैं, जैसे- *snake, cat, river, बच्चा, पुस्तक* और शब्दों के जिस रूप से अनेकता का बोध होता हो, उसे बहुवचन कहते हैं। अंग्रेजी में संज्ञा के बहुवचन रूप के लिए (-S) प्रत्यय का प्रयोग किया जाता है। जैसे- *snakes, cats, rivers* तो कुछ शब्द में -es (*buses, wishes*) प्रत्यय प्रयोग कर बहुवचन शब्द बनाए जाते हैं। और जहाँ अंग्रेजी एकवचन संज्ञा के व्यंजन के अंत में -y प्रत्यय होता है उसे हटाकर -ies (*babies, spies, cities*) प्रत्यय जोड़कर बहुवचन संज्ञा का निर्माण होता है। कुछ अनियमित संज्ञाएँ (Irregular nouns) *women, men, people* आदि का प्रयोग भी बहुवचन के रूप में अंग्रेजी में होता है। हिंदी में बहुवचन शब्द *कपड़े, टोपियाँ, गुरुजन* आदि शब्द हैं। हिंदी में

संज्ञाओं का एकवचन से बहुवचन में रूपांतर करने के लिए प्रत्ययों का प्रयोग होता है। हिंदी में पुल्लिंग रूप '-आ' का रूपांतर बहुवचन '-ए' में होता है, जैसे- लड़का-लड़के । एकवचन और बहुवचन में प्रत्ययों का प्रयोग टेबल में दिया गया है।

प्रत्यय	एकवचन	प्रत्यय	बहुवचन
-अ	पुस्तक	-एँ	पुस्तकें
-आ	घोड़ा	-ए	घोड़े
-आ	कला	-एँ	कलाएँ
-या	गुड़िया	-याँ	गुड़ियाँ
-उ	धातु	-एँ	धातुएँ
-ऊ	वधू	-एँ	वधुएँ

टेबल 2.6 : एकवचन और बहुवचन के प्रत्यय

कुछ शब्दों के रूप 'एकवचन' और 'बहुवचन' में समान होते हैं, जैसे-जल-जल, प्रेम-प्रेम । जब संज्ञाओं के साथ परसर्ग (ने, से, को आदि) का प्रयोग होता है, तो संज्ञाओं का बहुवचन बनाने के लिए '-ओं' का प्रयोग होता है। जैसे-लड़के को बुलाओ/लड़कों को बुलाओ । परसर्ग के साथ प्रयुक्त होनेवाले रूप को तिर्यक् या विकारी रूप कहते हैं। संबोधन में भी 'ओ' जोड़कर बहुवचन बनाया जाता है। जैसे- बच्चों ! ध्यान से सुनो ।

### 2.2.1.3. कारक (case)

अंग्रेजी में कोई ऐसी विशिष्ट कारक व्यवस्था (Case system) नहीं है लेकिन तीन कारक पाए जाते हैं जो सर्वनाम के रूप में भी प्रयोग किए जाते हैं। i) Subjective case कर्ताकारक (I, you, he, she, it, we, they, who, whoever) इसमें I, he, we कर्ता की भूमिका निभाने के लिए इनका प्रयोग होता है। ii) objective case कर्मकारक (me, you, him, her, it, us, them, whom, whomever) इसमें me, him, us का प्रयोग कर्म की भूमिका के लिए होता

है। iii) possessive case संबंधकारक में (*your, yours, my, mine, his, hers, its, their, theirs, our, ours, whose, whose ever*) इन सर्वनाम संज्ञा का प्रयोग वाक्य में संबंध दर्शाने के लिए किया जाता है।<sup>17</sup>

हिंदी में भी तीन प्रकार के कारक पाए जाते हैं।

i) प्रत्यक्ष कारक (Direct case) कर्ता और प्रत्यक्ष कर्म को इंगित करते हैं जिस शब्द के साथ कारक चिन्ह का प्रयोग नहीं होता, जैसे- *लड़का* साइकिल चला रहा है। ii) अप्रत्यक्ष कारक (Oblique case) इसमें संज्ञा के साथ प्रत्यय और सामान्यतः कर्म के बाद परसर्ग (postposition) का प्रयोग किया जाता है। जैसे- मैं कल सभी *लड़कों से* मिला। iii) संबोधन कारक (Vocative case) जिसका प्रयोग किसी को संबोधित करने के लिए किया जाता है। जैसे- *ओ सोहन, ए लड़के* आदि।

कारक	पुल्लिंग		स्त्रीलिंग	
	एकवचन	बहुवचन	एकवचन	बहुवचन
प्रत्यक्ष	∅	∅	∅	∅
अप्रत्यक्ष	-ए -e	-ओं -ō	-इ -i	-ओं -ō
संबोधन	-ए -e	-ओ -o	-इ -i	-ओ -o

टेबल 2.7 : कारक में प्रत्यय प्रयोग

संबोधन कारक में *ओ, हे, अरे* आदि शब्द का प्रयोग पुल्लिंग और स्त्रीलिंग शब्दों के साथ होता है। हिंदी में हमेशा कारक-प्रत्यय प्रयोग के बाद परसर्ग का प्रयोग संज्ञा पदबंध और क्रिया पदबंध के संबंध को दर्शाता है। दोनों भाषाओं में संज्ञा पदबंध स्तर पर निर्धारक, पूर्व विशेषक, पश्च विशेषक का स्थान महत्वपूर्ण है, विशेषतः अंग्रेजी के निर्धारक *a, an, the* जिनके लिए हिंदी में किसी शब्द का प्रयोग नहीं होता।

<sup>17</sup> [https://en.wikipedia.org/wiki/Grammatical\\_case](https://en.wikipedia.org/wiki/Grammatical_case) (retrieved on 15.04.2017)

### 2.2.2 सर्वनाम (Pronouns)

सर्वनाम संज्ञा के स्थान पर प्रयुक्त होते हैं। सर्वनाम संज्ञा शब्दों से भिन्न हैं उनका संबोधन में प्रायः प्रयोग नहीं होता। सर्वनाम उस विकारी शब्द को कहते हैं जो पूर्वपार संबंध से किसी भी संज्ञा के बदले में आते हैं। जैसे- मैं, तुम, आप, वह, यह इत्यादि। सर्वनाम में तीन पुरुषवार- उत्तम पुरुष, मध्यम पुरुष, अन्य पुरुष तथा दो वचन एकवचन और बहुवचन और दो कारक प्रत्यक्ष और अप्रत्यक्ष में वर्गीकृत किया गया है। सामान्यतः अंग्रेजी और हिंदी में सर्वनाम के 6 प्रकार पाए जाते हैं।

क) Personal (पुरुषवाचक): *I, you, he, they, mine* (मैं, हम, तू, तुम, आप, यह, ये, वह, वे)

ख) Demonstrative (संकेतवाचक): *This, that, these, those* (यह, वह, ये, वे)

ग) Interrogative (प्रश्नवाचक) : *who, which, what, where* (क्या, कौन, कहाँ, कैसा)

घ) Relative (संबंधवाचक): *which, that, who, whom* (जो, जब, जैसा)

ङ) Reflexive (निजवाचक): *-self, themselves* (स्वयं, खुद, अपने-आप)

च) Indefinite (अनिश्चयवाचक): *some, all, none, few* (कोई, कुछ, अनेक, कई)

सर्वनामों के कर्ता और कर्म के स्थान पर प्रयुक्त होने पर इनके रूप अकसर बदल जाते हैं इसी प्रकार पुरुषवार इन सर्वनामों के भी रूप बदलते हैं।

क) उत्तम पुरुष (1<sup>st</sup> person): मैं, मेरा, मुझे/मुझको, मैंने; हम, हमारा, हमें/हमको, हमने; *I, my, me, mine; we, our, ours, us.*

ख) मध्यम पुरुष (2<sup>nd</sup> person): तुम, तुम्हारा, तुम्हें/ तुमको, तुमने; तू तेरा, तुझे/तुझको, तूने; आप, आपका, आपको, आपने *you, your, yours.*

ग) अन्य पुरुष (3<sup>rd</sup> person): वह, उसका, उसे/ उसको, उसने; वे, उनका, उन्हें/उनको, उन्होंने, यह, इसका, इसे/इसको, इसने; इनका, इन्हें/ इनको, इन्होंने *he, his, him; she, her, it, it's; they, their, theirs, them; this, these, those, that* इत्यादि। इनके अलावा संबंधवाचक और प्रश्नवाचक सर्वनामों के भी तीर्थक (oblique) रूप होते हैं। अंग्रेजी में 5 व्यक्तिवाचक सर्वनामों के तीर्थक (oblique) रूप - *me, him, her, us, them* होते हैं जबकि हिंदी में तीर्थक सर्वनामों की संख्या कहीं अधिक है, जैसे- मुझे, हमें, उसे, उन्हें, तुम्हें आदि। इनके

अलावा संबंधवाचक और प्रश्नवाचक सर्वनामों के भी तिर्यक रूप होते हैं।<sup>18</sup> हिंदी सर्वनामों का आदरार्थ रूप में भी प्रयोग किया जाता है। जिसमें हिंदी सर्वनाम *आप (you)* शब्द हमेशा आदरसूचक के रूप में प्रयोग किया जाता है।

### 2.2.3 पूर्वसर्ग और परसर्ग (Preposition and postposition)

पूर्वसर्ग (preposition) का प्रयोग अंग्रेजी में और परसर्ग (postposition) का प्रयोग हिंदी में होता है। संज्ञा के साथ इन घटकों का प्रयोग होता है। संज्ञा तथा सर्वनाम के साथ भी इनका प्रयोग किया जाता है। नीचे टेबल में दिए गए पूर्वसर्ग और परसर्ग को देखा जा सकता है।

भूमिका	पूर्वसर्ग (preposition)	परसर्ग (postposition)
कर्ताकारक (agentive)	Ø, by (only in passive constructions)	ने, Ø
कर्मकारक (Objective/Patient or Experiencer)	To	को
करणकारक (Instrumental)	With, by, through	से, के द्वारा
संप्रदान (Dative, Beneficiary)	to, for	को, के लिए
अपादान (ablative)	From	से
अधिकरण (Locative)	In, at, on	में, पर
संबंधकारक (Genitive)	Of, 's'	का, की, के
संबोधन (vocative)	O!	हे, अरे, ऐ

टेबल 2.8 : अंग्रेजी पूर्वसर्ग और हिंदी परसर्ग का प्रयोग

<sup>18</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 94

हिंदी में संयुक्त परसर्ग में के लिए (for), के पास (near), के बाहर (outside), के अंदर (inside), के सामने (in front of), के पीछे (behind), के बगल में (beside), के उपर (above and over), के नीचे (beneath, underneath), के द्वारा (by through), के जरिए (through), के बजाए (instead of), के सिवाए (except) आदि परसर्गों का प्रयोग होता है।<sup>19</sup> दोनों भाषाओं में पूर्वसर्ग तथा परसर्ग का प्रयोग संज्ञा तथा सर्वनाम के बाद किया जाता है।

#### 2.2.4 विशेषण (Adjectives)

जो संज्ञा या सर्वनाम की विशेषता बताए, उसे विशेषण कहते हैं। जिसकी विशेषता बताई जाए, वह विशेष्य कहलाता है। जैसे- Beautiful girl-सुंदर लड़की, Big room-बड़ा कमरा।

अंग्रेजी में विशेषण के सामान्यतः 5 प्रकार पाए जाते हैं<sup>20</sup>- 1. Adjectives of quality (गुणवाचक विशेषण) - *Honest man*- ईमानदार आदमी, *foolish* - मूर्ख, *large*- बड़ी

2. Adjectives of quantity (परिमाणवाचक विशेषण) - *some*- कुछ, *little*- थोड़ा, *enough*- पर्याप्त।

3. Numeral adjectives (संख्यावाचक विशेषण) में तीन प्रकार हैं- (i) Definite numeral (निश्चित संख्यावाचक) - *one*-एक, *two*-दो (ii) Indefinite numeral (अनिश्चित संख्यावाचक) - *first*-पहला, *second*- दूसरा, *third*- तिसरा (iii) Distributive numeral- *each*, *every* 4. Demonstrative adjectives (संकेतवाचक विशेषण)-*this*, *that*, *these*, *those*, *such*.

5. Interrogative adjectives (प्रश्नवाचक विशेषण) - *what*, *which*, *whose*, *where*. हिंदी में सामान्यतः 1. सार्वनामिक विशेषण- मैं, तू, वह, वे, उन 2. गुणवाचक विशेषण- नया, पुराना, बुरा 3. संख्यावाचक विशेषण- एक, दो, पहला, दूसरा, तिगुना 4. परिमाणवाचक विशेषण- सेर भर दूध, दस हाथ, सब धन आदि विशेषण के प्रकारों का समावेश होता है।<sup>21</sup>

<sup>19</sup> Kacharu, Yamuna, *Hindi*, 2006, p.103-106

<sup>20</sup> Wren & Martin, *High School English Grammar and compostin*, 2012, p. 17-18

<sup>21</sup> गुरु, कामताप्रसाद, *हिंदी व्याकरण*, 1920, पृ.95-99

अंग्रेजी विशेषणों में विशेष्य, लिंग, वचन के कारण प्रभावित नहीं होते, जैसे- large room/rooms, लेकिन हिंदी में आकारांत (आ में अंत होनेवाले) विशेषण शब्द विशेष्य के लिंग-वचन से प्रभावित होते हैं, जैसे- एक छोटा कमरा, दो छोटे कमरे। अन्य विशेषण शब्द जो आकारांत नहीं होते वे अविकारी होते हैं, उनमें कोई परिवर्तन नहीं होता। अंग्रेजी में द्विरूक्ति (reduplication) का प्रयोग व्याकरणसम्मत नहीं माना जाता (good-good, red-red) लेकिन हिंदी में अक्सर विशेषणों में द्विरूक्ति (दो बार) प्रयोग किया जाता है, जैसे- अच्छे-अच्छे पकवान, बड़े-बड़े घर आदि का प्रयोग किया जाता है। विशेषण का दो वस्तुओं या अस्तित्व की तुलना के लिए भी प्रयोग किया जाता है। अंग्रेजी में इस कोटि को व्यक्त करने के लिए प्रायः विशेषणों के बाद *-er+than* का तो कुछ विशेषणों में *more* और *than* का प्रयोग होता है। हिंदी में इस कोटि को व्यक्त करने के लिए अक्सर परसर्ग *से* का प्रयोग होता है। अंग्रेजी में विशेषण के बाद संज्ञा शब्द नहीं प्रयुक्त होता और उससे पहले अक्सर निश्चयवाची (*definite*) आर्टिकल *the* का प्रयोग होता है।<sup>22</sup>

अंग्रेजी में संज्ञा, क्रिया, क्रिया-विशेषण शब्दों को प्रत्यय लगाकर विशेषण का निर्माण होता है। जैसे- *-able/-ible, -al, -an, -ar, -ent, -ful, -ic/-ical, -ine, -ile, -ive, -less, -ous, -some*. उसी तरह हिंदी में भी-*आ, ईक, -ईय, -वान, -ई, -ईला, -मान* (प्रसाद:2005) आदि प्रत्ययों के प्रयोग से विशेषण का निर्माण होता है।<sup>23</sup> भाषा में कई बार संज्ञा की विशेषता बताने के लिए मात्र विशेषण का ही प्रयोग नहीं होता बल्कि कई बार संज्ञा, सर्वनाम, कृदंत (particle) और उपवाक्य भी वाक्य में विशेषण का कार्य करते हैं।

### 2.2.5 क्रिया (Verb)

क्रिया उस शब्द या शब्दों का समूह को कहते हैं, जो किसी व्यक्ति या वस्तु के सम्बन्ध में कुछ बताए या उनके द्वारा किए गए किसी कार्य को प्रकट करते हों। वाक्य में क्रिया महत्वपूर्ण घटक होती है, जो पूरे वाक्य की संरचना को स्पष्ट करती है। जैसे- *Rinku laughs*. दोनों भाषाओं में क्रिया का रूप-रचना, अर्थ और प्रकार्य की दृष्टि से

<sup>22</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 100-105

<sup>23</sup> <http://grammar.yourdictionary.com/parts-of-speech/adjectives/what-is-an-adjective.html> (retrived on 16.04.2017)

वर्गीकरण किया गया है। लेकिन उनमें अकर्मक (intransitive), सकर्मक (transitive), द्विकर्मक (Ditransitive) क्रियाएँ महत्वपूर्ण हैं। सूरजभान सिंह ने अपनी पुस्तक<sup>24</sup> में क्रियाओं का वर्गीकरण किया है। उनके आनुसार पर यहाँ क्रिया को स्पष्ट किया गया है।

अंग्रेजी में रूप रचना की दृष्टि से दो क्रियाएँ होती हैं। (क) एकल क्रियाएँ (single verbs)- जो क्रियाएँ एक शब्द से बनती हैं, जैसे- *take, lend, give* आदि, (ख) पदबंधीय क्रियाएँ (phrasal verbs)- जो दो या दो से अधिक शब्दों को मिलाकर बनती हैं, जैसे- *give up, set out* आदि।

हिंदी में रूप रचना की दृष्टि से निम्न क्रियाएँ होती हैं। (क) सरल धातु (simple root) - ये क्रियाएँ एक शब्द की होती हैं, जैसे- *कट, काट, देख, पढ़, जा, खा, लिख, पी* आदि। अंग्रेजी में इस तरह की क्रियाएँ *cut, sleep, go* आदि शब्द होते हैं। (ख) सामासिक धातु (composite root)- जो कुछ क्रियाएँ दो धातुओं से मिलकर बनती हैं, जैसे- *पढ़-लिख, खा-पी, चल-फिर* आदि।

(ग) मिश्र क्रिया (complex verb)- संज्ञा या विशेषण शब्दों के बाद *कर* या *हो* आदि क्रियाकरों (verblizers) के योग से बनती हैं, जैसे- *मना करना, मालूम होना* आदि। मिश्र क्रिया का दूसरा घटक (*कर, हो*) का काम संज्ञा या विशेषण को क्रिया में परिणत करना है। अंग्रेजी में -en और -ize से क्रिया का निर्माण होता है, जैसे- *deep: deepen, nation: nationalize* आदि।

(घ) संयुक्त क्रिया (compound verb)- इस क्रिया की रचना दो क्रियाओं के योग से होती है, जैसे- *सुमित रो पड़ा*। इनमें प्रयोग की दृष्टि से रंजक क्रियाएँ *जाना* (भाग जाना), *लेना* (कर लेना), *देना* (फेंक देना), *आना* (घिर आना), *पड़ना* (रो पड़ना), *उठना* (कॉप उठना), *डालना* (मार डालना), *बैठना* (कर बैठना) आदि का प्रयोग होता है।

---

<sup>24</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, पृ.167-180

(ड) यौगिक क्रिया (conjunct verb) - जिस क्रिया की रचना एक से अधिक तत्वों से होती है, उसे यौगिक क्रिया कहते हैं। जैसे- *हो* (be, become), *कर* (do), *दे* (give), *आ* (come), लिखवाना, आते जाते रहना, पढ़वाना, बताना, बड़बड़ाना आदि।

(च) प्रेरणार्थक क्रिया (causative verb)- जिस क्रिया के व्यापार में कर्ता पर किसी दूसरे की प्रेरणा जानी जाती है उसे प्रेरणार्थक क्रिया कहते हैं। जैसे- शिक्षक ने विद्यार्थी से पुस्तक *पढ़वायी*। इससे स्पष्ट होता है कि हिंदी में प्रेरणार्थक क्रियाएँ शब्दों में आंतरिक परिवर्तन करके बनायी जाती हैं। जैसे- *जाग (ना) जगाना जगवाना, नाच (ना) नचाना नचवाना, रो (ना) रुलाना रुलवाना* आदि। लेकिन अंग्रेजी में इस तरह वाक्य बनाने के लिए क्रिया शब्दों में कोई आंतरिक परिवर्तन नहीं होता, बल्कि, *keep, have, get, cause, make* आदि शब्दों का प्रयोग कर प्रेरणार्थक अर्थ व्यक्त किए जाते हैं।

अंग्रेजी और हिंदी में रूपरचना के बाद अर्थ और प्रकार्य की दृष्टि से प्रकार लगभग समान है।

क) सकर्मक क्रिया (Transitive verb)- वाक्य में एक कर्म के प्रयोग होने की सम्भावना होती है, जैसे- *cut, eat, scold, kill* आदि। यह कर्म संज्ञा के रूप में भी हो सकते हैं, भावार्थक संज्ञा to -infinitive, कृदंत gerund (-ing), कि-उपवाक्य that-clause भी। हिंदी में 'में, पर' का समावेश इसमें किया जाता है।

ख) अकर्मक क्रिया (Intransitive verb)- जिन क्रियाओं का व्यापार कर्ता पर हो और जिसमें कर्म की आकांक्षा नहीं होती, अंग्रेजी में जैसे- *sit, go, live, swim, decay* आदि। हिंदी में वे अकर्मक क्रिया होती हैं जो अपनी अर्थपूर्ति के लिए स्थान, समय या रीतिवाची सह-अव्ययों की माँग करती हैं। जैसे- *बच्चा रो रहा है*।

ग) द्विकर्मक क्रिया (Ditransitive verb)- यह क्रिया वाक्य में दो कर्मों (objects) की माँग करती हैं, प्रत्यक्ष कर्म और अप्रत्यक्ष कर्म। जैसे- *I gave Late a gold ring* वाक्य में *gold ring* प्रत्यक्ष कर्म है और *Late* अप्रत्यक्ष कर्म है। हिंदी में दो कर्म किसी द्विपक्षीय कार्य-व्यापार को संपन्न करने के लिए अनिवार्य होते हैं। जैसे- *पिताजी ने ड्राइवर को* पैसे दिए।

घ) योजक क्रिया (Linking verb) के अंतर्गत अंग्रेजी के to be क्रिया के सभी रूप शामिल हैं, जैसे- *is, are, was, were* इनके अलावा *look, appear, get, turn, become, grow, remain, feel, prove, happen* आदि का समावेश योजक क्रिया में होता है। हिंदी में योजक क्रिया के अंतर्गत मूल रूप से 'हो' क्रिया और इसके अन्य रूप (है, था) शामिल हैं, इसके अतिरिक्त *लगना, दिखाई देना, सिद्ध होना* आदि शामिल हैं।

उपरोक्त मुख्य क्रिया के अतिरिक्त वाक्य में सहायक क्रिया से कार्य व्यापार के काल (tense), वचन (number), पक्ष (aspect), वृत्ति (mood), वाच्य (voice) आदि से संबंधित व्याकरणिक सूचना का बोध होता है। सहायक क्रियाओं का मुख्य क्रियाओं के साथ प्रयोग होता है। वह कभी एक क्रिया और कभी एक से अधिक क्रियाओं के रूप में सहायक होती है। लिंग, वचन, काल, पक्ष, वृत्ति, वाच्य का क्रिया पर प्रभाव पड़ने से क्रिया में विकार उत्पन्न होता है। इसलिए यहाँ काल, पक्ष, वृत्ति, वाच्य को भी स्पष्ट किया गया है।

### 2.2.5.1 काल (Tense)

काल क्रिया का वह रूप है जिससे कार्य-व्यापार के घटना और समय की संकल्पना (भूत, वर्तमान, भविष्य) के बीच संबंध का बोध होता है। सामान्य रूप से काल को वर्तमान काल, भूत काल, और भविष्य काल तीन खंडों विभाजित किया गया है। लेकिन पक्ष और काल दोनों के अर्थ-क्षेत्र आपस में इतने, मिले-जुले हैं कि अक्सर व्याकरणों में इन दोनों के क्रिया-रूपों को संयुक्त रूप में प्रस्तुत करने की परंपरा रही है।<sup>25</sup> क) वर्तमान काल (Present tense)- इसमें क्रियाओं के व्यापार की निरन्तरता को दिखाया जाता है। जिसमें सामान्य वर्तमान में (*-s, -ता/-ती*), अपूर्ण वर्तमान में (*is-ing, रहा है*), पूर्ण वर्तमान में (*have -ed/en, -या/आ/ई, हूँ/हैं*), सतत् पूर्ण वर्तमान में (*has/have been-ing, रहा, रही, रहे हैं*)

---

<sup>25</sup> <http://www.dataspec.info/english/tense/1102PrTen.htm> (retrived on 17.04.2017)

ख) भूत काल (Past tense)- जिससे क्रिया से कार्य की समाप्ति का बोध हो। सामान्य भूत में (-ed (used to), -ता था, -या/-आ), अपूर्ण भूत में (was/were +ing, रहा/रहे/रही, था/थे/थी), पूर्ण भूत में (had -ed, --या/-आ था, -ता था), सतत् पूर्ण भूत में (had been -ing, कर रहा था)

ग) भविष्यत काल (Future tense)- भविष्य में होनेवाली क्रिया को दर्शाया जाता है। सामान्य भविष्य में (will, shall, -गा; is going to, -ने वाला है; is -ing रहा है; will be -ing, -एगा), अपूर्ण भविष्य में (will/shall be -ing, रहा होगा), पूर्ण भविष्य में (will/shall have -ed, -या/आ होगा, चुका होगा), सतत् अपूर्ण भविष्य (shall /will have been -ing, रहा होगा, रही होगी, रहा हूँगा, रहे होंगे) आदि क्रिया रूपों का प्रयोग इस काल में होता है।<sup>26</sup>

#### 2.2.5.2 पक्ष (Aspect)

अंग्रेजी और हिंदी में तीन पक्ष पाए जाते हैं।

क) नित्य पक्ष (habitual) - (-ता है, -s/-es लौटता है (returns) इस पक्ष से कार्य व्यापार के बार-बार घटित होने या उसके एक अनिश्चित काल तक होने का बोध होता है। हिंदी में -ता, -या, -रह प्रत्यय का प्रयोग नित्य पक्ष में होता है और अंग्रेजी में -s/-es प्रत्यय का प्रयोग क्रिया के साथ होता है। जैसे- सूरज पूरब से निकलता है। (The sun rises from the east)।

ख) सातत्य पक्ष (continuous aspect)- (रहा है, be +ing) खेल रहा है (is playing)- इस पक्ष से उक्ति (utterance) या संदर्भ के समय कार्य-व्यापार के जारी या चालू रहने की अस्थायी (temporary) अवस्था का बोध होता है। इस पक्ष को सूचित करने के लिए हिंदी में प्रत्यय-ता+रहा, -चल, जा-, आ-, चला आ-, चला जा-, -या+ (रह, कर-, चाह-), -ने+(लग, वाला है, को है, जा रह-) तथा अंग्रेजी में be+ing प्रत्यय प्रयोग से सातत्यता स्पष्ट होती है। जैसे- बाहर बारिश हो रही है। (It is raining outside)

---

<sup>26</sup> सिंह, सूरजभान, अंग्रेजी-हिंदी अनुवाद व्याकरण, पृ.125-131

ग) पूर्ण पक्ष (perfective aspect)- (-या/आ, -ed/en) लौटा (returned) इस पक्ष से कार्य-व्यापार (action) के पूर्ण हो जाने की अवस्था का बोध होता है। हिंदी में पूर्ण पक्ष से किसी एक निश्चित अवस्था (definite state), एकल (single) घटना की पूर्णता का बोध होता है। जैसे- श्याम कल देर से घर लौटा। अंग्रेजी में-ed क्रिया रूप से भूतकालिक नित्यता का भी बोध होता है जो हिंदी में संभव नहीं। हिंदी में इसके लिए नित्य भूत (habitual past) की सूचना देने के लिए -ता था का प्रयोग होता है। जैसे- मैं दिल्ली में रहता था।<sup>27</sup>

### 2.2.5.3 वृत्ति (Mood)

क्रिया विकार में वृत्ति से कार्य व्यापार के प्रति वक्ता की अभिवृत्ति (attitude) या कार्य-व्यापार के संबंध में उसके अपने दृष्टिकोण का बोध होता है, जैसे- *खाना खाइए; सर आते रहें होंगे।* वाक्य में वक्ता की यह अभिवृत्ति कई प्रकार से व्यक्त होती है, लेकिन अंग्रेजी में तीन प्रकार की वृत्ति पाई जाती है।

क) कथनात्मक (Indicative)- वस्तु-स्थिति के वर्णन के लिए, प्रश्न पूछने के लिए इसका प्रयोग होता है। *Rama goes to a school daily* (राम रोज स्कूल जाता है) (-es, is -ing, -ता)

ख) आज्ञार्थक (imperative)- आज्ञा, शिक्षा, उपदेश, अनुरोध तथा प्रार्थना के लिए आज्ञार्थक रूप का प्रयोग होता है। जैसे- *come here* (इधर आओ/आइए/आना/आइएगा) अंग्रेजी के एक क्रिया रूप के साथ हिंदी में अक्सर क्रिया के साथ -ओ, -इए, -ना, -इएगा, -ए, -एगा आदि प्रत्ययों का प्रयोग होता है।

ग) संभावनार्थक (subjunctive)- इच्छा, उद्देश्य, सत्य के विपरीत, कोई शर्त या कल्पना की भावना को प्रकट करने वाली वृत्ति का बोध होता है। जैसे- *It may rain today* (आज बारिश हो सकती है)। इसमें *may, should, can, wish* (सक, पा), *were* (-ए, एँ, -ऊँ, ता), *perhaps, probably, may be, possibly* (शायद, संभवतः, काश) आदि क्रिया वृत्ति के साथ प्रत्यय का प्रयोग होता है।<sup>28</sup>

<sup>27</sup> वही, पृ.121-124

<sup>28</sup> <http://www.dataspec.info/english/Verb/0403Mood.htm> (retrived on 18.04.2017)

संकेतार्थ वृत्ति में अंग्रेजी के if का प्रयोग होता है और इस वृत्ति में हिंदी प्रत्यय -  
 ता, ता, -या होता का प्रयोग होता है। सामर्थ्यसूचक में सकना, पाना, -या जा- प्रत्यय  
 का प्रयोग क्रिया के साथ होता है। *Can, may, might (सक); will, shall, would (-एगा);*  
*could (could not go) (सका, नहीं जा पाया/सका); should, ought to, must (-ना, चाहिए, -ना*  
*है); used to (-ता था, -या करता था); need, dare (मुख्य क्रिया)* आदि वृत्तिवाचक सहायक  
 क्रियाएँ के रूप में प्रयोग करते हैं तथा प्राथमिक सहायक क्रिया के रूप में do, have be  
 जो अंग्रेजी के कुछ खास तरह के प्रश्नवाचक, नकारात्मक, बलसूचक, तथा पूर्णतावाची  
 वाक्यों की रचना में सहायक रूप में प्रयोग होते हैं।

#### 2.2.5.4 वाच्य (Voice)

क्रिया के उस परिवर्तन को वाच्य कहते हैं, जिसके द्वारा इस बात का बोध होता  
 है कि वाक्य के अन्तर्गत कर्ता, कर्म अथवा भाव इनमें से किसकी प्रधानता है और उनमें  
 किसके अनुसार क्रिया के लिंग, वचन तथा पुरुष का प्रभाव पड़ा है। वाक्य में कर्ता, कर्म,  
 क्रिया तीन घटक होते हैं। लेकिन कभी-कभी वाक्य में कर्ता प्रधान बनकर प्रयोग होता है  
 तो कभी कर्म। अंग्रेजी में केवल दो वाच्य होते हैं कर्तृवाच्य और कर्मवाच्य, लेकिन हिंदी  
 में इसके अलावा भाववाच्य का प्रयोग होता है।

क) कर्तृवाच्य (active voice)- जिनमें कर्ता की प्रधानता का बोध होता है तथा कर्ता  
 उद्देश्य के रूप में प्रयुक्त होता है। जैसे- The servant is opening the door (नौकर दरवाजा  
 खोल रहा है।) The children sang songs (बच्चों ने गीत गाए।), हिंदी की प्रकृति में  
 कर्तृवाच्य वाक्यों का प्रयोग अधिक सहज है। इसीलिए अंग्रेजी के कई कर्मवाच्य वाक्यों  
 का सहज अनुवाद हिंदी में कर्तृवाच्य संरचना में किया जाता है।

ख) कर्मवाच्य (passive voice)- जिन वाक्यों में कर्म व्याकरणिक कर्ता बनकर वाक्य के  
 उद्देश्य के स्थान पर प्रयोग होता है और कर्ता गौण होकर कर्म के स्थान पर प्रयोग होता  
 हो या लुप्त हो जाता हो उसे कर्मवाच्य कहते हैं। जैसे- Song were sung by the children  
 (बच्चों द्वारा गीत गाए गए)। हिंदी में कर्ता+से और पैसिव क्रिया रूप -या जा- प्रत्यय  
 प्रयोग से एक खास तरह का कर्मवाच्य बनता है जिसमें असमर्थता का बोध होता है और

जो हमेशा नकारात्मकता प्रकट करता है। जैसे- *मुझसे दरवाजा नहीं खोला जाता* लेकिन अंग्रेजी में *I am unable to open the door* यह वाक्य कर्तृवाच्य (Active voice) को दर्शाता है। हमेशा अंग्रेजी कर्मवाच्य वाक्य में be+ past participle और कर्म से पहले कभी-कभी by का प्रयोग होता है तो कभी-कभी नहीं होता।

ग) भाववाच्य (impersonal voice)- क्रिया के उस रूपान्तर को भाववाच्य कहते हैं, जिससे वाक्य में क्रिया अथवा भाव की प्रधानता का बोध होता है। इस वाच्य का प्रयोग हिंदी में होता है, अंग्रेजी में नहीं। हिंदी में जो अकर्मक क्रियाओं से पैसिव वाक्य बनते हैं, उन्हें भाववाच्य कहते हैं। भाववाच्य का कोई कर्म नहीं होता। क्रिया ही वाक्य के केंद्र में होती है। जैसे- *मुझसे बहुत दूर चला नहीं जाता। (I am unable to walk very far)*<sup>29</sup> अतः हिंदी के भाववाच्य की कोटि के वाक्यों के लिए अंग्रेजी में संदर्भ के अनुसार अलग-अलग प्रकार के वाक्यों का प्रयोग करना पड़ता है।

### 2.2.6 क्रिया-विशेषण (Adverb)

जिस शब्द से क्रिया, विशेषण या दूसरे क्रिया-विशेषण की विशेषता प्रकट हो, उसे क्रिया-विशेषण कहते हैं। जैसे- *राम धीरे-धीरे टहलता है।* क्रिया-विशेषण का प्रयोग सिर्फ क्रिया की विशेषता प्रकट करने के लिए नहीं होता, बल्कि विशेषण, अन्य क्रिया-विशेषण और संपूर्ण वाक्य की भी विशेषता बताने के लिए होता है। जैसे- वह मेरा *बहुत अच्छा* दोस्त है। (He is my *very good* friend) अंग्रेजी के क्रिया-विशेषण वाक्य में कई रूपों में प्रयुक्त होते हैं- संज्ञा पदबंध Noun phrase) के रूप में, पूर्वसर्गीय पदबंध के रूप में (prepositional phrase), उपवाक्य के रूप में (clause), क्रियार्थक संज्ञा के रूप में (infinitive), तो कभी कृदंत (Particle) के रूप में प्रयोग किया जाता है। अंग्रेजी में विशेषण शब्दों तथा कृदंतों में-ly प्रत्यय से क्रिया-विशेषण बनाने की प्रवृत्ति बहुत आम है। जैसे- easily, loudly, hurriedly, quickly आदि। निम्न अंग्रेजी-हिंदी क्रिया-विशेषण के प्रकार दोनों भाषाओं में पाए जाते हैं।<sup>30</sup>

<sup>29</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, पृ.141-151

<sup>30</sup> Koul, Omkar N., *Modern Hindi grammar*, 2008, p.130

- क) समयवाचकता (time) के लिए : आज 'today', कल 'yesterday', सुबह 'morning' आदि।
- ख) स्थान या दिशा (place or direction) के लिए: अंदर 'in/ inside', बाहर 'out/outside' आदि।
- ग) रीतिवाची (manner) के लिए: आसानी से 'easily', धीरे- धीरे 'slowly' आदि।
- घ) कारण (reason) क्रिया-विशेषण के लिए: गरीबी के कारण 'for the reason of poverty', कमजोरी के कारण 'for the reason of weakness' आदि।
- ङ) करण (instrument) क्रिया-विशेषण के लिए: कलम से 'with pen', चाकू से 'with knife' आदि।
- च) उद्देश्य (purpose) क्रिया-विशेषण के लिए: पढ़ने के लिए 'for reading', काम के लिए 'for work' आदि शब्द।
- छ) मात्रावाची (degree) क्रिया-विशेषण में, बहुत 'very', काफी 'enough', लगभग 'approximately' आदि।
- ज) परिमाणसूचक क्रिया-विशेषण में लगभग 'nearly', निश्चित रूप से 'definitely, वास्तव में 'really' आदि क्रिया-विशेषण शब्दों का प्रयोग होता है।

अन्य क्रियाविशेषणों में भोलानाथ तिवारी<sup>31</sup> ने तीन प्रकारों का समावेश किया हैं-

क) समुच्चयबोधक (conjunction) में - और, एवम (and); या, अथवा, तथा (or), पर, मगर, किन्तु, लेकिन, परंतु (but), वरना, नहीं तो (otherwise), इसलिए (so), अगर, यदि (if); हालांकि, यद्यपि (although); चूंकि, क्योंकि (because) आदि शब्द। ख) संबंधबोधक/संबंधसूचक (indeclinable) - जो संज्ञा के बाद आकर उस संज्ञा का संबंध वाक्य के दूसरे शब्द से दिखाता है। जैसे- का, के, की, तक, भर, सहित, समेत आदि। ग) विस्मयादिबोधक (interjection)- जिन अव्ययो से हर्ष-शोक आदि प्रकट होता हो, जैसे- अहा!, अजी!, वाह-वाह!, अच्छा! वाह/खुब/शाबाश! (Oh good), ओह!/अरे!/ओहो! (O you), हे भगवान! (O God!), ओ लड़के (O boy!), अरे वाह! (Wow!), हाय!, अहा!, उफ़! आदि का प्रयोग विस्मयादिबोधक शब्दों में होता है।

<sup>31</sup> तिवारी, भोलानाथ, हिंदी भाषा, 2005, पृ.168

## 2.3 अंग्रेजी और हिंदी की वाक्य संरचना (English-Hindi sentence structure)

भाषिक संरचना में ध्वनि, व्याकरण और अर्थ का महत्वपूर्ण स्थान होता है और व्याकरणिक संरचना की दृष्टि से वाक्य भाषा की सबसे बड़ी इकाई मानी जाती है। इसलिए किसी भी भाषा की संरचना में वाक्य की भूमिका महत्वपूर्ण होती है। वाक्य पदों के समूह की उस इकाई को कहते हैं जो व्याकरणिक दृष्टि से पूर्ण हो तथा जिसमें एक क्रिया अवश्य हो। सामान्यतः उद्देश्य और विधेय ये दो वाक्य के महत्वपूर्ण अंग माने जाते हैं। वाक्य की संरचना में जिस तरह कर्ता, क्रिया, कर्म को वाक्य के प्रमुख घटक माना जाता है, उसी तरह वाक्यों में कुछ आंतरिक घटक होते हैं, जिनका विस्तृत वर्णन विभिन्न विद्वानों तथा राजमणि शर्मा ने 'हिंदी भाषा का इतिहास और स्वरूप' पुस्तक<sup>32</sup> में किया है।

### 2.3.1 वाक्य संरचना के आधार

वाक्य में जिस तरह शब्दों और पदबंधों का समावेश होता है, उसी तरह वाक्य में विभिन्न घटकों के बीच एक प्रकार की क्रमबद्धता (Sequence) होती है, जिसे हम पदक्रम या शब्दक्रम (Word Order) कहते हैं। पदबंध (Phrase) भी वाक्य के पदों (शब्दों) से बना होता है। जब एक से अधिक पद एक में बंधे हों तथा वे सभी मिलकर एक व्याकरणिक इकाई (संज्ञा, विशेषण, क्रिया-विशेषण आदि) का काम कर रहे हैं तो उस इकाई को पदबंध कहते हैं। पदबंध एक से अधिक पद के होते हैं और ये पद इस तरह से सम्बद्ध होते हैं कि उनसे एक इकाई बन जाती है और वह वाक्य के एक अंश के रूप में प्रयुक्त होते हैं। उदाहरण के लिए- *सुरेश के बँगले के चारों ओर पेड़-पौधे हैं।* इस वाक्य में '*सुरेश के बँगले के चारों ओर*' में कई पद हैं, जो स्थान को दर्शाता है। इस कई पदों के समूह को ही पदबंध कहते हैं।

#### 2.3.1.1 शब्दक्रम (Word order)

सभी भाषाओं की वाक्यात्मक संरचना में पद (शब्द) का स्थान महत्वपूर्ण होता है। पदक्रम से तात्पर्य वाक्य में प्रयुक्त पदों या पदबंधों के उस क्रम से है, जिस क्रम में वे

---

<sup>32</sup> शर्मा, राजमणि, *हिंदी भाषा का इतिहास और स्वरूप*, 2010, पृ. 292

वाक्य में प्रयुक्त होते हैं। इसके अंतर्गत विशेष रूप से कर्ता, कर्म, पूरक (complement), क्रिया-विशेषण (adverbs) और क्रिया (verb) आदि पदबंधों के प्रयोग स्थान शामिल है। अंग्रेजी का पदक्रम या शब्दक्रम भारतीय भाषाओं से भिन्न है। अंग्रेजी का पदक्रम कर्ता-क्रिया-कर्म (SVO) पर आधारित होने के साथ ही इस भाषा का पदक्रम स्थिर रहता है। हिंदी तथा अधिकांश भारतीय भाषाओं में वाक्यों का पदक्रम कर्ता-कर्म-क्रिया (SOV) पर आधारित है।

जैसे - *Pankaj is going to home* (subject, verb, object) (अंग्रेजी वाक्य संरचना)  
*पंकज घर जाता है।* (कर्ता, कर्म, क्रिया) (हिंदी वाक्य संरचना)

अंग्रेजी भाषा की तुलना में हिंदी भाषा का पदक्रम अस्थिर है, लेकिन संस्कृत, रूसी आदि भाषाओं का पदक्रम स्वतंत्र है। संस्कृत विभक्ति प्रधान भाषा है, इसलिए इस भाषा में पदक्रम का कोई कठोर नियम नहीं है।<sup>33</sup> अंग्रेजी की वाक्य संरचना में पदक्रम का व्याकरणिक महत्व होता है और उनकी वाक्य संरचना में बहुत कम परिवर्तन होता है। लेकिन हिंदी में सामान्य तौर पर पदक्रम अस्थिर होने से उसमें वाक्य के कई तरह के रूप मिल सकते हैं। जिससे भारतीय भाषाओं के अनुवाद में संदिग्धता निर्मित होती है।

### 2.3.1.2 अन्विति (Agreement)

अन्वय (Agreement) का अर्थ है- व्याकरणिक संरचना में एकरूपता। विभिन्न भाषाओं में विशेषण-विशेष्य, कर्ता-क्रिया, कर्म-क्रिया आदि विभिन्न व्याकरणिक कोटियों में लिंग, वचन, पुरुष, मूल और विकृत रूप की अनुरूपता होती है। जिसमें प्रत्येक भाषा के अन्वय के नियम अलग-अलग होते हैं। जैसे की वाक्य- *अच्छा छात्र - Good Student, अच्छी छात्रा - Good Student, अच्छे छात्र - Good Students, अच्छी छात्राएँ - Good Students*

हिंदी में अन्विति वचन, पुरुष और लिंग तीनों स्तरों पर मिलती है और अंग्रेजी में अन्विति कर्ता और क्रिया के बीच ही होती है, लेकिन हिंदी में अन्विति का प्रभाव संज्ञा, विशेषण और क्रिया पर भी पड़ता है। हिंदी की क्रिया कभी कर्ता के अनुरूप होती है, जैसे-

<sup>33</sup> शर्मा, रामकिशोर, *भाषाविज्ञान हिंदी और भाषा और लिपि*, 2007, पृ.150

‘राम गया’ ‘नीलिमा गई’ तो कभी कर्म के अनुरूप- ‘राम ने रोटी खाई’ ‘नीलिमा ने आम खाया’ आदि। इसलिए कहा जाता है कि हिंदी में अन्विति का क्षेत्र अंग्रेजी से अधिक व्यापक है। अंग्रेजी में अन्विति केवल वचन और पुरुष में ही मिलती है।<sup>34</sup> यह अन्विति किसी भी भाषा की व्याकरणिक एकरूपता के लिए महत्वपूर्ण होती है। जिससे भाषिक अनुवाद में कम समस्याएँ पाई जाती हैं।

### 2.3.1.3 कारक संबंध (Case relationship)

वाक्य के विभिन्न पदबंध आपस में कुछ व्याकरणिक संबंधों (Grammatical relations) से जुड़े होते हैं, जिससे उनका सही अर्थ बोध प्रतीत होता है। क्रिया और संज्ञा पदबंधों के बीच यह व्याकरणिक संबंध कारक कहलाता है। वाक्य में कारक संबंध का विशेष महत्त्व होता है। वाक्य में कारक की सूचना दो प्रकार से मिलती है- कारक विभक्तियों से (जैसे- ने, को, में, से, के लिए आदि) और पद के प्रयोग-स्थान से। अंग्रेजी की संरचना में कारक- *of, from, at, with, on, about, in, of* आदि का प्रयोग होता है। हिंदी व्याकरण में कारक के 6 प्रकार हैं। कर्ता (०,ने), कर्म(०,को), करण (से,के द्वारा), संप्रदान (को, के लिए), अपादान (से), अधिकरण (में,पर) आदि। जिनका प्रयोग दोनों भाषाओं में होता है।<sup>35</sup> लेकिन आज वाक्य विश्लेषण के संदर्भ में कारक की संकल्पना (concept) में काफ़ी परिवर्तन आया है।

अमेरिकी भाषाविद चार्ल्स फिल्मोर ने अलग से एक कारक व्याकरण (1968)<sup>36</sup> की रचना की। इस व्याकरण में वाक्य का विस्तार प्रकारता + प्रतिज्ञप्ति (Modality + Preposition) में किया है। उनके अनुसार, वाक्य एक क्रिया तथा एक से अधिक संज्ञा पदबंधों की ऐसी रचना होती है, जिसमें प्रत्येक संज्ञा और क्रिया के साथ, एक विशेष कारक का संबंध जुड़ा होता है। जिसमें अभिकर्ता (Agent), करण (Instrument),

<sup>34</sup> तिवारी, भोलानाथ, *हिंदी भाषा*, 2005, पृ. 262

<sup>35</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 201-202

<sup>36</sup> Fillmore, Charles J., *The Case for Case*, 1968, p. 1-88

अनुभावक (Experiencer), स्थान और कर्म (Place and object), अधिकरण (Locative) आदि का कारक में समावेश होता है।

#### 2.3.1.4 अर्थ-संगति (Semantic concordance)

व्याकरण और अन्विति की दृष्टि से सही होते हुए भी वाक्य अर्थ-संगति (semantic concordance) की दृष्टि से दोषपूर्ण हो सकता है, जैसे- *कल मैंने दूसरे के विचारों को खाया।* ये वाक्य अर्थ की दृष्टि से असंगत है। इस वाक्य में शब्द जहाँ एक व्याकरणिक रचना है वहीं अर्थ के स्तर पर किसी वस्तु या व्यापार (action) का प्रतीक (symbol) भी है। जिस प्रकार वाक्य में व्याकरण की रचना संबंधी अपनी कुछ आकांक्षाएँ होती हैं उसी तरह वस्तु के कार्य-व्यापार संबंधी अपनी कुछ आकांक्षाएँ हो सकती हैं। उपर्युक्त वाक्य में 'खाना' कार्य व्यापार (action) का कर्म कोई खाद्य पदार्थ हो सकता है, लेकिन 'विचार' नहीं। इसी तरह वाक्य के हर घटक को अर्थ की दृष्टि से भी अन्य घटकों के अनुकूल होना चाहिए, अर्थात् उनके बीच अर्थ-संगति होनी चाहिए।<sup>37</sup>

#### 2.3.1.5 अध्याहार या लोप (Ellipsis)

सामान्यतः वाक्य के गठन या निर्मिति के समय में बहुत से शब्द अनावश्यक मानकर छोड़ दिए जाते हैं। वास्तविक जीवन में अनेक बार वाक्य के कुछ अंशों का लोप कर दिया जाता है और कभी-कभी केवल एक-दो शब्दों का ही प्रयोग किया जाता है। फिर भी अर्थ-बोध (comprehension) होता है। भाषा के इस लक्षण को अध्याहार कहते हैं। जैसे- *Good morning सुप्रभात, Who is there? कौन?* आदि को अध्याहार वाक्य कहते हैं। कर्ता, क्रिया वाक्यांश का अध्याहार होता है। एक ही भाषा में शब्दों का कहीं लोप होता है तो किसी वाक्य में नहीं होता।<sup>38</sup> अंग्रेजी से हिंदी अनुवाद में अधिकतर वाक्य के शब्द का लोप होता है, लेकिन हिंदी में शब्द का लोप अशं:त पाया जाता है।

---

<sup>37</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 202

<sup>38</sup> वही, पृ. 202

संज्ञा से पहले प्रयुक्त होने वाले विशेषकों को पूर्व विशेषक (Pre-modifiers) कहते हैं और संज्ञा के बाद प्रयुक्त होने वाले विशेषकों को पश्च विशेषक (Post modifiers) कहा जाता है। अंग्रेजी में संज्ञा से पहले और बाद में जिस तरह विशेषकों का प्रयोग होता है। उस तरह हिंदी में केवल संज्ञा के पहले ही विशेषणों/विशेषकों का प्रयोग होता है, संज्ञा के बाद नहीं। लेकिन अंग्रेजी में संज्ञा के बाद विशेषक कई रूपों में प्रयुक्त होते हैं। विशेषक, विशेषण, क्रिया-विशेषण, परसर्गीय पदबंध, संबंधवाचक उपवाक्य, भावार्थक संज्ञा और कृदंत, क्रिया-विशेषण आदि के रूप में प्रयुक्त होते हैं।<sup>39</sup> सामान्यतः अंग्रेजी में निर्धारक के रूप में *a, an, the* का प्रयोग होता है लेकिन हिंदी में कोई शब्द प्रयुक्त नहीं होता।

इस वाक्य संरचना के आधार पर वाक्य की अंतर्भूत संरचना को यहाँ स्पष्ट करने के बाद ज्ञात होता है कि वाक्य केवल शब्द का जाल नहीं है, उसमें व्याकरणिक घटक के साथ अर्थ की संरचना का भी समावेश होना आवश्यक है, तभी वाक्य अर्थ की दृष्टि से पूर्ण हो सकता है।

### 2.3.2 वाक्य के प्रकार

सामान्यतः वाक्य को तीन प्रकारों में विभाजित किया गया है। संरचना के आधार पर- सरल वाक्य, संयुक्त वाक्य, और मिश्र वाक्य। जिसे नीचे विश्लेषित किया गया है।

#### 2.3.2.1 सरल वाक्य (Simple Sentence)

वाक्य रचना की मूल इकाई सरल वाक्य है। जिस वाक्य में एक संज्ञा उद्देश्य (कर्ता) और क्रिया एक विधेय होती है, उस वाक्य को सरल वाक्य (Simple Sentence) तथा साधारण वाक्य कहते हैं। जैसे- 'श्याम रोता है।' इसमें 'श्याम' उद्देश्य है और 'रोता है' विधेय है। कभी-कभी वाक्य में एक से अधिक उद्देश्य और विधेय का प्रयोग होता है पर उस वाक्य की संरचना संयुक्त या मिश्रित नहीं होती, जैसे- 'सुहास हँसता-खेलता है', 'रजत गाता-बजाता है' ये वाक्य भी सरल वाक्य के अंतर्गत आते हैं। वाक्य का आंतरिक विश्लेषण सरल वाक्य के आधार पर ही किया जाता है। सरल वाक्यों के उद्देश्य और

<sup>39</sup> Verma, M.K., *The Structure of Noun Phrase in English and Hindi*, 1971, p.12

विधेय का कभी-कभी विस्तार भी हो सकता है तथा इसका प्रश्नवाचक, निषेधात्मक तथा प्रश्नवाचक या निषेधात्मक रूप में रूपान्तर भी हो सकता है।<sup>40</sup>

उपवाक्य को भी प्रकार के रूप में राजमणि शर्मा ने 'हिंदी भाषा का इतिहास और स्वरूप' पुस्तक में विश्लेषित किया है और इस वाक्य को संयुक्त वाक्य और मिश्र वाक्य से अलग माना है। जब दो या दो से अधिक सरल वाक्यों को मिलकर एक वाक्य बनता है तो उस एक वाक्य में जो अन्य वाक्य मिले होते हैं, उन्हें उपवाक्य कहते हैं। वस्तुतः इस वाक्य के जोड़ने से बने बड़े वाक्य को जब 'वाक्य' कहते हैं तो उसके भीतर के छोटे वाक्यों को 'उपवाक्य' कहते हैं। उदाहरण के लिए 'श्याम आया और घनश्याम गया।' वाक्य में दो उपवाक्य हैं- 'श्याम आया' 'घनश्याम गया'।

### 2.3.2.2 मिश्र वाक्य (Complex Sentence)

जब एक से अधिक सरल वाक्यों को इस प्रकार जोड़ा जाता है कि उसमें एक प्रधान तथा शेष आश्रित उपवाक्य हों तो इस वाक्यात्मक संरचना को मिश्र वाक्य कहते हैं। 'जो सज्जन होता है, उसका सभी लोग आदर करते हैं।' इस वाक्य में 'उसका सभी लोग आदर करते हैं' प्रधान उपवाक्य है तथा 'जो सज्जन होता है' आश्रित उपवाक्य है। प्रधान उपवाक्य और आश्रित उपवाक्य के बीच परस्पर आश्रय-आश्रित का संबंध होता है। प्रधान उपवाक्य वाक्य का मुख्य कथन होता है और आश्रित उपवाक्य का आश्रय होता है। आश्रित उपवाक्य कुछ समुच्चयबोधक (Conjunction) शब्द (जो, जहाँ, कि, ताकि, who, that, where, which, so that आदि) द्वारा मुख्य उपवाक्य से जुड़े होते हैं, जो अकसर इन उपवाक्यों के शुरू में जुड़ते हैं।<sup>41</sup> इसलिए मिश्र वाक्य में प्रधान उपवाक्य एक ही रहता है, पर आश्रित उपवाक्य एक से अधिक रह सकते हैं। आश्रित उपवाक्य भी संज्ञा उपवाक्य, विशेषण उपवाक्य और क्रिया-विशेषण उपवाक्य में विभाजित है।

---

<sup>40</sup> सिंह, सूरजभान, *अंग्रेजी-हिंदी अनुवाद व्याकरण*, 2003, पृ. 205

<sup>41</sup> शर्मा, राजमणि, *हिंदी भाषा का इतिहास और स्वरूप*, 2010, पृ. 208

### 2.3.2.3 संयुक्त वाक्य (Compound Sentence)

जब एक से अधिक सरल वाक्यों को मिलाकर एक वाक्य बनाया जाता है और इन दो या अधिक वाक्यों को समानाधिकरण समुच्चयबोधक (Conjunction) (जैसे पर, किन्तु, और, लेकिन या, and, but, or आदि) द्वारा जोड़ा जाता है, तो उस वाक्य को संयुक्त वाक्य कहते हैं। संयुक्त वाक्य में एक से अधिक प्रधान उपवाक्य रहते हैं और इन प्रधान उपवाक्यों के साथ इनके आश्रित उपवाक्य भी होते हैं। लेकिन कोई भी वाक्य एक-दूसरे पर आश्रित नहीं होते, इससे यह नहीं समझ लेना चाहिए कि उनमें पारस्परिक संबंध कुछ भी नहीं होता। बल्कि वाक्य में अर्थ संगति की दृष्टि से इस तरह का प्रयोग होता है। 'पिताजी दुकान में हैं और माताजी खाना बना रही हैं' (Father is in the shop and mother is cooking food.) उक्त वाक्य में 'पिताजी दुकान में हैं' और 'माताजी खाना बना रही हैं' दोनों सरल वाक्य हैं जिन्हें समुच्चयबोधक द्वारा जोड़ा गया है। इस दोनों वाक्यों में कोई भी मुख्य या गौण वाक्य नहीं है, दोनों का संबंध समान स्तरीय है। संयुक्त वाक्य एक दूसरे पर आश्रित न होते हुए भी संयुक्त वाक्य के दो उपवाक्य आपस में किसी न किसी प्रकार के संबंध से जुड़े होते हैं।<sup>42</sup>

इन दोनों भाषाओं की संरचना में वाक्य संरचना के आधार और वाक्य के प्रकार से यह ज्ञात होता है, वाक्य की संरचना विभिन्न घटकों से होती है जिसमें संज्ञापद, कारक, सर्वनाम, विशेषण, क्रिया-विशेषण, क्रियापद का स्थान महत्वपूर्ण है। अनुवाद के लिए दो भाषाई संरचना के अध्ययन से सहज अनुवाद किया जा सकता है। इस भाषाई संरचना का उपयोग मशीनी अनुवाद के लिए भी किया जा सकता है। जब किन्हीं दो भाषाओं का अनुवाद करना होता है तो अनुवादक को दोनों भाषाओं का ज्ञान, शब्दावली, मुहावरे और कहावतें तथा व्याकरणिक संरचना का ज्ञान होना चाहिए। उसी तरह मशीनी अनुवाद के लिए भी भाषाई संसाधनों के साथ अन्य उपकरणों का होना आवश्यक है अन्यथा दो भाषाओं के अनुवाद में विभिन्न समस्याएँ पाई जाती हैं। इसलिए यहाँ मशीनी अनुवाद में आनेवाली समस्याओं को स्पष्ट किया गया है।

---

<sup>42</sup> गुरु, कामता प्रसाद, हिंदी व्याकरण, 1920, पृ. 433

## 2.4 अंग्रेजी-हिंदी मशीनी अनुवाद की समस्याएँ

प्राकृतिक भाषा मानव जीवन का अभिन्न अंग है। मानवों के बीच परस्पर संवाद स्थापित करने और सूचनाओं को लिपिबद्ध करके अभिलिखित करने का यह एक प्रमुख साधन है। इसके माध्यम से मानव जटिल, सूक्ष्म तथा व्यापक विचारों को अभिव्यक्त करने में सफल हो जाता है। अगर यह कार्य कंप्यूटर को दिया जाए तो वह कुछ कार्य आसानी से करता है तो कुछ कार्य में ज्यादा गलतियाँ करता है। आज एक भाषा से दूसरी भाषा में अनुवाद करने के लिए कंप्यूटर को दिया जाता है जिसमें मानव बुद्धि की तरह कृत्रिम बुद्धि का समावेश किया गया है। मशीनी अनुवाद तंत्र इसके माध्यम से निर्गत अनुवाद करता है लेकिन दो भाषाओं में होनेवाले अनुवाद में काफ़ी गलतियाँ मिलती हैं, जिसे अनुवाद की दृष्टि से मूल्यांकित करना आवश्यक है।<sup>43</sup>

अंग्रेजी-हिंदी भाषा के विजातीय होने के कारण मशीनी अनुवाद तंत्रों में भाषिक स्तर पर विभिन्न समस्याएँ पायी जाती हैं। इन समस्याओं में भाषा का शब्दक्रम, उपयुक्त पर्याय का चयन, द्विअर्थकता, परसर्ग और पूर्वसर्ग का प्रयोग, लिंग-वचन-काल-वाच्य का संज्ञा तथा क्रिया पर प्रभाव, सर्वनाम प्रयोग, विशेषण प्रयोग, कर्तृवाच्य और कर्मवाच्य का प्रयोग, प्रत्यक्ष और अप्रत्यक्ष कथन, लिप्यंतरण की समस्याएं, आदरार्थ शब्दों का प्रयोग, मुहावरों और कहावतों की समस्याएँ पायी जाती है जिसका विवेचन नीचे किया गया है।

1. शब्दक्रम: दोनों भाषाओं की संरचना भिन्न होने के कारण अक्सर मशीनी अनुवाद में समस्या उत्पन्न होती है। सामान्यतः अंग्रेजी भाषा का शब्दक्रम स्थिर और कर्ता-क्रिया-कर्म (SVO) पर आधारित है और हिंदी भाषा विभक्तिप्रधान होने की वजह से इसका शब्दक्रम अस्थिर और कर्ता-कर्म-क्रिया (SOV) पर आधारित है, इस क्रम संरचना में क्रिया का स्थान वाक्य के अंत में आता है। जैसे- *I learn Hindi (मैं हिंदी सिखता हूँ)* इस शब्दक्रम से भी दो भाषाओं में समस्या पैदा होती है।

---

<sup>43</sup> हाडके, सुमेध, *कॉर्पोरा आधारित मशीनी अनुवाद: हिंदी-अंग्रेजी संज्ञापद के संदर्भ में*, 2016, पृ.120-124

2. संज्ञा: संज्ञा का अंग्रेजी से हिंदी में अनुवाद करते समय कुछ अंग्रेजी व्यक्तिवाचक नाम का मशीनी अनुवाद में शब्दगत अनुवाद किया जाता है या तो उसे लिप्यंतरित किया जाता है। अंग्रेजी में *He/she* का प्रयोग करने पर 'वह, यह और उसको' का प्रयोग होता है तथा अंग्रेजी में *you* के लिए 'तु, तुम' के साथ हिंदी में आदरार्थ शब्द 'आप' का भी प्रयोग होता है। अन्य शब्दों का भी आदरार्थकता के लिए संदर्भ के अनुसार शब्द प्रयोग होता है। उदाहरण के लिए- *His father has come.* 'उनके पिता आए हैं' इस वाक्य में 'उनके' शब्द आदरार्थ रूप में प्रयुक्त हुआ है। अंग्रेजी के सर्वनाम के साथ हिंदी में भी सर्वनाम रूप के साथ परसर्ग का प्रयोग होता है, जैसे- *to me-मुझको, to us-हमको, to him-इसको/उसको, To you-तुझको/तुमको* आदि सर्वनाम का प्रयोग होता है।
3. अंग्रेजी में विशेषक (Modifier) संज्ञा से पहले और बाद, दोनों स्थानों पर प्रयुक्त होते हैं। हिंदी में विशेषक (Modifier) संज्ञा के बाद ही प्रयुक्त होते हैं। अंग्रेजी में *by, for, from, to, of, on, at, in, with* आदि पूर्वसर्गों का प्रयोग संज्ञा के साथ होता है और हिंदी परसर्गों में *ने, से को, पर, में, को, का, की, के लिए* आदि परसर्ग का प्रयोग संज्ञा के बाद किया जाता है। अंग्रेजी में जब पूर्वसर्ग (preposition) का प्रयोग नहीं होता लेकिन हिंदी में परसर्ग (postposition) का प्रयोग होता है तो अक्सर मशीनी अनुवाद के लिए समस्या पैदा होती है। जैसे- *Studies have proved that* 'अध्ययनों ने यह प्रमाणित कर दिया है।
4. अंग्रेजी-हिंदी में दो वचन होते हैं- एकवचन (Singular) और बहुवचन (Plural)। अंग्रेजी में वचन, लिंग का प्रभाव संज्ञा (*boy-boys*), पुरुषवाचक सर्वनाम (*he-they*), निर्देशवाचक सर्वनाम (demonstrative pronoun- *this-these*) और क्रियाओं (*is-are, lives-live*) पर पड़ता है, लेकिन विशेषण पर नहीं। हिंदी में वचन का प्रभाव संज्ञा (*एक लड़का-दो लड़के*), सर्वनाम (*वह-वे, उसे-उन्हें*), क्रिया (*लड़का हँस रहा है - लड़के हँस रहे हैं*), विशेषण (*छोटा कमरा-छोटे कमरे*) आदि पर पड़ता है। जैसे- *लंबा लड़का, लंबे लड़के (the tall boy/boys)* आदि।

5. कुछ शब्द अंग्रेजी में हमेशा बहुवचन रूप में प्रयुक्त होते हैं। लेकिन हिंदी में एकवचन और बहुवचन शब्द के लिए हमेशा अलग रूप होता है, जैसे- 1. The company manufactures scissors (many) 2. The scissors are very sharp. (one or many) लेकिन हिंदी में scissors का अर्थ कैंची और कैंचीयाँ (बहुवचन) होता है। इससे स्पष्ट होता है कि अंग्रेजी में कुछ शब्द के एकवचन में एकल अर्थ निकलता है लेकिन अंग्रेजी बहुवचन में -s प्रत्यय के प्रयोग से हिंदी में कई अर्थ निकलते हैं।
6. अंग्रेजी में तीन लिंग हैं- पुल्लिंग, स्त्रीलिंग और नपुंसक लिंग। जबकि हिंदी में पुल्लिंग और स्त्रीलिंग दो लिंग हैं। हिंदी में लिंग के अंतर्गत प्राणिवाचक (animate) संज्ञाओं का समावेश लड़का (पु.) और लड़की (स्त्री.) में किया जाता है उसी तरह अप्राणिवाचक या अचेतन (inanimate) संज्ञाओं को भी यादृच्छिक (arbitrary) रूप से पुल्लिंग और स्त्रीलिंग वर्गों में ही रखा जाता है। अंग्रेजी में लिंग का प्रभाव केवल अन्य पुरुष (3<sup>rd</sup> person) सर्वनामों पर पड़ता है। जबकि हिंदी में लिंग का प्रभाव विशेषण और क्रिया रूपों पर पड़ता है, लेकिन सर्वनाम पर नहीं। जैसे- *I will meet you.*, मैं (m/f) कल आपसे (m/f) मिलूँगा/मिलूँगी। उसी तरह अंग्रेजी में *child* शब्द का समावेश भी नपुंसक लिंग में होता है तो हिंदी में इसे पुल्लिंग में रखा जाता है।
7. अंग्रेजी में विशेषण पर लिंग-वचन के कारण कोई विशेष परिवर्तन नहीं होता लेकिन हिंदी में विशेषण पर लिंग-वचन-कारक का परिणाम होता जैसे- *Large room/ rooms* 'बड़ा कमरा/बड़े कमरे', *Good boys* 'अच्छे लड़के', *good girls* 'अच्छी लड़कियाँ' दूसरी ओर हिंदी में द्विरुक्ति विशेषण का प्रयोग भी होता है जैसे- 'अच्छे-अच्छे पकवान' के लिए अंग्रेजी में (*good-good food*) का प्रयोग नहीं होता बल्कि (*good food*) का ही प्रयोग होता है।
8. अंग्रेजी में निर्धारक के रूप में *a. an, the* का प्रयोग होता है, लेकिन हिंदी में इसके लिए कोई शब्द प्रयुक्त नहीं होता। हिंदी में कभी कभी संख्यावाचक विशेषणों का प्रयोग संज्ञा के रूप में किया जाता है। जैसे- *a spoon of salt* का 'एक चम्मच नमक' अनुवाद हुआ है। जिसमें *a* निर्धारक के लिए हिंदी में 'एक' का प्रयोग हुआ है।

9. क्रियापदों के अनुवाद की समस्याएँ : अंग्रेजी के अनेक क्रियापद ऐसे हैं जिनका यथातथ्य अनुवाद करना हिंदी में कठिन हो जाता है। अंग्रेजी के दो शब्द प्रायः समानार्थक हैं 'tell' और 'say' किंतु उनकी अर्थछाया में सूक्ष्म अंतर है। 'tell' में व्यक्तिगत रूप से कहने की व्यंजना है और 'say' में सामान्य रूप से कहने का अर्थ प्रकट होता है। जैसे- He told me 'उसने मुझे बताया', he said to me 'उसने मुझसे कहा' अंग्रेजी के कई क्रियापद ऐसे हैं जिनमें हिंदी में निश्चित पर्याय नहीं मिलते जो मशीनी अनुवाद के लिए समस्याएँ पैदा करते हैं।<sup>44</sup>
10. कर्तृवाच्य (active voice) और कर्मवाच्य (passive voice) के अनुवाद में भी समस्या उत्पन्न होती है। कर्मवाच्य का जिस तरह अंग्रेजी और संस्कृत में मुक्त प्रयोग होता है, उस तरह हिंदी में नहीं होता। जैसे- The train has been delayed, का सही अनुवाद 'रेलगाड़ी देर से आ रही है' न कि 'रेलगाड़ी देर कर दी गई है।' दूसरा वाक्य- I am obliged 'मैं आभारी हूँ', न कि- 'मैं आभारी कर दिया गया हूँ' इस तरह वाक्य मशीनी अनुवाद द्वारा प्राप्त होते हैं जिससे अनुवाद की समस्या उत्पन्न होती है।<sup>45</sup>
11. प्रत्यक्ष और अप्रत्यक्ष कथन (Direct and Indirect speech) अंग्रेजी में प्रत्यक्ष कथन को अप्रत्यक्ष कथन में बदलते समय कई व्याकरणिक परिवर्तन पाए जाते हैं, जैसे अप्रत्यक्ष कथन में उद्धरण चिन्ह (") का हट जाना, I का she हो जाना, will का would हो जाना और this का that हो जाना आदि। हिंदी में प्रत्यक्ष कथन और अप्रत्यक्ष कथन के बीच बहुत कम संरचनात्मक अंतर होता है। 1. She said, I will not marry this year (प्रत्यक्ष कथन), 2. She said that she would not marry that year (अप्रत्यक्ष कथन) आदि। हिंदी के अप्रत्यक्ष कथन में उद्धरण चिन्ह हट जाता है और दो उपवाक्यों के बीच 'कि' का प्रयोग किया जाता है, जो वैकल्पिक रूप से लुप्त भी हो सकता है। जैसे कि उसने कहा, 'मैं इस साल शादी नहीं करूंगी।' और अप्रत्यक्ष कथन में (उसने कहा कि मैं इस साल शादी नहीं करूंगी।)<sup>46</sup>

<sup>44</sup> डॉ.नगेन्द्र, *अनुवाद विज्ञान*, 1993, पृ.159

<sup>45</sup> वही. पृ. 159

<sup>46</sup> वही, पृ. 160

12. नाम के शीर्षकों की पहचान करना : अक्सर मशीनी अनुवाद में नाम के शीर्षक के साथ समस्याएँ आती हैं। जैसे- प्रो, श्रीमती, श्री' लगाते समय, इस अनुवाद में लिंग की पहचान करना मशीनी अनुवाद के लिए मुश्किल होता है, विशेषतः व्यक्ति वाचक नाम में। उपनाम (surname) की पहचान करना जैसे- अग्रवाल (Agarwal), ओबोरॉय (oberoy) इस तरह के अन्य नामों में भी मशीनी अनुवाद की समस्या पायी जाती है। इसके लिए अनुवाद से अच्छा लिप्यंतरण करना आसान होगा।<sup>47</sup>
13. संक्षिप्त रूपों और संक्षिप्तियों के अनुवाद की समस्या : प्रत्येक भाषा में उच्चारण और प्रयोग की सुविधा के लिए संक्षिप्त रूप और संक्षिप्तियां बना ली जाती हैं और उनका उस भाषा में प्रयोग किया जाता है, जिसे उस भाषा के भाषा-भाषी समझते, बोलते और प्रयोग करते हैं। दूसरी भाषा में इनकी स्थिति भिन्न प्रकार की होती है।
14. उच्चारण की समस्या : कई भाषाएँ ऐसी हैं, जिनके शब्दों के उच्चारण में स्वर के आरोही-अवरोही तथा बलाघात से शब्दों के अर्थ बदल जाते हैं। ऐसे शब्दों का जब लिखित रूप सामने आता है, तो उसमें अर्थ की भिन्नता स्पष्ट नहीं होती। यह अनुवाद अगर मशीन से प्राप्त होगा तो उसमें गलतियाँ होंगी, जिनके समाधान के लिए कोई आंकिक प्रणाली निर्दिष्ट करनी होगी और उसी के आधार पर मशीनी अनुवाद संभव हो सकेगा।
15. व्याकरणिक समस्याएँ : मशीनी अनुवाद के लिए जिस प्रकार के व्याकरणिक मॉडल की आवश्यकता होती है वह स्वरूप परंपरागत व्याकरणिक मॉडल से भिन्न होते हैं। जिससे मशीनी अनुवाद के प्रकारों के अनुसार ही मशीनी अनुवाद के उपयोग में व्याकरण के प्रकार भी अलग-अलग दिखाई देते हैं। पाणिनी व्याकरण, प्रकार्यात्मक एकीकरण व्याकरण (Functional Unification Grammar), शाब्दिक प्रकार्यात्मक व्याकरण (Lexical Functional grammar), वृक्ष संलग्नक व्याकरण- अरविंद जोशी (Tree adjoining Grammar 1992) आदि व्याकरण की विभिन्न समस्याओं की इन कोटियों में भाषिक और व्याकरणिक समस्याएँ सबसे अधिक जटिल होती हैं।<sup>48</sup>

<sup>47</sup> Goyal, Vishal, *Development of a Hindi to Punjabi Machine Translation System*, PhD Thesis, 2010, p. 653

<sup>48</sup> Jha, Girish Nath, "Computational Linguistics and Sanskrit" 4th International Symposium, New Delhi, India, December 10-12, 2010

व्याकरणिक अशुद्धता (Incorrectness) में वर्तनी त्रुटि, शब्द व्युत्क्रम, अशुद्ध वाक्य रचना, अपूर्ण वाक्य और गलत विरामादि चिह्न होने पर भी मनुष्य पाठ को प्रायः समझ लेता है। लेकिन कंप्यूटर इन अशुद्धियों को कुछ हद तक ही समझने की क्षमता रखता है। इसलिए मशीन के लिए यह एक समस्या है।<sup>49</sup>

16. भाषिक संसाधनों के उपयोग और निर्माण की समस्याएँ : किसी एक मशीनी अनुवाद तंत्र निर्माण में लाए गए सभी भाषिक साधन (Language Tools) किसी दूसरी भाषा के मशीनी अनुवाद तंत्र निर्माण के उपयोग में आएंगे यह निश्चित नहीं कहा जा सकता है। भाषिक संसाधनों की यह समस्या भाषाओं की असमान संरचना के कारण उत्पन्न होती है। इसलिए भाषिक संसाधनों का निर्माण कार्य देश-विदेश में तेजी से हो रहा है ताकि मशीनी अनुवाद तंत्र के विकास से मानव का समय और श्रम की बचत की जा सके।

17. संदिग्धार्थकता (ambiguity): किसी भी भाषा के अनुवाद के समय प्रत्येक भाषा के वाक्य या शब्द संरचना में संदिग्धताएँ उत्पन्न होती हैं जिसका समाधान करना आवश्यक है। वाक्यगत द्विअर्थकता में जब वाक्य में शब्द ही इस तरह प्रयुक्त किए जाते हैं जिससे संपूर्ण वाक्य ही द्विअर्थक प्रतीत होता है। जैसे- *Time flies like an arrow.* 'समय जल्दी बीत जाता है' अनुवाद होगा लेकिन उसका दूसरा अर्थ होगा 'समय तीर की भाँति बीत जाता है।' अंग्रेजी में व्यंजनात्मक शब्द का प्रयोग होने के चलते हिंदी के अनुवाद में संदिग्धता उत्पन्न हुई है जिसके लिए द्विभाषिक कोश से पर्याय चयन की समस्या का हल निकाला जा सकता है। अर्थगत संदिग्धता में जैसे- *He went to the bank* इसमें वाक्य का अर्थ स्पष्ट नहीं हो पा रहा है कि *river bank* या *money bank* जिसके हिंदी अनुवाद में भी समस्या आएगी। इसके साथ शब्दगत द्विअर्थकता भी वाक्य में अक्सर पाई जाती है। जैसे- *solution* शब्द है, जिसका अर्थ

---

<sup>49</sup> विकास, ओम, 'मशीनी अनुवाद की समस्याएँ', अनुवाद विज्ञान सिद्धांत और अनुप्रयोग-संपा. डॉ.नगेन्द्र, 1993, पृ. 339

1. घोल, विलयन 2. हल, उत्तर, समाधान आदि, यहाँ समस्या पर्याय चयन की भी आ रही है जिसे मशीन कैसे हल करेगी यह समस्या है।<sup>50</sup>

18. लिप्यंतरण की समस्या: वाक्य के प्रत्येक घटक कि भाषा की प्रकृति के अनुकूल अनुवाद किया जाता है। परंतु भाषा में कुछ शब्द ऐसे होते हैं, जिनके अनुवाद से स्रोतभाषा पाठ में संचित अर्थ का खंडन होता है; जैसे- सांस्कृतिक शब्द, नामवाची शब्द आदि। इसलिए आवश्यकता होती है कि ऐसे शब्दों का लिप्यंतरण किया जाए। उदाहरण स्वरूप सिंदूर, बुरका, जनेऊ, गंगाजल, चरणामृत आदि ऐसे कई शब्द हैं, जिनका अनुवाद करने से इन शब्दों का वास्तविक अर्थ-सौंदर्य नष्ट हो जाएगा। नामवाची सभी शब्दों का अनुवाद नहीं किया जा सकता; जैसे- *राजा, कमल, रानी, मुस्कान* आदि नामों का अनुवाद करने से इनके अर्थ का अनर्थ हो जाता है। ऐसे में लिप्यंतरण की आवश्यकता होती है।<sup>51</sup>

19. मुहावरों और कहावतों की समस्या : मुहावरे और कहावतें किसी भी देश की संस्कृति-व्यवहार, रीति-रिवाज, रहन-सहन आदि से संबंधित होती हैं और यह प्रत्येक भाषा के अभिन्न अंग होते हैं। इन भिन्नताओं के चलते मशीन से अनुवाद करना प्रायः संभव नहीं होता, यहाँ तक कि मानव अनुवादक को भी अनुवाद करने में बहुत कठिनाई आती है। इन समस्याओं को दूर करने के लिए द्विभाषी कहावतें और मुहावरा कोश निर्मित करना चाहिए। मशीनी अनुवाद के शब्दकोश को अधिक विस्तृत बनाया जाना चाहिए। Give a man the air. इस वाक्य की अभिव्यक्ति मुहावरेदार है। इस संपूर्ण अभिव्यक्ति का अर्थ है, 'किसी को चले जाने के लिए कहना'। किसी को कहने कि आवश्यकता नहीं है, कि 'वह चला जाए या हवा जाए।' इस प्रकार के मुहावरेदार प्रयोगों को भी प्रश्नोत्तर रूटीनों द्वारा बांटा जाता है।<sup>52</sup>

50

[http://shodhganga.inflibnet.ac.in/bitstream/10603/4239/10/10\\_chapter%203.pdf](http://shodhganga.inflibnet.ac.in/bitstream/10603/4239/10/10_chapter%203.pdf)  
(retrieved on 27.04.2017)

<sup>51</sup> Goyal, Vishal, *Development of a Hindi to Punjabi Machine Translation System*, PhD Thesis, 2010, p.653

<sup>52</sup> कुमार, नरेश, 'मशीनी अनुवाद', अनुवाद सिद्धांत और समस्याएँ- संपा.रवीन्द्रनाथ श्रीवास्तव, कृष्णकुमार गोस्वामी, 2010, पृ.137-138

20. द्विभाषी कोश का निर्माण : मशीनी अनुवाद के लिए द्विभाषी कोश का विकास होना जरूरी है जिसे कंप्यूटर के डाटाबेस के माध्यम से आसानी से तैयार किया जा सकता है। जिससे किसी भी शब्द को देखना तथा अनुवाद करना आसान हो जाता है। अभी इस दिशा में बहुत प्रयास करने की आवश्यकता है जिसके लिए भारत की अधिकांश भाषाओं के द्विभाषी कोश अभी तैयार करने हैं। इस दृष्टि से अंग्रेजी, रूसी, जर्मनी, फ्रेंच, चीनी और हिंदी में कुछ कार्य अवश्य हुआ हैं, परंतु अन्य भारतीय भाषाओं में यह कार्य करना अत्यंत आवश्यक है।
21. उपयुक्त शब्द पर्याय का चयन : संगणकिय कोश में शब्दों के अनेक अर्थ का समावेश होता है। जिससे शब्द के पर्याय चयन में मशीन के लिए समस्या आती है। क्योंकि कंप्यूटर के डाटाबेस में एक शब्द के अनेक पर्यायों का समावेश किया होता है जिसमें एक भाषा के शब्द के लिए लक्ष्य भाषा में अनेक शब्द उपलब्ध होते हैं। इसमें मशीन के लिए उनमें से सही शब्द का चुनाव करना चुनौती होती है।
22. सार्वभौमिक व्याकरण का अभाव : प्रो. नोम चॉम्स्की द्वारा अनेक व्याकरणों का निर्माण किया गया परंतु आज भी मशीनी अनुवाद के लिए आवश्यक सार्वभौमिक अर्थात् सभी भाषाओं पर आधारित नियमों से संबद्ध एक व्याकरण का निर्माण नहीं हो पाया है। इसके चलते मशीनी अनुवाद के लिए आवश्यक अनुप्रयोगों के विकास में कठिनाई आ रही है।<sup>53</sup>
23. भाषा वैज्ञानिकों का अभाव : भाषा के नियमों के निर्माण के लिए मुख्यतः भाषा वैज्ञानिकों की आवश्यकता होती है। मशीनी अनुवाद के क्षेत्र में काम करने के लिए कंप्यूटर वैज्ञानिक का अधिक इस्तेमाल होता है जिससे भाषा की संरचना तथा भाषा के व्याकरण में कमियां रह जाती है और मशीन इससे अनुवाद में अक्सर गलतियाँ करती है। इसलिए दोनों कंप्यूटर वैज्ञानिक और भाषा वैज्ञानिक के सहयोग से बेहतर मशीनी अनुवाद तंत्रों का निर्माण किया जा सकता है।

---

<sup>53</sup> शर्मा, राकेश, *मशीनी अनुवाद: स्थिति एवं संभावनाएँ*, अनुवाद अंक:158, जनवरी-मार्च 2014, पृ.70

24. कृत्रिम बुद्धि (Artificial Intelligence) का विकास : इसके विकास से मशीन को मानव मस्तिष्क की तरह विकसित करने का कार्य चल रहा है जिसमें अनेक कंप्यूटर वैज्ञानिक तथा भाषा वैज्ञानिक कार्य कर रहे हैं। इसके विकास से कंप्यूटर को मानव की तरह ही सृजनशील बनाया जाएगा और उससे अनेक कार्य आसानी से किए जाएंगे। केवल मशीनी अनुवाद ही करना इसका उद्देश्य नहीं होगा तो उसमें रोबोटिक्स प्रणाली, मशीनरी का संचालन करना, चिकित्सा पध्दती में इस कृत्रिम बुद्धि का प्रयोग कर कार्य संपादन करना आदि कार्य इसके माध्यम से किए जाएंगे।

उपरोक्त मशीनी अनुवाद की समस्याओं के विवेचन से स्पष्ट होता है कि किसी भी एक भाषा से दूसरे भाषा के अनुवाद में मशीन द्वारा समस्याएँ निर्माण होती हैं। जिनपर विशेष ध्यान देने की आवश्यकता है। आज विश्व के प्रत्येक भाषा क्षेत्र (domain) में मशीन अनुवाद के विकास हेतु निरंतर शोध कार्य चल रहे हैं। इस संबंध में नए-नए सॉफ्टवेयर का निर्माण भी हुआ है और पुराने सॉफ्टवेयर को लगातार सुधारित किया जा रहा है। कृत्रिम बुद्धि के विकास द्वारा मशीन को अनुवाद के उद्देश्य से अधिक विकसित किया जा रहा है फिर भी इसमें आनेवाली समस्या का समाधान के लिए भाषा वैज्ञानिक तथा कंप्यूटर वैज्ञानिक संयुक्त रूप से प्रयास कर रहे हैं।

भारतीय भाषाओं में मशीनी अनुवाद तंत्रों के निर्माण में आनेवाली समस्याओं में भाषिक संरचना, वाक्यगत भाषिक अस्पष्टता, संदिग्धता, अलंकारो एवं व्यंग की अस्पष्टता, विकसित नए शब्दों के अनुवाद की समस्या, शब्द के अपूर्णता की समस्या, लिप्यंतरण की समस्या आदि हैं। इन समस्याओं के समाधान के लिए हमें भाषिक संसाधनों को विकसित करने की आवश्यकता है। साथ ही समानांतर कॉर्पोरा, द्विभाषी संगणकीय कोश, व्याकरणिक संरचना, व्याकरण जाँचक, रूपवैज्ञानिक विश्लेषक, प्रजनक, वर्डनेट का विकास तथा प्रोग्राम या अल्गोरिदम आदि संसाधनों तथा उपकरणों का विकास मशीनी अनुवाद तंत्रों के निर्माण के लिए करने की आवश्यकता है।

## तृतीय अध्याय

### शोध का उद्देश्य और प्रविधि

3.1 शोध से संबंधित पूर्ववर्ती शोध कार्य

3.2 शोध का उद्देश्य

3.3 शोध के लिए स्रोत सामग्री का प्रयोग

3.4 शोध प्रविधि

## तृतीय अध्याय

### शोध का उद्देश्य और प्रविधि

#### (RESEARCH OBJECTIVE AND METHODOLOGY)

---

अनुवाद एक भाषा के लिखित एवं वाचिक पाठ को अन्य भाषा के लिखित या वाचिक (readable) पाठ में रूपांतरित करने की प्रक्रिया है। जिसे मानव अनुवादक स्रोत और लक्ष्य भाषा की संरचना, अभिव्यक्ति, संस्कृति और व्याकरणिक नियमों द्वारा अनुवाद करता है। ऐसे में मशीन का प्रयोग करना अनुवाद के क्षेत्र में नई पहल कहा जा सकता है। मानव को अनुवाद में सहायता करने के फलस्वरूप मशीनी अनुवाद की इस नई संकल्पना ने संपूर्ण सूचना और संप्रेषण जगत में अभूतपूर्व क्रांति लाई है। जिसका प्रयोग आज अनेक भाषिक अनुप्रयोगों के लिए मुख्य रूप से हो रहा है।

मशीनी अनुवाद स्वचालित अनुवाद की प्रक्रिया है जिसे कंप्यूटर द्वारा किया जाता है। इसमें स्रोत भाषा के किसी भी वाक्य या पाठ का अनुवाद लक्ष्य भाषा में होता है। यह प्रक्रिया कंप्यूटर में संचालित किए गए कार्पस, शब्दकोश, प्रोग्राम और कृत्रिम बुद्धि आदि संचित किए गए भाषिक उपकरणों की सहायता से होती है। इससे पाया जानेवाले अनुवाद में अक्सर त्रुटियाँ होती हैं। क्योंकि, मानव अपनी मस्तिष्क क्षमता और उससे जुड़े हुए संदर्भ जानकारी के आधार पर भाषा की अभिव्यक्तियों (Expressions) का सही निर्वचन (Interpretation) कर उनके सही अर्थों का ग्रहण करता है। उसके पास भाषा की समझ और सांसारिक ज्ञान (World Knowledge) पहले से निहित होता है, लेकिन कंप्यूटर केवल शब्द का वही अर्थ ग्रहण करता है, जो उसकी स्मृति (Memory) में संचित रहता है और उसे प्रोग्रामिंग द्वारा संचालित किया जाता है।

मशीनी अनुवाद का विकास 'संगणकीय भाषाविज्ञान' (Computational Linguistics) के अनुप्रयोगिक (Applied) क्षेत्र के रूप में हुआ है। संगणकीय भाषाविज्ञान

यह भाषाविज्ञान विषय की अंतर-विषयी (Interdisciplinary) शाखाएँ हैं। जिसमें पिछले चार दशकों से प्राकृतिक भाषा संसाधन और मशीनी अनुवाद के क्षेत्र में शोध एवं अध्ययन हो रहे हैं। प्राकृतिक भाषा संसाधन (Natural Language Processing) का प्रयोग आधुनिक रूप में भाषाओं के संगणकीय अध्ययन के लिए किया जाता है। जिसमें पाठ या ध्वनि के शुद्धीकरण से लेकर शब्द रूप-रचना विश्लेषण (Morphological analysis), मूल शब्द निष्कर्षण (Stemming), शब्द-भेद वर्गीकरण (Parts of speech tagging), पद-भेद वर्गीकरण (chunking) व वाक्यात्मक पदबंध वर्गीकरण (Parsing) आदि भाषा संसाधन के कार्यों का समावेश किया जाता है।<sup>1</sup> प्राकृतिक भाषा संसाधन के प्रमुख कार्यों में विभिन्न प्रकार के कॉर्पोरा की आवश्यकता होती है जिससे किसी भी भाषा के लिए मशीनी अनुवाद तंत्रों का विकास करना आसान होता है। कॉर्पोरा से तात्पर्य डिजिटल रूप में उपलब्ध भाषाई पाठ के संग्रह से है। कॉर्पोरा का संग्रह भाषा के अनुसार अलग-अलग प्रकार से किया जाता है।

आज किसी भी भाषा का अध्ययन करना कंप्यूटर के माध्यम से सरल हुआ है। कारण विभिन्न भाषाओं के लिए कॉर्पोरा का निर्माण किया जा रहा है। यह कॉर्पोरा किसी भी भाषा के प्रायोगिक स्वरूप का संचयन तथा संकलन होता है जिसका प्रयोग किसी भाषा के अध्ययन, विश्लेषण और अनुवाद के लिए किया जाता है। इसका प्रयोग मशीनी अनुवाद के संदर्भ में करने से मशीनी अनुवाद आधारित वाक्यों का मूल्यांकन एवं त्रुटियों के विश्लेषण के लिए भी किया जाता है।

दो भाषाओं में मशीनी अनुवाद से होनेवाली गलतियों के कारण विभिन्न विद्वानों, भाषावैज्ञानिक तथा कंप्यूटर विज्ञानियों ने इस पर गहन अध्ययन किया है। उसमें

---

<sup>1</sup> दीक्षित, नेहा, *वाक्यात्मक परिप्रेक्ष्य में हिंदी क्रिया वर्ग अभिज्ञानक*, पीएच. डी. शोध-प्रबंध, 2014, पृ.14-15

प्रमुखतः भाषिक संरचना में आनेवाली भिन्नताओं का अध्ययन-विश्लेषण बोनी जे. डोर ने अपने कई शोध कार्यों में किया है।

बोनी जे. डोर (Bonnie J. Dorr) के अनुसार “जब स्रोत भाषा के समातुल्य वाक्य संरचना का अनुवाद लक्ष्य भाषा की समान वाक्य संरचना में नहीं होता है तब भिन्नता निर्माण होती है।” [Divergence occurs when structurally similar sentences of the source language do not translate into sentences that are similar in structures in the target language.]<sup>2</sup> डोर ने यूरोपीय भाषाओं के अनुवाद में आनेवाली भिन्नता को देखते हुए शाब्दिक और अर्थगत भिन्नता का 7 प्रकारों में वर्गीकरण किया है। उसमें संरचनात्मक (structural), समसामयिक (conflational), श्रेणीगत (categorical), स्थानापन्न संबंधी (promotional), निम्न स्थानापन्न संबंधी (demotional), विषयक (thematic) और शाब्दिक (lexical) भिन्नताओं को स्पष्ट किया गया है।

डोर के वर्गीकरण को आधार मानते हुए अंग्रेजी से हिंदी भाषाओं में पायी जानेवाली भिन्नताओं का वर्गीकरण सचिन दवे और दिपा गुप्ता ने भी अपने शोध कार्यों में किया है। जिसमें दिपा गुप्ता<sup>3</sup> ने संरचनात्मक, श्रेणीगत, नामिक, सर्वनाम-संबंधी और निम्न स्थानापन्न संबंधी भिन्नता को स्पष्ट किया है। इन भिन्नताओं के कारण मशीनी अनुवाद में शब्दक्रम, वाक्यसंरचना, द्विरुक्ति (reduplication), नाम प्रतीति (named entities), संदिग्धार्थक शब्द (ambiguous words) और विभक्ति प्रत्यय (inflections) आदि में समस्याएँ पायी जाती हैं।

प्रस्तुत शोध-प्रबंध में मशीनी अनुवाद मूल्यांकन के लिए प्रमुखतः अंग्रेजी से हिंदी दो भाषाओं के कॉर्पोरा के वाक्यों का गहन अध्ययन द्वितीय अध्याय में किया गया है।

---

<sup>2</sup> Dorr Bonnie J., *Machine Translation Divergences: A Formal Description and Proposed Solution*, 1994, p. 597-616

<sup>3</sup> Gupta, Deepa, *Contributions to English to Hindi Machine Translation Using Example Based Approach*, PhD Thesis, 2005, p.96-131

मशीनी अनुवाद में कॉर्पोरा के वाक्यों का प्रयोग करते समय दोनों भाषाओं की प्रकृति का गहन अध्ययन-विश्लेषण करना आवश्यक है। क्योंकि दोनों भाषाओं की संरचना में काफ़ी भिन्नता पायी जाती है। जैसे- अंग्रेजी भाषा की वाक्य संरचना कर्ता-क्रिया-कर्म (SVO) पर आधारित है। जबकि भारतीय भाषाओं की वाक्यसंरचना शब्दक्रम मुक्त (free word order) और कर्ता-कर्म-क्रिया (SOV) पर आधारित है। यह दो भाषाओं की भिन्नता अनुवाद में संज्ञा, सर्वनाम, विशेषण, क्रिया-विशेषण, क्रिया तथा अन्य व्याकरणिक घटकों के स्तर पर पायी जाती है। इन भिन्नताओं के चलते मशीनी अनुवाद में अक्सर त्रुटियाँ होती हैं। जिनमें वर्तनी की त्रुटियाँ, उच्चारण की त्रुटियाँ, संक्षिप्तीकरण, क्रिया रूप-रचना, लिंग की रूप-रचना, अतिरिक्त शब्द, अनअनूदित शब्द, लिप्यंतरण, शाब्दिक अस्पष्टता, अर्थगत त्रुटियाँ, कहावतों एवं मुहावरों की त्रुटियाँ, शब्दक्रम, पूर्वसर्ग तथा परसर्ग, सरल, संयुक्त और मिश्रित वाक्य में त्रुटियाँ अनुवाद के बाद पायी जाती हैं। इसलिए मशीनी अनुवाद तंत्र से अनूदित वाक्यों का मूल्यांकन करना चुनौतीपूर्ण कार्य माना जाता है।

मशीनी अनुवाद की गुणवत्ता का मूल्यांकन करने के लिए विभिन्न पद्धतियों का प्रयोग किया जाता है जिससे मशीनी अनुवाद तंत्र से पाए जानेवाले अनुवाद के गुण-दोष की जाँच की जा सकती है। अगर अनुवाद में कई गलतियाँ होती हैं तो उसे परिष्कृत करने के लिए विभिन्न भाषाई सामग्री तथा उपकरणों का निर्माण किया जाता है। इन तंत्रों के विकास एवं निर्माण के लिए कंप्यूटर वैज्ञानिक और भाषा वैज्ञानिक दोनों की आवश्यकता होती है। किसी भी मशीनी अनुवाद तंत्र का विकास विकासकर्ता (Developer) पर निर्भर कर करता है कि तंत्र का विकास किस उद्देश्य के लिए किया जा रहा है तथा उसके व्यवहार-क्षेत्र कौनसे हैं और किन प्रयोजनों के लिए उसका प्रयोग किया जानेवाला है। यह किसी भी मशीनी अनुवाद तंत्र के विकास एवं निर्माण करते समय ध्यान रखना आवश्यक होता है। क्योंकि यदि तंत्र का विकास नियम आधारित दृष्टिकोण पर किया जाता है तो उसमें दो भाषाओं की वाक्य संरचना के अनुसार समानांतर नियमों का समावेश करना होता है। यदि वह उदाहरण आधारित दृष्टिकोण पर है तो उसकी रूपरेखा

दूसरी तरह की होगी। साथ ही मशीनी अनुवाद में अर्थ संरचना की दृष्टि से समृद्ध कोश और शब्दतंत्र (Wordnet) जैसे समान कोशगत संसाधनों और उपकरणों का भी विकास करना होगा। ताकि वाक्य स्तर पर शब्द में आनेवाली संदिग्धता तथा अनेकार्थकता की समस्या को शब्द के उचित पर्याय चयन से दूर किया जा सकता है। अतः यह शोध प्रमुख रूप से मशीनी अनुवाद तंत्रों के मूल्यांकन और त्रुटि विश्लेषण पर आधारित होने की वजह से इस अध्याय में शोध से संबंधित पूर्ववर्ती शोध कार्य, शोध का उद्देश्य एवं प्रविधि का विवेचन प्रमुख रूप से किया गया है।

### 3.1 शोध से संबंधित पूर्ववर्ती शोध कार्य

सन 1980 के बाद भारतीय भाषाओं में मशीनी अनुवाद तंत्रों के निर्माण के लिए प्रयास शुरू हुए थे। उसके बाद विश्व में तथा भारतीय भाषाओं में विभिन्न मशीनी अनुवाद तंत्रों का निर्माण हुआ है। इन तंत्रों के निर्माण से मानव के समय, श्रम और पैसे की बचत हो रही है तथा इससे मानव को तत्काल अनुवाद करने के लिए अधिक सहायता मिली है। आज भारतीय भाषाओं के लिए इनमें आंग्लभारती, अनुसारक, अनुभारती, आंग्ल-हिंदी, मंत्र-राजभाषा, मात्रा, यू.सी.एस.जी.मैट, शक्ति, तेलुगु-तमिल, पंजाबी-हिंदी, अंग्रेजी-कन्नड़, तमिल-हिंदी, संपर्क, बांग्ला-हिंदी, अनुवाद, हिंग्लीश आदि तंत्रों का निर्माण हुआ है। इसमें से अधिकांश तंत्र अंग्रेजी से हिंदी में सरकारी दस्तावेजों, रिपोर्ट, समाचार, कहानी तथा अन्य पाठ (Texts) का अनुवाद अनेक विषय-क्षेत्रों (Domain) में करते हैं। लेकिन मशीन से प्राप्त अनुवाद में काफ़ी गलतियाँ पायी जाती हैं। मशीन अनुवाद से यह उम्मीद की जाती है कि लक्ष्य भाषा में प्राप्त अनुवाद बोधगम्य, सही, स्वीकार्य और गुणवत्ता आधारित हो जिससे मूल भाषा की तरह लक्ष्य भाषा पाठ में अर्थ स्पष्ट हो। इसलिए अनुवाद की गुणवत्ता जाँचने के लिए अनेक मूल्यांकन पद्धतियों का भी निर्माण किया गया है। जिसका मुख्य उद्देश्य मशीन द्वारा अनूदित आउटपुट, मानव अनुवाद की तुलना में कितना सही अनुवाद करते हैं, इससे ज्ञात होता है। और प्राप्त अनुवाद के गुण-

दोषों का विवेचन त्रुटि विश्लेषण में किया जाता है। जिससे मशीनी अनुवाद तंत्रों में सुधार करने तथा परिष्कृत करने के लिए सहायता मिलती है।

मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए सर्वप्रथम 1956 में मिलर और बीबेर-सेंटर (Miller and Beeber-center) तथा पफ्फिन (Pfaffine) ने मानव द्वारा मशीनी अनुवाद तंत्र के मूल्यांकन के लिए सुझाव दिए थे। लेकिन मानव द्वारा मूल्यांकन कार्य काफी श्रमसाध्य और अधिक समय लेने वाला कार्य था। इसलिए स्वचालित मशीनी अनुवाद तंत्र के मूल्यांकन के लिए विविध पद्धतियों का विकास किया गया। पफिनेनी (2001) ने ब्लू पद्धति (BLUE), आई.बी.एम मेट्रीक द्वारा (BLUE), डीएआरएपी द्वारा निस्ट (NIST), लावी और डेनकोवस्की द्वारा मेटेयर (METEOR) पद्धतियों को प्रस्तावित किया गया है जिसका प्रयोग सभी मशीनी अनुवाद तंत्र के विकासकर्ता एवं शोधकर्ता मूल्यांकन के लिए करते हैं।<sup>4</sup> इस शोध कार्य में भी इन पद्धतियों का अध्ययन किया है तथा पूर्ववर्ती शोध कार्य के लिए भारतीय भाषाओं में विकसित मशीनी अनुवाद तंत्रों के अध्ययन और विवेचन को प्रथम अध्याय में दिया गया है। साथ ही मशीनी अनुवाद तंत्रों के मूल्यांकन पर किए गए शोध कार्य तथा शोध पत्र का नीचे विवरण दिया गया है।

- गणितशास्त्र विभाग, आई.आई.टी., (IIT) नई दिल्ली, में दीपा गुप्ता ने पीएच.डी. का शोध प्रबंध 'अंग्रेजी से हिंदी में मशीनी अनुवाद के लिए उदाहरण आधारित दृष्टिकोण का योगदान' (Contributions to English to Hindi Machine Translation Using Example-Based Approach, 2005) पर अध्ययन किया है। इस शोध को डॉ. नीलाद्रि चटर्जी ने निर्देशित किया है। इस शोध में अंग्रेजी से हिंदी में मशीनी अनुवाद के लिए उदाहरण आधारित मशीनी अनुवाद तंत्र के विकास पर अध्ययन किया है। साथ ही

---

<sup>4</sup> Kalyani, Aditi, Sajja, Priti, S., *A Review of Machine Translation Systems in India and different Translation Evaluation Methodologies*, 2015, p.16-23

इन दो भाषाओं के अनुवाद में आनेवाली भिन्नता का अध्ययन भी इस शोध में किया गया है।

- आर. अनंतकृष्णन, पुष्पक भट्टाचार्य, एम. ससीकुमार और रितेश एम. शाह द्वारा 2007 में लिखित शोध पत्र 'अंग्रेजी-हिंदी मशीनी अनुवाद तंत्र के स्वचालित मूल्यांकन में कुछ मुद्दे' (Some Issues in Automatic Evaluation of English-Hindi MT: More Blues for BLEU) में उन्होंने ब्लू (BLUE) की मूल्यांकन पद्धति पर विस्तृत चर्चा की है। साथ ही मानव द्वारा मूल्यांकन के लिए पैरामीटर बताएँ हैं, जिससे मशीनी अनुवाद तंत्र का मूल्यांकन सही रूप में किया जा सके।
- विशाल गोयल और गुरप्रीत सिंह लीहल (पंजाबी विश्वविद्यालय, पटियाला) ने 2009 में अपने शोध पत्र 'हिंदी से पंजाबी मशीनी अनुवाद तंत्रों का मूल्यांकन' (Evaluation of Hindi to Punjabi Machine Translation System) में हिंदी से पंजाबी भाषा में मशीनी अनुवाद तंत्र से आउटपुट प्राप्त करने के बाद और मानव आधारित मूल्यांकन करने के बाद यह पाया कि ये तंत्र 95 प्रतिशत सटीक अनुवाद करने के लिए सक्षम हैं। इसी के साथ विशाल गोयल ने अपने पीएच.डी. के शोध कार्य में 'हिंदी से पंजाबी मशीनी अनुवाद तंत्र' का विकास किया है।
- नकुल शर्मा ने 2011 में मास्टर ऑफ इंजीनियरिंग के शोध-प्रबंध विषय 'अंग्रेजी से हिंदी सांख्यिकीय मशीन अनुवाद प्रणाली' (English to Hindi Statistical Machine Translation System) विषय पर थापर विश्वविद्यालय, पटियाला में शोध कार्य किया है। उसमें मशीनी अनुवाद के दृष्टिकोण तथा पद्धतियों में नियम आधारित, कॉर्पस आधारित, ज्ञान आधारित और सांख्यिकीय आधारित मशीनी अनुवाद पद्धति का विश्लेषण किया गया है। साथ ही अनूदित वाक्य के मूल्यांकन के लिए (1-4) रैंकिंग का प्रयोग किया गया है।

- नीरज तोमर, दीपा सिन्हा, पीयूष कांत राय ने 2012 में 'अंग्रेजी से हिंदी भाषा में मशीनी अनुवाद के लिए एफ़ मेट्रीक परिमाण' (F-Measure Metric for English to Hindi Language Machine Translation), विषयक शोध पत्र में अनुवादक्ष मशीनी अनुवाद तंत्र में पर्यटन क्षेत्र के कॉर्पोरा का प्रयोग कर उस तंत्र का मूल्यांकन एफ़-मेट्रीक परिमाण से किया गया है। साथ ही इससे प्राप्त मूल्यांकन के परिणाम की मानव मूल्यांकन के साथ तुलना की गयी। अंत में पाया गया कि दोनों की मूल्यांकन पद्धति में अधिक समानता है। इसी के साथ लेखक ने अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए ब्लू, निस्ट आदि पद्धति का भी प्रयोग मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए किया है। नीरज तोमर ने 'Evaluating Machine Translation (MT) Evaluation Metrics for English to Indian Language Machine Translation' इस विषय पर पी.एच.डी भी वनस्थली विद्यापीठ से की है।
- निशीध जोशी, इति माथुर (वनस्थली विद्यापीठ, राजस्थान) और हेमंत दरबारी (सी-डैक, पुणे, महाराष्ट्र) ने 'अंग्रेजी-हिंदी मशीनी अनुवाद का मानव और स्वचालित मूल्यांकन' (Human and Automatic Evaluation of English-Hindi Machine Translation) विषय पर 2013 में प्रकाशित अपने शोध पत्र में मानव आधारित और स्वचालित मशीनी अनुवाद मूल्यांकन की ब्लू (BLUE) तथा मेटेयोर (METEOR) प्रणाली की तुलना की गई है। अपने दूसरे शोध पत्र 'Heval: Yet Another Human Evaluation Metric' में उन्होंने अंग्रेजी-हिंदी समानांतर कॉर्पोरा के मूल्यांकन के लिए मानव मूल्यांकन के साथ मेट्रीक पद्धति का भी प्रयोग किया है।
- पूजा मलिक, अभिलाष गुप्ता, अनुराग बघेल द्वारा 2013 में लिखित शोध पत्र 'अंग्रेजी से भारतीय भाषाओं में मशीनी अनुवाद मूल्यांकन के प्रमुख मुद्दे (Key Issues in Machine Translation Evaluation of English-Indian Languages) विषय पर प्रकाशित शोध पत्र में उन्होंने मशीनी अनुवाद मूल्यांकन के लिए वाक्य स्तर, कॉर्पस स्तर के मूल्यांकन का विवेचन किया है। इसी के साथ मानव और स्वचालित मशीनी अनुवाद

मूल्यांकन की पद्धतियों की चर्चा भी की है जिसमें अंग्रेजी से भारतीय भाषाओं में मशीनी अनुवाद मूल्यांकन के लिए ब्लू मूल्यांकन पद्धति के प्रयोग को उचित बताया है।

- अदिति कल्याणी, हेमंत कुमुद (बनस्थली विश्वविद्यालय, राजस्थान), शशिपाल सिंह, अजय कुमार (सी-डैक, पुणे) द्वारा 2014 में 'हिंदी से अंग्रेजी अनुवाद के लिए मशीनी अनुवाद तंत्र के गुणवत्ता का आकलन' (Assessing the Quality of MT Systems for Hindi to English Translation) विषय पर शोध पत्र प्रकाशित किया गया। यह शोध पत्र IJCA Journal में प्रकाशित हुआ है। हिंदी से अंग्रेजी में मशीनी अनुवाद तंत्रों द्वारा प्राप्त आउटपुट का ब्लू (BLUE), निस्ट (NIST), जीटीएम (GTM), मेटेयोर (METEOR) और टीईआर (TER) प्रणाली द्वारा मूल्यांकन किया गया है और प्राप्त मूल्यांकन के परिणाम का मानव मूल्यांकन से तुलना कर मूल्यांकन की रैंकिंग दी गई है।
- हिंदी-अंग्रेजी मशीनी अनुवाद तंत्रों का मूल्यांकन (Evaluation of Hindi-English MT Systems, 2014) विषयक शोध पत्र में अतुल ओझा, आकांक्षा बंसल, सुमेध हाडके और गिरीश नाथ झा (जे.एन.यू., नई दिल्ली) ने बिंग (माइक्रोसॉफ्ट) और गूगल अनुवादक में अनुवाद के लिए स्वास्थ्य और सामान्य पाक-कला (General cooking) के वाक्य दिए और उनसे अनुवाद प्राप्ति के बाद उसका मूल्यांकन मानव द्वारा 0-4 स्केलिंग में किया। जिससे यह ज्ञात हुआ कि दोनों में से कौनसा तंत्र बोधगम्य और सही अनुवाद करता है। इसमें पाया गया कि गूगल अधिक सही अनुवाद करता है। साथ ही इसमें इन तंत्रों द्वारा अनूदित वाक्यों की त्रुटियों का विश्लेषण किया गया है।
- अनुवाद प्रौद्योगिकी विभाग, महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा (महाराष्ट्र) में चिप्पाडा अंबेडकर "तेलुगु - हिंदी यंत्र अनुवाद प्रणाली का मूल्यांकन"

विषय पर (2012-2017 )शोध कार्य कर रहे हैं। इनकी शोध निर्देशिका - डॉ. सी. अन्नपूर्णा हैं।

### 3.2 प्रस्तुत शोध का उद्देश्य

- मशीनी अनुवाद दो भाषाओं में होनेवाली प्रक्रिया है, जिसमें किसी एक भाषा का अनुवाद कंप्यूटर के माध्यम से दूसरी भाषा में किया जाता है।
- इस शोध के प्रमुख दो उद्देश्य हैं- मशीनी अनुवाद तंत्रों का मूल्यांकन करना और प्राप्त अनुवाद का त्रुटि विश्लेषण करना।
- इस शोध से अंग्रेजी से हिंदी में अनुवाद करने वाले प्रमुख मशीनी अनुवाद तंत्र गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद करने की प्रक्रिया का विश्लेषण किया गया है।
- इस शोध के लिए विभिन्न स्रोतों से अंग्रेजी से हिंदी में संमानांतर कॉर्पोरा का निर्माण किया गया है।
- अंग्रेजी से हिंदी मशीनी अनुवाद तंत्र में दोनों भाषाओं की संरचना में आनेवाली भिन्नता का विश्लेषण शब्द, वाक्य और अर्थ संरचना के स्तर पर किया गया है।
- विभिन्न मशीन अनुवाद तंत्रों के अनुवाद करने की प्रक्रिया, नियम आधारित, सांख्यिकीय आधारित, उदाहरण आधारित और मिश्रित आधारित मशीनी अनुवाद दृष्टिकोण पर आधारित है। इसलिए इसका भी शोध में विवेचन किया गया है।
- मशीन अनुवाद तंत्रों का मूल्यांकन करनेवाली ब्लू पद्धति (BLUE), मेट्रीक (BLUE), निस्ट (NIST) और मेटियोर (METEOR), मेटियोर-हिंदी इत्यादि पद्धतियों का गहन अध्ययन किया गया है।

- गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी तंत्रों से कॉर्पोरा के वाक्यों के अंग्रेजी से हिंदी में अनुवाद प्राप्त के बाद उनका मूल्यांकन किया गया है।
- मूल्यांकन को वर्तनी विषयक, रूपविषयक, शब्दविषयक, अर्थविषयक और वाक्य स्तर पर वर्गिकृत किया गया है।
- मशीन अनुवाद तंत्रों से प्राप्त अनुवाद में त्रुटियों का विश्लेषण किया गया है।
- त्रुटि विश्लेषण में लक्ष्य भाषा के स्तर पर आनेवाली समस्याएँ संज्ञा, सर्वनाम, विशेषण, क्रिया विशेषण, क्रियाओं तथा लिंग, वचन, काल, पक्ष, वृत्ति, कारक तथा प्रत्ययों में आनेवाली समस्याओं का विश्लेषण किया गया है।
- मूल्यांकन और त्रुटि विश्लेषण से यह ज्ञात हुआ कि भारतीय और विदेशी भाषाओं में अनुवाद करने वाले तंत्र किन-किन विषय क्षेत्रों में सही तथा बोधगम्य अनुवाद करने में सक्षम हैं।
- इस मूल्यांकन से मशीनी अनुवाद तंत्रों को अद्यतन या सुधारित करने के लिए विभिन्न भाषाई उपकरणों का निर्माण करने के लिए सुझाव दिए गए हैं।

### 3.3 शोध के लिए स्रोत सामग्री का प्रयोग

किसी शोध-कार्य के प्राथमिक उद्द्यमों में सर्वाधिक महत्त्वपूर्ण सामग्री-संकलन का काम होता है। विषय-वस्तु से सम्बन्धित तथ्यों के संकलन की सावधानी और विश्वसनीयता ही किसी शोध-कार्य को उत्कर्ष एवं महत्ता देती है।

#### प्राथमिक स्रोत सामग्री (Primary Sources)

- प्रस्तुत शोध अध्ययन में प्रमुख रूप से विभिन्न स्रोतों से भाषाई सामग्री का संकलन किया गया है। उसके निम्नलिखित स्रोत हैं-

- भारतीय भाषा कॉर्पोरा उपक्रम' परियोजना, 2012-2016,
- फारवर्ड प्रेस पत्रिका, 2015-2016,
- एअर इंडिया की 'शुभ यात्रा' पत्रिका, 2015-2016
- आर्थिक सर्वेक्षण 2014-2015, A flagship annual document of the Ministry of Finance, Government of India.
- श्रम एवं रोजगार मंत्रालय, वार्षिक रिपोर्ट, 2014-2015. Ministry of Labour & Employment, Government of India.
- स्वास्थ्य एवं परिवार कल्याण मंत्रालय, वार्षिक रिपोर्ट, 2015-2016. Ministry of Health and Family Welfare, Government of India.
- रक्षा अनुसंधान एवं विकास संगठन, वार्षिक रिपोर्ट 2015. मैटकॉफ हाउस, दिल्ली।
- राजभाषा विभाग, गृह मंत्रालय, भारत सरकार, वार्षिक रिपोर्ट, 2015-2016.

मशीनी अनुवाद और अनुवाद मूल्यांकन संबंधी किताबें, शोध पत्र, लेख, पत्रिकाएं तथा ऑनलाइन स्रोत इत्यादि सहायक स्रोत के रूप में प्रयुक्त किए गए हैं।

### 3.4 शोध प्रविधि (Research Methodology)

इस शोध के लिए प्रयुक्त शोध प्रविधि निम्नलिखित बिंदुओं पर आधारित हैं।

मशीनी अनुवाद के शोध में प्रयोगात्मक तथा सैद्धांतिक पद्धति का प्रयोग किया जाता है। प्रस्तुत शोध में भी मात्रात्मक, प्रयोगात्मक और सैद्धांतिक आधारित दृष्टिकोण का प्रयोग किया गया है। इसी के साथ विभिन्न मशीनी अनुवाद मूल्यांकन पद्धतियों का अध्ययन किया गया है और मानव एवं स्वचालित मूल्यांकन पद्धतियों के बारे में चर्चा की गयी है।

## ➤ कॉर्पोरा निर्माण एवं संकलन

इस शोध में अंग्रेजी से हिंदी मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए विभिन्न स्रोतों से कॉर्पोरा के 18000 वाक्यों का संकलन किया गया है। इस भाषाई सामग्री के संकलन के विषय-क्षेत्र मनोरंजन, कृषि, राजनीति, विज्ञान और प्रौद्योगिकी, कला और संस्कृति, पर्यटन, समाज और शिक्षा, स्वास्थ्य, खेल कूद, अर्थव्यवस्था, इतिहास, दर्शन, धर्म, राष्ट्रीय सुरक्षा और रक्षा, कानून, समाज और समुदाय, साहित्य और भूगोल इत्यादि हैं। इस कॉर्पोरा के संकलन के बाद अंग्रेजी से हिंदी में मशीनी अनुवाद के लिए गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी आदि तंत्रों में अनुवाद के लिए वाक्य दिए और अनुवाद प्राप्त किया गया। जिनका प्रयोग चतुर्थ अध्याय में मशीनी अनुवाद मूल्यांकन के लिए किया गया है।

## ➤ मानव मूल्यांकन

मूल्यांकन के लिए गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी, इन 4 मशीनी अनुवाद तंत्रों से प्राप्त अंग्रेजी से हिंदी में अनूदित वाक्यों का मानव आधारित मूल्यांकन (Evaluation) किया गया है। इस मूल्यांकन के लिए BLEU, METEOR, NIST, METEOR-HINDI, F-MEASURE आदि स्वचालित मूल्यांकन पद्धतियों का अध्ययन किया गया है। इस मूल्यांकन के लिए मानक डाटा के रूप में 500 वाक्यों में स्रोत वाक्य (अंग्रेजी) और लक्ष्य वाक्य (हिंदी) को देखते हुए 5 पॉइन्ट स्केल (0-4) पद्धति का प्रयोग किया गया है। अनुवाद की बोधगम्यता की दृष्टि से मानव मूल्यांकन के लिए निम्न स्केलिंग स्कोर का प्रयोग किया गया है।<sup>5</sup>

---

<sup>5</sup> Atul K.O. Bansal A. Hadke S. Jha GN. *Evaluation of Hindi-English MT Systems*, 2014, p.94-101

Evaluation on 5 point scale (0-4)	
4	पूर्णतः बोधगम्य (All meaning)
3	लिप्यंतरित शब्द के साथ वाक्य बोधगम्य (Most meaning)
2	वाक्य समझने में कुछ गलतियाँ (Much meaning)
1	कुछ समझना, कुछ गलतियाँ, लेकिन बोधगम्य नहीं (little meaning)
0	कुछ भी बोधगम्य नहीं (None)

टेबल 3.1: मूल्यांकन की स्केलिंग पद्धति

मानव द्वारा मूल्यांकन के लिए प्रत्येक वाक्य के लिए इस स्केलिंग स्कोर का प्रयोग किया गया है। यह मशीनी अनुवाद तंत्र के मूल्यांकन के लिए मानक स्केलिंग पद्धति मानी जाती है। मूल्यांकन स्केलिंग की इस पद्धति के प्रयोग से सटीक परिणाम प्राप्त हुए हैं। जिससे यह ज्ञात हुआ है कि कौनसे मशीनी अनुवाद तंत्र कितना सटीक अनुवाद करते हैं। इस प्रक्रिया में भाषिक विश्लेषणात्मक तथा मशीनी अनुवाद के अंतर्गत भाषिक संरचना के व्याकरणिक कोटियों में आनेवाली समस्याओं को स्पष्ट किया गया। अंतः इस प्रक्रिया से यह ज्ञात हुआ है कि गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्रा-राजभाषा और आंग्ला-हिंदी ये तंत्र सीमित विषय-क्षेत्र में सही और शीघ्र अनुवाद करते हैं।

#### ➤ त्रुटि विश्लेषण

मूल्यांकन प्रक्रिया के बाद दोनों भाषाओं की संरचना स्तर पर आनेवाली भिन्नताओं को देखते हुए तथा मशीन से अनूदित वाक्यों के त्रुटि विश्लेषण में वर्तनी विषयक, रूपविषयक, शब्दविषयक, अर्थविषयक, वाक्य स्तर पर, लिप्यंतरण और व्याकरणिक संरचना में प्राप्त त्रुटियों के विश्लेषण के साथ उसका समाधान किया गया है।

निष्कर्ष में यह कहा जा सकता है कि विश्व तथा भारत में विभिन्न संस्थानों तथा विश्वविद्यालयों ने मशीनी अनुवाद के तंत्र विकसित किए गए हैं। भारत में मशीनी अनुवाद के विकास कार्य के लिए विभिन्न परियोजनाएँ विश्वविद्यालयों में तथा संस्थानों में चल रही हैं जिन्हें टीडीआईएल द्वारा भारत सरकार से वित्तीय सहायता मिल रही है। भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (TDIL) के अंतर्गत मशीनी अनुवाद के लिए कई उपकरणों का निर्माण किया गया है। इनका निर्माण करते समय विभिन्न सामग्री तथा उपकरणों में कॉर्पोरा, शब्दतंत्र, द्विभाषी शब्दकोश, कृत्रिम बुद्धि, रूपवैज्ञानिक विश्लेषक, टैगर, पार्सर, प्रजनक, अंतरण व्याकरण आदि घटकों का विकास किया गया होगा और इन सभी में भाषा, कंप्यूटर और मानव का सहयोग महत्वपूर्ण रहा होगा। इसमें संदेह नहीं कि भारत में भारतीय भाषाओं के लिए मशीनी अनुवाद के अनेक तंत्र विकसित किए गए हैं, लेकिन इनके अनुवाद की गुणवत्ता में आज भी कमी महसूस होती है, उन्हें अधिक सुधार करने की आवश्यकता है।

अतः इस अध्याय में मशीनी अनुवाद के महत्व के साथ इस शोध के पूर्ववर्ती शोध कार्य की समीक्षा, शोध का उद्देश्य तथा प्राथमिक स्रोत सामग्री में कॉर्पोरा संकलन के बाद उसके मानव द्वारा मूल्यांकन में 5 पॉइन्ट स्केल (0-4) पद्धति के प्रयोग को स्पष्ट किया गया है। त्रुटि विश्लेषण में मशीन अनूदित वाक्यों में पाई जानेवाली त्रुटियों की जानकारी दी गयी है साथ ही शोध प्रविधि में शोध के लिए प्रयुक्त की गयी पद्धति की जानकारी दी गयी है।

# चतुर्थ अध्याय

## अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन

4.1 मशीनी अनुवाद मूल्यांकन का परिचय

4.1.1 मानव मूल्यांकन

4.1.2 स्वचालित मूल्यांकन

4.2 मशीनी अनुवाद तंत्रों का मूल्यांकन

4.2.1 कॉर्पोरा निर्माण

4.2.2 निर्गत अनुवाद

4.2.3 मानव मूल्यांकन

4.2.4 कॉर्पोरा का क्षेत्रों के अनुसार मूल्यांकन

## चतुर्थ अध्याय

### अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन (EVALUATION OF ENGLISH-HINDI MACHINE TRANSLATION SYSTEMS)

---

विश्व में अनुवाद का महत्व प्रतिदिन बढ़ता जा रहा है। पूरे विश्व में अनेक भाषाएँ बोली जाती हैं और प्रत्येक व्यक्ति के लिए इन सभी भाषाओं में पारंगत होना या जानना संभव नहीं है। इसलिए अनुवाद ही ऐसा माध्यम है जो किसी भी भाषा में लिखी गई साहित्य सामग्री को दूसरी भाषा में अनूदित करने का कार्य करता है। कोई भी व्यक्ति किसी दूसरी भाषा में लिखित साहित्य को जानने या बोलने के लिए अनुवाद के माध्यम का ही प्रयोग करता है तभी वह दूसरी भाषा की सामग्री को जान पाता है। प्रत्येक भाषा में बहुत ज्ञान-सामग्री तथा ग्रंथ समाहित हैं उन्हें अनुवाद के माध्यम से ही अनेक भाषाओं में अनूदित कर हर भाषा-भाषी के लिए सुलभ बनाया गया है। आज ज्ञान-विज्ञान की बहुत सी सामग्री का अनुवाद कर एक दूसरे के लिए सुलभ बनाई जा रही है। अनेक देशों में किसी भी भाषा की श्रेष्ठ रचना का अनुवाद दूसरी भाषाओं में कर लिया जाता है। वर्तमान युग में जीवन के अनेक क्षेत्रों में कंप्यूटर के अनुप्रयोग से क्रांतिकारी परिवर्तन आया है। कंप्यूटर की सहायता से बहुत सा अनुवाद कार्य शीघ्र किया जा रहा है जिससे मानव के श्रम और समय की बचत हो रही है।

भारत एक बहुभाषी देश है। यहाँ एक भाषा से दूसरी भाषाओं में ग्रंथ तथा दस्तावेजों के अनुवाद के लिए बहुत मांग है, इसके कारण कई भाषाओं में उपलब्ध ज्ञान, विज्ञान की सामग्री का अनुवाद प्रमुख भाषाओं तथा स्थानीय भाषाओं में करने के लिए अनेक देशों ने मशीनी अनुवाद तंत्र विकसित किए हैं। मशीनी अनुवाद से एक प्राकृतिक भाषा का दूसरी प्राकृतिक भाषा में कंप्यूटर के द्वारा अनुवाद किया जाता है। आज विश्व स्तर पर विभिन्न भाषाओं में अनुवाद के लिए कई मशीनी अनुवाद तंत्रों का निर्माण किया गया है। भारत ने भी कई वर्षों से अनुसंधान करने के बाद विभिन्न विश्वविद्यालयों, संस्थानों और संगठनों की सहायता से भारतीय भाषाओं के लिए अनेक

तंत्रों का निर्माण किया है तथा अन्य भाषाओं के लिए भाषिक उपकरण और संसाधनों का निर्माण किया जा रहा है। हालांकि अनेक मशीनी अनुवाद तंत्र भारतीय भाषाओं के लिए उपलब्ध होने के बावजूद उनसे अनुवाद में होने वाली त्रुटियाँ प्रमुख समस्या रही हैं। इन त्रुटियों के विश्लेषण और मूल्यांकन के लिए अनेक मशीनी अनुवाद मूल्यांकन की पद्धतियों का विकास किया गया है ताकि भविष्य में इन तंत्रों को परिष्कृत किया जा सके।

अनुवादक किसी पाठ या वाक्य का अनुवाद करता है तो वह विश्लेषण, अंतरण और पुनर्गठन की प्रक्रिया से गुजरते हुए लक्ष्य भाषा में अनुवाद करता है। उसके पश्चात अनुवाद पुनरीक्षण और मूल्यांकन की प्रक्रिया प्रारंभ होती है। पुनरीक्षण में परीक्षण अथवा जाँच के साथ-साथ अनुवाद को परिष्कृत, संशोधित और संपादित भी किया जाता है। इसी के साथ मूल्यांकन में मुख्यतः जाँच अथवा पुनरीक्षण तो होता है किंतु उसमें अनुवाद के स्वरूप, उसकी गुणवत्ता, उसकी सफलता और असफलता का भी विवेचन किया जाता है। अनुवादक चाहे कितना भी सफल अनुवाद करे, उसकी सफलता का मूल्यांकन करना आवश्यक होता है। वास्वव में अनुवाद का मूल्यांकन मूल पाठ के परिप्रेक्ष्य में अनूदित पाठ के गुण-दोषों का विवेचन होता है। यह विवेचन अनुवाद कार्य की निष्पत्ति से है जो अनुवाद संशोधन और संवर्धन में सहायक होता है।<sup>1</sup> इसी प्रकार मशीनी अनुवाद तंत्रों के मूल्यांकन में भी अनूदित पाठ के गुण-दोषों की जाँच की जाती है।

मशीनी अनुवाद मूल्यांकन का प्रमुख उद्देश्य अनुवाद में आनेवाली त्रुटियों का विश्लेषण करने के लिए, तंत्र को बेहतर बनाने के लिए, स्रोत भाषा से लक्ष्य भाषा में प्रभावी अनुवाद करने के लिए तथा तंत्रों से प्राप्त अनुवाद की मानव अनुवाद के साथ तुलना करने के लिए किया जाता है। सामान्यत यह अंग्रेजी से हिंदी या भारतीय भाषाओं में मूल्यांकन निम्न बिन्दुओं के आधार पर किया जाता है। 1. संज्ञाओं के लिंग एवं वचन का लक्ष्य अनुवाद में सटीक प्रयोग, 2. मूल वाक्य में प्रयुक्त काल का अनुवाद में प्रयोग, 3. मूल वाक्य में प्रयुक्त वाच्य (voice) का अनुवाद में प्रयोग, 4. व्यक्तिवाचक संज्ञा का

---

1 गोस्वामी, कृष्ण कुमार, *अनुवाद विज्ञान की भूमिका*, 2012 पृ.105

अनुवाद, 5. मूल वाक्य में प्रयुक्त संज्ञा व क्रिया के अनुरूप विशेषण एवं क्रिया विशेषण का प्रयोग, 6. अनुवाद में सही शब्दों के पर्याय का चयन, 7. अनुवाद में संज्ञा, क्रिया एवं सहायक क्रियाओं का शब्दक्रम (word order), 8. अनुवाद में विराम चिन्हों का प्रयोग, 9. मूल अनुवाद के महत्वपूर्ण पाठ को अधिक महत्व देना, 10. अनूदित वाक्य में मूल वाक्य के अर्थ को बनाए रखना।<sup>2</sup> इन बिन्दुओं के आधार पर मूल्यांकन करने के बाद उनका त्रुटि विश्लेषण भी किया जाता है।

#### 4.1 मशीनी अनुवाद मूल्यांकन का परिचय

मशीनी अनुवाद तंत्रों का मूल्यांकन करना कठिन और चुनौतीपूर्ण कार्य है। आज मशीनी अनुवाद के मूल्यांकन के लिए अनेक विधियों का प्रयोग किया जाता है। जब मार्केट में कोई नया उत्पाद आता है तो कंपनी भी अपने उत्पाद की जाँच कर उसका परीक्षण करती है वैसे ही, मशीनी अनुवाद के मूल्यांकन के लिए तंत्रों की गुणवत्ता की जाँच करने की आवश्यकता होती है। इस प्रक्रिया में मूल्यांकन के माध्यम से उसके अनुवाद की सहजता और बोधगम्यता की जानकारी मिलती है। मशीनी अनुवाद का मूल्यांकन दो स्तरों पर किया जा सकता है, एक वाक्य के स्तर पर और दूसरा कार्पस आधारित वाक्यों के स्तर पर।

प्रारंभ में, मशीनी अनुवाद का मूल्यांकन मानव द्वारा ही किया जाता था। लेकिन इस कार्य के लिए अधिक समय तथा धन खर्च होता था। इसके अलावा एक ही वाक्य के मूल्यांकन के परिणाम एक-दूसरे मूल्यांकनकर्ता से भिन्न होते थे। इसलिए आज मूल्यांकन के लिए स्वचालित मूल्यांकन पद्धतियों का प्रयोग अधिक किया जाता है। हाल के वर्षों में मशीनी अनुवाद के स्वचालित मूल्यांकन के लिए विभिन्न पद्धतियों का विकास किया है। प्रमुखतः मशीनी अनुवाद के मूल्यांकन को मानव मूल्यांकन (Human

---

<sup>2</sup> Joshi N., Darbari H., Mathur I. (2012): *Human and Automatic Evaluation of English to Hindi Machine Translation Systems*, Advances in Intelligent and Soft Computing, vol. 166. Springer, p.427

evaluation) और स्वचालित मूल्यांकन (Automatic evaluation) इन दो श्रेणियों में विभाजित किया गया है।<sup>3</sup>

#### 4.1.1 मानव मूल्यांकन (Human Evaluation)

मशीन द्वारा अनूदित वाक्यों का मूल्यांकन जब मनुष्यों द्वारा किया जाता है तो उसे मानव मूल्यांकन या मानव आधारित मूल्यांकन कहते हैं। मानव मूल्यांकन के प्रयोजन के लिए मानव अनुवादक या भाषावैज्ञानिक की आवश्यकता होती है जो दोनों भाषाओं में जानकार होता है। अनेक कारणों के कारण हम हर समय मानव मूल्यांकन का प्रयोग नहीं कर सकते। मानव मूल्यांकनकर्ता मशीनी अनुवाद तंत्रों का मूल्यांकन करते समय विभिन्न पहलुओं को ध्यान में रखते हुए वह अनुवाद में सहजता और अर्थ की उपयुक्तता को देखते हुए मूल्यांकन करता है। इसलिए कहा जाता है कि मानव मूल्यांकन सभी मूल्यांकन पद्धतियों से अधिक उपयुक्त पद्धति मानी जाती है। लेकिन इस मूल्यांकन में कुछ समस्याएँ भी हैं जैसे- यह समय और श्रम की दृष्टि से कठिन कार्य है जिसमें एक सप्ताह या एक महीना भी लग सकता है। साथ ही मूल्यांकन के सांख्यिकीय परिणाम के लिए मानव मूल्यांकन निष्पक्ष रूप से होना चाहिए। इसमें मूल्यांकन व्यक्तिपरक होने से निष्पक्षता पर संदेह होता है। प्रत्येक मशीनी अनुवाद के निर्गत अनुवाद का मानव मूल्यांकन एक से अधिक मूल्यांकनकर्ताओं द्वारा किया जाना चाहिए ताकि मूल्यांकन के परिणाम सही प्राप्त हों।<sup>4</sup>

---

<sup>3</sup> Malik P., Gupta A., Baghel A. (2013): *Key Issues in Machine Translation Evaluation of English-Indian Languages*, International Journal of Engineering Research & Technology (IJERT), 2013, p.3244-3245

<sup>4</sup> Malik P., Baghel A., (2016): *An Improvement in BLEU Metric for English-Hindi Machine Translation Evaluation*, International Conference on Computing, Communication and Automation, p.331-332

मानव मूल्यांकन के लिए अनेक मूल्यांकन पद्धतियों और अनेक श्रेणियों (Grading) का प्रयोग किया जाता है। उसमें 7 पॉइन्ट ग्रेडिंग स्केल, 5 पॉइन्ट ग्रेडिंग स्केल, अल्पैक द्वारा 0-9 ग्रेडिंग स्केल और एडवांस्ड रिसर्च प्रोजेक्ट्स एजेंसी (एआरपीए) द्वारा 0-5 ग्रेडिंग स्केल आदि शामिल हैं। इसमें प्रमुख रूप से मानव मूल्यांकन के लिए शुरूवात में सन 1966 में अल्पैक समिति [Automatic Language Processing Advisory Committee (ALPAC)] की रिपोर्ट में मानव अनुवाद के साथ मशीनी अनुवाद का तुलनात्मक अध्ययन किया गया था जिसमें मशीनी अनुवाद से प्राप्त निर्गत अनुवाद का मूल्यांकन करने के लिए मानव मूल्यांकनकर्ता का प्रयोग किया गया था। इस अध्ययन में मूल्यांकन के लिए रूसी से अंग्रेजी मशीनी अनुवाद तंत्रों से प्राप्त निर्गत अनुवाद की तुलना मानव अनुवाद से की गई थी। इसमें मानव मूल्यांकन के लिए दो मानदंडों को निर्धारित कर ग्रेडिंग स्केल से स्कोर निकाले गए थे। वह बोधगम्यता (Intelligibility) और परिशुद्धता (Fidelity) के लिए (0-9) ग्रेडिंग स्केल का प्रयोग किया गया था। बाद में इस अध्ययन से ज्ञात हुआ कि मूल्यांकन के लिए प्रयुक्त ग्रेडिंग स्केल से मानव मूल्यांकन के साथ अधिक समरूप परिणाम मिले। लेकिन शोधकर्ताओं ने सिफारिश की थी कि कम से कम 3 या 4 ग्रेडिंग स्केल का प्रयोग मूल्यांकन के लिए किया जाना चाहिए। अंत में समिति के अध्ययन से ज्ञात हुआ कि मानव और मशीनी अनुवाद की गुणवत्ता का अत्यधिक विश्वसनीय मूल्यांकन इस पद्धति से किया जा सकता है।<sup>5</sup>

एडवांस्ड रिसर्च प्रोजेक्ट्स एजेंसी ने [Advanced Research Projects Agency (ARPA)] मानव भाषा प्रौद्योगिकी कार्यक्रम के भाग के रूप में, उन्नत अनुसंधान परियोजना एजेंसी (एआरपीए) ने मशीनी अनुवाद तंत्र के मूल्यांकन के लिए एक पद्धति विकसित की जिसकी शुरुआत 1991 में की। इसमें मूल्यांकन के लिए सांख्यिकीय (Statistical), नियम आधारित और मानव-सहायता (Human assisted) आधारित मशीनी अनुवाद तंत्रों का परीक्षण किया गया। इसमें मूल्यांकन के लिए सहजता (fluency) और पर्याप्तता (adequacy), इन दो मानदंड के लिए (0-5) ग्रेडिंग स्केल का प्रयोग किया गया

---

<sup>5</sup>[https://en.wikipedia.org/wiki/Evaluation\\_of\\_machine\\_translation#Automatic Language Processing Advisory Committee](https://en.wikipedia.org/wiki/Evaluation_of_machine_translation#Automatic_Language_Processing_Advisory_Committee) (Retrieved on 16.06.2017)

था। मशीनी अनुवाद से निर्गत अनुवाद का मानव द्वारा मूल्यांकन के लिए ARPA मूल्यांकन कार्यक्रम ने इसे मानक पद्धति के रूप में प्रयोग किया था।<sup>6</sup>

मानव मूल्यांकन की मानक श्रेणियों में (0-5) स्केलिंग स्कोर का प्रयोग किया जाता है। 5-उत्कृष्ट, 4-अच्छा, 3-समझने योग्य, 2- समझने में कठिन, 1-अस्वीकार्य अनुवाद, इस ग्रेडिंग स्केल का प्रयोग सामान्यतः मानव मूल्यांकर्ता मशीनी अनुवाद के मूल्यांकन के लिए करते हैं। मानव मूल्यांकनकर्ता प्रथम अनुवाद को देखता है और उसे जाँचने के बाद तय करता है कि अनुवाद सही है या नहीं। मानव मूल्यांकन की सबसे महत्वपूर्ण विलक्षणता यह है जब दो मानव मूल्यांकनकर्ता एक ही पाठ को देखते हुए दो अलग-अलग मूल्यांकन करते हैं तो उसके मूल्यांकन के स्कोर अलग-अलग आते हैं। इसका अर्थ यह है कि मशीन से निर्गत अनुवाद का मूल्यांकन व्यक्तिपरक निर्भर करता है। फिर भी मानव द्वारा मूल्यांकन स्वचालित मूल्यांकन से अधिक विश्वसनीय और सटीक मूल्यांकन किया जाता है।<sup>7</sup> इसमें कुछ समस्याओं के चलते तथा इस मूल्यांकन का पुनर्प्रयोग न होने की वजह से वर्तमान में स्वचालित मूल्यांकन पद्धतियों का प्रयोग अधिक किया जाता है जो अधिक शीघ्रता से मशीनी अनुवाद का मूल्यांकन करती है। इसमें समय और श्रम की बचत होती है और मूल्यांकन में प्राप्त सांख्यिकीय आँकड़ों का प्रयोग कई बार किया जा सकता है।

#### 4.1.2 स्वचालित मूल्यांकन (Automatic Evaluation)

जब अनूदित पाठ का मूल्यांकन मशीन की सहायता से किया जाता है तो उसे स्वचालित मूल्यांकन या मशीनी मूल्यांकन कहते हैं। स्वचालित मूल्यांकन वाक्य स्तर और कॉर्पस स्तर पर किया जाता है। वाक्य स्तर पर मूल्यांकन में अनूदित वाक्य के समूह का एल्गोरिथम द्वारा स्कोर निकालकर गणना की जाती है फिर उसकी मानव

---

<sup>6</sup>[https://en.wikipedia.org/wiki/Evaluation\\_of\\_machine\\_translation#Advanced Research Projects Agency .28ARPA.29](https://en.wikipedia.org/wiki/Evaluation_of_machine_translation#Advanced_Research_Projects_Agency_.28ARPA.29) (Retrieved on 16.06.2017)

<sup>7</sup> Kalyani A., Kumud H., Singh S.P., Kumar A., Darbari H.(2014): *Evaluation and Ranking of Machine Translated Output in Hindi Language using Precision and Recall Oriented Metrics*, International Journal of Advanced Computer Research, p.54-55

मूल्यांकन के साथ तुलना की जाती है। उसी तरह कॉर्पस स्तर पर मूल्यांकन में वाक्यों से अधिक स्कोर के लिए मानव निर्णय (Human Judgment) और मीट्रिक निर्णय (Metric judgment) को एकत्रित किया जाता है और फिर कुल स्कोर से मूल्यांकन के परिणाम निकाले जाते हैं। इससे मशीनी अनुवाद तंत्रों का स्वचालित मूल्यांकन होता है। मशीनी अनुवाद का स्वचालित मूल्यांकन प्रमुख रूप से अनुवाद में बोधगम्यता, सटीकता, सहजता और पर्याप्तता के लिए किया जाता है। इन मानदंडों के द्वारा मशीनी अनुवाद का मूल्यांकन किया जाता है

स्वचालित मूल्यांकन यह मानव मूल्यांकन की तुलना में अधिक शीघ्र और सस्ता होता है जो किसी भी मशीनी अनुवाद तंत्रों का मूल्यांकन करने के लिए आसान है। इसका प्रयोग वाक्य स्तर पर त्रुटियों के विश्लेषण, मशीनी अनुवाद तंत्र की तुलना करने और तंत्र को अनुकूल (System optimization) करने के लिए किया जाता है। कहा जाता है कि इसकी ग्रेडिंग मानव मूल्यांकन की तरह सही नहीं होती। लेकिन इसके मूल्यांकन के परिणाम मानव मूल्यांकन से अधिक निष्पक्ष रूप में मिलते हैं।<sup>8</sup>

प्रथमतः मशीनी अनुवाद का मूल्यांकन करने के लिए 1956 में मिल्लर और बीबे-सेंटर<sup>9</sup> ने (Miller G.A. & Beebe-Center J.G.) और 1965 में पफ्लिन<sup>10</sup> ने (Pfafflin S.M.) मानव द्वारा मूल्यांकन को प्रस्तावित किया था। शुरुआत में मशीनी अनुवाद मूल्यांकन केवल मानव द्वारा ही किया जाता था। उसके बाद स्वचालित मूल्यांकन पद्धतियों का प्रयोग अधिक होने लगा। वर्तमान में मशीनी अनुवाद के मूल्यांकन के लिए अनेक

---

<sup>8</sup> Avinesh P.V.S., *Transfer Grammar Engine and Automatic Learning of Reorder Rules in Machine Translation*, M.Sc Computer Science Thesis, September- 2010, p.14-15

<sup>9</sup> Miller G.A.& Beebe-Center J.G., *Some Psychological Methods for Evaluating the Quality of Translation*, Mechanical Translations, vol 3, December 1956, pp. 73-80

<sup>10</sup> Pfafflin S.M., *Evaluation of Machine Translations by Reading Comprehension Tests and Subjective Judgments*, Mechanical Translation and Computational Linguistics, 1956, vol. 8, pp 2-8

स्वचालित मूल्यांकन पद्धतियाँ हैं जिनका प्रयोग विभिन्न मशीनी अनुवाद तंत्र करते हैं। इसमें प्रमुख रूप से ब्लू (BLEU), निस्ट (NIST), मेटियोर (METEOR), एफ़-मूल्यांकन पद्धति (F-measure) और मेटियोर-हिंदी आदि मूल्यांकन की पद्धतियाँ शामिल हैं जिनका नीचे विश्लेषण किया गया है।

#### 4.1.2.1 ब्लू (BLEU)

ब्लू मूल्यांकन पद्धति को आईबीएम ने विकसित किया है। यह पहली मीट्रिक पद्धति है जो मशीनी अनुवाद के मूल्यांकन की मानव मूल्यांकन की गुणवत्ता के साथ तुलना करने के लिए प्रयोग की जाती है। वर्तमान में इसका स्वचालित मूल्यांकन के लिए सबसे अधिक प्रयोग किया जाता है। इसमें एन-ग्राम के आधार पर अलग-अलग वाक्यों के स्कोर की गणना करने के लिए ज्यामितीय साधन का प्रयोग करती है और स्कोर की गणना करने के बाद फिर कुल स्कोर के लिए पूरे कॉर्पस से गणितीय आंकड़ों द्वारा औसत स्कोर निकाला जाता है, जिससे मूल्यांकन के परिणाम निकलते हैं। इस मीट्रिक पद्धति से गणना निम्नलिखित सूत्र द्वारा की जाती है।<sup>11</sup>

$$BLEU = \min\left(1, \frac{MT\ output-length}{reference-length}\right) \times \exp\left(\sum_{n=1}^N w_n \log(p_n)\right)$$

उपरोक्त सूत्र में दिया है कि मशीनी अनुवाद से निर्गत अनुवाद और संदर्भ अनुवाद के माध्यम से ब्लू के परिणाम निकलते हैं। इसका प्रयोग अंग्रेजी से हिंदी मशीनी अनुवाद के तंत्र मूल्यांकन के लिए अधिक करते हैं जिससे मूल्यांकन के सटीक परिणाम प्राप्त होते हैं।

#### 4.1.2.2 निस्ट (NIST)

निस्ट ब्लू मीट्रिक का संशोधित रूप है। इसका विकास राष्ट्रीय मानक एवं प्रौद्योगिकी संस्थान, अमेरिका ने किया है। इसी संस्थान का ही नाम इस मूल्यांकन

---

<sup>11</sup> Papineni K., Roukos S., Ward T., & Zhu W.J. *Bleu: a method for automatic evaluation of machine translation*, 2001, RC22176 Technical Report, IBM T.J. Watson Research Center.

पद्धति को दिया गया है। यह मूल्यांकन पद्धति BLEU मीट्रिक पर आधारित है, लेकिन इसमें कुछ परिवर्तन किया गया है। इसमें एन-ग्राम की गणना करते समय प्रत्येक एन-ग्राम को समान महत्व देने के बजाय उनके अनुवाद की बोधगम्यता के अनुसार महत्व दिया जाता है। ब्रेविटी पेनल्टी (Brevity penalty) का प्रयोग कर निर्गत अनुवाद से प्रति शब्द के लिए एक एन-ग्राम का प्रयोग करते हुए फिर उसे संदर्भ अनुवाद के एन-ग्राम के साथ मिलान (match) किया जाता है और अंत में औसत स्कोर निकाला जाता है। इसमें स्कोर के लिए 1 से 5 संख्या का प्रयोग किया जाता है। इसमें गणना निम्नलिखित सूत्र द्वारा की जाती है।<sup>12</sup>

$$NIST = BP * \sum_{n=1}^N \left( \frac{\sum_{n\text{-gram} \in \text{matching}} \text{match}(n\text{-gram})}{\sum_{n\text{-gram}' \in \text{translation}} \text{match}(n\text{-gram}')} \right)$$

#### 4.1.2.3 मेटियोर (METEOR)

मेटियोर मूल्यांकन पद्धति को कार्नेगी मेलॉन विश्वविद्यालय ने विकसित किया है। यह यूनियोग्राम सुव्यतता (precision) और यूनियोग्राम स्मरण क्षमता (recall) पर आधारित ग्रेडिंग के लिए 0-4 स्कोर का प्रयोग किया जाता है। इसे ब्लू की कमियों को दूर करने के लिए बनाया गया है। इस मीट्रिक पद्धति में मशीनी अनुवाद एवं मानवीय संदर्भ अनुवाद में कई तरह से मिलान (Matching) किया जाता है। संदर्भ पाठ या अनुवाद पाठ के प्रत्येक स्तर पर मूल शब्दों के साथ मिलान किया जाता है जो मशीनी अनुवाद तथा संदर्भ अनुवाद में समान हों। संदर्भ अनुवाद से पर्यायवाची शब्दों के प्रयोग की गणना करने के लिए वर्डनेट (Wordnet) का प्रयोग किया जाता है। इस मीट्रिक पद्धति का कार्यान्वयन (Implementation) मॉड्यूलर के रूप में होता है क्योंकि एल्गोरिदम में जो शब्द मेल खाते हैं उन शब्द को मॉड्यूल में परिवर्तित किया जाता है और मिलान न होने

---

<sup>12</sup> Doddington, G., *Automatic evaluation of machine translation quality using n-gram co occurrence statistics*, Proceedings of the Human Language Technology Conference (HLT), San Diego, 2002, CA p. 128-132

पर उसमें किसी शब्द को जोड़ने की सुविधा भी होती है। अंत में इन सभी ग्रेडिंग का औसत निकाल कर कुल स्कोर प्राप्त किया जाता है।<sup>13</sup>

#### 4.1.2.4 एफ-परिमाण पद्धति (F-Measure Metric Method)

इसे न्यूयॉर्क विश्वविद्यालय ने विकसित किया है। मानव द्वारा मूल्यांकित वाक्यों और संदर्भ वाक्यों के मिलान के लिए एन-ग्राम (N-gram) स्कोर की गणना की जाती है जिसमें एन-ग्राम के अधिकतम मिलान (Maximum matching) के लिए ग्राफ थ्योरी (Graph theory) का प्रयोग किया जाता है। इसमें मूल अनुवाद के साथ मानव द्वारा अनुवाद कितना सही है इसे जाँचने के लिए एन-ग्राम परिणाम की दोनों के स्कोर के साथ तुलना की जाती है। साथ ही दोनों के बीच समान घटकों का मूल्यांकन कर विश्लेषण किया जाता है। शुद्धता (precision) और स्मरण क्षमता (Recall) का महत्वपूर्ण योगदान इस पद्धति में किया जाता है। इस मूल्यांकन पद्धति का प्रयोग अंग्रेजी से हिंदी मशीनी अनुवाद के मूल्यांकन के लिए भी किया जाता है।<sup>14</sup>

#### 4.1.2.5 मेटियोर-हिंदी परिमाण पद्धति (Meteor-Hindi Metric)

यह मेटियोर का एक संशोधित संस्करण है, जिसमें मूल्यांकन की दृष्टि से हिंदी के लिए विशेष घटकों का समावेश किया गया है। अंग्रेजी से हिंदी मशीनी अनुवाद तंत्र के मूल्यांकन के लिए इसकी सहायता ली जाती है। इसके मूल्यांकन के स्कोर की मानव मूल्यांकन स्कोर के साथ तुलना की जाती है। मेटोयोर-हिंदी का स्कोर (0.703) ब्लू (0.271) से मिलान करने से ज्ञात होता है कि काफी अधिक है। अंग्रेजी से हिंदी अनुवाद के मूल्यांकन के लिए इसे अधिक गुणवत्ता आधारित बनाने के लिए मेटोयोर-हिंदी में

---

<sup>13</sup> Denkowski M. and Lavie A. *Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems*, Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation. 2011, p.85-90

<sup>14</sup> Tomer N., Sinha D. and Rai P.K., *Evaluating Machine Translation Evaluation's F-Measure Metric for English to Hindi Language Machine Translation*, International Journal of Academy Research Computer Engineering and Technology, 2012, Vol. 1(7), p.151-156

निम्न मॉड्यूल का समावेश किया गया है।<sup>15</sup> 1) स्थानीय शब्द समूह (एलडब्ल्यूजी): जिसमें मूल शब्द और उसके संबंधित प्रकार्य शब्द शामिल हैं। 2) उपवाक्य मिलान (Clause Match): उपवाक्य को एक वाक्यांश के रूप में परिभाषित किया जाता है, जिसमें कम से कम एक क्रिया और एक कर्ता शामिल हैं। 3) शब्दभेद मिलान (POS Matcher) से प्रत्येक शब्द की गणना की जाती है साथ ही पीओएस टैग से भी गणना की जाती है। इस संशोधित मेटोयोर-हिंदी मूल्यांकन पद्धति का प्रयोग हिंदी में उचित मूल्यांकन के लिए किया जाता है।

स्वचालित मूल्यांकन पद्धतियों का प्रयोग मशीनी अनुवाद मूल्यांकन के लिए समय की मांग के अनुसार अधिक प्रयोग किया जाता है। जिससे अनुवाद की गुणवत्ता का जाँच संख्यात्मक मीट्रिक पद्धति का प्रयोग कर की जाती है, इससे मानव अनुवाद के साथ तुलना कर अंतिम मूल्यांकन के परिणाम प्राप्त किए जा सकते हैं। इसके लिए अधिक से अधिक गणितीय सूत्रों या एल्गोरिथम का प्रयोग किया जाता है।

## 4.2 मशीनी अनुवाद तंत्रों का मूल्यांकन

इस अध्याय में मूल्यांकन के लिए संकलन किए गए कॉर्पोरा के वाक्यों का निर्गत अनुवाद प्राप्त करने के बाद मानव द्वारा मूल्यांकन किया गया और यह मूल्यांकन विषय-क्षेत्र के अनुसार किया गया उसके बाद मूल्यांकन के परिणाम में बोधगम्यता का कुल औसत निकाला गया। जिससे ज्ञात हुआ है कि कौनसे मशीनी अनुवाद तंत्र किस विषय क्षेत्रों के लिए अच्छा अनुवाद करते हैं।

### 4.2.1 कॉर्पोरा निर्माण (Corpora creation)

अंग्रेजी से हिंदी मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए मानव द्वारा अनूदित 18,000 वाक्यों का संकलन किया गया है। इस संकलन के लिए विभिन्न संदर्भ स्रोत जिसमें फारवर्ड प्रेस पत्रिका, भारतीय भाषा कॉर्पोरा उपक्रम, एअर इंडिया की 'शुभ यात्रा'

---

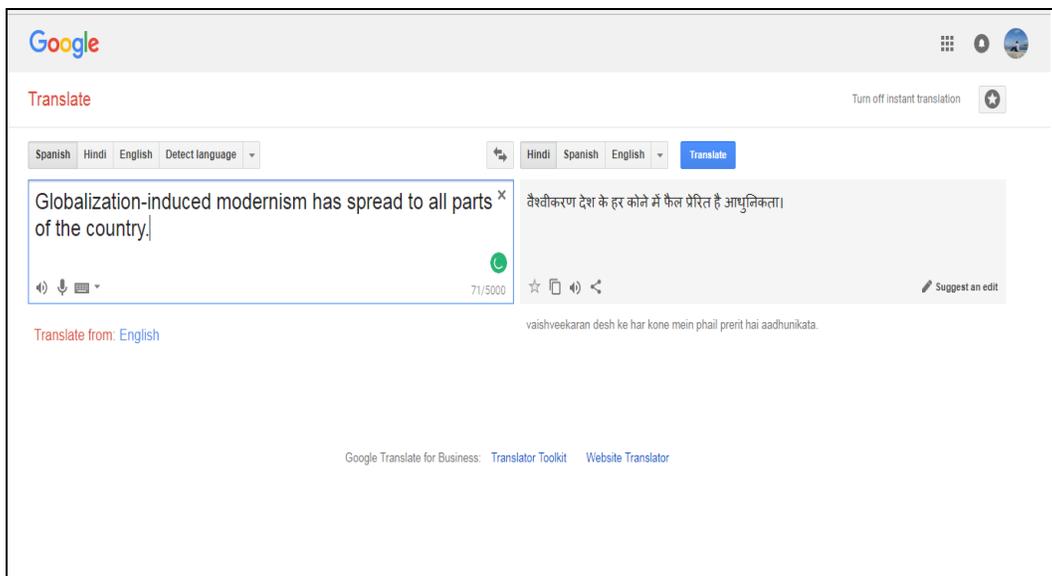
<sup>15</sup> Ankush G., Sriram V. and Rajeev S., *METEOR-Hindi: Automatic MT Evaluation Metric for Hindi as a Target Language*, Proceedings of ICON-2010: 8th International conference on Natural language processing, Macmillan Publishers, India. p.5-6

पत्रिका, विभिन्न मंत्रालयों की रिपोर्ट, आर्थिक सर्वेक्षण 2014-2015, डी.आर.डी.ओ. की मासिक गृह पत्रिका, राजभाषा विभाग की वार्षिक रिपोर्ट और पुस्तकें आदि स्रोतों से भाषाई सामग्री का संकलन किया गया है। इस भाषाई सामग्री के विषय-क्षेत्र (Domain) मनोरंजन, कृषि, राजनीति, विज्ञान और प्रौद्योगिकी, कला और संस्कृति, पर्यटन, समाज और शिक्षा, स्वास्थ्य, खेल कूद, अर्थव्यवस्था, इतिहास, दर्शन, धर्म, राष्ट्रीय सुरक्षा और रक्षा, कानून, समाज और समुदाय तथा साहित्य और भूगोल इत्यादि हैं।

#### 4.2.2 निर्गत अनुवाद (Translation output)

कॉर्पोरा के संकलन के बाद, अंग्रेजी से हिंदी में मशीनी अनुवाद प्राप्त करने के लिए 4 मशीनी अनुवाद तंत्रों का चयन किया गया जिसमें अनुवाद के लिए विभिन्न विषय-क्षेत्र के वाक्य दिए गए और उनसे अनुवाद प्राप्त किया। इसमें निम्न मशीनी अनुवाद के तंत्र शामिल हैं-

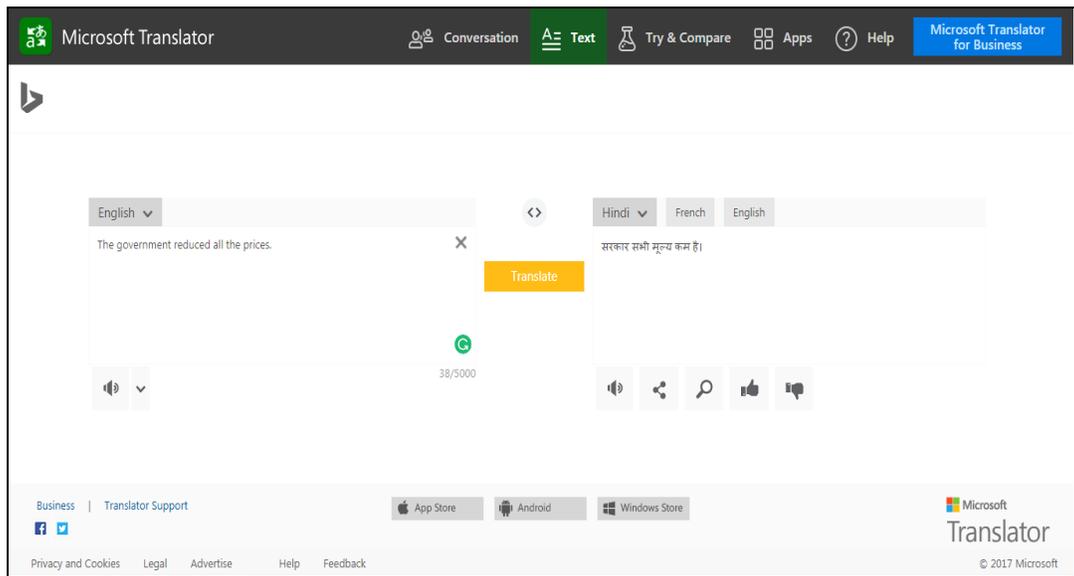
**1. गूगल अनुवादक :** इस तंत्र का प्रयोग विश्व की अनेक भाषाओं के अनुवाद के लिए अधिक किया जाता है। यह सबसे लोकप्रिय तंत्र है जो विश्व की 103 भाषाओं के लिए निःशुल्क अनुवाद करता है। इसे गूगल कार्पोरेशन ने विकसित किया है। इससे कॉर्पोरा के वाक्यों का अनुवाद अंग्रेजी से हिंदी में किया गया, जिसे नीचे स्क्रीनशॉट में दिया गया है।



स्क्रीनशॉट 4.1 : गूगल अनुवादक से निर्गत अनुवाद  
(<https://translate.google.com/> Retrieved on 17.05.2017)

इस स्क्रीनशॉट द्वारा दिखाई देता है कि अंग्रेजी वाक्य का हिंदी में गूगल अनुवादक द्वारा कैसे अनुवाद किया जाता है। ऐसे ही अनेक वाक्यों का हिंदी अनुवाद इस तंत्र द्वारा किया गया।

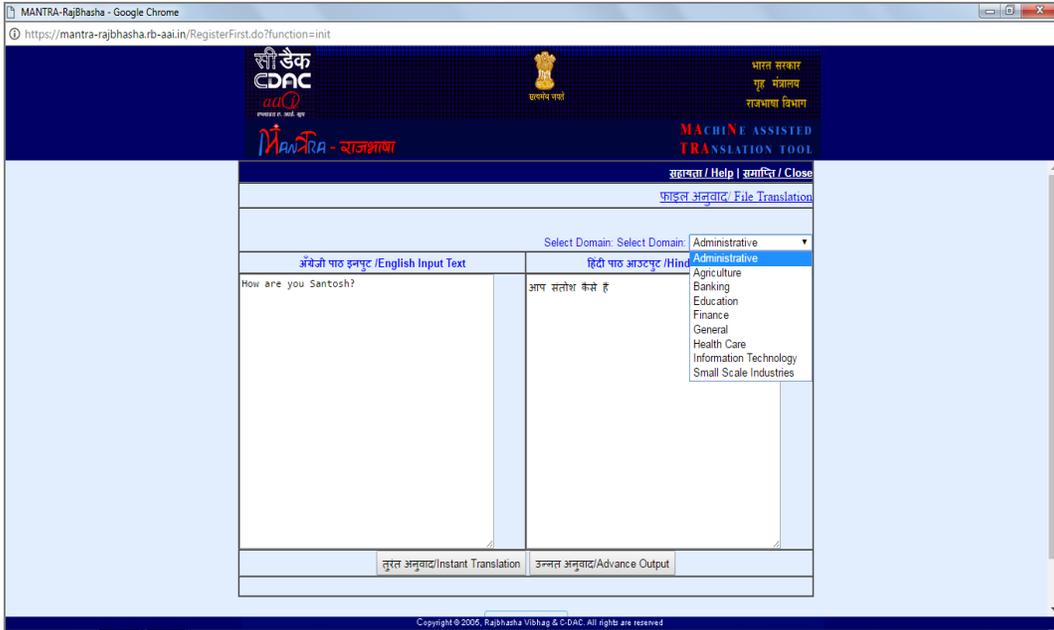
**2. माइक्रोसॉफ्ट अनुवादक :** यह विश्व की 60 भाषाओं के लिए निःशुल्क अनुवाद प्रदान करता है। इसे माइक्रोसॉफ्ट कॉर्पोरेशन ने विकसित किया है। यह तंत्र विभिन्न भाषाओं में स्वचालित रूप से कंप्यूटर द्वारा अनुवाद करता है। ये विश्व के प्रमुख मशीनी अनुवाद तंत्रों में से एक है। यह फरवरी 2017 से विश्व की 60 भाषाओं में अनुवाद प्रदान करता है। भारत की हिंदी और उर्दू भाषा का भी समावेश इस तंत्र में किया गया है। माइक्रोसॉफ्ट अनुवादक के द्वारा एक भाषा के पाठ (Text) का या वेबपेज (webpage) का दूसरी भाषा में अनुवाद किया जाता है। साथ ही पाठ (Text) से वाक् (Speech) प्रणाली के अंतर्गत निर्गत (Output) अनुवाद सुन सकते हैं। इस तंत्र से भी अंग्रेजी से हिंदी वाक्यों का अनुवाद किया गया है। जिसे नीचे स्क्रीनशॉट में दिया गया है।



स्क्रीनशॉट 4.2 : माइक्रोसॉफ्ट अनुवादक से निर्गत अनुवाद  
<https://www.bing.com/translator> (Retrieved on 18.05.2017)

इस तंत्र से भी कॉर्पोरा के वाक्यों का अनुवाद अंग्रेजी से हिंदी में किया और उसका मानव द्वारा मूल्यांकन किया गया।

**3. मंत्र-राजभाषा :** प्रशासनिक दस्तावेजों के अनुवाद के लिए इस तंत्र का प्रयोग किया जाता है। इसे सी-डैक, पुणे के एप्लाइड आर्टिफिशियल इंटेलीजेंस ग्रुप ने विकसित किया है। अब यह तंत्र प्रशासनिक दस्तावेजों के अलावा वित्तीय, कृषि, लघु उद्योग, सूचना प्रौद्योगिकी, स्वास्थ्य, रक्षा, शिक्षा एवं बैंकिंग क्षेत्रों के दस्तावेजों का अनुवाद करता है। इस तंत्र की रूपरेखा तथा इससे अनुवाद करने की प्रक्रिया को नीचे स्क्रीनशॉट में दिया गया है।



स्क्रीनशॉट 4.3 : मंत्र-राजभाषा से निर्गत अनुवाद

<https://mantra-rajbhasha.rb-aai.in/RegisterFirst.do?function=init>  
(Retrieved on 19.05.2017)

इसमें अनुवाद संबंधित सभी सहायक उपकरण जैसे कि थिसॉरस और शब्दकोश का समावेश इसमें किया गया है। जिससे यह अंग्रेजी से हिंदी में अनुवाद करता है। इससे भी कॉर्पोरा के वाक्यों का अनुवाद मूल्यांकन के लिए किया गया है।

4. **आंग्ल-हिंदी** : इस तंत्र को प्रो.आर.एम.के. सिन्हा और अजय जैन ने भारतीय प्रौद्योगिकी संस्थान, कानपुर में विकसित किया है। यह तंत्र नियम आधारित प्रणाली पर आधारित है जो अंग्रेजी से हिंदी के अलावा अन्य भारतीय भाषाओं के लिए अनुवाद करता है। आंग्ल-हिंदी मशीनी अनुवाद तंत्र का विकास व्यावहारिक सहायता प्रदान करने के उद्देश्य से किया गया है। जिसमें मशीन द्वारा 90 प्रतिशत और मानव द्वारा 10 प्रतिशत कार्य पश्च संपादन से किया जाता है। इस तंत्र की अंग्रेजी से हिंदी में अनुवाद प्रक्रिया को नीचे स्क्रीनशॉट में दिया गया है।



स्क्रीनशॉट 4.4 : आंग्ल-हिंदी तंत्र से निर्गत अनुवाद

[http://tdil-dc.in/components/com\\_mtsystem/CommonUI/homeMT.php](http://tdil-dc.in/components/com_mtsystem/CommonUI/homeMT.php) (Retrieved on 19.05.2017)

उपरोक्त स्क्रीनशॉट में अंग्रेजी वाक्य का हिंदी में अनुवाद किया गया है। इसमें ऐसी प्रणाली का प्रयोग किया है जो अधिक जटिल संरचना का अनुवाद करने के लिए सक्षम है। इसमें अंग्रेजी से भारतीय भाषाओं में पाठ प्रजनक मॉड्यूल का भी समावेश किया गया है। जिसमें स्रोत भाषा से लक्ष्य भाषा में अनुवाद के लिए कॉर्पोरा का प्रयोग कर नियमों के आधार पर पाठ का प्रजनन लक्ष्य भाषा में किया जाता है। इस तरह इन तंत्रों से अनुवाद प्राप्त होने के बाद मूल्यांकन की प्रक्रिया पूर्ण की गयी।

### 4.2.3 मानव मूल्यांकन

मानव द्वारा मूल्यांकन के लिए 24 विषय-क्षेत्रों (domains) के 18 हजार वाक्यों में से 500 वाक्यों मूल्यांकन करने के लिए 5 मूल्यांकनकर्ता का चयन किया गया। जो अंग्रेजी और हिंदी भाषा के जानकार हैं। प्रत्येक मूल्यांकनकर्ता को इन 4 तंत्रों से प्राप्त अनुवाद का मूल्यांकन करने के लिए दिया गया। साथ ही इसमें मशीनी अनुवाद की गुणवत्ता आधारित मूल्यांकन करने के लिए अनुवाद में बोधगम्यता, शुद्धता और सटीकता को ध्यान में रखा गया है। इसमें मूल वाक्य (अंग्रेजी) और अनूदित वाक्य (हिंदी) को देखते हुए 5 पॉइंट स्केल (0-4) पद्धति का प्रयोग कर मूल्यांकन किया गया है। अनुवाद की बोधगम्यता की दृष्टि से मानव मूल्यांकन के लिए निम्न स्केलिंग स्कोर का प्रयोग किया गया है।<sup>16</sup>

	<b>Evaluation on 5 point scale (0-4)</b>	
4	पूर्णतः बोधगम्य (All meaning )	for Flawless or Perfect:
3	लिप्यंतरित शब्द के साथ वाक्य बोधगम्य (Most meaning)	for Good or Comprehensible but has quite a few errors
2	वाक्य समझने में कुछ गलतियाँ (Much meaning)	for Non-native or Comprehensible but has quite a few errors
1	कुछ समझना, कुछ गलतियाँ, लेकिन बोधगम्य नहीं (little meaning)	for Diffluent or Some parts make sense but is not comprehensible over all
0	कुछ भी बोधगम्य नहीं (None)	Incomprehensible or Non-Sense

---

<sup>16</sup> Atul K.O., Bansal A., Hadke S., Jha GN., *Evaluation of Hindi-English MT Systems*, proceedings of the "2nd Workshop on Indian Language Data Resource and Evaluation" held under LREC, 26-31 May 2014, p.97

#### टेबल 4.1: मूल्यांकन के लिए 5 पॉइन्ट स्केलिंग पद्धति

मानव द्वारा मूल्यांकन के लिए प्रत्येक वाक्य के लिए इस स्केलिंग स्कोर का प्रयोग किया गया है। यह मशीनी अनुवाद तंत्र के मूल्यांकन के लिए मानक स्केलिंग पद्धति मानी जाती है। इससे भारतीय भाषाओं के लिए अनुवाद करने वाले तंत्र का मूल्यांकन करना आसान हो जाता है। अगर इस स्केलिंग स्कोर के लिए 5 पॉइन्ट स्केल से अधिक पॉइन्ट स्केल का प्रयोग करते हैं तो परिणाम परिशुद्धता और गुणवत्ता की दृष्टि से सही नहीं मिलते। इसलिए मानव मूल्यांकन के लिए इस 5 पॉइन्ट स्केल पद्धति का प्रयोग किया जाता है।

#### 4.2.4 कॉर्पोरा का क्षेत्रों के अनुसार मूल्यांकन

प्रत्येक मशीनी अनुवाद तंत्रों से प्राप्त अनूदित वाक्य के लिए उनके विषय-क्षेत्रों के अनुसार उन्हें स्केलिंग स्कोर दिए गए हैं। यहाँ उदाहरण के तौर पर नीचे टेबल में देखा जा सकता है कि गूगल अनुवादक द्वारा वाक्यों का स्कोर 5 मानव मूल्यांकनकर्ताओं ने अलग-अलग क्षेत्र के अनुसार दिया है। उसके बाद सभी क्षेत्रों के अनुसार वाक्यों का औसत (Average) निकाला गया है। गूगल अनुवादक के कृषि क्षेत्र का औसत मूल्यांकन स्कोर 2.8 है, व्यापार क्षेत्र का औसत स्कोर 2.6 है और अर्थव्यवस्था का औसत 2.8 है। इसका अर्थ गूगल अनुवादक सभी क्षेत्रों के लिए अनुवाद प्रदान करने में सक्षम है। इसी के तरह सभी क्षेत्रों के अनुसार माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी का औसत स्कोर निकाला गया है।

English Sentences	Hindi outputs	Domains	User 1 Score	User 2 Score	User 3 Score	User 4 Score	User 5 Score	Average
The study of agriculture is known as agricultural science.	कृषि के अध्ययन के कृषि विज्ञान के रूप में जाना जाता है।	Agriculture	3	3	2	3	3	2.8
Kalpana Construction completed its first project, the Kohinoor Plaza, on this plot of land.	कल्पना निर्माण अपनी पहली परियोजना है, कोहिनूर प्लाजा, देश के इस भूखंड पर पूरा किया।	Business	2	3	3	3	2	2.6
This budget is disappointing on both these counts.	इस बजट में इन दोनों मोर्चों पर निराशाजनक है।	Economy	2	3	3	3	3	2.8

टेबल 4.2 : गूगल अनुवादक के मूल्यांकन का औसत स्कोर

मशीनी अनुवाद के मूल्यांकन में बोधगम्यता को प्रमुख आधार माना गया है। सामान्यतः स्रोत भाषा अंग्रेजी से हिंदी में अनुवाद प्राप्त के बाद मूल्यांकन में व्याकरणिक त्रुटियाँ, अस्पष्ट और अनेकार्थी शब्द (ambiguous) और अनअनुदित शब्दों (un-translated words) की समस्याओं पर ध्यान दिया गया है।

कॉर्पोरा के सभी क्षेत्रों के मशीनी अनुवाद तंत्रों के अनुसार औसत स्कोर निकालने के बाद उनका कुल औसत स्कोर निकाला गया है, जिसे नीचे टेबल में दिया गया है।

	क्षेत्र / Domains	गूगल अनुवादक (औसत स्कोर)	माइक्रोसॉफ्ट अनुवादक (औसत स्कोर)	मंत्र-राजभाषा (औसत स्कोर)	आंग्ल-हिंदी (औसत स्कोर)
1	कृषि	2.4	2.48	1.92	1.9
2	कला और संस्कृति	2.36	2.6	1.92	2.26
3	व्यापार	2.36	2.32	1.96	2.24
4	प्रशासनिक सेवा	2.5	2.2	1.5	2.2
5	रक्षा	2.333333	2.333333333	2.0666667	2.4
6	अर्थव्यवस्था	2.476923	2.4	2.0846154	2
7	शिक्षा	2.309090909	2.47272727	1.6181818	2.054545
8	रोज़गार	2.36	1.96	2.04	1.84
9	पर्यावरण	2.6	2.4	2.2666667	2.2
10	फिल्म और मनोरंजन	2.16	2.12	1.47	2.04
11	खाद्य/food	2	2.257143	2	2.142857
12	भूगोल	2.65	2.5	1.85	2.45
13	स्वास्थ्य	2.605	2.715	1.975	2.1
14	इतिहास	2.470588235	2.470588235	1.929411765	2.047058824

15	भाषा और साहित्य	2.306666667	2.266666667	1.906666667	2.293333333
16	कानून/विधि	2.566666667	2.466666667	2.366666667	1.866666667
17	विविध	2.24	2.47	1.74	2.14
18	दर्शनशास्त्र	2.48	2.04	1.76	1.72
19	योजना	2.55	2.45	1.95	2
20	राजनीति	2.272727273	2.436363636	1.945454545	2.290909091
21	समाज और समुदाय	2.244444444	2.207407407	1.814814815	2.059259259
22	खेल	2.371428571	2.342857143	1.742857143	1.885714
23	प्रौद्योगिकी	2	1.733333	1.866667	2.333333
24	पर्यटन	2.65945946	2.486486	2.048649	2.178378
	<b>कुल औसत / Total Average</b>	<b>2.385927314</b>	<b>2.332546624</b>	<b>1.905318183</b>	<b>2.119219747</b>

टेबल 4.3 : विषय-क्षेत्रों के अनुसार म. अनु. तंत्रों के मूल्यांकन का औसत स्कोर

1. **कृषि क्षेत्र** : मूल्यांकन के लिए प्रथम कृषि क्षेत्र के वाक्य लिए गए और मशीनी अनुवाद तंत्रों के अनुसार औसत स्कोर निकाला गया। इसमें अनुवाद के लिए दिए गए वाक्यों का औसत स्कोर 2.4 है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.48, मंत्र-राजभाषा का औसत स्कोर 1.92 और आंग्ल-हिंदी मशीनी अनुवाद तंत्र का औसत स्कोर 1.9 है। ये सभी औसत स्कोर इन तंत्रों के द्वारा अनूदित वाक्यों की बोधगम्यता को दर्शाते हैं। गूगल अनुवादक में कई विषय-क्षेत्रों का समावेश किया गया है। उपरोक्त कृषि क्षेत्र के औसत स्कोर से ज्ञात होता है यह सामान्य अनुवाद करने

के लिए सही है, लेकिन इसके अनुवाद के दौरान कई ऐसे शब्द हैं जिनका अनुवाद हुआ नहीं। उन्हें हिंदी अनुवाद में वैसे ही रखा गया है। दूसरा तंत्र माइक्रोसॉफ्ट अनुवादक के स्कोर से भी जात होता है कि यह तंत्र भी इस क्षेत्र में अनुवाद करने के लिए सही है। इसी के साथ मंत्र-राजभाषा का प्रयोग प्रशासनिक दस्तावेजों के अनुवाद के लिए किया जाता है, लेकिन वर्तमान में इससे कई विषय-क्षेत्रों का अनुवाद किया जाता है। उसी तरह आंग्ल-हिंदी तंत्र का भी औसत स्कोर 1.9 है। इससे प्रतीत होता है कि यह तंत्र संदिग्धार्थक (ambiguous) शब्दों का अनुवाद करता है तथा कुछ शब्दों को अनुवाद नहीं होता तो उन्हें वैसे ही रहने देता है।

2. **कला और संस्कृति** : इस क्षेत्र के मूल्यांकन के लिए 4 मशीनी अनुवाद तंत्रों से 10 वाक्यों का अनुवाद प्राप्त किया। उसमें गूगल अनुवादक का औसत स्कोर 2.36 है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.6, मंत्र-राजभाषा का औसत स्कोर 1.92 और आंग्ल-हिंदी मशीनी अनुवाद तंत्र का औसत स्कोर 2.26 है। इसका अर्थ गूगल अनुवादक कला और पुरातत्व संबंधी वाक्यों का अनुवाद हिंदी में कृषि क्षेत्र के समान ही अनुवाद करता है। वही माइक्रोसॉफ्ट इस क्षेत्र में 2.6 औसत से सबसे अधिक बोधगम्य अनुवाद प्रदान करता है। मंत्र-राजभाषा इस तरह के क्षेत्र का अनुवाद कम बोधगम्य करता है क्योंकि इसके अनुवाद में शब्दक्रम की समस्या आती है। वाक्य में कोई शब्दक्रम न होने की वजह से वाक्य का अनुवाद अस्पष्ट होने से अर्थ समझना मुश्किल होता है। इसलिए इसका औसत स्कोर सभी तंत्रों से कम है। उसके बाद आंग्ल-हिंदी के स्कोर से प्रतीत होता है की यह तंत्र सामान्य अनुवाद करने के लिए सक्षम है। ये इसके औसत स्कोर 2.26 से स्पष्ट होता है।
3. **व्यापार** : इस क्षेत्र के मूल्यांकन में गूगल अनुवादक का औसत स्कोर 2.36 है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.32, मंत्र-राजभाषा का औसत स्कोर 1.96 और आंग्ल-हिंदी मशीनी अनुवाद तंत्र का औसत स्कोर 2.24 है। इन सभी तंत्रों के औसत स्कोर में मंत्र-राजभाषा का स्कोर सबसे कम है। क्योंकि यह तंत्र अंग्रेजी से हिंदी अनुवाद में अक्सर कुछ शब्दों को वैसे का वैसे ही अंग्रेजी में रखता है तथा

इसके अनुवाद के शब्दक्रम में एकरूपता न होने की वजह से अनुवाद बोधगम्य नहीं होता इस कमी की वजह से इसका औसत स्कोर 1.96 है। गूगल, माइक्रोसॉफ्ट और आंग्ल-हिंदी मशीनी अनुवाद तंत्र व्यापार क्षेत्र में लगभग समान अनुवाद करते हैं। इसलिए इनका मूल्यांकन स्कोर लगभग समान है।

4. **प्रशासन** : इस क्षेत्र की सामग्री भारतीय प्रशासनिक सेवा की रिपोर्ट के होने की वजह से इसके भी 4 वाक्य मशीनी अनुवाद तंत्रों को दिए और उनसे अनुवाद प्राप्त करने के बाद उनका मूल्यांकन किया गया। जिसमें गूगल अनुवादक का औसत स्कोर 2.5 है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.2, मंत्र-राजभाषा का औसत स्कोर 1.5 और आंग्ल-हिंदी का औसत स्कोर 2.2 है। गूगल और माइक्रोसॉफ्ट अनुवादक इस विषय-क्षेत्र का अनुवाद अच्छा करता है, लेकिन किसी व्यक्ति या स्थान के शब्द का अनुवाद करते समय ये तंत्र उस शब्द को हिंदी अनुवाद में वैसे ही रखते हैं, जबकि उसका हिंदी में लिप्यंतरण होना चाहिए। मंत्र-राजभाषा का स्कोर सबसे कम है, जिसमें 1.5 स्कोर प्राप्त हुआ है।
5. **रक्षा क्षेत्र** : इस क्षेत्र के वाक्यों के अनुवाद का मूल्यांकन गूगल द्वारा औसत स्कोर 2.333333 है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.333333333 है, मंत्र-राजभाषा का औसत स्कोर 2.0666667 है और आंग्ल-हिंदी का औसत स्कोर 2.4 है। रक्षा क्षेत्र में विशेष शब्दावली का प्रयोग किया जाता है। इसलिए उसके अनुवाद के लिए इन तंत्रों में इस तरह की शब्दावली का समावेश होना आवश्यक है। इस मूल्यांकन के औसत स्कोर से ज्ञात होता है कि गूगल और माइक्रोसॉफ्ट की अनुवाद करने की प्रक्रिया इस क्षेत्र के लिए लगभग समान है और मंत्र-राजभाषा में अनुवाद के कुछ समस्याओं की वजह से इसका स्कोर कम है। उसी तरह आंग्ल-हिंदी के मूल्यांकन स्कोर से ज्ञात होता है कि यह अन्य तंत्रों की तरह सामान्य अनुवाद करता है।
6. **अर्थव्यवस्था** : मशीनी अनुवाद तंत्रों से 25 वाक्यों का अनुवाद प्राप्त करने के बाद उनका मूल्यांकन किया गया। उसमें गूगल अनुवादक का औसत स्कोर 2.4769 है।

माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.4 है, मंत्र-राजभाषा का औसत स्कोर 2.08462 है और आंग्ल-हिंदी का औसत स्कोर 2 है। अर्थव्यवस्था के विषय-क्षेत्र में गूगल अनुवादक अच्छा अनुवाद करता है लेकिन शब्दक्रम की संरचना सही न होने की वजह से वाक्य में कुछ शब्द क्लिष्ट लगते हैं। माइक्रोसॉफ्ट का मूल्यांकन स्कोर भी गूगल की तरह है। मंत्र-राजभाषा ने इस क्षेत्र का समावेश अपने अनुवाद तंत्र में करने के कारण इस क्षेत्र के लिए कुछ सीमा तक सही अनुवाद किया है। आंग्ल-हिंदी का स्कोर सबसे कम है, क्योंकि इस क्षेत्र के अनुवाद में काफ़ी गलतियाँ मिली हैं।

7. **शिक्षा** : इस क्षेत्र में गूगल अनुवादक का औसत स्कोर 2.3090 के लगभग है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.4727 है, मंत्र-राजभाषा का औसत स्कोर 1.61818 है और आंग्ल-हिंदी का औसत स्कोर 2.0545 है। इसमें माइक्रोसॉफ्ट अन्य सभी तंत्रों से अच्छा अनुवाद करता है यह औसत स्कोर से ज्ञात होता है। उसके बाद गूगल भी अनुवाद करने के लिए सही तंत्र माना जा सकता है। मंत्र-राजभाषा का औसत स्कोर बहुत कम है। इसका अर्थ है कि इस तंत्र को अधिक भाषाई सामग्री का समावेश कर अपने तंत्र को अद्यतन (update) करने की आवश्यकता है। साथ ही आंग्ल-हिंदी को भी अधिक परिष्कृत करने की आवश्यकता है। इसमें प्रायः लिप्यंतरण की समस्या प्रमुख पाई गई है।
8. **रोज़गार** : मशीनी अनुवाद तंत्रों से प्राप्त अनुवाद का मूल्यांकन करने के बाद गूगल अनुवादक का औसत स्कोर 2.36 है, माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 1.96 है, मंत्र-राजभाषा का औसत स्कोर 2.04 है और आंग्ल-हिंदी का औसत स्कोर 1.84 है, जिसमें गूगल और मंत्र-राजभाषा लगभग समान अनुवाद करते हैं। माइक्रोसॉफ्ट और आंग्ल-हिंदी के औसत स्कोर से ज्ञात होता है कि ये दोनों तंत्र अनुवाद में वाक्य स्तर पर गलतियाँ करते हैं। उसकी वजह से अनुवाद पढ़ने में अस्पष्ट लगता है।
9. **पर्यावरण** : इस क्षेत्र में अनुवाद के लिए 10 वाक्य देने के बाद उनके मूल्यांकन के परिणाम तंत्र के अनुसार अलग-अलग आए हैं। गूगल अनुवादक का औसत स्कोर 2.6 है, माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.4 है, मंत्र-राजभाषा का औसत स्कोर

2.26667 है और आंग्ल-हिंदी का औसत स्कोर 2.2 है। इसका अर्थ है कि ये सभी तंत्र इस क्षेत्र में अच्छा अनुवाद प्रदान करते हैं। लेकिन इस विषय की शब्दावली और अनुवाद पाठ अलग होने की वजह से इन तंत्रों में भी इस शब्दावली का समावेश करने की आवश्यकता है। साथ ही उसमें शब्दक्रम (word order) में सुधार करने की आवश्यकता है क्योंकि अंग्रेजी और हिंदी की भाषा संरचना अलग होने की वजह से अनुवाद के लिए भी वैसे प्रोग्राम संचालित करने की आवश्यकता है।

**10. फिल्म और मनोरंजन :** यह ऐसा क्षेत्र है जहाँ प्रतिदिन नए शब्दों का निर्माण होता है जिससे अनुवाद के लिए समस्याएँ निर्मित होती हैं। इसलिए नए शब्दों का समावेश सभी मशीनी अनुवाद तंत्रों में होना चाहिए। इस क्षेत्र के वाक्य का निर्गत अनुवाद प्राप्त करने के बाद उनके मूल्यांकन स्कोर में गूगल अनुवादक का औसत स्कोर 2.16 है। इससे अनुवाद प्राप्त करने वाले काफी प्रयोक्ता होने की वजह से इसे अद्यतन किया जाता है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.12 है, जो साधारण स्कोर है। इस क्षेत्र के लिए इससे अधिक अनुवाद करने के लिए अधिक भाषाई सामग्री का समावेश किया जा रहा है तथा इसे अद्यतन प्रत्येक माह में किया जाता है। मंत्र-राजभाषा का औसत स्कोर 1.47 है, जो सबसे कम है और आंग्ल-हिंदी का औसत स्कोर 2.04 है। इसके अनुवाद में कुछ संस्कृतनिष्ठ शब्दों का अनुवाद किया जाता है जिससे पाठ या वाक्य पढ़ने में अस्पष्ट और गलत लगता है।

**11.आहार (food) :** इस क्षेत्र में गूगल अनुवादक का औसत स्कोर 2 है, जो सामान्यतः कम ही है अन्यथा यह अन्य तंत्रों से अच्छा अनुवाद करता है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.2571 है, जो गूगल से भी अधिक स्कोर है। इससे प्रतीत होता है कि इस विषय-क्षेत्र के लिए इससे सही अनुवाद होता है। मंत्र-राजभाषा का औसत स्कोर 2 है और आंग्ल-हिंदी का औसत स्कोर 2.1429 है। दोनों तंत्रों में हिंदी अनुवाद की वाक्य संरचना के स्तर पर द्विअर्थकता की समस्या निर्मित होती है तथा अनुवाद में सरल शब्दों का प्रयोग करने के बजाय अनूदित वाक्य में क्लिष्ट शब्द प्राप्त होते हैं जिससे अनुवाद की बोधगम्यता कम होती है।

12. **भूगोल** : इस क्षेत्र से सामग्री अनुवाद के लिए देने के बाद उनके मूल्यांकन का परिणाम गूगल द्वारा औसत स्कोर 2.65 प्राप्त हुआ है, जो अनुवाद की गुणवत्ता की दृष्टि से सही है क्योंकि मशीनी अनुवाद से पूर्णतः सही (Perfect) अनुवाद प्राप्त नहीं कर सकते, अनुवाद में कुछ त्रुटियाँ अवश्य मिलती हैं। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.5 है, मंत्र-राजभाषा का औसत स्कोर 1.85 है और आंग्ल-हिंदी का औसत स्कोर 2.45 है। इसका अर्थ यह है कि इस क्षेत्र से संबंधित सामग्री और शब्दावली का समावेश अंग्रेजी से हिंदी अनुवाद करने वाले तंत्र में किया गया है। इसलिए इन सभी तंत्रों से अनुवाद काफ़ी मात्रा में सही पाया गया ऐसा इस क्षेत्र के औसत स्कोर से ज्ञात होता है।

13. **स्वास्थ्य** : इस क्षेत्र से संबंधी वाक्य भारतीय भाषा कॉर्पोरा उपक्रम<sup>17</sup> और अन्य पत्रिकाओं से लिए गए हैं। 40 वाक्य अनुवाद के लिए सभी तंत्रों में देने के बाद हिंदी में अनुवाद प्राप्त किया और उनका मूल्यांकन वाक्य स्तर पर किया गया। जिसमें गूगल अनुवादक का औसत स्कोर 2.605 है। इस तंत्र द्वारा स्वास्थ्य संबंधी पाठ तथा वाक्य का अनुवाद सहज होता है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.715 है, जो गूगल से भी अधिक है। इसका कारण इसमें स्वास्थ्य संबंधी सामग्री तथा शब्दावली का समावेश किया गया है, इसलिए इसके परिणाम सभी तंत्रों से अलग हैं। मंत्र-राजभाषा का औसत स्कोर 1.975 है, जो अन्य तंत्रों से कम है। इसमें वाक्य संरचना के स्तर पर होने वाली त्रुटियों को दुरुस्त करने की आवश्यकता है। अंत में आंग्ल-हिंदी का औसत स्कोर 2.1 से स्पष्ट होता है कि यह तंत्र सामान्य अनुवाद करता है।

14. **इतिहास** : मशीनी अनुवाद का मूल्यांकन में गूगल अनुवादक और माइक्रोसॉफ्ट अनुवादक का औसत स्कोर लगभग 2.48 है, दोनों तंत्रों द्वारा अनुवाद के लिए अधिक सामग्री का समावेश किया गया है इसलिए दोनों के मूल्यांकन परिणाम समरूप आ रहे हैं। मंत्र-राजभाषा का औसत स्कोर 1.929 है और आंग्ल-हिंदी का

---

<sup>17</sup> <http://sanskrit.jnu.ac.in/projects/ilci.jsp?proj=ilci> (retrieved on 22.06.2017)

औसत स्कोर 2.0470 है। मंत्र-राजभाषा और आंग्ल-हिंदी तंत्रों के स्कोर से स्पष्ट होता है कि अनुवाद में दोनों से अधिक गलतियाँ होती हैं। विशेषतया हिंदी अनुवाद में शब्द स्तर पर संदिग्धता (Ambiguity) प्राप्त होती है जिससे संपूर्ण वाक्य का अर्थ समझने में कठिन लगता है।

**15. भाषा और साहित्य :** इस क्षेत्र में मशीनी अनुवाद का मूल्यांकन स्कोर, गूगल अनुवादक द्वारा 2.3066 के लगभग है और माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.2666 है, इन दोनों तंत्रों के औसत स्कोर में अंशतः भिन्नता है। मंत्र-राजभाषा का औसत स्कोर 1.9066 है और आंग्ल-हिंदी का औसत स्कोर 2.2933 है। इस क्षेत्र के अनुवाद में वाक्य, भाषा और साहित्य से संबंधित होने की वजह से वाक्य संरचना की अलग शैली होती है अधिकतर वाक्य पद्य या लोकगीत से संबंधित होने की वजह से हिंदी अनुवाद में शब्द के अर्थांतरण की समस्या निर्मित हुई। साथ ही किसी व्यक्ति या स्थान के नाम के अनुवाद में गलतियाँ प्राप्त हुईं। जिसकी वजह से मंत्र-राजभाषा और आंग्ल-हिंदी का स्कोर सामान्य प्राप्त हुआ है।

**16. कानून/विधि :** इस क्षेत्र से मशीनी अनुवाद तंत्रों को दिए गए वाक्यों के मूल्यांकन का औसत स्कोर गूगल अनुवादक द्वारा 2.5666 है, माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.4666 के लगभग है उसके बाद मंत्र-राजभाषा का औसत स्कोर 2.3666 है और आंग्ल-हिंदी का औसत स्कोर 1.8666 है। सभी तंत्रों से अनूदित वाक्य सही थे। अधिकतर गलतियाँ आंग्ल-हिंदी के अनुवाद में मिलीं। अनुवाद में आंकड़ों का अनुवाद देवनागरी लिपि में प्राप्त हुआ और हिंदी अनुवाद में अधिकतर अंग्रेजी शब्दों को वैसे ही रखा गया है। इसलिए इसके औसत स्कोर अन्य अनुवाद तंत्रों से काफी कम है। मंत्र-राजभाषा तंत्र विधि क्षेत्र के वाक्य का अधिक सही और सरल अनुवाद करता है। इस विषय-क्षेत्र की भी अलग शब्दावली होती है जिसे अनुवाद तंत्रों में शामिल करना होता है, अन्यथा कानूनी शब्द का सामान्य अर्थ में अनुवाद हो सकता है।

17. **दर्शनशास्त्र** : इस क्षेत्र से मशीनी अनुवाद के निर्गत वाक्यों का मूल्यांकन स्कोर गूगल अनुवादक द्वारा 2.48 है, जिससे अन्य तंत्रों से अधिक स्कोर मिला है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.04 है जो सामान्य स्कोर है जिसमें शब्दक्रम की समस्याएँ मिली हैं। इसे अधिक अद्यतन करने की आवश्यकता है। उसके बाद मंत्र-राजभाषा का औसत स्कोर 1.76 है और आंग्ल-हिंदी का औसत स्कोर 1.72 है। इन दोनों तंत्रों से अनुवाद में गलतियाँ होने की वजह से इनका औसत स्कोर कम है। वाक्य स्तर पर अनुवाद में शाब्दिक अनेकार्थकता और लिप्यंतरण की समस्याएँ पाई गई हैं, जिसे अनुवाद की बोधगम्यता और सहजता के लिए दुरूस्त करना आवश्यक है।

18. **योजनाएँ** : इस क्षेत्र से संबंधित सामग्री भारत सरकार की योजनाएँ और नीतियों से संकलित की गई है और उसे 4 मशीनी अनुवाद तंत्रों में अनुवाद के लिए देने के बाद उनसे निर्गत अनुवाद का मूल्यांकन किया गया। उनमें गूगल अनुवादक द्वारा औसत स्कोर 2.55, माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.45 है, मंत्र-राजभाषा का औसत स्कोर 1.95 है और आंग्ल-हिंदी का औसत स्कोर 2 है। इसमें गूगल ने भी अनुवाद में गलतियाँ की हैं। विशेषतः योजनाओं के नाम का अनुवाद करते समय सभी तंत्रों ने गलतियाँ की हैं इन सभी तंत्रों में लिप्यंतरण की भी समस्या पाई जाती है। अक्सर किसी नाम का अनुवाद करते समय ये मशीनी अनुवाद तंत्र उसका शाब्दिक अनुवाद कर देते हैं जिससे अनुवाद में अनेकार्थकता की समस्या निर्मित होती है। जैसे गूगल द्वारा 'संतोष जा रहा है' वाक्य का अनुवाद 'Satisfaction is going on' किया है। इसका अर्थ नाम का शाब्दिक अनुवाद हुआ है। इस तरह की समस्या का समाधान करना मशीनी अनुवाद के सामने चुनौती है।

19. **राजनीति** : इस क्षेत्र से संबंधित वाक्यों के अनुवाद में गूगल अनुवादक का औसत स्कोर 2.2727 के लगभग है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.43636 है, जो अन्य तंत्रों से अच्छा स्कोर है, जिसमें इस विषय से संबंधित सामग्री का अनुवाद करने के लिए भाषाई सामग्री तथा द्विभाषिक शब्दावली का समावेश इस तंत्र में

किया गया है। मंत्र-राजभाषा का औसत स्कोर 1.94545 है, जो अन्य सभी मशीनी अनुवाद तंत्रों से सबसे कम है, इसलिए इसे अधिक परिवर्धित करने की आवश्यकता है। आंग्ल-हिंदी का औसत स्कोर 2.29090 है। इस तंत्र के अनुवाद में वर्तनी संबंधी अधिक गलतियाँ पाई गई हैं तथा कुछ शब्दों के अनुवाद में अनेकार्थी शब्द भी मिले हैं जिससे अनुवाद अर्थ की दृष्टि से अस्पष्ट प्रतीत होता है। इसे अधिक विकसित करने के लिए सी-डैक नोएडा द्वारा अनुसंधान कार्य चल रहा है।

**20. समाज और समुदाय :** मशीनी अनुवाद के लिए 25 वाक्य देने के बाद उनसे निर्गत अनुवाद प्राप्त किया, जिसमें गूगल अनुवादक का औसत मूल्यांकन स्कोर 2.24444 के लगभग है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.207407 है। यह क्षेत्र साहित्य, समाज और संस्कृति से संबंधित होने की वजह से मशीनी अनुवाद द्वारा गलत अनुवाद पाया गया। मंत्र-राजभाषा का औसत स्कोर 1.814814 के पास है। इस तंत्र के अनुवाद में अंग्रेजी शब्दों को हिंदी में वैसे ही रखने से वाक्य स्तर पर अनुवाद गलत हो जाता है तथा शब्द में संदिग्धता उत्पन्न होती है और अनुवाद पाठ बोधगम्य नहीं होता। आंग्ल-हिंदी का औसत स्कोर 2.059259 है। इस तंत्र में भी अंग्रेजी शब्द 'muscleman brother' का 'मुस्क्लेमन भाई' और 'leftist movements' का 'लेफ़िस्त आंदोलनों' आदि अनुवाद किया है जिससे यह ज्ञात होता है कि इस तंत्र से वर्तनी की अधिक समस्या आती है जिसमें हिंदी वर्तनी जाँचक (Spell checker) में सुधार करने की आवश्यकता है।

**21. खेल :** इस विषय क्षेत्र से संबंधित वाक्य मशीनी अनुवाद तंत्रों में देने के बाद उनका मूल्यांकन स्कोर निकाला गया उनमें गूगल अनुवादक का औसत स्कोर 2.371428 के करीब है तथा माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.34285 के करीब है। दोनों तंत्रों से प्राप्त अनुवाद के मूल्यांकन से स्पष्ट होता है कि अनुवाद की दृष्टि से दोनों सामान्य अनुवाद करते हैं। अन्य तंत्र में मंत्र-राजभाषा का औसत स्कोर 1.742857 है और आंग्ल-हिंदी का औसत स्कोर 1.8857 है। दोनों के स्कोर में लगभग समानता दिखाई देती है। मंत्र-राजभाषा ने कुछ अंग्रेजी वाक्यों का अनुवाद रिक्त छोड़ा है।

जिससे अनुवाद का मूल्यांकन करना कठिन हो जाता है। आंग्ल-हिंदी से अनुवाद में वर्तनी की अधिक समस्या होने के कारण अनुवाद पढ़ने में दूरूह लगता है इसलिए इनके स्कोर कम है।

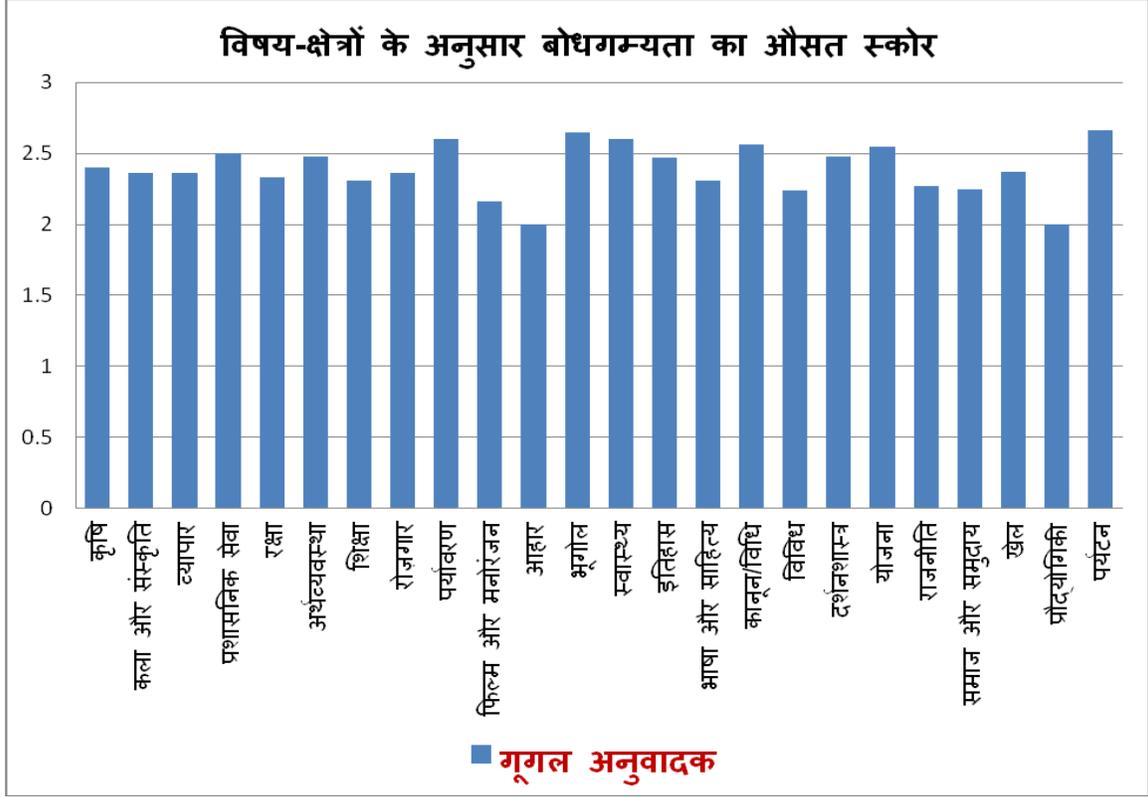
22. प्रौद्योगिकी : इस क्षेत्र में प्रतिदिन नए शब्दों का प्रयोग किया जाता है तथा नए शब्द गढ़े जाते हैं। मशीनी अनुवाद तंत्र अक्सर तकनीकी शब्दों का सामान्य अनुवाद करता है। इस विषय क्षेत्र के मशीनी अनुवाद के लिए दिए गए वाक्य का मूल्यांकन करने के बाद गूगल अनुवादक का औसत स्कोर 2 पाया गया है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 1.7333 है जो इस तंत्र के लिए बहुत कम है। मंत्र-राजभाषा का औसत स्कोर 1.8667 है। मंत्र-राजभाषा में सूचना प्रौद्योगिकी क्षेत्र का समावेश किया गया, इसलिए इसके अनुवाद में सुधार किया जा सकता है। आंग्ल-हिंदी का औसत स्कोर 2.3333 है। तुलनात्मक रूप से देखा जाए तो आंग्ल-हिंदी से इस क्षेत्र में सामान्य अनुवाद हुआ है लेकिन इससे अनुवाद में काफ़ी गलतियाँ प्राप्त हुई हैं। विशेषतः अंग्रेजी से हिंदी मशीनी अनुवाद में अंग्रेजी शब्दों का ही अनुवाद अधिक हुआ है जिससे अनुवाद में दुरुहता निर्मित होती है।

23. पर्यटन : इस क्षेत्र के वाक्य अंग्रेजी से हिंदी में मशीनी अनुवाद के लिए दिए गए और उनका मूल्यांकन स्कोर निकाला गया। उसमें गूगल अनुवादक का औसत स्कोर 2.66 के करीब पाया गया। जो अन्य सभी तंत्रों से अधिक है। इसका अर्थ यह है कि इस विषय का अनुवाद ये तंत्र सही करता है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.49 के करीब पाया गया। यह तंत्र अनुवाद में सामान्य गलतियाँ करता है। इसलिए गूगल के बाद अनुवाद करने के क्रम में दूसरा है। वाक्य की संरचना में इससे शब्दक्रम, स्थान और नाम के अनुवाद में समस्याएँ पाई जाती हैं। फिर भी अनुवाद करने की दृष्टि से अंतरराष्ट्रीय स्तर पर अनेक भाषाओं के लिए अनुवाद प्रदान करता है। भारत में विकसित मंत्र-राजभाषा का औसत स्कोर 2.048 से अधिक पाया गया। इसमें भी नाम और स्थान के अनुवाद की समस्याएँ पाई जाती है तथा हिंदी अनुवाद

करते समय अंग्रेजी शब्दों को वैसे ही रखा जाता है। आंग्ल-हिंदी का औसत स्कोर 2.18 के लगभग है। इसकी समस्याएँ भी अन्य अनुवाद तंत्रों की तरह हैं लेकिन इसमें शब्दों के वर्तनी की समस्या प्रमुख रूप में पाई जाती है।

**24. विविध (Miscellaneous):** इसमें अनेक उपक्षेत्र के वाक्यों का समावेश किया गया है इन वाक्यों का निर्गत अनुवाद प्राप्त करने के बाद मूल्यांकन किया गया उसमें गूगल अनुवादक का औसत स्कोर 2.24 है। माइक्रोसॉफ्ट अनुवादक का औसत स्कोर 2.47 है। यह तंत्र तेजी से विश्व की भाषाओं के अनुवाद के लिए अद्यतन (update) किया जा रहा है। इसलिए आनवाले दिनों में यह तंत्र अधिक अच्छा अनुवाद करने में सक्षम रहेगा। मंत्र-राजभाषा का औसत स्कोर 1.74 है, जिसे अनुवाद के लिए अधिक अद्यतन करने की आवश्यकता है। आंग्ल-हिंदी का औसत स्कोर 2.14 के लगभग है जिसमें अनुवाद के लिए विभिन्न भाषाई सामग्री तथा विभिन्न अनुवाद के उपकरणों का समावेश करने की आवश्यकता है, क्योंकि अनुवाद में इससे अधिक मात्रा में अनेकार्थक शब्दों का प्रयोग किया जाता है। इस तरह अनेकार्थकता की समस्या मशीनी अनुवाद में अधिक पाई जाती है जिससे वाक्य संरचना पढ़ने में अस्पष्ट लगती है।

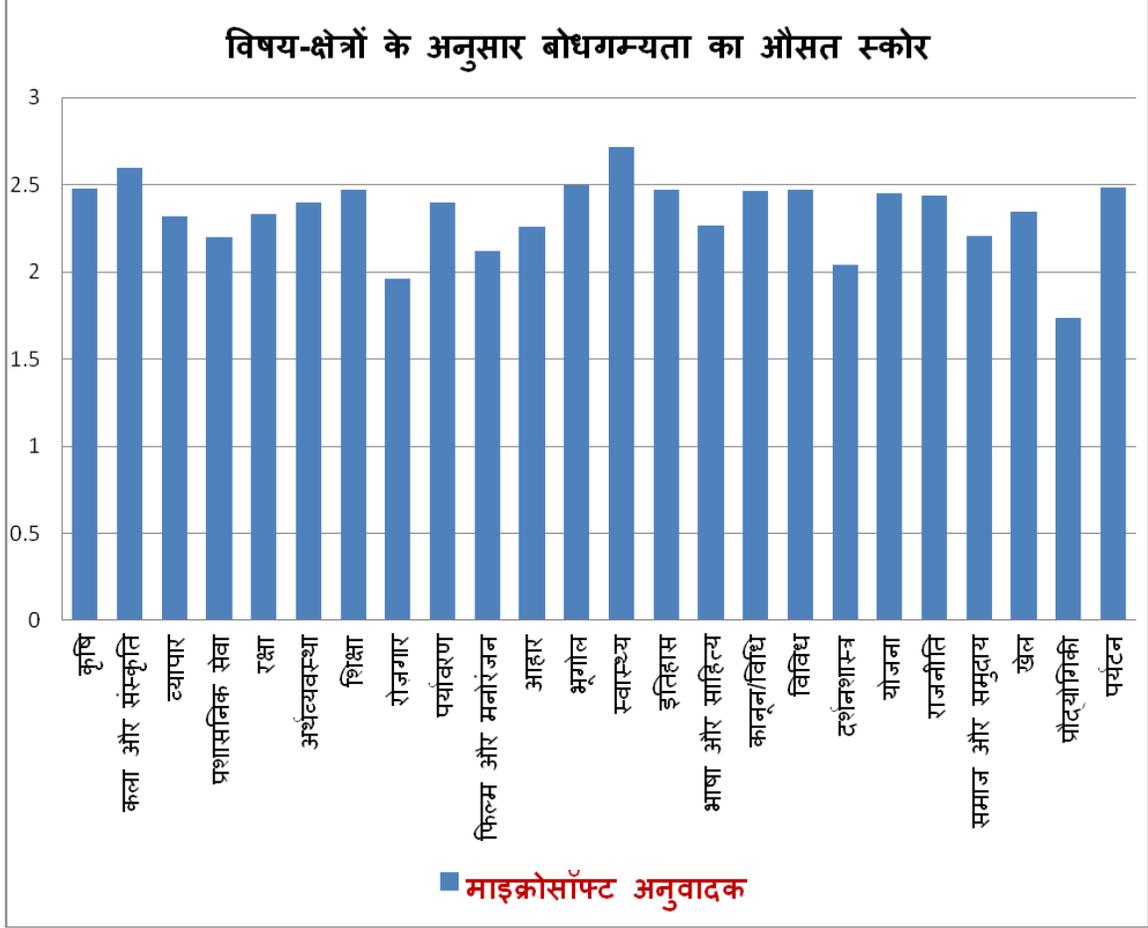
इन सभी क्षेत्रों के वाक्यों का अनुवाद इन 4 मशीनी अनुवाद तंत्रों में गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी से निर्गत अनुवाद के बाद क्षेत्रों के अनुसार अनुवाद की बोधगम्यता को देखते हुए मूल्यांकन से औसत स्कोर निकाले गए और उनका विश्लेषण किया गया है। इन औसत स्कोर को निम्न आकृति (बार चार्ट) द्वारा भी विवेचित किया गया है। जिसमें 0-4 स्केलिंग के स्कोर का प्रयोग कर उनके औसत स्कोर निकाले गए हैं। इससे यह स्पष्ट होता है कि किन विषय क्षेत्रों में ये तंत्र अनुवाद करने के लिए सक्षम हैं।



आकृति 4.1: गूगल अनुवादक का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर

उपरोक्त आकृति से स्पष्ट होता है कि पर्यटन, पर्यावरण, भूगोल, स्वास्थ्य, इतिहास, कानून, योजना, दर्शनशास्त्र और खेल आदि विषय-क्षेत्रों में गूगल अनुवादक सही और सरल अनुवाद करता है, क्योंकि इन सभी क्षेत्रों के मशीनी अनुवाद मूल्यांकन में बोधगम्यता का औसत स्कोर 2 से अधिक है। इन विषय-क्षेत्रों के लिए दिए जानेवाले वाक्यों का अनुवाद गूगल द्वारा अधिक बोधगम्य पाया जाता है। लेकिन प्रौद्योगिकी, आहार, विविध, भाषा और साहित्य, शिक्षा, रक्षा, राजनीति, समाज और समुदाय आदि विषय-क्षेत्रों में कुछ हद तक अनुवाद अबोधगम्य पाया जाता है।

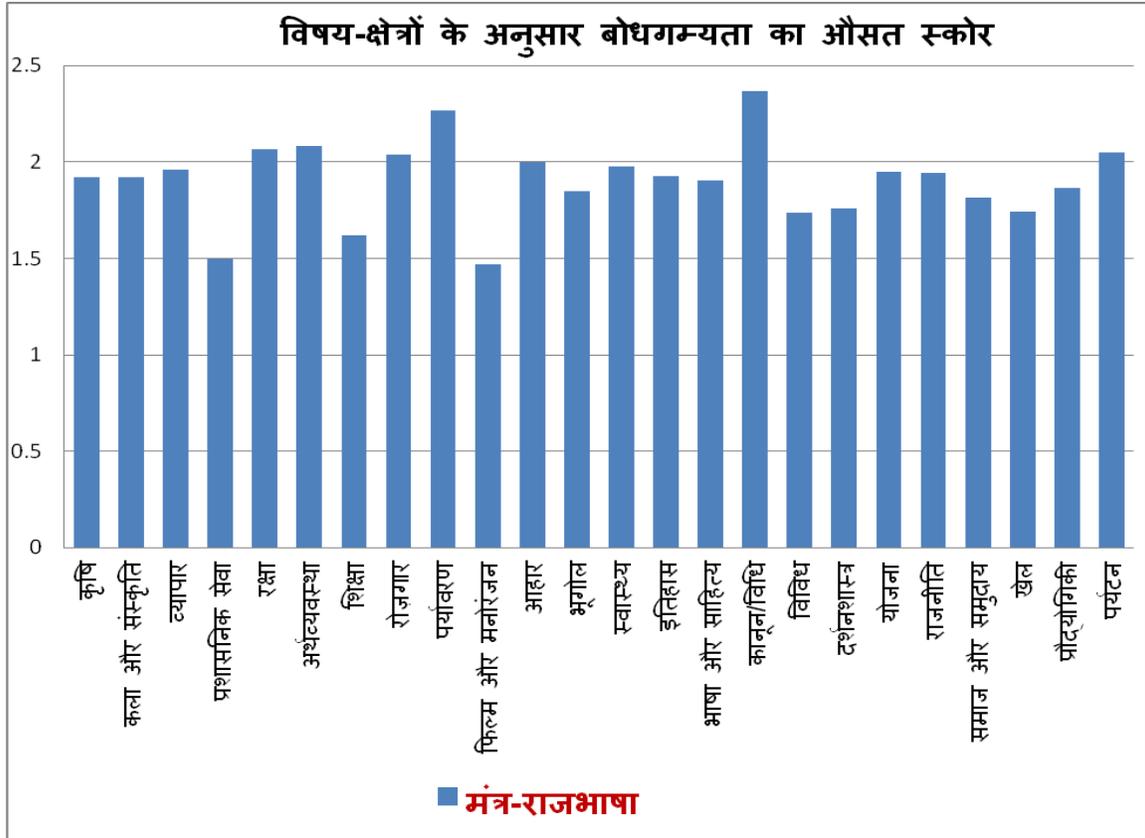
माइक्रोसॉफ्ट अनुवादक का विषय-क्षेत्रों के अनुसार औसत स्कोर निम्न आकृति से स्पष्ट हो रहा है।



आकृति 4.2: माइक्रोसॉफ्ट अनुवादक का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर

माइक्रोसॉफ्ट अनुवादक स्वास्थ्य, पर्यटन, कला एवं संस्कृति, कृषि, शिक्षा, भूगोल, इतिहास, विविध, योजना, विधि और राजनीति आदि विषय-क्षेत्रों में बोधगम्य अनुवाद करता है, लेकिन प्रौद्योगिकी, रोजगार, दर्शनशास्त्र, फिल्म और मनोरंजन, प्रशासन, आहार आदि विषय-क्षेत्रों के अनुवाद में अंशतः गलतियाँ होती हैं। इसलिए इनके औसत स्कोर 2 से कम और उसके लगभग पाए गए हैं। हालांकि इससे लक्ष्य भाषा के अनुवाद में अंग्रेजी शब्दों का प्रयोग अधिक किया जाता है तथा नाम एवं स्थान के अनुवाद में लिप्यंतरण की समस्याएँ पाई जाती हैं।

मंत्र-राजभाषा प्रशासनिक, वित्तीय, कृषि, लघु उद्योग, सूचना प्रौद्योगिकी, स्वास्थ्य रक्षा, शिक्षा एवं बैंकिंग क्षेत्रों के दस्तावेजों का अंग्रेजी से हिंदी में अनुवाद करता है। मंत्र टैक्नॉलाजी पर आधारित यह तंत्र सी-डैक, पुणे के एप्लाइड आर्टिफिशियल इंटेलीजेंस ग्रुप ने विकसित किया है। इस तंत्र से क्षेत्रों के अनुसार औसत निकाले गए हैं, वह स्कोर निम्न आकृति द्वारा स्पष्ट होते हैं।

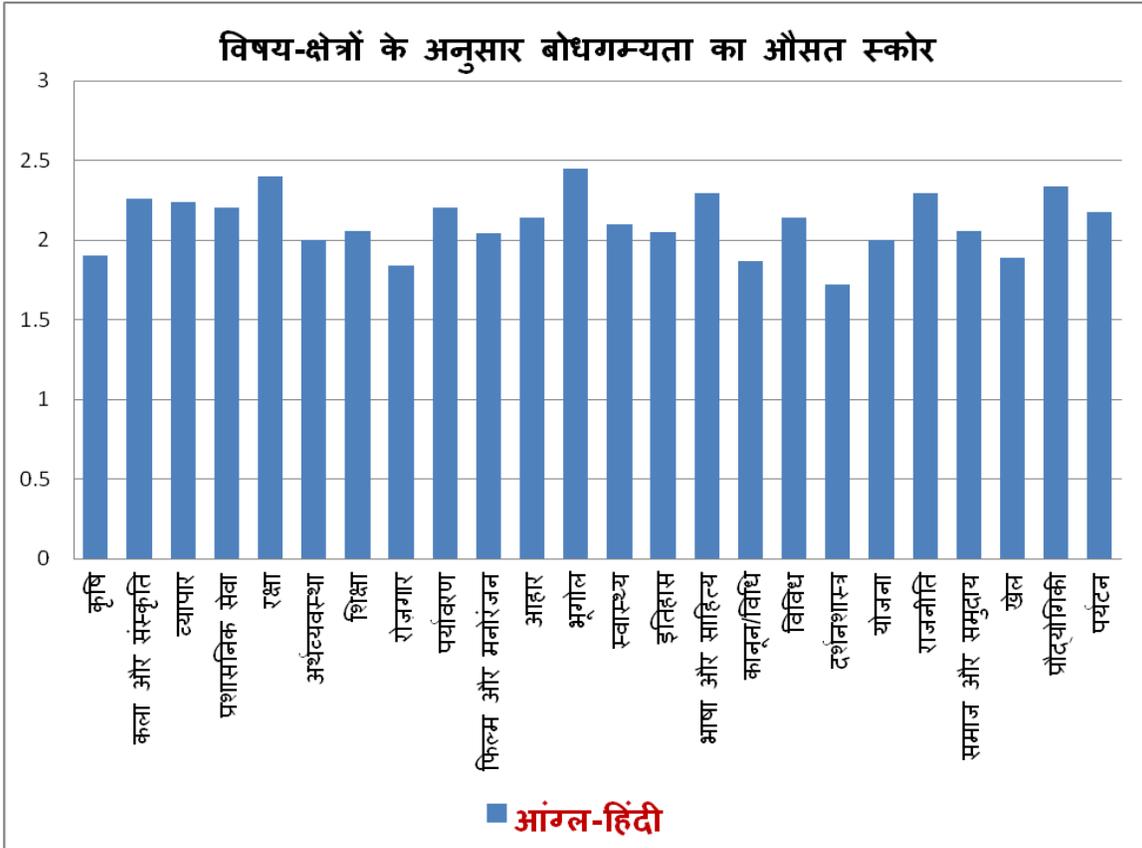


आकृति 4.3: मंत्र-राजभाषा तंत्र का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर

इस तंत्र ने कानून, पर्यावरण, रक्षा, अर्थव्यवस्था, रोजगार, आहार और पर्यटन आदि क्षेत्रों में अनुवाद सही और बोधगम्य किया है। वस्तुतः इस तंत्र का विकास प्रशासनिक दस्तावेजों का अनुवाद करने के लिए किया गया था लेकिन इसके औसत स्कोर से प्रतीत हो रहा है, कि प्रशासन से संबंधित कॉर्पोरा के वाक्य बड़े होने की वजह से इसके अनुवाद में गलतियाँ होती हैं। जिसके चलते सभी विषय-क्षेत्रों में कम औसत

स्कोर पाए गए हैं। इसके औसत स्कोर से ज्ञात होता है कि इसमें अधिक सुधार करने की आवश्यकता है।

आंग्ल-हिंदी का प्रयोग भी अंग्रेजी से भारतीय भाषाओं में अनुवाद के लिए किया जाता है। प्रस्तुत अध्याय में विभिन्न विषय-क्षेत्रों के अनुसार मूल्यांकन कर उनकी अनुवाद की बोधगम्यता को देखते हुए औसत स्कोर निकाला गया है, जिसे नीचे आकृति में दिया गया है।



आकृति 4.4: आंग्ल-हिंदी तंत्र का विषय-क्षेत्रों के अनुसार बोधगम्यता का औसत स्कोर

ये तंत्र प्रौद्योगिकी, राजनीति, भाषा और साहित्य, भूगोल, रक्षा, कला और संस्कृति, व्यापार, प्रशासन, पर्यटन, पर्यावरण, आहार, विविध आदि विषय-क्षेत्रों में अच्छा अनुवाद करता है। लेकिन दर्शनशास्त्र, खेल, रोजगार, कृषि, कानून, अर्थव्यवस्था, योजना आदि विषय-क्षेत्रों में अनुवाद बोधगम्य नहीं पाया गया इसलिए इन क्षेत्रों में औसत स्कोर

कम है। इसमें होने वाली वर्तनी की अधिक गलतियों को सुधार करने तथा इसे परिष्कृत करने के लिए सी-डैक, नोएडा द्वारा प्रयास चल रहे हैं।

इन सभी क्षेत्रों के अनुसार मशीनी अनुवाद तंत्रों के औसत स्कोर टेबल.न.4.3 निकालने के बाद अंत में मशीनी अनुवाद तंत्रों के अनुवाद की बोधगम्यता का कुल औसत स्कोर (Total average score) निकाला गया। जिसे नीचे टेबल न. 4.4 में दिया गया है।

<b>बोधगम्यता का कुल औसत स्कोर</b> <b>(Total average score of Comprehensibility)</b>			
गूगल अनुवादक	माइक्रोसॉफ्ट अनुवादक	मंत्र-राजभाषा	आंग्ल-हिंदी
2.385927314	2.332546624	1.905318183	2.119219747

टेबल 4.4 : मशीनी अनुवाद तंत्रों की बोधगम्यता का कुल औसत स्कोर

उपरोक्त दिए गए चारों मशीनी अनुवाद तंत्रों के कुल औसत स्कोर से स्पष्ट होता है कि गूगल अनुवादक सबसे अधिक बोधगम्य अनुवाद करता है। माइक्रोसॉफ्ट अनुवादक भी सभी क्षेत्रों में बोधगम्य तथा सही अनुवाद प्रदान करता है। उसके बाद मंत्र-राजभाषा का कुल औसत स्कोर कम होने से स्पष्ट होता है कि ये तंत्र वाक्य स्तर पर अनुवाद में अधिक गलतियाँ करता है। इसी के साथ आंग्ल-हिंदी का औसत स्कोर गूगल और माइक्रोसॉफ्ट के बाद अनुवाद की सही बोधगम्यता को दर्शाता है।

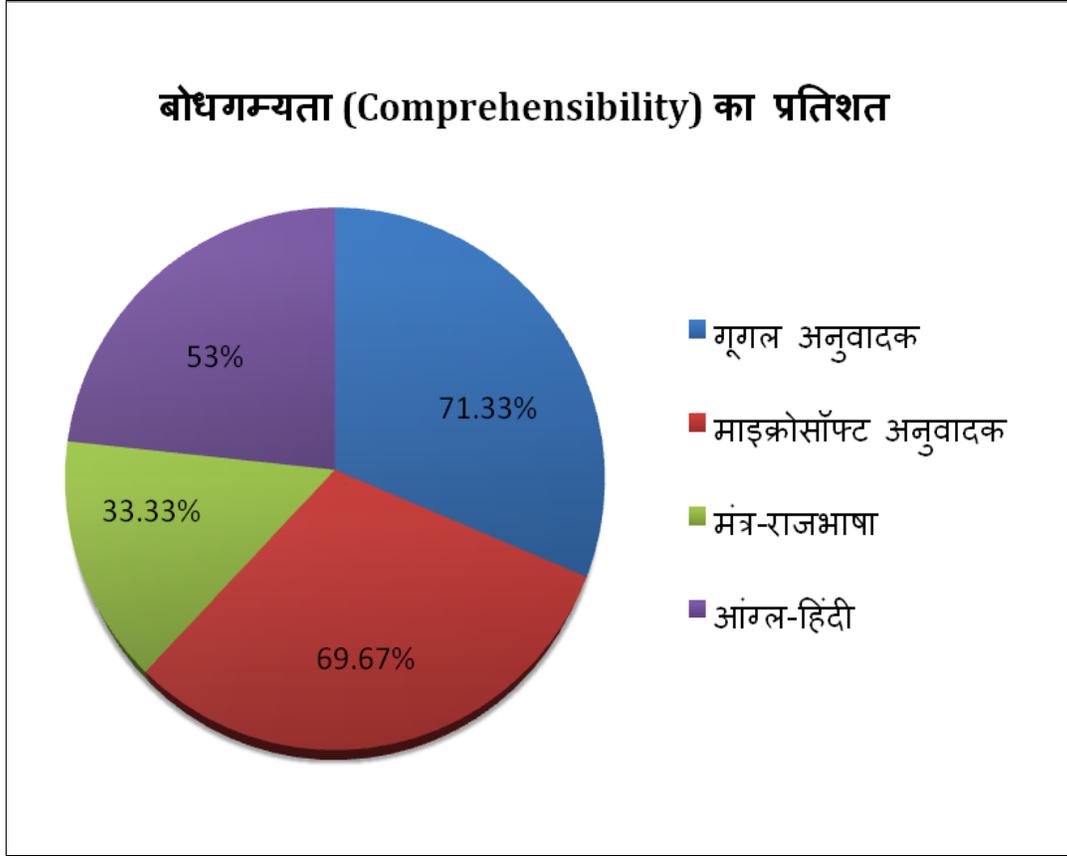
औसत स्कोर के आधार पर इन चारों मशीनी अनुवाद तंत्रों में गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी के औसत स्कोर से उनकी बोधगम्यता और अबोधगम्यता का प्रतिशत निकाला गया है। जिसे नीचे टेबल में दिया गया है।

मशीनी अनुवाद तंत्र	बोधगम्यता का (Comprehensibility) प्रतिशत	अबोधगम्यता का (Incomprehensibility) प्रतिशत
गूगल अनुवादक	71.33%	28.66%
माइक्रोसॉफ्ट अनुवादक	69.67%	30.33%
मंत्र-राजभाषा	33.33%	66.67%
आंग्ल-हिंदी	53%	47%

टेबल 4.5 : मशीनी अनुवाद तंत्रों की बोधगम्यता और अबोधगम्यता का प्रतिशत

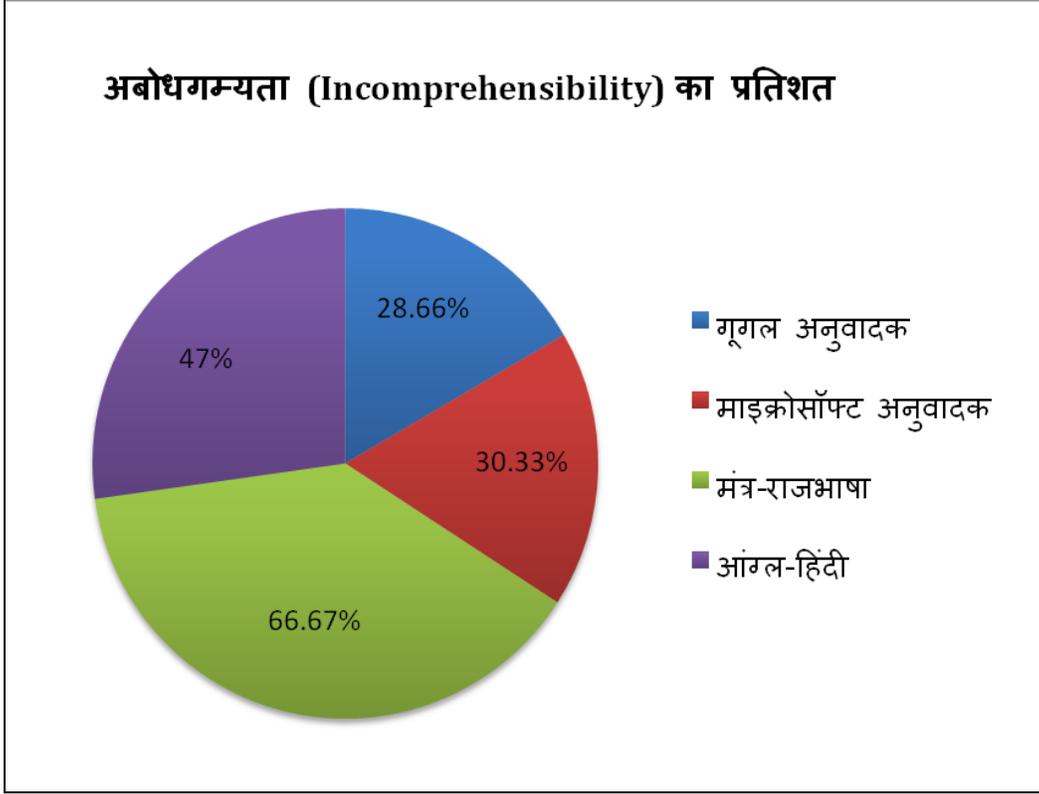
उपरोक्त टेबल में दिए गए सभी मशीनी अनुवाद तंत्रों की बोधगम्यता और अबोधगम्यता का प्रतिशत औसत स्कोर की सहायता से निकाला गया है। इसमें सभी मशीनी अनुवाद तंत्रों की बोधगम्यता को निर्धारित करने के लिए तंत्रों के पाँचों मूल्यांकनकर्ताओं के औसत स्कोर के आकड़े 2.1 से 4 तक के औसत स्कोर का समावेश किया गया है। इसी के साथ अबोधगम्यता को निर्धारित करने के लिए प्रत्येक मशीनी अनुवाद तंत्र के अनुसार 1 से 2 तक के संपूर्ण औसत स्कोर का समावेश किया गया है और अंत में उनके वाक्यों की संख्या के अनुरूप प्रतिशत निकाला गया है।

उपरोक्त टेबल.न. 4.5 के बोधगम्यता के प्रतिशत के आधार पर निम्न आरेखित आकृति में सभी मशीनी अनुवाद तंत्रों का प्रतिशत दिया गया है। जिसमें स्पष्ट होता है कि कौनसे मशीनी अनुवाद तंत्र अधिक बोधगम्य अनुवाद करने में सक्षम है और कौनसे मशीनी अनुवाद तंत्र अनुवाद करने में गलतियाँ करते हैं, जिससे उनकी अबोधगम्यता का पता चलता है।



आकृति 4.5: मशीनी अनुवाद तंत्रों की बोधगम्यता का प्रतिशत

उपरोक्त आकृति (पाइ चार्ट) द्वारा स्पष्ट होता है कि गूगल अनुवादक सर्वाधिक अर्थात् 71.33 प्रतिशत बोधगम्य अनुवाद करता है। इसके द्वारा स्रोत भाषा के सरल एवं छोटे वाक्यों का अनुवाद अधिक सही होता है। माइक्रोसॉफ्ट अनुवादक 69.67 प्रतिशत बोधगम्य अनुवाद करता है जो गूगल के बाद द्वितीय स्थान पर अनुवाद करने के क्रम में है। इसी के साथ मंत्र-राजभाषा केवल 33.33 प्रतिशत बोधगम्य अनुवाद करता है, इसका अर्थ इससे अनुवाद में अधिक गलतियाँ होती है। इसके बाद आंग्ल-हिंदी से अनुवाद की बोधगम्यता प्रतिशत 53 है, जो तृतीय क्रम में अनुवाद करता है। उपरोक्त सभी तंत्रों अंग्रेजी से हिंदी में अनुवाद किया जाता है लेकिन अनुवाद में त्रुटियाँ अनेक रूपों में पाई जाती है। इसलिए उनके अबोधगम्यता के प्रतिशत को निम्न आकृति में दिया गया है।



आकृति 4.6 : मशीनी अनुवाद तंत्रों की अबोधगम्यता का प्रतिशत

उपरोक्त आकृति (पाइ चार्ट) से स्पष्ट होता है कि गूगल अनुवादक से अंग्रेजी से हिंदी वाक्यों के अनुवाद में कम गलतियाँ हुई हैं। इसलिए इसकी अबोधगम्यता का प्रतिशत 28.66 है और माइक्रोसॉफ्ट अनुवादक की अबोधगम्यता का प्रतिशत 30.33 है। इसी के साथ मंत्र-राजभाषा की अबोधगम्यता का प्रतिशत 66.67 है, जो सबसे अधिक है। इस तंत्र से वाक्यों के अनुवाद में अनेक त्रुटियाँ पायी गयी हैं जिसका विश्लेषण पंचम अध्याय में दिया गया है। अंत में आंग्ल-हिंदी की अबोधगम्यता 47 प्रतिशत है। जो अबोधगम्यता के दृष्टि से कई अधिक है। इससे भी अनुवाद में वर्तनी, लिप्यंतरण और व्याकरणिक संरचना में अनेक गलतियाँ प्राप्त हुई हैं।

इस अध्याय में इस अंग्रेजी से हिंदी मशीनी अनुवाद तंत्रों के मूल्यांकन का प्रमुख उद्देश्य यही था कि ये सभी तंत्र कितना अबोधगम्य अनुवाद करते हैं तथा वाक्य संरचना के स्तर पर कौनसी गलतियाँ करते हैं, इस मूल्यांकन से स्पष्ट हो गया है। इसी के साथ

मूल्यांकन के बाद मशीनी अनुवाद तंत्रों से होने वाली त्रुटियों का विश्लेषण पांचवें अध्याय में विश्लेषित किया गया है।

निष्कर्ष में कहा जा सकता है कि प्रस्तुत अध्याय में मानव द्वारा मशीनी अनुवाद तंत्रों का मूल्यांकन सीमित कॉर्पोरा के वाक्यों के आधार पर किया गया है जिससे केवल अनुवाद तंत्रों की अनुवाद करने की बोधगम्यता और प्रवाहता स्पष्ट हुई है। मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए बृहद कॉर्पोरा में अनेक प्रकार के विषय-क्षेत्रों के वाक्यों का समावेश होना चाहिए, जिससे मूल्यांकन के परिणाम सही प्राप्त होंगे। विभिन्न भाषाओं के लिए अनुवाद करने वाले तंत्रों में गूगल अनुवादक और माइक्रोसॉफ्ट अनुवादक प्रमुख हैं। जो अनेक विषय-क्षेत्रों के लिए अनुवाद करते हैं लेकिन इनमें भी त्रुटियाँ पाई जाती हैं। अन्य दो अनुवाद तंत्र मंत्र-राजभाषा और आंग्ल-हिंदी भी अनुवाद करने के लिए सक्षम हैं लेकिन उनसे अनूदित पाठ में वर्तनी और अननूदित शब्दों की अनेक गलतियाँ पाई जाती हैं, जिन्हें सुधार करने की आवश्यकता है।

अतः भारतीय भाषाएँ अत्यधिक विभक्ति प्रत्ययों वाली भाषाएँ हैं। जो रूप-रचना की दृष्टि से शब्दक्रम मुक्त और कर्ता-कर्म-क्रिया की संरचना पर आधारित है। इसके अलावा इसमें कई शैलीगत भेद भी हैं। जबकि अंग्रेजी भाषा की संरचना कर्ता-क्रिया-कर्म पर आधारित है। यूरोपीय भाषाओं के लिए निर्माण की गयीं मशीनी अनुवाद मूल्यांकन की पद्धतियाँ अलग संरचना पर आधारित होने से ये भारतीय भाषाओं के लिए सही से कार्य नहीं करतीं। साथ ही भारतीय भाषाओं के अनुवाद में अक्सर अनेकार्थता की समस्या आती है जिसे केवल मानव ही मूल्यांकित कर सकता है। इसलिए भारतीय भाषाओं के मशीनी अनुवाद तंत्रों का मूल्यांकन मानव द्वारा करना ही अधिक विश्वसनीय है। हालांकि इन सभी मशीनी अनुवाद तंत्रों की गुणवत्ता में सुधार करने के लिए अद्यतन द्विभाषिक शब्दावली (Bilingual dictionary) और अनूदित पाठ में आने वाली शाब्दिक अस्पष्टता के लिए शब्द-आशय विसंदिग्धीकरण मॉड्यूल (Word sense disambiguation module) का प्रयोग किया जाना चाहिए। इसी के साथ ही अनुवाद में सुधार के लिए पश्च संपादन (post editing) की सुविधा भी होनी चाहिए। जिससे मशीनी अनुवाद तंत्रों की अनुवाद करने की गुणवत्ता बढ़ेगी और अनुवाद पढ़ने में बोधगम्य प्राप्त होगा।

# पंचम अध्याय

## मशीनी अनुवाद तंत्रों से निर्गत वाक्यों का त्रुटि विश्लेषण

5.1 त्रुटि विश्लेषण का परिचय

5.2 त्रुटियों का वर्गीकरण

5.3 मशीन अनूदित वाक्यों का त्रुटि विश्लेषण

5.3.1. वर्ण विचार संबंधी त्रुटियाँ

5.3.2. शब्द संरचना संबंधी त्रुटियाँ

5.3.3. शाब्दिक त्रुटियाँ

5.3.4. अर्थगत त्रुटियाँ

5.3.5 वाक्यात्मक त्रुटियाँ

5.4 मशीनी अनुवाद की त्रुटियों के लिए सुधारात्मक उपाय

## पंचम अध्याय

### मशीनी अनुवाद तंत्रों से निर्गत वाक्यों का त्रुटि विश्लेषण

#### (ERROR ANALYSIS OF MACHINE TRANSLATION SYSTEM OUTPUTS)

---

मशीनी अनुवाद की बढ़ती माँग को देखते हुए आज अनुवाद के लिए अनेक मशीनी अनुवाद तंत्रों का विकास किया गया है चाहे वह फिर भारतीय भाषाओं के लिए हो या फिर विदेशी भाषाओं के लिए। सभी भाषाओं के लिए मशीनी अनुवाद तंत्र अनुवाद कार्य कर रहे हैं, लेकिन कोई भी तंत्र पूर्ण रूप से अनुवाद करने में सक्षम नहीं है। जब कंप्यूटर से एक भाषा से दूसरी भाषा में अनुवाद किया जाता है तो उसमें बहुत गलतियाँ मिलती हैं और इन गलतियों की वजह से अनुवाद पढ़ने में कठिन लगता है। इन गलतियों को सही करने के लिए मशीनी अनुवाद तंत्रों का मूल्यांकन करना आवश्यक है ताकि इससे यह ज्ञात होगा कि कंप्यूटर से गलतियाँ कहाँ हो रही हैं। साथ ही अनुवाद की गुणवत्ता की दृष्टि से उनका मूल्यांकन और त्रुटि विश्लेषण करना आवश्यक है। इससे मशीनी अनुवाद तंत्र में सुधार करने में सहायता मिलेगी।

### 5.1 त्रुटि विश्लेषण का परिचय

त्रुटि विश्लेषण के प्रस्तावक एस.पी. कार्डर और उनके सहकर्मियों ने 1960 में इस संकल्पना का प्रयोग द्वितीय भाषा अधिगम के संदर्भ में किया था। उसके बाद में इसका प्रयोग अनुवाद के संदर्भ में किया जाने लगा।<sup>1</sup> आज इसका प्रयोग मशीनी अनुवाद से प्राप्त अनुवाद की त्रुटियों के विश्लेषण के लिए किया जाता है। किसी भी मशीनी अनुवाद तंत्र के निर्गत अनुवाद में पाई जानेवाली शुद्धता की कमी उस तंत्र की विश्वसनीयता और उपयोग को प्रतिकूल रूप से प्रभावित करती है। मूल्यांकन कार्य से यह पता लगता है कि

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Error\\_analysis\\_\(linguistics\)](https://en.wikipedia.org/wiki/Error_analysis_(linguistics)) (retrieved on 28..6.2017)

इन तंत्रों के परिणामों में क्या कमी है और किस तरह से इसमें अभाव हैं। उसी तरह त्रुटि विश्लेषण में मशीनी अनुवाद से अनूदित वाक्यों में पाई जानेवाली त्रुटियों का विश्लेषण किया जाता है और ज्ञात किया जाता है कि त्रुटियाँ कहाँ हो रही हैं। मशीनी अनुवाद तंत्र के विकास की प्रक्रिया में यह अपरिहार्य है कि निर्गत अनुवाद का मूल्यांकन शुद्धता, सरलता और स्वीकार्यता के मानदंडों के आधार पर सही होना चाहिए।

यह त्रुटि विश्लेषण की प्रक्रिया धीमी और नाजुक है, जिसमें मानव श्रम और समय की अधिक आवश्यकता होती है। इसके लिए उच्च स्तरीय विश्लेषण के लिए मशीनी अनुवाद तंत्रों की संरचना का सटीक ज्ञान होना चाहिए। मशीनी अनुवाद की त्रुटियों के विश्लेषण के लिए मानव के अलावा स्वचालित पद्धतियों का भी प्रयोग किया जाता है। इस व्यक्तिपरक और स्वचालित मूल्यांकन में मशीनी अनुवाद के गहन विश्लेषण के लिए अनुवाद में होने वाली त्रुटियों के स्रोत को जानना आवश्यक है। जहाँ वाक्य स्तर पर अनुवाद गलत होता है वहाँ त्रुटियों की जांच कर उसका विश्लेषण किया जाना चाहिए। यह त्रुटि विश्लेषण मशीनी अनुवाद तंत्रों के सुधारात्मक उपाय के लिए महत्वपूर्ण है।

## 5.2 त्रुटियों का वर्गीकरण

कंप्यूटर से स्रोत वाक्य के अनुवाद अनेक प्रकार के पाए जाते हैं। यह अनुवाद संदर्भ अनुवाद (Reference translation) के साथ अधिक या कम अनुरूप हो सकता है। इस अनुवाद में प्राप्त त्रुटियों का विश्लेषण संदर्भ अनुवाद और निर्गत अनुवाद को देखते हुए मानव द्वारा तथा स्वचालित रूप से किया जा सकता है। शब्दों और भाव का प्रतिस्थापन, समानार्थक शब्द, शब्दक्रम, अप्राप्त शब्द, अतिरिक्त शब्द के स्तर पर त्रुटि विश्लेषण किया जाता है। संदर्भ अनुवाद का प्रयोग अनूदित वाक्य में मूल अर्थ को बनाए रखने के लिए तथा मशीनी अनुवाद के निर्गत अनुवाद की तुलना करने के लिए किया जाता है। सामान्यतः मशीनी अनुवाद की त्रुटियों का विश्लेषण करना आसान नहीं है यह अनूदित पाठ की सामग्री पर निर्भर करता है। विशेष रूप से कंप्यूटर से प्राप्त अनुवाद में

गलत शब्दों का अनुवाद, अप्राप्त शब्दों (Missing words), अतिरिक्त शब्दों (Extra words) तथा संदिग्धार्थकता (Ambiguity) को सुस्पष्ट कर पाना मुश्किल कार्य होता है।<sup>2</sup>

मशीनी अनुवाद के त्रुटि विश्लेषण के लिए विभिन्न विद्वानों ने अलग-अलग त्रुटियों का वर्गीकरण किया है। कुछ शोधकर्ताओं ने भाषाई आधार पर त्रुटियों का वर्गीकरण किया है। उनमें वर्ण विचार संबंधी (Orthographic), रूपात्मक संबंधी (morphological), शब्द के स्तर पर (lexical), अर्थ के स्तर पर (semantic) और वाक्यात्मक (syntactic) संबंधी त्रुटियों का वर्गीकरण किया है।<sup>3</sup> तो कुछ शोधकर्ताओं में डेविड विलर ने उनके शोध पत्र में 5 त्रुटियों का वर्गीकरण किया है- 1. अप्राप्त शब्द (Missing Words), 2. शब्दक्रम (Word Order), 3. गलत शब्द (Incorrect Words), 4. अज्ञात शब्द (Unknown Words) 5. विराम चिह्न त्रुटियाँ (Punctuation errors) आदि हैं।<sup>4</sup> इन सभी वर्गीकरण का प्रयोग मानव द्वारा तथा स्वचालित रूप से त्रुटियों के विश्लेषण के लिए किया जाता है। त्रुटि विश्लेषण के वर्गीकरण से अनुवाद में होने वाली गलतियों की पहचान कर उनमें सुधार के लिए तथा अनुवाद की गुणवत्ता को बढ़ाने के लिए सहायता होती है।

इस अध्याय में मशीनी अनुवाद के तंत्र गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी इन 4 तंत्रों से अंग्रेजी वाक्यों का हिंदी में अनुवाद प्राप्त किया गया और उन वाक्यों का विषय-क्षेत्र (Domain) के अनुसार मूल्यांकन करने के बाद वाक्य में पाई जाने वाली त्रुटियों का विश्लेषण इस अध्याय में किया गया है। यह त्रुटि विश्लेषण विभिन्न सामग्री से संकलित वाक्यों के कॉर्पोरा पर आधारित है। त्रुटि विश्लेषण

---

<sup>2</sup> Maja Popovi ´c and Aljoscha Burchardt. (2011), *From human to automatic error Classification for machine translation output*, p. 265-272

<sup>3</sup> Costa-Jussà, M.R., Farrús, M., & Mariño, J.B. (2010), *Linguistic-based Evaluation Criteria to identify Statistical Machine Translation Errors*. p. 52-57

<sup>4</sup> David Vilar, Jia Xu, Luis Fernando D’Haro, and Hermann Ney (2006), *Error analysis of machine translation output*, p. 697-702

करने के लिए भी विभिन्न स्वचालित पद्धतियों का प्रयोग किया जाता है लेकिन इससे प्राप्त विश्लेषण के परिणाम अधिक विश्वसनीय होने चाहिए। मूलतः मशीनी अनुवाद से प्राप्त अनुवाद का त्रुटि विश्लेषण हो या मूल्यांकन मानव द्वारा अधिक विश्वसनीय होता है तथा इसकी समस्या का समाधान मानव द्वारा सही रूप में किया जा सकता है। इसलिए इस अध्याय में मानव द्वारा त्रुटियों का विश्लेषण किया गया है।

### 5.3 मशीन अनूदित वाक्यों का त्रुटि विश्लेषण

मशीनी अनुवाद से प्राप्त वाक्यों में त्रुटियों का वर्गीकरण करना मुश्किल कार्य माना जाता है। क्योंकि कॉर्पोरा आधारित अनुवाद में विभिन्न प्रकार के वाक्यों का संकलन होता है और जब अनुवाद दो से अधिक मशीनी अनुवाद तंत्रों से किया जाता है तो उनसे अनुवाद भी अलग-अलग प्रकार का प्राप्त होता है साथ ही उनके त्रुटियों की भी कोई निश्चित मानकता नहीं होती। इसलिए त्रुटियों का वर्गीकरण करना मुश्किल हो जाता है। इस अध्याय में उन त्रुटियों का विश्लेषण किया गया है जो मशीनी अनुवाद तंत्रों के मूल्यांकन के दौरान पाई गई हैं। जैसे- वर्तनी की त्रुटियाँ, अप्राप्त शब्द, उच्चारण, कैपिटल शब्द, संक्षिप्तीकरण, क्रिया रूपरचना, लिंग की रूप रचना, अतिरिक्त शब्द, अनअनूदित शब्द, लिप्यंतरण की त्रुटियाँ, गलत शब्द चयन, संदिग्धता अथवा अस्पष्टता, अर्थगत त्रुटियाँ, वाक्यांश, कहावतों एवं मुहावरों की त्रुटियाँ, शब्दक्रम, पूर्वसर्ग तथा परसर्ग, वाक्यगत त्रुटियाँ, स्रोत वाक्य में त्रुटियाँ, सर्वनाम, विशेषण, आर्टिकल संबंधी त्रुटियाँ, सरल, संयुक्त और मिश्रित वाक्य में प्राप्त त्रुटियाँ आदि हैं। जिनका नीचे भाषाई स्तर पर (Linguistic level) 5 प्रकार से वर्गीकरण किया गया है।

#### 5.3.1. वर्ण विचार संबंधी त्रुटियाँ (Orthographic errors)

विराम चिह्न, गलत उच्चारण, कैपिटल शब्द, संयुक्त शब्द, अक्षर लोप, संयोजक शब्द, विदेशी शब्दों में त्रुटियाँ, वर्तनी में त्रुटियाँ, संक्षिप्तीकरण आदि त्रुटियों का समावेश इसमें किया जाता है।

➤ **विराम चिह्न की त्रुटियाँ (Punctuation errors)**

सभी मशीनी अनुवाद तंत्रों को अनुवाद के लिए वाक्य देने के बाद वाक्य में विराम चिह्न के स्तर पर निम्न वाक्यों में त्रुटियाँ पाई गई हैं।

1. अंग्रेजी वाक्य : I was commissioner, Tirhut Division, Bihar, from 1991 to 1995.

गूगल : मैं आयुक्त तिरहुत डिवीजन, बिहार, 1991 से 1995 तक था।

माइक्रोसॉफ्ट : आयुक्त, तिरहुत प्रमंडल, बिहार, 1995 को 1991 से मैं था।

मंत्र-राजभाषा : मैं 1995 के लिए 1991 से आयुक्त, तिहुत प्रभाग, बिहार था

आंग्ल-हिंदी : मैं १९९५ को १९९१ से कोमिशनर तिहुत दिविसायन बिहार था ।

सही अनुवाद : मैं सन् 1991 से 1995 तक बिहार के तिरहुत संभाग का आयुक्त था।

स्रोत अंग्रेजी वाक्य में अल्पविराम का प्रयोग हुआ है जिसके चलते इन सभी मशीनी अनुवाद तंत्रों के अनूदित वाक्यों में भिन्नता पाई जाती है। विशेषतः इसमें मंत्र-राजभाषा के अनुवाद में पूर्ण विराम का प्रयोग ही नहीं हुआ है। साथ ही आंग्ल-हिंदी अनुवाद में भारतीय अंकों का प्रयोग किया गया है बल्कि अंतरराष्ट्रीय अंकों का प्रयोग होना चाहिए। इसका अर्थ स्रोत वाक्य में अनुवाद के लिए अधिक अल्प विराम चिह्नों का यदि प्रयोग किया जाता है तो अनूदित वाक्य के शब्दक्रम में बदलाव पाया जाता है।

➤ **वर्तनी की त्रुटियाँ (Spelling errors)**

वर्तनी की त्रुटियाँ सभी मशीन अनूदित वाक्यों में पाई गई हैं। विशेषतः तब जब नए शब्द प्रतिदिन प्रयुक्त होते हैं तो अक्सर अनुवाद तंत्रों से गलतियाँ होती हैं।

2. अंग्रेजी वाक्य : There is no treatment for Chikungunya disease.

गूगल : वहाँ चिकनगुनिया की बीमारी के लिए कोई इलाज नहीं है।

माइक्रोसॉफ्ट : Chikungunya रोग के लिए कोई उपचार है।

मंत्र-राजभाषा : \_ चिकुंगुजा बीमारी के लिए किसी भी उपचार है नहीं

आंग्ल-हिंदी : वहाँ चिकुन्गुन्या बीमारी के लिये कोई उपचार नहीं है ।

सही अनुवाद : चिकुनगुन्या रोग का कोई उपचार नहीं है।

उपरोक्त स्रोत वाक्य अंग्रेजी का हिंदी में अनुवाद प्राप्ति के बाद देखा गया है कि अनुवाद में वर्तनी की त्रुटियाँ हैं। जैसे कि मंत्र-राजभाषा ने 'चिकुंगुजा' अनुवाद किया है और उसमें विराम चिह्न का भी प्रयोग नहीं किया है। इसी के साथ आंग्ल-हिंदी ने भी 'Chikungunya' का गलत अनुवाद किया है 'चिकुन्गुन्या'। इस तरह के शब्दों का अनुवाद तब होता है जब मशीनी अनुवाद तंत्रों के डाटाबेस में नए शब्दों का समावेश नहीं होता है।

3. अंग्रेजी वाक्य : For example, the Contract Labour Act legitimizes contract labour.

गूगल : उदाहरण के लिए, ठेका श्रम अधिनियम ठेका श्रम legitimizes।

माइक्रोसॉफ्ट : उदाहरण के लिए, ठेका श्रम अधिनियम संविदा श्रम legitimizes वह।

मंत्र-राजभाषा : उदाहरण के लिए, संविदा मजदूर अधिनियम संविदा मजदूर को लेगिटिमिज़ेस करता है

आंग्ल-हिंदी : उदाहरण के लिए कोन्त्राक्त लाबाउर अक्त लेगिटिमिज़ेस अनुबन्ध पत्र मजदूर ।

सही अनुवाद : उदाहरणार्थ ठेका श्रम अधिनियम, ठेका श्रम व्यवस्था को मान्यता देता है।

उपरोक्त अनूदित वाक्य में गूगल और माइक्रोसॉफ्ट ने 'legitimizes' शब्द का अनुवाद नहीं किया है। शब्द में 's' का प्रयोग होने से अनुवाद तंत्र अनुवाद कर पाने में असफल रहा है। तो मंत्र-राजभाषा ने उसी शब्द का 'लेगिटिमिज़ेस' कर वर्तनी में गलती की है। इसी के साथ आंग्ल-हिंदी से 'कोन्त्राक्त लाबाउर अक्त लेगिटिमिज़ेस' इस संपूर्ण वाक्यांश ही गलत अनूदित हुआ है और इसमें वर्तनी की अधिक गलतियाँ हुई हैं। इसके

लिए उपरोक्त 'सही अनुवाद' को देखने से ज्ञात होगा की अनुवाद स्तर पर कौनसी त्रुटियाँ हुई हैं।

➤ **बड़े अक्षर की त्रुटियाँ (Capital words errors)**

स्रोत वाक्य के शब्द में कैपिटल शब्दों के प्रयोग होने से भी अनुवाद का अर्थ बदल जाता है तथा अनुवाद में अनेक गलतियाँ पाई जाती हैं।

4. अंग्रेजी वाक्य : THEY ARE YOUR CHILDREN'S CHILDREN.

गूगल : वे अपने बच्चों के बच्चे हैं

माइक्रोसॉफ्ट : वे अपने बच्चों के बच्चों रहे हैं।

मंत्र-राजभाषा : वे हैं आपके चिल्डरेनस बच्चे।

आंग्ल-हिंदी : टीएचईवाई ऐआरई वाईओयूआर सीएचआईएलडीआरईएनएस  
सीएचआईएलडीआरईएन ।

सही अनुवाद : वे आपकी संतान की संतानें हैं।

अंग्रेजी वाक्य में संपूर्ण कैपिटल शब्दों का प्रयोग किया गया है। उस वाक्य का अनुवाद गूगल और माइक्रोसॉफ्ट द्वारा बोधगम्य तो है लेकिन सही नहीं है। मंत्र-राजभाषा तंत्र ने 'CHILDREN'S' का 'चिल्डरेनस' अनुवाद किया है और आंग्ल-हिंदी अनुवाद में संपूर्ण वाक्य का अक्षर के रूप में अनुवाद किया है। इससे ज्ञात होता है कि मशीनी अनुवाद तंत्रों के लिए पूर्णतः कैपिटल शब्दों का प्रयोग करने से गलत अनुवाद होता है।

5. अंग्रेजी वाक्य : THE GAME OF EVICTION.

गूगल : खेल का अनुमान

माइक्रोसॉफ्ट : बंगला खाली करने का खेल है।

मंत्र-राजभाषा : खेल की निष्कासन।

आंग्ल-हिंदी : टीएचई जीएएमई ओएफ ईवीआईसीटीआईओएन ।

सही अनुवाद : बेदखली का खेल।

उपरोक्त स्रोत वाक्य के गूगल अनुवाद में विराम चिह्न का प्रयोग नहीं हुआ है और अनुवाद भी अर्थ की दृष्टि से सार्थक नहीं है। माइक्रोसॉफ्ट ने अनुवाद के अतिरिक्त शब्दों का प्रयोग कर अनुवाद गलत किया है। मंत्र-राजभाषा का अनुवाद बोधगम्यता के स्तर पर अंशतः सही लगता है लेकिन आंग्ल-हिंदी ने स्रोत वाक्य को अक्षर के रूप में लिप्यंतरित किया है। इसका अर्थ यह है कि इस तंत्र से कैपिटल अक्षर के शब्दों का अनुवाद सही नहीं किया जाता।

### ➤ संक्षिप्तीकरण की त्रुटियाँ (Abbreviation errors)

संक्षिप्त नाम के प्रयोग से संपूर्ण वाक्य की आर्थी संरचना (Semantic structure) में बदलाव आता है। जिससे वाक्य में संदिग्धता एवं अस्पष्टता उत्पन्न होती है।

6. अंग्रेजी वाक्य : H.I.V. attacks the immune system of the body .

गूगल : H.I.V. शरीर की प्रतिरक्षा प्रणाली पर हमला करता है।

माइक्रोसॉफ्ट : कि एचआईवी शरीर की प्रतिरक्षा प्रणाली हमलों।

मंत्र-राजभाषा : एच.आई.वी. शरीर की प्रतिरोधक क्षमता को आक्रमण करता है

आंग्ल-हिंदी : एच.आई.वी. शरीर का प्रतिरक्षित सिस्टम पर आक्रमण करता है।

सही अनुवाद : एच.आई.वी. शरीर की प्रतिरक्षण प्रणाली को आघात पहुँचाता है।

उपरोक्त स्रोत वाक्य में H.I.V. शब्द में संक्षिप्त नाम का प्रयोग किया गया है लेकिन इसके प्रयोग से अनुवाद की वाक्य संरचना में काफी बदलाव होता है। गूगल ने उस शब्द का अनुवाद अंग्रेजी में वैसे ही किया है। माइक्रोसॉफ्ट द्वारा अनुवाद अस्पष्ट दिखाई देता है जिससे वाक्य पूर्ण स्पष्ट नहीं हो रहा है। मंत्र-राजभाषा से अनुवाद तो सही है लेकिन पूर्ण विराम का प्रयोग वाक्य के अंत में नहीं हुआ है। आंग्ल-हिंदी का

अनुवाद बोधगम्य है। उपरोक्त 'सही अनुवाद' देखने से ज्ञात होता है कि अनुवाद में कौनसी गलतियाँ हुई हैं।

7. अंग्रेजी वाक्य : Appointment of supplementary A.N.M. on contract.

गूगल : अनुपूरक A.N.M. की नियुक्ति अनुबंध पर।

माइक्रोसॉफ्ट : अनुबंध पर अनुपूरक A.N.M. की नियुक्ति।

मंत्र-राजभाषा : संविदा पर पूरक ए.एन.एम. की नियुक्ति

आंग्ल-हिंदी : अनुबन्ध पत्र पर अतिरिक्त ए.एन.एम. का आबन्ध ।

सही अनुवाद : संविदा पर अतिरिक्त ए.एन.एम. की तैनाती ।

सभी मशीनी अनुवाद तंत्रों के वाक्यों से प्रतीत होता है कि स्रोत वाक्य में संक्षिप्त नाम का प्रयोग करने से अनुवाद में अनेकार्थकता उत्पन्न होती है। जब संक्षिप्त नाम का समावेश तंत्रों की भाषायी सामग्री में नहीं होता है तब गूगल और माइक्रोसॉफ्ट जैसे अनुवाद तंत्र अक्सर उसी नाम का अनुवाद लक्ष्य भाषा के वाक्य में करते हैं।

### 5.3.2. शब्द संरचना संबंधी त्रुटियाँ (Morphological errors)

वाक्य में जब लिंग या वचन की कमी हो, क्रिया में विभक्ति या रूपरचना में बदलाव पाया गया हो तथा शब्द में भी रूपात्मक दृष्टि से जब बदलाव होता है तो उसका प्रभाव वाक्य संरचना में होता है इसलिए इसका वर्गीकरण रूप प्रक्रिया में किया गया है।

#### ➤ लिंग, वचन का क्रिया पर प्रभाव

8. अंग्रेजी वाक्य : I slept at my house.

गूगल : मैं अपने घर में सोए थे।

माइक्रोसॉफ्ट : मैं अपने घर में सोए थे।

मंत्र-राजभाषा : मैं मेरा सदन में सोया

आंग्ल-हिंदी : मैं अपने घर में सोया ।

सही अनुवाद : मैं अपने घर में सोया/सोई।

उपरोक्त स्रोत वाक्य में संज्ञा 'I' का प्रयोग होने से अनूदित वाक्य में क्रिया के प्रयोग में लिंग स्पष्ट नहीं हो रहा है। इसलिए गूगल और माइक्रोसॉफ्ट के दोनों अनुवादों में बहुवचन रूप 'सोए थे' का प्रयोग किया है जो वाक्य संरचना के स्तर पर गलत है। मंत्र-राजभाषा और आंग्ल-हिंदी ने अनुवाद तो सही किया है, लेकिन क्रिया से लिंग का स्पष्टीकरण नहीं हो रहा है। बल्कि 'सही अनुवाद' से स्पष्ट होता है कि 'सोया या सोई' क्रिया रूप का प्रयोग हो सकता है।

9. अंग्रेजी वाक्य : The Nilgiri cow is becoming extinct now .

गूगल : नीलगिरि गाय अब विलुप्त होता जा रहा है।

माइक्रोसॉफ्ट : नीलगिरि गाय अब विलुप्त होता जा रहा है।

मंत्र-राजभाषा : निलिगिरि गाय अब लुप्त को बन जा रहा है

आंग्ल-हिंदी : निलिगिरि गाय अब अनुरूप लुप्त है ।

सही अनुवाद : अब नीलगिरि गाय लुप्त प्रायः होती जा रही है।

स्रोत वाक्य का गूगल द्वारा अनुवाद में लिंग का प्रभाव क्रिया पर नहीं पड़ा है। स्रोत वाक्य में 'गाय' स्त्रीलिंग का बोध कराती है। लेकिन अनुवाद में पुल्लिंग का क्रिया पर प्रभाव पड़ा है ऐसा प्रतीत होता है। इसलिए व्याकरणिक दृष्टि से अनुवाद गलत हो जाता है। अन्य दोनों तंत्र द्वारा वर्तनी में गलतियाँ होने के साथ क्रिया पर पुल्लिंग के प्रभाव को दर्शाते हैं। बल्कि अनुवाद संदर्भ (सही) अनुवाद की तरह होना चाहिए।

10. अंग्रेजी वाक्य : She wants control over her life. She decides whom to love and marry.

गूगल : वह अपने जीवन पर नियंत्रण चाहता है। वह जिसे प्यार करते हैं और शादी करने का फैसला किया।

माइक्रोसॉफ्ट : वह उसके जीवन पर नियंत्रण करना चाहता है। वह जिनके प्यार और शादी का फैसला किया।

मंत्र-राजभाषा : वह उनका जीवन से ज्यादा नियंत्रण को चाहता है वह जिसको प्यार करने के लिए सुनिश्चित करता है

आंग्ल-हिंदी : वह अपने जीवन पर नियन्त्रण चाहती है । वह निश्चय करती है कि किससे पसंद किया जाय और विवाह करती है ।

सही अनुवाद : वो अपने जीवन की डोर अपने हाथों में रखना चाहती है, प्रेम और विवाह का फैसला खुद करती है।

इसमें स्रोत वाक्य बृहद या बड़ा है, जिसका सभी तंत्रों ने कई प्रकार से अनुवाद किया है। इस वाक्य संरचना में दिखाई देता है कि क्रिया पर लिंग के आधार पर बदलाव होना चाहिए था। इसमें सिर्फ आंग्ल-हिंदी तंत्र का अनुवाद लिंग और क्रिया की संरचना के अनुसार बोधगम्य लगता है।

11. अंग्रेजी वाक्य : The little boy got the sweets.

गूगल : छोटा लड़का स्वीट्स मिला है।

माइक्रोसॉफ्ट : छोटा लड़का मिठाई मिल गया।

मंत्र-राजभाषा : छह ोटा लड़का ने मिठाई को प्राप्त किया

आंग्ल-हिंदी : छोटा लड़का मिठाई लाया ।

सही अनुवाद : छोटे लड़के को मिठाइयाँ मिलीं।

स्रोत वाक्य के अंग्रेजी से हिंदी अनुवाद में उपरोक्त सभी मशीनी अनुवाद तंत्रों द्वारा अनुवाद में क्रिया के प्रयोग के कई रूप मिलते हैं। अनुवाद में वचन के अनुसार क्रिया पर बदलाव होना चाहिए, लेकिन कंप्यूटर से ये त्रुटियाँ अक्सर होती हैं।

12. अंग्रेजी वाक्य : Where does Gita live?

गूगल : गीता कहाँ रहता है?

माइक्रोसॉफ्ट : गीता कहाँ रहते है?

मंत्र-राजभाषा : गीता कहाँ रहता है

आंग्ल-हिंदी : गीता जीवित कहाँ रहती है ?

सही अनुवाद : गीता कहाँ रहती है?

उपरोक्त सभी वाक्यों के हिंदी अनुवाद में अनेक भिन्नताएं पाई गई हैं। गूगल और मंत्र-राजभाषा द्वारा अनुवाद में क्रिया का प्रयोग, पुल्लिंग की संरचना के अनुसार किया गया है जबकि 'गीता' स्त्रीलिंग संज्ञा शब्द है। माइक्रोसॉफ्ट ने भी क्रिया का बहुवचनीय अनुवाद किया है और आंग्ल-हिंदी के अनूदित वाक्य में अस्पष्टता उत्पन्न हुई है। इसलिए लिंग, वचन, काल, पक्ष और वृत्ति का क्रिया पर प्रभाव पड़ने से लक्ष्य भाषा वाक्य संरचना में बदलाव होता है। इसलिए उसी रूप में अनुवाद होना चाहिए।

### 5.3.3. शाब्दिक त्रुटियाँ (Lexical errors)

स्रोत भाषा के शब्द के लिए लक्ष्य भाषा के शब्दों में सटीक पर्याय चयन न होने की वजह से अनुवाद में अप्राप्त शब्द (Missing Words), गलत शब्द (Incorrect Words), अज्ञात शब्द (Unknown Words), अनअनूदित शब्द (untranslated words), अतिरिक्त शब्द (extra words), गलत शब्द चयन (incorrect words choice), अनअनूदित व्यक्तिवाचक संज्ञा आदि शब्दों में लक्ष्य भाषा (हिंदी) के अनुवाद में त्रुटियाँ पाई गई हैं। जिन्हें निम्नलिखित वाक्यों में स्पष्ट किया गया है।

#### ➤ अतिरिक्त और अनअनूदित शब्दों में त्रुटियाँ

13. अंग्रेजी वाक्य : A thing is remembered for a long time only when attention is paid to it .

गूगल : एक बात एक लंबे समय के लिए केवल जब यह ध्यान करने के लिए भुगतान किया जाता है के लिए याद किया जाता है।

माइक्रोसॉफ्ट : एक बात एक लंबे समय के लिए याद किया जाता है केवल जब ध्यान करने के लिए यह भुगतान किया जाता है।

मंत्र-राजभाषा : वस्तु जब ध्यान इसके लिए भुगतान कर दी जाती है सिर्फ लंबे समय से के लिए याद किया जाता है

आंग्ल-हिंदी : यह अनुवाद नहीं किया जा सकता

सही अनुवाद : कोई बात तभी लंबे समय तक याद रहती है जब उस पर ध्यान दिया जाए।

उपरोक्त क्रम में तीनों मशीनी अनुवाद तंत्रों से अनूदित वाक्यों में अतिरिक्त शब्दों का अनुवाद किया गया है। स्रोत वाक्य के शब्द 'attention is paid to it' के लिए 'भुगतान किया जाता है या भुगतान कर दी जाती है' अनुवाद किया है जिससे वाक्य में अनेकार्थता की समस्या निर्मित होती है। आंग्ल-हिंदी से तो अनुवाद ही नहीं हुआ है। इसका अर्थ बृहद वाक्यों के अनुवाद में आंग्ल-हिंदी से अक्सर समस्या निर्माण होती है।

#### 14. अंग्रेजी वाक्य : Drink plenty of water.

गूगल : खूब पानी पिए।

माइक्रोसॉफ्ट : पानी का खूब सेवन करें।

मंत्र-राजभाषा : जल की पेय पदार्थ अधिक मात्रा में

आंग्ल-हिंदी : आप पानी की प्रचुर मात्रा पीजिये।

सही अनुवाद : खूब पानी पीएँ।

उपरोक्त स्रोत वाक्य का हिंदी में अतिरिक्त शब्दों का अनुवाद हुआ है। इससे अनूदित वाक्य अर्थ के दृष्टि से अस्पष्ट लगता है। मंत्र-राजभाषा में शाब्दिक अनुवाद करने से वाक्य के अर्थ में अस्पष्टता दिखाई देती है एवं यह तंत्र अनुवाद में अक्सर विराम चिह्न का प्रयोग नहीं करता यह स्पष्ट होता है।

15. अंग्रेजी वाक्य : The washerman often used to wash saris.

गूगल : धोबी अक्सर साड़ी धोने के लिए इस्तेमाल किया।

माइक्रोसॉफ्ट : धोबी साड़ी धोने के लिए अक्सर इस्तेमाल किया।

मंत्र-राजभाषा : साड़ियाँ धो लेने के लिए प्रायः प्रयोग वाशेरमान

आंग्ल-हिंदी : बाशर का आदमी अक्सर साड़ियाँ धोता था ।

सही अनुवाद : धोबी अक्सर साड़ियाँ धोता था।

उपरोक्त स्रोत वाक्य का मशीनी अनुवाद तंत्रों द्वारा अप्राप्त और गलत शब्द में अनुवाद हुआ है। गूगल और माइक्रोसॉफ्ट दोनों तंत्रों द्वारा अनुवाद अर्थ की दृष्टि से अस्पष्ट है। तो मंत्र-राजभाषा ने 'washerman' शब्द के लिए 'वाशेरमान' शब्द का प्रयोग किया है जिससे अनुवाद अस्पष्ट लग रहा है। उसी तरह आंग्ल-हिंदी ने भी अनुवाद किया है।

➤ **गलत शब्द चयन (Incorrect words choice)**

स्रोत वाक्य से लक्ष्य भाषा में अनुवाद के लिए भी मशीनी अनुवाद तंत्र गलत शब्दों का चयन करते हैं जिससे लक्ष्य भाषा का अनुवाद त्रुटिग्रस्त हो जाता है।

16. अंग्रेजी वाक्य : There is no real treatment for yellow fever.

गूगल : वहाँ पीले बुखार के लिए कोई वास्तविक उपचार है।

माइक्रोसॉफ्ट : पीले बुखार के लिए कोई वास्तविक उपचार है।

मंत्र-राजभाषा : \_ पीत ज्वर के लिए किसी भी वास्तविक उपचार है नहीं

आंग्ल-हिंदी : वहाँ पीले बुकार के लिये बिल्कुल वास्तविक उपचार नहीं है ।

सही अनुवाद : पीतज्वर का कोई वास्तविक उपचार नहीं है ।

उपरोक्त स्रोत वाक्य के मशीनी अनुवाद में विभिन्न अनुवाद पाए गए हैं, लेकिन सभी अनुवाद में त्रुटियाँ पाई जाती हैं। गूगल और माइक्रोसॉफ्ट ने अनुवाद में नकारात्मक वाक्य को सकारात्मक में बदल दिया है तथा 'yellow fever' का अनुवाद 'पिले बुखार' किया है। जिसके लिए 'पीत ज्वर या पीला बुखार' सही शब्द है।

17. अंग्रेजी वाक्य : Ambedkar's monograph *Annihilation of Caste* was a lecture he had written to deliver at annual convention of Jat Pat Todak Mandal.

गूगल : जाति के अम्बेडकर के मोनोग्राफ विनाश एक व्याख्यान वह जाट पैट Todak मंडल के वार्षिक सम्मेलन में वितरित करने के लिए लिखा था।

माइक्रोसॉफ्ट : अम्बेडकर के मोनोग्राफ जाति का विनाश वह जाट पैट Todak मंडल के वार्षिक सम्मेलन में वितरित करने के लिए लिखा था एक व्याख्यान था।

मंत्र-राजभाषा : अम्बेडकरस मोनोग्राफनिहिलाशन कास्तईवास व्याख्यान वह की जत थपकी तोडक मंडल की वार्षिक परम्परा में निर्वाह करने के लिए लिखा था

आंग्ल-हिंदी : अम्बेडकार्स मोनोग्राफनिहिलाशन का/की/के कास्तेवास इक व्याख्यान वह रखा लिखित तक वितरित कर पर वार्षिक परम्परा का/की/के जात पात तोदक मण्डल।

सही अनुवाद : आंबेडकर ने 'एनिहिलेशन ऑफ कास्ट' इस निबंध को जात पात तोडक मंडल के वार्षिक अधिवेशन में भाषण देने के लिए तैयार किया था।

उपरोक्त स्रोत वाक्य का हिंदी अनुवाद में सभी मशीनी अनुवाद तंत्रों ने अलग-अलग अनुवाद किया है। गूगल और माइक्रोसॉफ्ट के अनुवाद में 'जाट पैट Todak' इस तरह के शब्दों का चयन किया गया है जिससे अनुवाद पूर्णतः त्रुटिग्रस्त दिखाई देता है। मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद में वर्तनी की अधिक गलतियाँ पायी गई हैं जिससे अनुवाद अस्पष्ट लग रहा है, इसमें संदर्भ अनुवाद से ही वाक्य का अर्थ स्पष्ट होता है।

18. अंग्रेजी वाक्य : There are several palm leaves scriptures Preserved in Indian School Of Marshal Art.

गूगल : कई हथेली मार्शल आर्ट इंडियन स्कूल में संरक्षित शास्त्रों पत्ते होते हैं।

माइक्रोसॉफ्ट : वहाँ कई ताड़ के पत्ते शास्त्रों Preserved इंडियन स्कूल ऑफ मार्शल आर्ट में कर रहे हैं।

मंत्र-राजभाषा : \_ कई करतल हैं भारतीय विद्यालय की मार्शल कला में संरक्षित धर्मग्रंथों छह ाड़ता है

आंग्ल-हिंदी : वहाँ भारतीय स्कूल ओफ मार्शल आर्ट में कई हथेली छुट्टियाँ धार्मिक ग्रंथ प्रेसेर्वेद हैं ।

सही अनुवाद : इंडियन स्कूल ऑफ मार्शल आर्ट्स में अनेक ताड़पत्रीय ग्रन्थ संरक्षित हैं ।

उपरोक्त स्रोत वाक्य के अनुवाद में गूगल ने 'हथेली' और माइक्रोसॉफ्ट द्वारा 'शास्त्रों' शब्द के गलत चयन से दोनों वाक्यों का अनुवाद त्रुटिग्रस्त पाया गया है तथा वाक्य में कोई शब्दक्रम (Word order) प्रयुक्त नहीं हुआ है। मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद में भी वाक्य संरचना में अतिरिक्त शब्दों के चयन के साथ शब्दक्रम का अभाव है, जिससे वाक्य पढ़ने में सरल नहीं लग रहा है।

19. अंग्रेजी वाक्य : Dengue fever occurs after a sting of aedes aegypti.

गूगल : डेंगू बुखार एडीज एजिप्टी की एक स्टिंग के बाद होता है।

माइक्रोसॉफ्ट : डेंगू बुखार एडीज aegypti का एक स्टिंग के बाद होती है।

मंत्र-राजभाषा : डेंगू बुखार ज्वर एडेस एग्यप्टि की डंक मार के बाद होता है

आंग्ल-हिंदी : डेंगू का बुकार एदेस एग्यिप्ट की चुभन के बाद घटित होता है ।

सही अनुवाद : डेंगू बुखार एडिस एजिप्टाई मच्छर के काटने से होता है ।

उपरोक्त स्रोत वाक्य के अनुवाद में सभी तंत्रों द्वारा विशेषतः वर्तनी की समस्या पाई जाती है तथा अतिरिक्त शब्द के अनुवाद से वाक्य में शाब्दिक अस्पष्टता उत्पन्न हुई है।

➤ **अनअनुदित व्यक्तिवाचक संज्ञा (Non-translated proper nouns)**

मशीनी अनुवाद तंत्रों द्वारा अक्सर व्यक्तिवाचक नाम का अनुवाद सही नहीं होता, यह समस्या मशीनी अनुवाद तंत्रों के डाटाबेस में उसका समावेश न करने से यह उत्पन्न होती है।

20. अंग्रेजी वाक्य : Prince Chandradeva got the area of Chandragiri in inheritance.

गूगल : राजकुमार Chandradeva विरासत में चंद्रगिरी के क्षेत्र मिला है।

माइक्रोसॉफ्ट : राजकुमार Chandradeva चंद्रगिरी के क्षेत्र विरासत में मिला है।

मंत्र-राजभाषा: राजकुमार चंद्रदेवा ने उत्तराधिकार में चंद्रगिरि की क्षेत्र को प्राप्त किया

आंग्ल-हिंदी : प्रिंस चान्द्रादेवा ने विरासत में चान्द्रागिरि का क्षेत्र प्राप्त किया।

सही अनुवाद : राजकुमार चन्द्रदेव को विरासत में चन्द्रगिरी का क्षेत्र मिला।

अंग्रेजी स्रोत वाक्य के हिंदी अनुवाद में व्यक्तिवाचक नाम 'Prince Chandradeva' का अनुवाद गूगल और माइक्रोसॉफ्ट द्वारा अनअनुदित (untranslated) हुआ है। तो मंत्र-राजभाषा और आंग्ल-हिंदी तंत्र ने नाम के अनुवाद में वर्तनी की गलती की है।

21. अंग्रेजी वाक्य : Ratna Stambha was built in the memory of Shreyansh Kumar .

गूगल : रत्न स्तंभ Shreyansh कुमार की याद में बनाया गया था।

माइक्रोसॉफ्ट : रत्ना स्तम्भ Shreyansh कुमार की स्मृति में बनाया गया था।

मंत्र-राजभाषा : रत्ना स्तम्भा श्रेयंश कुमार की अंतर्निर्मित स्मरण शक्ति था

आंग्ल-हिंदी : रत्ना स्ताम्ब्हा श्रेयान्श कुमार की स्मृति में बनाया गया था ।

सही अनुवाद : श्रेयांश कुमार की स्मृति में रत्न स्तम्भ बनवाया था ।

उपरोक्त स्रोत वाक्य का गूगल और माइक्रोसॉफ्ट के अनुवाद में 'Shreyansh' नाम का अनुवाद नहीं हुआ है। लेकिन मंत्र-राजभाषा और आंग्ल-हिंदी ने इसका अनुवाद किया है। इनके अनुवाद में वर्तनी की गलतियाँ अधिक पाई जाती हैं, लेकिन शब्द के स्तर पर अनुवाद करने में ये तंत्र सक्षम हैं। मंत्र-राजभाषा में पूर्ण विराम का प्रयोग अनुवाद में बहुत कम स्थान पर प्रयुक्त होता है। व्यक्तिवाचक संज्ञा का अनुवाद ठीक से नहीं होता तब वाक्य संरचना में अस्पष्टता निर्माण होती है।

### ➤ लिप्यंतरण की समस्या (Problems of Transliteration)

एक भाषा से दूसरी भाषा में अनुवाद करते समय अनेक शब्दों का अनुवाद करने में मशीन असमर्थ होती है। इसलिए कुछ शब्दों का लिप्यंतरण किया जाता है तो कुछ शब्दों को वैसे ही अनअनूदित रूप में रखा जाता है। लिप्यंतरण में एक ही पाठ को एक लिपि से दूसरी लिपि में परिवर्तित किया जाता है। इस लिप्यंतरण में अधिकतर व्यक्ति नाम और स्थल होते हैं।

22. अंग्रेजी वाक्य : And telecast in between the aforementioned discussion at the restaurant.

गूगल : और रेस्तरां में ऊपर उल्लिखित चर्चा के बीच में प्रसारित करती हैं।

माइक्रोसॉफ्ट : और रेस्तरां में aforementioned चर्चा के बीच में प्रसारण।

मंत्र-राजभाषा : और रेस्तराँ में बहस उपर्युक्त के दरमियान प्रसारण

आंग्ल-हिंदी : और प्रसारण के मध्य पूर्वोक्त विचार विमर्श पर रेस्तराँ ।

सही अनुवाद : जिसका प्रसारण रेस्टोरेंट में उपरोक्त चर्चा के बीच किया गया।

उपरोक्त स्रोत अंग्रेजी वाक्य का अनुवाद सभी मशीनी अनुवाद तंत्रों ने हिंदी में किया है जिसमें 'restaurant' शब्द को अलग-अलग प्रकार से सभी तंत्रों ने लिप्यंतरित

किया है। तो माइक्रोसॉफ्ट ने 'aforementioned' शब्द का अनुवाद लक्ष्य भाषा वाक्य में नहीं किया है, जिसका अनुवाद 'उपरोक्त या ऊपर उल्लिखित' होना चाहिए था।

23. अंग्रेजी वाक्य : Kumbhaswami or Kumbhashyam was originally a Vaishnav temple .

गूगल : कुम्भस्वामी या कुंभशम मूल रूप से एक वैष्णव मंदिर था।

माइक्रोसॉफ्ट : Kumbhaswami या Kumbhashyam मूल रूप से एक वैष्णव मंदिर था।

मंत्र-राजभाषा : कुम्भस्वामि या कुम्भशयम वेशनव मंदिर प्रारम्भ में था

आंग्ल-हिंदी : कुम्भास्वामि या कुम्भाशयम था मूलतः वेशं व मंदिर ।

सही अनुवाद : कुंभस्वामी या कुंभशयाम मंदिर मूलतः वैष्णव मंदिर था ।

उपरोक्त स्रोत वाक्य के अनुवाद में सभी मशीनी अनुवाद तंत्रों ने 'Kumbhaswami or Kumbhashyam' के नाम को अनेक प्रकार से लिप्यंतरित अनुवाद किया है। लेकिन अनुवाद में वर्तनी की अधिक गलतियाँ पाई जाती हैं। अनुवाद के लिए संदर्भ (सही) अनुवाद से ज्ञात होता है कि नाम का लिप्यंतरण किस तरह होना चाहिए।

24. अंग्रेजी वाक्य : Kanjar waterfall remains filled with water whole year.

गूगल : कंजर झरना पानी पूरे साल के साथ भरा रहता है।

माइक्रोसॉफ्ट : Kanjar झरना के साथ पूरे साल पानी भरा रहता है।

मंत्र-राजभाषा : कंजर जलपात जल सभी वर्ष सहित भरा में अवशेष है

आंग्ल-हिंदी : कान्जर जल प्रपात के अवशेष ने पानी से सम्पूर्ण वर्ष भरा ।

सही अनुवाद : कांजार जलप्रपात पूरे वर्ष जलयुक्त रहता है।

उपरोक्त स्रोत वाक्य के हिंदी अनुवाद में भी 'Kanjar' का अनुवाद 4 मशीनी अनुवाद तंत्रों ने अलग-अलग प्रकार से लिप्यंतरण किया है, इसलिए लिप्यंतरण में एकरूपता नहीं है। अक्सर नए शब्द मशीनी अनुवाद तंत्रों के कॉर्पोरा में न होने की वजह से यह समस्या उत्पन्न होती है जिससे वाक्य भी पढ़ने में अस्पष्ट लगता है।

#### 5.3.4. अर्थगत त्रुटियाँ (Semantic errors)

स्रोत भाषा के वाक्य का लक्ष्य भाषा वाक्य में जब अनुवाद कंप्यूटर से होता है तो अनुवाद में स्रोत भाषा के सामान्य शब्दों, सांस्कृतिक शब्दों, कहावतों और मुहावरों का अनुवाद नहीं होता, तब अनुवाद में समस्या उत्पन्न होती है। इसमें समनामिता (Homonymy) अनेकार्थी (Polysemy) शब्दों का भी समावेश है। यह समस्या प्रमुख रूप से शब्द के अर्थ स्पष्ट न होने के कारण होती है।

25. अंग्रेजी वाक्य : The follow consequences show up if a person is caught up with AIDS .

गूगल : का पालन परिणामों को दिखाने के लिए यदि एक व्यक्ति को एड्स के साथ पकड़ा है।

माइक्रोसॉफ्ट : का पालन परिणाम दिखाएँ यदि एक व्यक्ति एड्स के साथ पकड़ा है।

मंत्र-राजभाषा : परिणामों दर्श यदि व्यक्ति बाद में भेजा जाते हैं एड्स सहित पकड़ यू.पी. करते हैं

आंग्ल-हिंदी : अनुसरण कर परिणाम दिखाई पड़ यदि इक व्यक्ति है पकड़ ले के साथ ऐआईडीएस ।

सही अनुवाद : किसी व्यक्ति को एड्स होने पर निम्न दुष्परिणाम सामने आते हैं।

उपरोक्त स्रोत वाक्य का हिंदी अनुवाद के लिए दिए गए वाक्यों में गूगल और माइक्रोसॉफ्ट द्वारा 'caught up with AIDS' शब्द का 'एड्स के साथ पकड़ा है' यह शाब्दिक अनुवाद हुआ है जिससे अर्थ की प्राप्ति नहीं होती और अनूदित वाक्य के कई

अर्थ निकलते हैं। आंग्ल-हिंदी और मंत्र-राजभाषा के अनुवाद में भी वाक्य का सही अनुवाद न होने से अर्थ स्पष्ट नहीं हो रहा है।

26. अंग्रेजी वाक्य : Gandhi threw out the British.

गूगल : गांधी ब्रिटिश बाहर फेंक दिया।

माइक्रोसॉफ्ट : गांधी ब्रिटिश बाहर फेंक दिया।

मंत्र-राजभाषा : गाँधी ब्रिटेन का फेंक बाहर करते हैं

आंग्ल-हिंदी : गाँधी अंग्रेज भर् पहेका।

सही अनुवाद : गांधीजी ने अंग्रेजों को बाहर निकाला।

स्रोत वाक्य अंग्रेजी का हिंदी अनुवाद में गूगल और माइक्रोसॉफ्ट ने 'threw out' का शाब्दिक अनुवाद 'बाहर फेंक दिया' किया है। जिससे वाक्य की आर्थी संरचना में संदिग्धता उत्पन्न हुई है। आंग्ल-हिंदी के अनुवाद में वर्तनी की समस्या अधिक दिखाई दे रही है।

27. अंग्रेजी वाक्य : Use mosquito nets while sleeping wherever possible.

गूगल : मच्छरदानी का प्रयोग करते हुए जहां भी संभव सो रहा है।

माइक्रोसॉफ्ट : जहाँ भी संभव हो सोते समय मच्छरदानी का उपयोग करें।

मंत्र-राजभाषा : जबकि उपयोग मच्छर जालीयाँ नींद जब कभी संभव

आंग्ल-हिंदी : उपयोग का मच्छर समय सोने कहीं भी सम्भव ढकता है ।

सही अनुवाद : सोते समय जहाँ तक संभव हो मच्छरदानी का प्रयोग करें ।

उपरोक्त स्रोत वाक्य के हिंदी अनुवाद में गूगल और मंत्र-राजभाषा, आंग्ल-हिंदी के द्वारा अनुवाद से ज्ञात होता है कि अर्थ के स्तर पर किस तरह की भिन्नता पाई जाती है।

28. अंग्रेजी वाक्य : Frequent loose motion for a month.

गूगल : एक महीने के लिए लगातार ढीला प्रस्ताव।

माइक्रोसॉफ्ट : एक महीने के लिए बार-बार ढीला गति।

मंत्र-राजभाषा : महीने के लिए बार बार खुले प्रस्ताव

आंग्ल-हिंदी : आप ढीला प्रस्ताव एक महीने के लिये बार बार जाइये।

सही अनुवाद : एक माह से लगातार दस्त आना।

उपरोक्त अंग्रेजी से हिंदी अनूदित वाक्यों में गूगल और आंग्ल-हिंदी द्वारा 'loose motion' का 'ढीला प्रस्ताव' तथा माइक्रोसॉफ्ट ने 'ढीला गति' और मंत्र-राजभाषा ने 'खुले प्रस्ताव' शब्द में अनुवाद किया है, जो शब्द तथा अर्थ के स्तर पर गलत है। जिसका अनुवाद 'दस्त लगना या आना' होना चाहिए।

➤ **संदिग्धार्थकता की समस्याएँ (Ambiguity problems)**

किसी भी भाषा में संदिग्धार्थकता प्रमुख रूप से शब्द, अर्थ और वाक्य के स्तर पर पाई जाती है।

29. अंग्रेजी वाक्य : It touches your heart at its most tender part.

गूगल : यह अपने सबसे निविदा भाग में अपने दिल को छू लेती है।

माइक्रोसॉफ्ट : यह अपने सबसे कोमल भाग पर अपना दिल छू लेती है।

मंत्र-राजभाषा: यह अपना व्यापक स्तर पर निविदा भाग में आपके हृदय को स्पर्श करता है

आंग्ल-हिंदी : यह इसकी अत्यधिक नरम भूमिका पर आपका हृदय छूता है ।

सही अनुवाद : यह आपके मन की सबसे नाजुक जगह को छू लेती है।

उपरोक्त स्रोत वाक्य का मशीनी अनुवाद तंत्र द्वारा हिंदी अनुवाद में गूगल और मंत्र-राजभाषा ने 'tender part' का 'निविदा भाग' अनुवाद किया है जिससे संपूर्ण वाक्य में संदिग्धता उत्पन्न हुई है। आंग्ल-हिंदी में पर्यायी शब्द के रूप में 'नरम भूमिका' का अनुवाद किया है जो अर्थ की दृष्टि से गलत है। इसके लिए सही अनुवाद को देखा जा सकता है।

**30. अंग्रेजी वाक्य :** The leaders of the Sangh were advocates of caste equality and harmony.

गूगल : संघ के नेताओं ने जाति समानता और सद्भाव की वकालत कर रहे थे।

माइक्रोसॉफ्ट : संघ के नेता जाति समानता और सद्भाव के पैरोकार थे।

मंत्र-राजभाषा : नेताओं संघ की जाति समता और समरसता की अधिवक्ताओं थे

आंग्ल-हिंदी : सान्ध के नेता जाति की समानता और सहमति के अधिवक्ता थे ।

सही अनुवाद : संघ के नेता जातीय समानता और समरसता के समर्थक थे।

उपरोक्त अंग्रेजी वाक्यों के अनुवाद में गूगल ने 'advocates' शब्द का 'वकालत' और माइक्रोसॉफ्ट ने 'पैरोकार' तथा मंत्र-राजभाषा और आंग्ल-हिंदी ने 'अधिवक्ता' अनुवाद किया है। इस शाब्दिक अनुवाद से वाक्य में विभिन्न अर्थ निकलते हैं और वाक्य अर्थ की दृष्टि से अस्पष्ट लगता है।

**31. अंग्रेजी वाक्य :** Iraqi head seeks arms.

गूगल : इराकी सिर हथियार चाहता है

माइक्रोसॉफ्ट : इराकी सिर हथियार चाहता है।

मंत्र-राजभाषा : ईराकी शीर्ष आयुध को प्राप्त करता है।

आंग्ल-हिंदी : इराकि जो हैड खोजता है शस्त्र - सज्जित

सही अनुवाद : ईराकी प्रमुख हथियारों की तलाश कर रहा है।

उपरोक्त स्रोत वाक्य का विभिन्न मशीनी अनुवाद तंत्रों ने अलग-अलग अनुवाद किया है। गूगल और माइक्रोसॉफ्ट ने 'head' का 'सिर' अनुवाद किया है तथा मंत्र-राजभाषा ने 'शीर्ष' अनुवाद किया है। जो वाक्य में आर्थी संरचना के अनुसार गलत है। आंग्ल-हिंदी में भी संदिग्ध अनुवाद के साथ पूर्ण विराम के प्रयोग की कमी है। स्रोत वाक्य में 'head' एक संज्ञा के रूप में है जिसका अर्थ या तो 'प्रमुख' या 'शरीर के सिर' के संबंध में हो सकता है, लेकिन यहाँ संदर्भ के अनुसार 'head' शब्द का अनुवाद 'प्रमुख' होगा।

32. अंग्रेजी वाक्य : "Educate, Agitate and Organize", Dr. Ambedkar used to say.

गूगल : शिक्षित, आंदोलन और व्यवस्थित करें, डॉ. अंबेडकर कहते थे।

माइक्रोसॉफ्ट : 'शिक्षित, Agitate और व्यवस्थित करें', डॉ. आंबेडकर कहते थे।

मंत्र-राजभाषा : एडुकात, घबरा और संगठन, डॉ. अम्बेड्कर कहता था

आंग्ल-हिंदी : आप कहना उपयोग किया हुआ अगिताते और ओर्गानाईजे, डॉ. अम्बेडकर की शिक्षा दीजिये।

सही अनुवाद : 'शिक्षित बनो, संघर्ष करो और संगठित रहो' डॉ.आंबेडकर कहते थे।

उपरोक्त स्रोत वाक्यांश का हिंदी अनुवाद में सभी मशीनी अनुवाद तंत्रों ने अनेक गलतियाँ की हैं। मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद में अधिक गलतियाँ पाई गई हैं। जिससे अनूदित वाक्य में संदिग्धार्थकता निर्मित होती है। अनुवाद में संदिग्धार्थकता शब्द, अर्थ और वाक्य तीनों स्तर पर पाई जाती हैं। इस समस्या के समाधान के लिए समानार्थी तथा पर्यायवाची शब्दों का प्रयोग कर संदर्भ के समानता के आधार पर शब्द के विशिष्ट अर्थ को प्रतिस्थापित किया जाता है। इसमें शब्द-आशय विसंदिग्धीकरण (word sense disambiguation) का प्रयोग मशीनी अनुवाद में आनेवाली संदिग्धार्थकता की समस्या के समाधान के लिए किया जाता है।

### 5.3.5 वाक्यात्मक त्रुटियाँ (Syntactic errors)

एक प्राकृतिक भाषा के दूसरी प्राकृतिक भाषा के अनुवाद में अक्सर वाक्य स्तर पर त्रुटियाँ पाई जाती हैं। उनमें शब्दक्रम, बृहद वाक्य संरचना, पूर्वसर्ग और परसर्ग में, संबंधवाचक उपवाक्य में, इनपुट वाक्य में त्रुटियाँ, विशेषण, आर्टिकल, सरल वाक्य, संयुक्त और मिश्र वाक्य आदि में त्रुटियाँ पाई जाती हैं।

#### ➤ शब्दक्रम (Word order)

एक भाषा से दूसरी भाषा में अनुवाद करने के लिए दोनों भाषाओं के पदक्रम का समावेश मशीनी अनुवाद में होना आवश्यक है जिससे सटीक अनुवाद के लिए प्रत्येक शब्द का रूपांतरण एक क्रम में होगा। अंग्रेजी भाषा का शब्दक्रम कर्ता-क्रिया-कर्म (SVO) पर आधारित है जबकि हिंदी भाषा का कर्ता-कर्म-क्रिया (SOV) पर आधारित है। वाक्य के स्तर पर इन दोनों भाषाओं के शब्दक्रम में भिन्नता पाई जाती है।

33. अंग्रेजी वाक्य : Pilgrimage is of utmost importance in Hinduism.

गूगल : तीर्थयात्रा हिंदू धर्म में अत्यंत महत्व का है।

माइक्रोसॉफ्ट : हिन्दू धर्म में अत्यंत महत्व का तीर्थ है।

मंत्र-राजभाषा : तीर्थ यात्री हिंदू धर्म में अत्यंत महत्व की है

आंग्ल-हिंदी : तीर्थयात्रा हिंदुइस्म में परम महत्व का है।

सही अनुवाद : तीर्थ का हिंदू धर्म में बड़ा महत्व है।

उपरोक्त स्रोत वाक्य अंग्रेजी का हिंदी अनुवाद सभी मशीनी अनुवाद तंत्रों ने अलग-अलग प्रकार से अनुवाद किया है। लेकिन सभी अनूदित वाक्य में शब्दक्रम का अभाव है इसलिए अनुवाद सहजता के दृष्टि से सही नहीं है। सही अनुवाद में देखा जा सकता है कि अंग्रेजी से हिंदी में किस तरह भाषा के शब्दक्रम के अनुसार अनुवाद किया है।

34. अंग्रेजी वाक्य : Malayalam language is spoken in Kerala .

गूगल : मलयालम भाषा केरल में बोली जाती है।

माइक्रोसॉफ्ट : केरल में मलयालम भाषा बोली जाती है।

मंत्र-राजभाषा : मलयालम भाषा केरल में बोला दिया गया है

आंग्ल-हिंदी : मलयालम की भाषा केरल में बोली जाती है ।

सही अनुवाद : केरल में मलयालम भाषा बोली जाती है ।

अंग्रेजी वाक्य के अनुवाद का निश्चित पदक्रम या शब्दक्रम होता है, लेकिन हिंदी भाषा का पदक्रम लचीला होने के बावजूद उसके वाक्य का अलग-अलग तरह से अनुवाद किया जा सकता है। जैसा की ऊपर सभी मशीनी अनुवाद तंत्रों ने अनुवाद किया है।

➤ **अनुवाद में परसर्ग (postposition) शब्दों का न होना**

35. अंग्रेजी वाक्य : The boys returned the books to me.

गूगल : लड़कों मेरे लिए किताबें लौट आए।

माइक्रोसॉफ्ट : लड़कों किताबें मेरे लिए वापस आ।

मंत्र-राजभाषा : लड़के ने ग्रंथ मेरे लिए को वापस किया

आंग्ल-हिंदी : लड़कों ने मुझे पुस्तकें वापस कीं।

सही अनुवाद : लड़को ने मुझे किताबें वापस कीं।

अंग्रेजी से हिंदी में मशीन अनूदित वाक्यों में परसर्ग का अनुवाद का प्रयोग वाक्य में अनूदित नहीं होता जिसके चलते वाक्य संरचना में त्रुटियाँ दिखाई देती हैं। उपरोक्त वाक्य में गूगल और माइक्रोसॉफ्ट तंत्र द्वारा अनुवाद में 'ने' परसर्ग का प्रयोग नहीं

किया है। लेकिन मंत्र-राजभाषा और आंग्ल-हिंदी द्वारा मशीनी अनुवाद में 'ने' का प्रयोग किया है।

36. अंग्रेजी वाक्य : The students listened to the teacher.

गूगल : छात्रों को शिक्षक की बात सुनी।

माइक्रोसॉफ्ट : शिक्षक को छात्रों की बात सुनी।

मंत्र-राजभाषा : विद्यार्थियों ने शिक्षक के लिए सुना

आंग्ल-हिंदी : विद्यार्थियों ने शिक्षक को सुना ।

सही अनुवाद : छात्रों ने शिक्षक को सुना।

अंग्रेजी वाक्य में संज्ञा से पहले और बाद में पूर्वसर्ग (Preposition) का प्रयोग होता है और हिंदी में संज्ञा के बाद परसर्ग (postposition) का प्रयोग होता है। उपरोक्त अनूदित वाक्य में गूगल और माइक्रोसॉफ्ट तंत्र द्वारा अनुवाद में 'ने' परसर्ग का प्रयोग नहीं हुआ है। लेकिन अन्य दो तंत्रों में 'ने' का प्रयोग हुआ है

37. अंग्रेजी वाक्य : The little boy got the sweets.

गूगल : छोटा लड़का स्वीट्स मिला है।

माइक्रोसॉफ्ट : छोटा लड़का मिठाई मिल गया।

मंत्र-राजभाषा : छह ोटा लड़का ने मिठाई को प्राप्त किया

आंग्ल-हिंदी : छोटा लड़का मिठाई लाया ।

सही अनुवाद : छोटे लड़के को मिठाइयाँ मिलीं।

उपरोक्त स्रोत वाक्य का हिंदी में अलग-अलग मशीनी अनुवाद तंत्रों ने अनुवाद किया है। लेकिन मंत्र-राजभाषा के अलवा किसी भी मशीनी अनुवाद तंत्र ने परसर्ग 'को'

का अनुवाद नहीं किया है। सभी तंत्रों के अनुवाद में आर्थी संरचना के दृष्टि से वाक्य अस्पष्ट है। सही अनुवाद 'छोटे लड़के को मिठाइयाँ मिलीं' होना चाहिए जिसमें 'को' परसर्ग प्रयोग से स्पष्ट होता है।

### ➤ बृहद वाक्यों में त्रुटियाँ

जब अनुवाद के लिए बड़े वाक्य मशीन में दिए जाते हैं तो अनुवाद में बहुधा अनेक त्रुटियाँ होती हैं। जिससे संपूर्ण अनुवाद की वाक्य संरचना पढ़ने में दुरूह लगती है।

38. अंग्रेजी वाक्य : The film is a warning to the gutsy couples opting for inter-caste marriages that they are not safe even in the metropolises.

गूगल : फिल्म अंतरजातीय विवाह के लिए चयन है कि वे महानगरों में भी सुरक्षित नहीं हैं साहसी जोड़ों के लिए एक चेतावनी है।

माइक्रोसॉफ्ट : फिल्म gutsy युगल अंतर जाति विवाह कि वे महानगर में भी सुरक्षित नहीं हैं के लिए ऑप्ट करने के लिए एक चेतावनी है।

मंत्र-राजभाषा : फिल्म इंटर जाति शादीयाँ के लिए विकल्प प्रस्तुत करके गुथी दम्पतियों के लिए चेतावनी है कि वे यहाँ तक कि राजधानीयाँ में सुरक्षित हैं नहीं

आंग्ल-हिंदी : फिल्म साहसी जोड़ों के प्रति एक चेतावनी है जो अंतः - जाति मारिआगेस चुनाव कर रही है जो वे भी मेत्रोपोलिसेस में सुरक्षित नहीं हैं ।

सही अनुवाद : फिल्म की चेतावनी है कि अंतरजातीय विवाह करने वाले बहादुर लोग महानगरों में भी महफूज़ नहीं हैं।

अक्सर बृहद वाक्य का अनुवाद मशीनी अनुवाद तंत्रों द्वारा गलत होता है उसमें काफ़ी त्रुटियाँ पाई जाती हैं और अनेक शब्दों का अनुवाद नहीं होता। वैसे ही उपरोक्त स्रोत वाक्य अंग्रेजी का गूगल द्वारा अनुवाद सही है लेकिन अन्य तंत्रों के अनुवाद में वाक्य संरचना और शब्दक्रम के स्तर पर बोधगम्य नहीं है। इसके लिए उपरोक्त सही हिंदी अनुवाद पढ़ा जा सकता है।

39. अंग्रेजी वाक्य : Kamble did admit though that the DICCI SME Fund had suffered due to the governments at the Centre announcing other similar funds.

गूगल : कांबले ने हालांकि स्वीकार किया कि डीआईसीसीआई एसएमई फंड को अन्य समान फंडों की घोषणा करते हुए केंद्र सरकार की वजह से नुकसान हुआ था।

माइक्रोसॉफ्ट : कांबळे मानते हैं हालांकि कि DICCI एसएमई फंड अन्य समान कोष की घोषणा केंद्र की सरकारों के कारण नुकसान उठाना पड़ा था।

मंत्र-राजभाषा : कम्बल ने दाखिल किया यद्यपि वह डी.आई.सी.सी.आई. एसएमई वित्तपोषण ने केंद्र में अन्य मिलते जुलते निधि घोषणा करके सरकारों के कारण भुगता था

आंग्ल-हिंदी : काम्बले ने माना तो था यद्यपि वह डीआईसीसीआई एसएमई फुन्द ने दूसरा घोषित करता हुआ सेन्त्रे को सरकारों के कारण समान निधियों ने हानि हो चुका था ।

सही अनुवाद : कांबले ने स्वीकार किया कि डिक्की के फंड की राह इसलिए कठिन हो गई, क्योंकि भारत सरकार ने इसी तरह के अन्य फंडों की स्थापना की घोषणा कर दी।

उपरोक्त स्रोत वाक्य का सभी मशीनी अनुवाद तंत्रों ने अनुवाद किया है। जिसमें कुछ तंत्रों ने 'DICCI' संक्षिप्त नाम को लिप्यंतरित किया है। लेकिन वाक्य के सभी अनुवाद शाब्दिक होने से अबोधगम्य लगते हैं।

#### ➤ संख्यात्मक त्रुटियाँ (Numerical error)

स्रोत भाषा से लक्ष्य भाषा में संख्यात्मक अंक के अनुवाद में भी त्रुटियाँ पाई जाती हैं।

40. अंग्रेजी वाक्य : Rohit goes to sleep at half 10.

गूगल : रोहित आधे 10 में सोने के लिए चला जाता है।

माइक्रोसॉफ्ट : रोहित में आधे 10 सोने के लिए चला जाता है।

मंत्र-राजभाषा : रोहित NOUN PREP NOUN PUNCT NOUN PREP आधा 10 NOUN  
PREP NOUN PREP PrPART NOUN TYPE TYPE CONJ NOUN PREP NOUN TYPE

आंग्ल-हिंदी : रोहित आधे १० पर सोना जाता है ।

सही अनुवाद : रोहित साढ़े दस बजे सो जाता है।

उपरोक्त स्रोत वाक्य के मशीनी अनुवाद तंत्रों द्वारा हिंदी अनुवाद में भिन्नता दिख रही है। उसमें अंकों का प्रयोग होने से अनुवाद की संरचना में बदलाव आया है। मंत्र-राजभाषा ने तो वाक्य के अनुवाद के बजाय शब्दवर्ग अंकन में अनुवाद किया है जिससे संपूर्ण अनुवाद गलत हुआ है। सही अनुवाद 'रोहित साढ़े दस बजे सो जाता है' होगा।

41. अंग्रेजी वाक्य : CDC has said that the death rate is considered to be 15 to 50 %.

गूगल : सीडीसी ने कहा है कि मृत्यु दर 15 से 50% करने के लिए माना जाता है।

माइक्रोसॉफ्ट : सीडीसी ने कहा कि मृत्यु दर 15 से 50% किया जा करने के लिए माना जाता है।

मंत्र-राजभाषा : सीडीसी कहा है कि मृत्यु दर 50 प्रतिशत तक 15 होने के लिए विचार किया जाता है

आंग्ल-हिंदी : सीडीसी कहा है कि मृत्यु - दर १५ से ५० तक प्रतिशत होना समझी जाती है।

सही अनुवाद : सी.डी.सी. ने बताया है कि मृत्यु दर 15 से 50 प्रतिशत तक मानी है।

उपरोक्त सभी मशीनी अनुवाद तंत्रों में से आंग्ल-हिंदी के अनुवाद को देखा जा सकता है जिसमें भारतीय अंकों का प्रयोग हुआ है। वाक्य में शाब्दिक अनुवाद होने से संपूर्ण अर्थ की प्राप्ति नहीं हो रही है। अनुवाद में अंतरराष्ट्रीय अंकों का प्रयोग होना चाहिए जिससे अनुवाद सरल और सटीक होगा।

➤ इनपुट वाक्य या शब्द में त्रुटियाँ (Error with input sentence or word)

42. अंग्रेजी वाक्य : Akbar's Red fort is also in Agra.

गूगल : Akbars लाल किला आगरा में भी है।

माइक्रोसॉफ्ट : Akbars लाल किला आगरा में भी है।

मंत्र-राजभाषा : अक्बर्स लाल कि ला भी आगरा में है

आंग्ल-हिंदी : अक्बार्स रेद किला भी आगरा में है।

सही अनुवाद : आगरा में अकबर का लाल किला भी है।

मशीनी अनुवाद से वाक्य का अनुवाद करने से पूर्व कुछ शब्दों की वर्तनी में गलतियाँ होने से अनुवाद गलत होता जाता है तथा कुछ शब्दों का अनुवाद नहीं होता। जिसकी वजह से संपूर्ण वाक्य का अर्थ संप्रेषित नहीं होता। उपरोक्त गूगल और माइक्रोसॉफ्ट से अनूदित वाक्य में 'Akbar' संज्ञा शब्द है जिसके साथ 's' प्रयुक्त होने से वह शब्द अनअनूदित (Untranslated) हुआ है तथा अन्य तंत्रों ने उसका अनुवाद कर लिप्यंतरित किया है।

43. अंग्रेजी वाक्य : (They) call it the Buland Darwaazaa.

गूगल : (वे) यह बुलंद Darwaazaa कहते हैं।

माइक्रोसॉफ्ट : (वे) इसे बुलन्द Darwaazaa कहते हैं।

मंत्र-राजभाषा : यह बुलांड दर्वाअज़ा फोन करें

आंग्ल-हिंदी : (ठ्हेय) आप यह बुलान्द दारवाज़ा को बुलाइये ।

सही अनुवाद : इसको बुलंद दरवाज़ा कहते हैं।

अनूदित वाक्यों में गूगल और माइक्रोसॉफ्ट ने 'Buland Darwaazaa' शब्द का वैसे ही अनुवाद किया है। यह स्रोत वाक्य की वर्तनी में गलती होने की वजह से अनुवाद नहीं

हुआ है। मंत्र-राजभाषा के अनूदित वाक्य से अलग तरह का अर्थ संप्रेषित हो रहा है। आंग्ल-हिंदी वाक्य में वर्तनी की अधिक गलतियाँ पाई जाती हैं तथा अनुवाद शब्द स्तर पर होने की वजह से अनुवाद अर्थ के धरातल पर गलत हुआ है। स्रोत वाक्य में उस शब्द का 'Buland Darwaza' होना चाहिए था, जिसका अनुवाद 'बुलंद दरवाजा' होता है।

44. अंग्रेजी वाक्य : Near the Imaambaaraa is the Aasafi Mosque.

गूगल : Imaambaaraa पास Aasafi मस्जिद है

माइक्रोसॉफ्ट : Aasafi मस्जिद के पास Imaambaaraa है

मंत्र-राजभाषा : इमाअम्बारा के पास आसफि मसजिद है

आंग्ल-हिंदी : पास इमाम्बारा आसाफि मोस्क्यू है ।

सही अनुवाद : इमामबाड़ा के पास आसफी मस्जिद है।

उपरोक्त स्रोत वाक्य में भी 'Imaambaaraa' में वर्तनी की गलती होने की वजह से गूगल और माइक्रोसॉफ्ट मशीनी अनुवाद तंत्र ने इस नाम का अनुवाद नहीं किया है। मंत्र-राजभाषा और आंग्ल-हिंदी ने इस शब्द का अनुवाद किया है, लेकिन उसमें वर्तनी की गलतियाँ मिलती हैं। 'Imambara' और 'Asfi' यह वर्तनी की दृष्टि से अंग्रेजी में मानक शब्द है तथा उसका अनुवाद 'इमामबाड़ा' और 'आसफी' होता है।

➤ विशेषण शब्द में त्रुटियाँ

अनुवाद में विशेषण शब्द में भी त्रुटियाँ पाई जाती हैं जिससे संज्ञा की विशेषता पूर्ण नहीं होती और वाक्य अर्थ की दृष्टि से अस्पष्ट दिखाई देता है।

45. अंग्रेजी वाक्य : The table is round.

गूगल : तालिका गोल है।

माइक्रोसॉफ्ट : तालिका दौर है।

मंत्र-राजभाषा : सारणी चरण है

आंग्ल-हिंदी : मेज घूमकर है।

सही अनुवाद : मेज गोल है।

उपरोक्त अंग्रेजी वाक्य में 'round' शब्द 'table' की विशेषता बता रहा है। जिसमें सभी मशीनी अनुवाद तंत्रों ने हिंदी में अनुवाद किया है। उसमें गूगल और माइक्रोसॉफ्ट ने 'table' का अनुवाद 'तालिका' किया है। माइक्रोसॉफ्ट ने भी 'round' शब्द के लिए 'दौर' शब्द का प्रयोग किया है। तो मंत्र-राजभाषा ने 'table और round' के लिए 'सारणी और चरण' का प्रयोग किया है, जो स्रोत वाक्य संरचना के अनुसार गलत है। आंग्ल-हिंदी तंत्र ने विशेषण शब्द का अनुवाद 'घूमकर' किया है जो वाक्य संरचना में गलत शब्द के पर्याय का चयन किया है। इसके लिए 'सही अनुवाद' को पढ़ा जा सकता है।

46. अंग्रेजी वाक्य : He is more gentle than intelligent.

गूगल : वह बुद्धिमान से ज्यादा कोमल है।

माइक्रोसॉफ्ट : वह बुद्धिमान से अधिक कोमल है।

मंत्र-राजभाषा : वह ज्यादा शरीफ से बुद्धिशाली है

आंग्ल-हिंदी : < { वह सि और विनम्र से बुद्धिमान ।

सही अनुवाद : वह बुद्धिमान कम है, शरीफ अधिक है।

स्रोत वाक्य का गूगल और माइक्रोसॉफ्ट द्वारा हिंदी अनुवाद शब्द के स्तर पर ठीक है। मंत्र-राजभाषा के अनुवाद में शब्दक्रम का अभाव है जिसके चलते अनुवाद बोधगम्य नहीं है और आंग्ल-हिंदी से अपूर्ण अनुवाद मिला है जिससे वाक्य का अर्थ स्पष्ट नहीं हो रहा है।

47. अंग्रेजी वाक्य : A man went for a walk with his dog.

गूगल : एक आदमी अपने कुत्ते के साथ चलने के लिए गया था

माइक्रोसॉफ्ट : एक आदमी अपने कुत्ते के साथ टहलने के लिए गया था।

मंत्र-राजभाषा : पुरुष उनका कुत्ता सहित चाल के लिए गया

आंग्ल-हिंदी : एक आदमी उसके कुत्ते के साथ एक चाल पर लागू हुआ।

सही अनुवाद : एक व्यक्ति अपने कुत्ते के साथ घुमने गया।

उपरोक्त अंग्रेजी वाक्य का सभी मशीनी अनुवाद तंत्रों द्वारा शाब्दिक अनुवाद हुआ है, जिसमें आर्थी संरचना के स्तर पर सभी वाक्य में अनेकार्थकता उत्पन्न हो रही है।

#### ➤ सरल वाक्य (Simple sentence)

मशीनी अनुवाद तंत्रों में सरल वाक्य देने से तंत्रों की अनुवाद संपादन करने की क्षमता बढ़ती है। लेकिन अगर इनपुट वाक्य जटिल और लंबे हों तो कंप्यूटर के लिए यह अनुवाद करना कठिन हो जाता है।

48. अंग्रेजी वाक्य : The sun sets in the west.

गूगल : सूर्य पश्चिम में अस्त होता है।

माइक्रोसॉफ्ट : सूर्य पश्चिम में सेट करता है।

मंत्र-राजभाषा : पश्चिम में सूर्यास्त

आंग्ल-हिंदी : धूप पश्चिम में छिपती है।

सही अनुवाद : सूर्य पश्चिम में अस्त हुआ है।

अंग्रेजी वाक्य का सभी मशीनी अनुवाद तंत्रों से हिंदी अनुवाद प्राप्त करने के बाद देखा जा सकता है कि गूगल और माइक्रोसॉफ्ट ने सही अनुवाद किया है, लेकिन एक में लिप्यंतरित शब्द का प्रयोग किया है। मंत्र-राजभाषा का अनुवाद सही है लेकिन इसमें पूर्ण विराम का वाक्य के अंत में प्रयोग नहीं हुआ है। आंग्ल-हिंदी शब्द का सही पर्यायी चयन न करने से गलत अनुवाद किया है। इसके लिए सही अनुवाद देखा जा सकता है।

49. अंग्रेजी वाक्य : Rajesh brought his friend's house.

गूगल : राजेश अपने दोस्त के घर लाया।

माइक्रोसॉफ्ट : राजेश अपने दोस्त के घर ले आया।

मंत्र-राजभाषा : राजेश उनका दोस्त सदन लाया

आंग्ल-हिंदी : राजेश उसका मित्र का घर लाया था ।

सही अनुवाद : राजेश ने अपने मित्र का घर खरीदा।

अंग्रेजी वाक्य का सभी मशीनी अनुवाद तंत्रों ने शाब्दिक अनुवाद किया है, जिसमें अर्थ के स्तर पर सभी तंत्रों से निर्गत अनुवाद में 'लाया, ले आया' क्रिया का प्रयोग सही नहीं है। यह मशीनी अनुवाद तंत्रों से तब होता है जब उसके कॉर्पोरा में सीमित शब्दावली का समावेश होता है। इसलिए उपरोक्त 'सही अनुवाद' पढ़ने से ज्ञात होगा कि तंत्रों के अनुवाद में गलतियाँ कहाँ हो रही हैं।

#### ➤ संयुक्त वाक्य (Compound sentence)

अक्सर मशीनी अनुवाद में संयुक्त और मिश्रित वाक्य के अनुवाद में अनेकार्थकता पाई जाती है तथा अनुवाद भी गलत होता है। इसमें सरल और मिश्र वाक्यों का मेल संयोजक अव्ययों द्वारा होता है। इसलिए इसे संयुक्त वाक्य कहते हैं। नीचे कुछ वाक्य दिए गए हैं।

50. अंग्रेजी वाक्य : The doctor diagnosed the patient and gave him the medicine.

गूगल : चिकित्सक ने रोगी की दवाई की और उसे दवाई दी।

माइक्रोसॉफ्ट : डॉक्टर मरीज का निदान किया और उसे दवा दे दी।

मंत्र-राजभाषा : चिकित्सक मरीज रोग निदान किया और उसको औषधि दिया है

आंग्ल-हिंदी : चिकित्सक ने रोगी का रोग ज्ञात किया और उसे दवा दी।

सही अनुवाद : डॉक्टर ने मरीज को देखा और उसे दवा दी।

उपरोक्त स्रोत वाक्य का गूगल ने दो बार वाक्य में 'दवाई की' शब्द का अनुवाद किया है, जिससे अनुवाद सही नहीं लग रहा है। माइक्रोसॉफ्ट का अनुवाद सही है। मंत्र-राजभाषा और आंग्ल-हिंदी का अनुवाद शब्द स्तर पर सही है।

51. अंग्रेजी वाक्य : One child was a playing and girl was reading.

गूगल : एक बच्चा एक खेल रहा था और लड़की पढ़ रही थी।

माइक्रोसॉफ्ट : एक खेल रहा एक बच्चा था और लड़की पढ़ रहा था।

मंत्र-राजभाषा : कोई भी संतान निभा और लड़की था पढ़ रहा था

आंग्ल-हिंदी : < { एक बच्चा था इक खेलनेवाला और लड़की था रीडिंग।

सही अनुवाद : एक बालक खेल रहा था और एक बालिका पढ़ रही थी।

इन वाक्यों के अनुवाद में शाब्दिक अनुवाद किया गया है, जिनमें गूगल का अनुवाद सही है। लेकिन माइक्रोसॉफ्ट के अनुवाद में संज्ञा शब्द 'लड़की' का क्रिया पर प्रभाव पड़ता है। इसलिए अनुवाद में 'लड़की पढ़ रही थी' होना चाहिए था। मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद में वर्तनी के साथ शब्द पर्याय चयन की समस्या पाई जाती है।

### ➤ मिश्रित वाक्य (Complex sentence)

इस वाक्य में एक प्रधान उपवाक्य और एक आश्रित उपवाक्य होता है जिससे मिश्रित वाक्य संरचना का निर्माण होता है। इन वाक्यों में मशीनी अनुवाद तंत्रों से वर्तनी, शब्द पर्याय चयन तथा लिप्यंतरण की समस्या पाई जाती है।

52. अंग्रेजी वाक्य : The house which is three stories high, is home of three family.

गूगल : यह घर तीन कहानियां ऊंचा है, तीन परिवार का घर है।

माइक्रोसॉफ्ट : है जो तीन कहानियों उच्च, है घर तीन परिवार के घर।

मंत्र-राजभाषा : सदन जो तीन कथाएँ उच्च, है तीन परिवार की घर है

आंग्ल-हिंदी : हाउसेविहचिस तीन कहानी उच्च है घर का/की/के तीन परिवार।

सही अनुवाद : जो मकान तीन मंजिला है, उसमें तीन परिवार रहते हैं।

स्रोत वाक्य का गूगल और माइक्रोसॉफ्ट ने हिंदी में शाब्दिक अनुवाद किया है, जिसमें मूल शब्द से अलग ही अर्थ निकल रहा है। मंत्र-राजभाषा ने 'घर' का 'सदन' और 'stories' का 'कथाएँ' अनुवाद किया है जो शाब्दिक अनुवाद है। आंग्ल-हिंदी के अनुवाद में वर्तनी की समस्या स्पष्ट हो रही है, जिससे अनुवाद बोधगम्य नहीं है।

53. अंग्रेजी वाक्य : People, who are hard worker, are successful.

गूगल : जो लोग कड़ी मेहनत करते हैं, वे सफल होते हैं

माइक्रोसॉफ्ट : लोग हैं, जो कठिन कार्यकर्ता कर रहे हैं, सफल हैं।

मंत्र-राजभाषा : आम आदमी, जो कठिन कार्यकर्ता हैं, सफल हैं

आंग्ल-हिंदी : लोग व्होआरे कठोर कार्यकर्ता हैं सफल।

सही अनुवाद : जो व्यक्ति परिश्रमी होती हैं, वे अवश्य सफल होते हैं।

स्रोत अंग्रेजी वाक्य का हिंदी अनुवाद में गूगल का अनुवाद सही है। माइक्रोसॉफ्ट और मंत्र-राजभाषा के वाक्य में उपयुक्त शब्दक्रम न होने की वजह से अनुवाद में संभ्रमता स्पष्ट हो रही है। आंग्ल-हिंदी के अनुवाद में भी शब्द चयन की समस्या आ रही है। जिसके लिए 'सही अनुवाद' को पढ़ा जा सकता है।

उपरोक्त मशीनी अनुवाद तंत्रों से प्राप्त अनूदित वाक्यों में विभिन्न प्रकार की त्रुटियाँ पायी गई हैं। उन त्रुटियों के विश्लेषण में कई प्रकार के वाक्यों का विश्लेषण किया गया है। ये त्रुटियाँ प्रमुख रूप से वर्तनी, उच्चारण, कैपिटल शब्द, संक्षिप्तीकरण, लिंग का क्रिया पर प्रभाव, अनअनूदित शब्द, लिप्यंतरण, गलत शब्द चयन, संदिग्धता अथवा

अनेकार्थकता, अर्थगत त्रुटियाँ, वाक्यांश, शब्दक्रम, सरल वाक्य, संयुक्त और मिश्रित वाक्यों में पायी जानेवाली त्रुटियों का विश्लेषण किया गया है। जिनके कारण अनूदित वाक्य की रूपरचना, शब्द रचना, वाक्य रचना और आर्थी संरचना में काफ़ी बदलाव पाया गया है जिससे अनुवाद की बोधगम्यता (comprehensibility) और प्रवाहता (fluency) प्रभावित हुयी है। इसलिए इन त्रुटियों के विश्लेषण के बाद इन समस्याओं का समाधान निकाला गया है।

#### 5.4 मशीनी अनुवाद की त्रुटियों के लिए सुधारात्मक उपाय

मशीनी अनुवाद की त्रुटियों के विश्लेषण के बाद उसमें आनेवाली समस्या का समाधान होना आवश्यक है। इसलिए यहाँ मशीन अनूदित वाक्यों की त्रुटियों का विश्लेषण मानव द्वारा करने के बाद त्रुटियों में सुधार के लिए प्रमुख सुधारात्मक उपाय दिए जा रहे हैं। भाषा शिक्षण में भी त्रुटि विश्लेषण में अक्सर होने वाली गलतियों के लिए निदानात्मक उपाय दिए जाते हैं। वैसे ही ये सुधारात्मक उपाय गूगल, माइक्रोसॉफ्ट, मंत्र-राजभाषा और आंग्ल-हिंदी मशीनी अनुवाद तंत्रों के संदर्भ में हैं। ये उपाय इसलिए दिए जा रहे हैं ताकि इन तंत्रों में सुधार करते समय या अद्यतन करते समय इन त्रुटियों पर विशेष ध्यान दिया जाए, जिससे इनकी अनुवाद करने की गुणवत्ता में सुधार होगा।

- शब्दक्रम - भाषाई युग्म के बीच संरचनात्मक भिन्नता को दूर करने के लिए शब्दक्रम संरचना का अनुगमन करना चाहिए। अंग्रेजी भाषा का शब्दक्रम कर्ता-क्रिया-कर्म (SVO) पर आधारित है तथा हिंदी भाषा का शब्दक्रम कर्ता-कर्म-क्रिया (SOV) पर आधारित है। इसलिए दोनों भाषाओं के शब्दक्रम का विशुद्ध रूप से मशीनी अनुवाद तंत्रों में समावेश करना चाहिए।
- द्विभाषी शब्दकोश - सूचनाओं के संचार के चलते प्रतिदिन नए शब्दों का प्रयोग बाजार में किया जाता है। इन शब्दों का मशीनी अनुवाद तंत्रों द्वारा अनुवाद करना कठिन हो जाता है। इसलिए इन शब्दों को अद्यतन कर अंग्रेजी-हिंदी द्विभाषी शब्दकोश में समावेश किया जाना चाहिए।

- कॉर्पोरा के वाक्यों का संकलन विशिष्ट (Specific) विषय क्षेत्र में अनुवाद करने के लिए समृद्ध होना चाहिए जिससे अनुवाद के लिए मशीनी अनुवाद तंत्रों में अधिक सामग्री उपलब्ध होगी और भाषाई नियमों के कारण अनुवाद सही होगा।
- वाक्य स्तर पर शाब्दिक अनुवाद के बजाय, संदर्भ के अनुसार शब्दावली में से उपयुक्त पर्याय चयन करने की क्षमता तंत्रों में विकसित करने की आवश्यकता है। जिससे अज्ञात शब्द और अनावश्यक शब्दों का अनुवाद होना कम होगा और अनुवाद सही और बोधगम्य होगा।
- गूगल द्वारा और माइक्रोसॉफ्ट द्वारा व्यक्ति और स्थल के नाम के अनुवाद में हमेशा समस्याएँ आती हैं। जिसके लिए भाषाई सामग्री में अधिक से अधिक इन नाम एवं स्थलों का समावेश करने की आवश्यकता है। मशीनी अनुवाद तंत्रों द्वारा नाम का शब्दशः अनुवाद भी नहीं होना चाहिए अन्यथा अनेकार्थकता की समस्या उत्पन्न होगी।
- बृहद वाक्यों के अनुवाद में आंग्ल-हिंदी से अक्सर त्रुटियाँ होती हैं। यह वाक्य अनुवाद के लिए बड़ा या बृहद हो तो अनुवाद तंत्र से काफी गलतियाँ होती हैं। इसलिए अनुवाद के लिए छोटे वाक्य देना ही सही है। जिससे अनुवाद सटीक होगा।
- अनुवाद में क्रियाओं का प्रयोग, वर्तनी की समस्या, विराम चिह्न प्रयोग तथा पूर्वसर्ग तथा परसर्ग प्रयोग के लिए मानक व्याकरणिक संरचना का समावेश होना चाहिए।
- अनुवाद में लिप्यंतरण के लिए दोनों भाषाओं की लिपि का समावेश होना चाहिए, जिससे लिप्यंतरित शब्द में वर्तनी की समस्या उत्पन्न नहीं होगी।

- हिंदी में अक्सर क्रिया पर लिंग का प्रभाव पड़ता है। जैसे 'गांधी एक गुजराती थी' के बजाय 'गांधी एक गुजराती थे' होना चाहिए। इसके लिए दोनों भाषाओं की व्याकरणिक संरचना में सुधार करने की आवश्यकता है।
- समनामिता (Homonymy), अनेकार्थी (Polysemy) और संदिग्धार्थकता (ambiguity) की समस्या से निपटने के लिए शब्द आशय विसंदिग्धीकरण (Word sense disambiguation) का प्रयोग मशीनी अनुवाद में किया जाता है। इससे प्रसंग विशेष में उस शब्द के सही आशय को निश्चित किया जाता है। इसमें विसंदिग्धीकरण की प्रक्रिया में संदर्भ की समानता के आधार पर शब्द के विशिष्ट अर्थ को स्थापित करने के लिए नियमों को सामान्यीकृत करने का प्रयास किया जाता है।
- विशेष रूप से वर्तनी की त्रुटियाँ, कैपिटल शब्दों के प्रयोग, गलत शब्द चयन और अंकों के प्रयोग की समस्या के लिए आंग्ल-हिंदी तंत्र को वर्तनी जाँचक में सुधार करने की तथा द्विभाषिक शब्दावली में नए शब्दों का समावेश करने की आवश्यकता है।
- मंत्र-राजभाषा में विराम चिन्हों का प्रयोग वाक्य के अंत में अनुवाद में नहीं होता। इसलिए इस तंत्र के लिए विराम चिह्न और हिंदी शब्द की वर्तनी में सुधार करने की आवश्यकता है।
- संदिग्ध शब्द, नाम पद, मुहावरे और लोकोक्तियों का अनुवाद तथा सांस्कृतिक तत्वों से भरपूर पाठ (जैसे कविता) का अनुवाद आदि मशीनी अनुवाद के लिए चुनौतियाँ उत्पन्न करते हैं। इसलिए इनका कंप्यूटर से सही अनुवाद कर पाना मुश्किल कार्य है इसके लिए मानव अनुवाद की ही सहायता लेनी चाहिए।
- मशीनी अनुवाद तंत्रों में सरल वाक्य देने से तंत्रों की अनुवाद करने की क्षमता बढ़ती है। लेकिन इनपुट मिश्र वाक्य बड़ा है तो कंप्यूटर से अनुवाद करना कठिन हो जाता है। इसलिए सरल वाक्यों का अधिक प्रयोग किया जाना चाहिए जिससे अनुवाद सरल होगा।

उपरोक्त मशीनी अनुवाद की त्रुटियों के समाधान के लिए सभी उपाय निश्चित तौर पर आवश्यक हैं जिससे मशीनी अनुवाद तंत्र से होने वाली त्रुटियों का निदान होगा और सही एवं बोधगम्य अनुवाद प्राप्त किया जा सकेगा।

निष्कर्ष में यह कहा जा सकता है कि दोनों भाषाओं की संरचना में मशीनी अनुवाद के स्तर पर काफी भिन्नताएँ पाई जाती हैं जिसकी वजह से मशीनी अनुवाद तंत्र से अक्सर गलतियाँ होती हैं। मशीनी अनुवाद तंत्रों से अंग्रेजी-हिंदी भाषा में अनुवाद प्राप्त करने के बाद उनसे वाक्य में होने वाली त्रुटियों का विश्लेषण इस अध्याय में किया गया है। यह त्रुटि विश्लेषण कॉर्पोरा में संकलित वाक्यों पर आधारित है जिसे मानव द्वारा विश्लेषित किया गया है। साथ ही इसमें त्रुटियों के वर्गीकरण को स्पष्ट किया गया है जिसमें वर्ण विचार (Orthographic), रूपात्मक संबंधी (morphological), शब्द स्तर पर (lexical), अर्थ के स्तर पर (semantic) और वाक्यात्मक (syntactic) स्तर पर होने वाली त्रुटियों का विश्लेषण किया गया है।

मशीनी अनुवाद द्वारा अंग्रेजी से हिंदी में अनूदित वाक्यों में अनेक त्रुटियों का विश्लेषण किया गया है। इन त्रुटियों में प्रमुख रूप से वर्तनी, संक्षिप्तीकरण, क्रिया रूपरचना, लिंग की रूप रचना, अनअनूदित शब्द, लिप्यंतरण, गलत शब्द चयन, संदिग्धार्थकता, अर्थगत त्रुटियाँ, वाक्यांश, कहावतें एवं मुहावरों की त्रुटियाँ, शब्दक्रम, परसर्ग और पूर्वसर्ग की त्रुटियाँ, आर्टिकल संबंधी त्रुटियाँ, सरल, संयुक्त और मिश्रित वाक्यों में आनेवाली त्रुटियों का विश्लेषण किया गया है। साथ ही मशीनी अनुवाद में सुधार करने के लिए वाक्य के पुनःक्रम (Reordering), द्विभाषी शब्दकोश (Bilingual Dictionary), भाषाई सामग्री का समावेश (Data Set), सामग्री का पूर्व संसाधन (Pre-processing) और तंत्र को प्रशिक्षण देना (Training) आदि सुझाव दिए गए हैं। अंत में केवल यह कह सकते हैं कि त्रुटि विश्लेषण से मशीनी अनुवाद की गुणवत्ता का मूल्यांकन करना आसान हो जाता है। जिससे मशीनी अनुवाद में सुधार के लिए नवीनतम सामग्री तथा उपकरणों का निर्माण किया जा सकता है।

# उपसंहार

✚ सारांश

✚ उपलब्धि

✚ सीमाएँ

✚ मशीनी अनुवाद तंत्रों में सुधार के लिए सुझाव

✚ भविष्य में शोध की योजनाएँ

## उपसंहार

---

वर्तमान में व्यावसायीकरण और मशीनीकरण के युग में विभिन्न कार्यों के लिए अनुवाद का महत्व बढ़ता जा रहा है। विश्व में अनेक भाषाएँ हैं, प्रत्येक भाषाओं में विविध ज्ञान-सामग्री मौजूद है। उस सामग्री को दूसरी भाषा में लाने के लिए अनुवाद ही ऐसा माध्यम है जो एक भाषा से दूसरी भाषाओं रूपांतरित करने का कार्य करता है। आज मानव अनुवाद (Human Translation) के अलावा मशीनी अनुवाद (Machine Translation) का प्रयोग अनुवाद के लिए तेजी से किया जा रहा है, जिससे मानव के समय और श्रम की बचत हो रही है। आज भारत में तथा विदेशों में अनेक मशीनी अनुवाद तंत्रों का विकास हुआ है। इनसे अनुवाद करनेवाले उपभोक्ता भी प्रतिदिन बढ़ रहे हैं, लेकिन तंत्रों के अनुवाद की गुणवत्ता में अभाव है। इसलिए इन तंत्रों का मूल्यांकन करना आवश्यक है जिससे इन तंत्रों के अनुवाद की गुणवत्ता में सुधार किया जा सकता है।

प्रस्तुत शोध-प्रबंध अंग्रेजी से हिंदी मशीनी अनुवाद तंत्रों के मूल्यांकन पर आधारित है। जिसमें अंग्रेजी से हिंदी भाषाओं में अनुवाद करने के लिए उपलब्ध मशीनी अनुवाद तंत्र - गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद की गुणवत्ता का मूल्यांकन किया गया है। इस मूल्यांकन से यह ज्ञात हुआ है कि ये तंत्र किस विषय क्षेत्रों के लिए सही अनुवाद करते हैं। प्रथमतः इस शोध के लिए विभिन्न स्रोतों से कॉर्पोरा का संकलन अंग्रेजी से हिंदी वाक्यों का किया गया और उसके बाद उपरोक्त मशीनी अनुवाद तंत्रों से निर्गत अनुवाद प्राप्त कर उनका मूल्यांकन, विषय क्षेत्र (Domain wise) के अनुसार किया गया। साथ ही प्राप्त अनुवाद में होनेवाली त्रुटियों का विश्लेषण किया गया है। इस मूल्यांकन और त्रुटि विश्लेषण के परिणाम से यह ज्ञात हुआ है कि गूगल और माइक्रोसॉफ्ट विभिन्न क्षेत्रों में अनुवाद करने के लिए अधिक सक्षम हैं, लेकिन मंत्र-राजभाषा और आंग्ल-हिंदी केवल सीमित विषय क्षेत्रों में अनुवाद करने में ही सक्षम हैं। इस शोध-प्रबंध को अध्ययन की सुविधा की दृष्टि से पांच अध्यायों में विभक्त किया गया है जिनमें मशीनी अनुवाद तंत्रों के मूल्यांकन का विस्तृत विवेचन किया गया है।

प्रथम अध्याय में मशीनी अनुवाद के परिचय में विश्व स्तर पर तथा भारत में विकसित मशीनी अनुवाद तंत्रों की चर्चा की गयी है तथा इनमें समावेशित घटकों में भाषाई सामग्री (Corpora), पद विच्छेदक (Parser), वर्तनी जाँचक (Grammar checker), रूपवैज्ञानिक विश्लेषक, अंतरण व्याकरण, प्रजनक और व्याकरणिक संरचना की चर्चा की गयी है। मशीनी अनुवाद की प्रक्रिया विश्लेषण, अंतरण और प्रजनक पर आधारित होती है इसलिए इनका भी विवेचन किया गया है। इसी के साथ मशीनी अनुवाद तंत्रों में प्रयुक्त पद्धतियाँ तथा देश-विदेश में विकसित मशीनी अनुवाद तंत्रों के बारे में चर्चा की गयी है। अंत में मशीनी अनुवाद तंत्रों के लिए विकास कार्य करनेवाले विभिन्न संस्थान, विश्वविद्यालय, गैर सरकारी संगठन तथा भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (TDIL) के कार्यों के योगदान का विश्लेषण किया गया है।

द्वितीय अध्याय में अंग्रेजी-हिंदी भाषा की संरचना में दोनों भाषाओं की संरचना में आनेवाली भिन्नता संज्ञा, सर्वनाम, विशेषण, क्रिया विशेषण, क्रिया, कारक और व्याकरणिक घटक में लिंग, वचन, काल, पक्ष, वृत्ति आदि को स्पष्ट किया गया है। साथ ही दोनों भाषाओं में आनेवाली मशीनी अनुवाद की समस्याओं में शब्दक्रम, संदिग्धार्थकता, लिप्यंतरण, शब्द पर्याय चयन, द्विभाषी कोश, मुहावरों और कहावतों की समस्या, सार्वभौमिक व्याकरण का अभाव, भाषा वैज्ञानिकों का अभाव, कृत्रिम बुद्धि (Artificial Intelligence) का विकास आदि समस्याओं का विश्लेषण किया गया है।

तृतीय अध्याय में प्रस्तुत प्रबंध में प्रयुक्त शोध का उद्देश्य और शोध प्रविधि का विवेचन किया गया है। इस शोध के उद्देश्य में अंग्रेजी से हिंदी में उपलब्ध मशीनी अनुवाद तंत्रों का प्रयोग अनुवाद की गुणवत्ता की दृष्टि से मूल्यांकन और त्रुटि विश्लेषण किया गया है। उसके लिए प्रयुक्त भाषाई सामग्री का संकलन, सामग्री स्रोत और पूर्ववर्ती शोध कार्यों में अंग्रेजी से हिंदी और अंग्रेजी से भारतीय भाषाओं के लिए प्रयुक्त स्वचालित मूल्यांकन पद्धतियों के बारे में चर्चा की गई है। इस प्रबंध में प्रयुक्त शोध प्रविधि प्रायोगिक तथा सैध्दांतिक स्तर पर आधारित है जिसमें प्रायोगिक रूप में मशीनी अनुवाद तंत्रों का परीक्षण (Testing) किया गया और सैध्दांतिक रूप में मानव द्वारा 4 मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए 5 पॉइन्ट स्केल (0-4) पद्धति का प्रयोग किया गया है।

इस मानक पद्धति के द्वारा इस शोध में प्रयुक्त भाषाई सामग्री के वाक्यों का मूल्यांकन किया गया है।

चतुर्थ अध्याय में अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों का मूल्यांकन किया गया है। प्रथमतः मशीनी अनुवाद के मूल्यांकन के लिए मानव मूल्यांकन और स्वचालित मूल्यांकन पद्धतियों के बारे में चर्चा की गई है। इस शोध के लिए 18,000 कॉर्पोरा के वाक्यों का संकलन विभिन्न स्रोतों से किया गया है जिनका गूगल अनुवादक, माइक्रोसॉफ्ट अनुवादक, मंत्र-राजभाषा और आंग्ल-हिंदी मशीनी अनुवाद तंत्रों से अनुवाद किया गया है। अनूदित वाक्यों का कॉर्पोरा के 24 क्षेत्रों (domain wise) के अनुसार मानव द्वारा मूल्यांकन कर औसत स्कोर निकाला गया है। इसमें सभी मशीनी अनुवाद तंत्रों की बोधगम्यता के आधार पर अंत में कुल औसत स्कोर (Total average score) निकाले गए हैं। जिसमें गूगल अनुवादक को 2.38, माइक्रोसॉफ्ट को 2.33, मंत्र-राजभाषा को 1.90 और आंग्ल-हिंदी को 2.11 कुल औसत स्कोर मिले हैं। कुल औसत स्कोर को प्रतिशत में बदल दिया है। जिससे यह स्पष्ट हुआ है कि गूगल और माइक्रोसॉफ्ट बोधगम्य अनुवाद करते हैं। इसी के साथ मंत्र-राजभाषा और आंग्ल-हिंदी भी अनुवाद करते हैं लेकिन इनके अनुवाद में अनेक समस्याएँ निर्मित होती हैं जिसके लिए पंचम अध्याय में इन त्रुटियों का विश्लेषण किया गया है।

पंचम अध्याय में मशीनी अनुवाद से निर्गत वाक्यों में पाई गई त्रुटियों का विश्लेषण किया गया है। इनमें त्रुटियों का भाषाई आधार पर वर्गीकरण किया गया है। इसमें वर्ण विचार संबंधी, रूपात्मक संबंधी, शब्द के स्तर पर, अर्थ के स्तर पर और वाक्यात्मक संबंधी वर्गीकरण कर वाक्यों का त्रुटि-विश्लेषण किया गया है। इस विश्लेषण में पाया गया है कि मशीन अक्सर एक भाषा से दूसरे भाषा के अनुवाद में वर्तनी, कैपिटल शब्द, संक्षिप्तीकरण, लिंग का क्रिया पर प्रभाव, अनअनूदित शब्द, लिप्यंतरण, गलत शब्द चयन, संदिग्धार्थकता अथवा अस्पष्टता, अर्थगत त्रुटियाँ, वाक्यांश, कहावतें एवं मुहावरों की समस्या, शब्दक्रम, पूर्वसर्ग तथा परसर्ग, सर्वनाम, विशेषण, आर्टिकल संबंधी त्रुटियाँ, सरल, संयुक्त और मिश्रित वाक्यों में पाई गई त्रुटियों का विस्तृत

विश्लेषण किया गया है। इस त्रुटि विश्लेषण के बाद मशीनी अनुवाद तंत्रों के लिए सुधारात्मक उपाय भी दिए गए हैं।

## उपलब्धि

- इस शोध कार्य की विशिष्टता यह है कि इसमें कॉर्पोरा के विभिन्न 24 क्षेत्रों के अनुसार मशीनी अनुवाद तंत्रों का मूल्यांकन किया गया है। इसमें गूगल, माइक्रोसॉफ्ट, मंत्र-राजभाषा और आंग्ल-हिंदी प्रमुख मशीनी अनुवाद तंत्रों का समावेश है।
- विश्व में अनुवाद के लिए प्रयोग किए जानेवाले तंत्र गूगल, माइक्रोसॉफ्ट तथा भारत के मशीनी अनुवाद तंत्र मंत्र-राजभाषा और आंग्ल-हिंदी के अनुवाद की गुणवत्ता के बारे में जानकारी प्राप्त हुई, जिससे यह ज्ञात हुआ है कि भारत के मशीनी अनुवाद तंत्रों को कॉर्पोरा, द्विभाषिक शब्दावली, व्याकरणिक संरचना, अंतरण व्याकरणिक नियम, प्रजनक, वर्डनेट, प्रोगामिंग आदि में अधिक अनुसंधान करने की आवश्यकता है।
- वर्तमान में समय और पैसे के खर्च के कारण अधिकांश अनुसंधानकर्ता स्वचालित मूल्यांकन पद्धतियों (automatic evaluation method) का प्रयोग करते हैं, जिससे मशीनी अनुवाद तंत्रों के मूल्यांकन के सटीक परिणाम प्राप्त नहीं होते। इसलिए इस शोध में मानव द्वारा मूल्यांकन कर परिणाम प्राप्त किए गए हैं।
- यूरोपीय भाषाओं के लिए निर्मित की गयीं मशीनी अनुवाद मूल्यांकन की पद्धतियाँ भारतीय भाषाओं के लिए सही रूप में कार्य नहीं करतीं। ये मूल्यांकन पद्धतियाँ अलग भाषाई संरचना पर आधारित होने की वजह से हमें मानव मूल्यांकन पर निर्भर रहना होता है और भारतीय भाषाओं में अक्सर अनेकार्थता की समस्या निर्मित होती है, जिसे केवल मानव ही मूल्यांकित कर सकता है। इसलिए भारतीय भाषाओं के मशीनी अनुवाद तंत्रों का मूल्यांकन मानव द्वारा करना अधिक विश्वसनीय होता है।

- मशीनी अनुवाद तंत्रों के मूल्यांकन के परिणाम, गूगल अनुवादक-2.38, माइकोसॉफ्ट-2.33, मंत्र-राजभाषा -1.90 और आंग्ल-हिंदी- 2.11 कुल औसत स्कोर यह दर्शाते हैं कि भारत के मशीनी अनुवाद तंत्रों में अधिक सुधार की आवश्यकता है।
- इस मूल्यांकन के परिणाम का प्रयोग तंत्र को अद्यतन करने तथा अनुवाद में होनेवाली गलतियों के सुधार के लिए विकासकर्ता कर सकते हैं।

## सीमाएँ

- मशीनी अनुवाद तंत्र वैज्ञानिक तथा तकनीकी, राजनीति, पर्यटन, स्वास्थ्य, प्रशासन, बैंकिंग, अर्थव्यवस्था, शिक्षा, प्रौद्योगिकी, रक्षा, कानून इत्यादि क्षेत्रों में अनुवाद के लिए सक्षम है लेकिन साहित्य के अनुवाद में सक्षम नहीं है। इसलिए इसका मूल्यांकन करना मुश्किल कार्य है।
- नाम पद, कैपिटल शब्द, संक्षिप्तीकरण, संज्ञा का क्रिया पर प्रभाव, लिंग का क्रिया पर प्रभाव, अननुवादित शब्द, लिप्यंतरित शब्द, संदिग्धार्थकता, शब्दक्रम मुहावरे और लोकोक्तियाँ, परसर्गीय पदबंध, मिश्र क्रियाएँ, प्रतिध्वन्यात्मक अभिव्यक्तियाँ, सहप्रयोग शब्द (collocation), पुनरुक्ति शब्द (reduplication), वर्तनी और पूर्व संसाधन में प्रयुक्त शब्दों के अनुवाद में मशीनी अनुवाद में हमेशा समस्या आती है जिसका स्वचालित मूल्यांकन करना कठिन है।
- प्रस्तुत शोध में मशीनी अनुवाद तंत्रों का मूल्यांकन मानव मूल्यांकनकर्ता द्वारा किया गया है। इसमें स्वचालित मूल्यांकन पद्धतियाँ ब्लू, मेटियोर मेट्रिकमैट्रिक, मेटियोर-हिंदी, निस्ट आदि से भी करना चाहिए था जिससे दोनों के मूल्यांकन परिणाम की तुलना की जा सकती थी।

## मशीनी अनुवाद तंत्रों में सुधार के लिए सुझाव

मशीनी अनुवाद का अधिक विकास करने और अनुवाद में सुधार के लिए सभी तंत्रों में बृहत कार्पोरा, वर्तनी जाँचक, पार्सर, टैगर, रूपवैज्ञानिक विश्लेषक, शब्दकोश,

प्रजनक, स्थानांतरण व्याकरणिक नियम (Transfer Grammar rule), वाक्य विश्लेषक (Syntactic analysis), वाक प्रौद्योगिकी (Speech Technology) आदि उपकरणों (Tools) का निर्माण कर उसे मशीन में संचालित करने की आवश्यकता है। इसलिए मशीनी अनुवाद में सुधार के लिए कुछ निम्न सुझाव दिए गए हैं।

- मशीन से त्रुटियों कम हों इसलिए भाषा की संरचना स्थिर होनी चाहिए इसके लिए व्याकरणिक नियमावलियों का स्पष्ट निर्देश एवं उल्लेख, विभिन्न प्रकार के कोश एवं पारिभाषिक शब्दावली के ग्रंथों का निर्माण, संदर्भ के अनुकूल संप्रेषण की संहिताओं (modes of communication) का निर्धारण करना आवश्यक है।
- मशीनी अनुवाद के विकास या निर्माण के लिए विभिन्न स्रोतों से भाषाई सामग्री (Corpora) का संकलन करने की आवश्यकता है। मशीनी अनुवाद के शब्दकोश (Lexicon) को अधिक विस्तृत बनाया जाना चाहिए। साथ ही, इसमें मुहावरों (Idioms) और लोकोक्तियों के लिए अलग से कोश बनाया जाना चाहिए।
- यदि नियम आधारित दृष्टिकोण (Rule Based Method) का प्रयोग कर मशीनी अनुवाद तंत्रों का निर्माण करना है तो इसमें स्थानांतरण व्याकरणिक नियमों का समावेश करना होगा। इससे द्विभाषिक व्याकरणिक संरचना और शब्दक्रम में आनेवाली भिन्नता कम होगी।
- सांख्यिकी आधारित अनुवाद प्रणाली (Statistical Based Translation System) का निर्माण करना हो तो विविध क्षेत्रों के कॉर्पोरा के समानांतर वाक्यों तथा व्याकरण का समावेश कर उसे एल्गोरिदम के माध्यम से कंप्यूटर में संचालित करना चाहिए।
- अनुवाद में बहुअर्थकता की समस्या के लिए संदर्भ को समझने योग्य प्रणाली विकसित करना चाहिए और एक से अधिक अर्थ चयन की उपयुक्तता का समावेश उसमें होना चाहिए।
- सभी मशीनी अनुवाद तंत्रों की गुणवत्ता में सुधार करने के लिए अद्यतन द्विभाषिक शब्दावली (Bilingual dictionary) और अनूदित पाठ में शाब्दिक अस्पष्टता में सुधार

के लिए शब्द-आशय विसंदिग्धकरण मॉड्यूल (Word sense disambiguation module) का प्रयोग किया जाना चाहिए।

- अनुवाद में सुधार के लिए पश्च संपादन (post editing) करने की सुविधा मशीनी अनुवाद तंत्रों में करने की आवश्यकता है जिससे मशीनी अनुवाद की गुणवत्ता बढ़ेगी और अनुवाद सटीक और बोधगम्य होगा।
- भारतीय भाषाओं के लिए अनेक भाषाई उपकरणों (Tools) का निर्माण किया जाना चाहिए जिससे विशिष्ट विषय क्षेत्रों के लिए तंत्रों का निर्माण करना आसान होगा। भाषाओं की व्याकरण संबंधी सभी विशेषताओं को ध्यान में रखते हुए दृढ़ और सुस्पष्ट नियम तैयार किए जाने चाहिए जिससे अनुवाद मूलनिष्ठ और सरल होगा।

इन मशीनी अनुवाद तंत्रों के विकास के लिए सुझावों के साथ यह कहना महत्वपूर्ण है कि भारत में इस क्षेत्र में शोध, अध्ययन और मशीनी अनुवाद तंत्र निर्माण के लिए कार्य कर रहे संस्थान 'भारतीय भाषाओं के लिए प्रौद्योगिकी विकास' (TDIL), आई.आई.टी. कानपुर, आई.आई.आई.टी. हैदराबाद, हैदराबाद विश्वविद्यालय, सी-डैक पुणे, सी-डैक बंगलोर, राष्ट्रीय सॉफ्टवेयर प्रौद्योगिकी केंद्र, मुंबई, अन्ना विश्वविद्यालय, चेन्नई; जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली, आई.बी.एम.रिसर्च इंडिया आदि का कार्य विशेष उल्लेखनीय तथा सराहनीय है।

### **भविष्य में शोध की योजनाएँ**

- भविष्य में अंग्रेजी-हिंदी भाषाओं के कॉर्पोरा के अलावा भारत की अन्य भाषाएँ हिंदी से मराठी, अंग्रेजी से मराठी भाषाओं में कॉर्पोरा का संकलन किया जाएगा और मशीनी अनुवाद के लिए उपलब्ध तंत्रों से उनका मूल्यांकन किया जाएगा।
- कॉर्पोरा का संकलन विभिन्न क्षेत्र का रहेगा। इस शोध में केवल 24 विषय क्षेत्रों (domain) के कॉर्पोरा का समावेश किया गया है। भविष्यगत शोध में विशेषतः नए क्षेत्रों का समावेश किया जाएगा। जैसे- अंतरिक्ष, लघु उद्योग संबंधी कॉर्पोरा का

संकलन। इस मूल्यांकन में त्रुटि विश्लेषण के बारे में उदाहरणों के साथ भाषाई तथा व्याकरणिक संरचनाओं में होनेवाली त्रुटियों का विश्लेषण किया जाएगा।

- मशीनी अनुवाद के मूल्यांकन के परिणाम प्राप्त करने के लिए मानव द्वारा मूल्यांकन और स्वचालित मूल्यांकन पद्धति का प्रयोग कर तुलना की जाएगी। इसमें मशीनी अनुवाद मूल्यांकन पद्धति का निर्माण किया जाएगा तथा इसके लिए सुझाव दिए जाएंगे।
- मूल्यांकन और त्रुटि विश्लेषण के प्रयोग से किसी भी मशीनी अनुवाद तंत्र के अनुवाद की गुणवत्ता में सुधार किया जा सकता है, जिससे अनुवाद सरल, सटीक और बोधगम्य प्राप्त होगा।
- मानव मूल्यांकन और स्वचालित मूल्यांकन पद्धति के परिणाम में अधिक सहसंबंध होना चाहिए। इसलिए मूल्यांकन में ब्लू, मेटियोर, मेटियोर-हिंदी स्वचालित मूल्यांकन पद्धतियों का भी प्रयोग किया जाएगा। जिससे मानव मूल्यांकन और स्वचालित मूल्यांकन पद्धतियों की चर्चा कर उचित परिणाम प्राप्त किए जाएंगे।
- भारतीय भाषाओं की संरचना यूरोपीय भाषाओं से भिन्न है। इसलिए उचित स्वचालित मूल्यांकन प्रणाली का विकास किया जाएगा। जिसका प्रयोग हिंदी, मराठी, पंजाबी, बंगाली आदि प्रमुख भाषाओं में उपलब्ध तंत्रों के मूल्यांकन के लिए किया जाएगा।

अतः कहा जा सकता है कि किसी भी मशीनी अनुवाद तंत्रों का मूल्यांकन करने से उस तंत्र के अनुवाद की गुणवत्ता की जानकारी मिलती है और उसमें सुधार करने के लिए भविष्य में कार्य योजनाएँ बनायी जा सकती हैं। इसलिए आशा की जा सकती है कि भविष्य में भारतीय भाषाओं के लिए अनेक मशीनी अनुवाद तंत्रों के साथ उनकी मूल्यांकन पद्धति का विकास किया जाएगा। जिससे मशीनी अनुवाद तंत्रों को विकसित तथा उनमें सुधार करना आसान हो जाएगा।

## संदर्भ ग्रंथ-सूची

- ✚ प्राथमिक स्रोत (कॉर्पोरा संकलन)
- ✚ द्वितीयक स्रोत : हिंदी पुस्तकें, अंग्रेजी पुस्तकें, शोध पत्र, शोध-प्रबंध, लघु-शोध प्रबंध, पत्रिकाएँ, शब्दकोश, विश्वकोश, Internet Website Links

## संदर्भ ग्रंथ-सूची (BIBLIOGRAPHY)

---

### ➤ प्राथमिक स्रोत (कॉर्पोरा संकलन)

1. 'भारतीय भाषा कॉर्पोरा उपक्रम' परियोजना, 2012-2016, निर्देशक, प्रो. प्रो.गिरीश नाथ झा, विशिष्ट संस्कृत अध्ययन केन्द्र जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली।
2. फारवर्ड प्रेस पत्रिका, 2015-2016, मुख्य संपा. इवान कोस्तका, FORWARD Press, New Delhi-110019 <https://www.forwardpress.in/>
3. एअर इंडिया की 'शुभ यात्रा' पत्रिका, 2015-2016, <http://shubh-yatra.in/emagazine/>
4. आर्थिक सर्वेक्षण 2014-2015, A flagship annual document of the Ministry of Finance, Government of India.  
<http://indiabudget.nic.in/budget2015-2016/survey.asp>
5. श्रम एवं रोजगार मंत्रालय, वार्षिक रिपोर्ट 2014-2015. Ministry of Labour & Employment, Govt. of India.  
<http://labour.gov.in/annual-reports>
6. स्वास्थ्य एवं परिवार कल्याण मंत्रालय, वार्षिक रिपोर्ट 2015-2016. Department of health and family welfare, Ministry of Health and Family Welfare, Government of India.  
<http://mohfw.nic.in/documents/publications/annual-report-department-health-family-welfare-year-2015-16/annual-report-department-health-family-welfare-year-2015-16>

7. रक्षा अनुसंधान एवं विकास संगठन, वार्षिक रिपोर्ट 2015. मैटकॉफ हाउस, दिल्ली।

<https://www.drdo.gov.in/drdo/English/index.jsp?pg=public.jsp>

<https://blog.mygov.in/ebooks/drdo/DRDO Annual Report.html#page/4>

8. राजभाषा विभाग, गृह मंत्रालय, भारत सरकार, वार्षिक रिपोर्ट 2015-2016.

[http://www.rajbhasha.nic.in/sites/default/files/ar\\_20152016hin.pdf](http://www.rajbhasha.nic.in/sites/default/files/ar_20152016hin.pdf)

अन्य पत्रिका, पुस्तकें आदि स्रोतों से भाषाई सामग्री (कॉर्पोरा) का संकलन।

### ➤ द्वितीयक स्रोत : हिंदी पुस्तकें

9. अय्यर, एन.ई. विश्वनाथ (2007): *अनुवाद कला*, प्रभात प्रकाशन, दिल्ली।

10. आचार्य वाजपेयी, किशोरीदास (2006): *हिंदी की वर्तनी तथा शब्द विश्लेषण*, वाणी प्रकाशन, नई दिल्ली।

11. काचरू, यमुना (1997) : *हिंदी रूपान्तरणात्मक व्याकरण के कुछ प्रकरण*, केंद्रीय हिंदी संस्थान, आगरा।

12. कुमार, सुरेश (2005): *अनुवाद सिद्धांत की रूपरेखा*, वाणी प्रकाशन, नई दिल्ली।

13. कैटफोर्ड, जे.सी. (1983): *अनुवाद का भाषिक सिद्धान्त*, हिंदी अनुवादक डॉ. रविशंकर दीक्षित, हिंदी ग्रंथ अकादमी, मध्य प्रदेश।

14. गुप्त, गार्गी., सिंहल, ओमप्रकाश (1991): *अनुवाद चिंतन के सैद्धान्तिक आयाम*, भारतीय (संपा.) अनुवाद परिषद, दिल्ली।

15. गुरु, कामता प्रसाद (1920) : *हिंदी व्याकरण*, काशी नागरी प्रचारिणी सभा, काशी।

16. गोपीनाथम, जी. (2008) : *अनुवाद सिद्धांत एवं प्रयोग*, लोकभारती प्रकाशन, इलाहाबाद।

17. गोस्वामी, कृष्ण कुमार (2006): *व्यावहारिक हिन्दी और रचना*, वाणी प्रकाशन, नई दिल्ली।
18. गोस्वामी, कृष्ण कुमार (2012) : *अनुवाद विज्ञान की भूमिका*, राजकमल प्रकाशन, दिल्ली।
19. जैन, वृषभ प्रसाद (1995) : *अनुवाद और मशीनी अनुवाद*, सारांश प्रकाशन, नई दिल्ली।
20. टण्डन, पूरनचन्द; सेठी, हरीश कुमार (2005): *अनुवाद के विविध आयाम* तक्षशिला प्रकाशन, नई दिल्ली।
21. तिवारी भोलानाथ (संपा.) (1984): *भारतीय भाषाओं से हिंदी अनुवाद की समस्याएँ*, शब्दकार प्रकाशन, दिल्ली।
22. तिवारी, उदयनारायण (2012): *हिंदी भाषा का उदगम और विकास*, भारती भंडार, प्रयाग।
23. तिवारी, बालेन्दु शेखार (संपा.), (2010): *अनुवाद विज्ञान*, प्रकाशन संस्थान, नई दिल्ली।
24. तिवारी, भोलानाथ (1989): *अनुवाद विज्ञान*, शब्दकार प्रकाशन, दिल्ली।
25. तिवारी, भोलानाथ (1999) : *हिंदी भाषा की संरचना*, वाणी प्रकाशन, नई दिल्ली।
26. तिवारी, भोलानाथ (2001): *हिंदी भाषा की लिपि संरचना*, साहित्य सहकार, दिल्ली।
27. तिवारी, भोलानाथ (2003): *भाषाविज्ञान*, किताब महल, इलाहाबाद।
28. तिवारी, भोलानाथ (2005) : *हिन्दी भाषा*, किताब महल, इलाहाबाद ।
29. तिवारी, भोलानाथ (2009): *हिंदी भाषा की वाक्य रचना*, साहित्य सहकार, नई दिल्ली।
30. तिवारी, भोलानाथ (2011): *अनुवाद विज्ञान सिद्धांत एवं प्रविधि*, किताबघर प्रकाशन, दिल्ली।

31. त्रिपाठी, रामदेव (1986): *हिन्दी भाषानुशासन*, बिहार हिन्दी ग्रन्थ अकादमी, पटना।
32. नगेन्द्र, (संपा.), (1993): *अनुवाद विज्ञान : सिद्धांत और प्रयोग*, हिंदी माध्यम कार्यान्वय निदेशालय, दिल्ली विश्वविद्यालय।
33. नजीर, आरिफ (2005): *हिन्दी में अनुवाद की भूमिका और द्विभाषी कम्प्यूटरीकरण*, अनंग प्रकाशन, नई दिल्ली।
34. पांडे, हेमचन्द्र (2006): *भाषा स्वरूप और संरचना*, ग्रंथलोक प्रकाशन, दिल्ली।
35. पाण्डेय, राम कमल (1985): *त्रुटि विश्लेषण: सिद्धांत और व्यवहार*, केन्द्रीय हिन्दी संस्थान, आगरा।
36. पालीवाल रीतारानी (2004): *अनुवाद प्रक्रिया और परिदृश्य*, वाणी प्रकाशन नई दिल्ली।
37. पालीवाल रीतारानी (2004): *अनुवाद सिद्धांत एल प्रयोग*, वाणी प्रकाशन नई दिल्ली।
38. प्रसाद, विनोद कुमार (2008): *भाषा और प्रोद्योगिकी*, वाणी प्रकाशन, नई दिल्ली।
39. बाजपेयी, किशोरीदास (1959): *हिंदी शब्दानुशासन*, नागरी प्रचारिणी सभा, वाराणसी।
40. बालचंद्रन, लक्ष्मीबाई (1988): *हिंदी का कारक व्याकरण*, केन्द्रीय हिन्दी संस्थान, आगरा।
41. बाहरी, हरदेव (2007): *व्यावहारिक हिन्दी व्याकरण तथा रचना*, लोकभारती प्रकाशन, इलाहाबाद।
42. भाटिया, कैलाश चन्द्र (1987): *अंग्रेजी-हिंदी शब्दों का ठीक प्रयोग*, प्रभात प्रकाशन, नई दिल्ली।

43. मल्होत्रा, विजय कुमार (2007): *कंप्यूटर के भाषिक अनुप्रयोग*, वाणी प्रकाशन, नई दिल्ली।
44. रस्तोगी, कविता (2010): *समसामायिक अनुप्रयुक्त भाषाविज्ञान*, अविराम प्रकाशन, दिल्ली।
45. राजगोपालन (1973): *हिंदी का भाषावैज्ञानिक व्याकरण*, केंद्रीय हिंदी संस्थान, आगरा।
46. रेड्डी, विजयराघव (1987): *हिंदी तेलुगु संज्ञा पदबंध*, केन्द्रीय हिन्दी संस्थान, आगरा।
47. रेड्डी, विजयराघव (1998): *हिंदी-शिक्षण और भाषा-विश्लेषण*, साहित्य सहकार, नई दिल्ली।
48. वर्मा, शिवेन्द्र किशोर (2005) : *भाषा अध्ययन के विविध पक्ष*, आलेख प्रकाशन, नई दिल्ली।
49. शर्मा, राजमणि (1996): *आधुनिक भाषा विज्ञान*, वाणी प्रकाशन, नई दिल्ली।
50. शर्मा, राजमणि (2004): *अनुवाद विज्ञान*, हरियाणा साहित्य अकादमी, पंचकूला।
51. शर्मा, रामकिशोर (2007): *भाषाविज्ञान हिन्दी और भाषा और लिपि*, लोकभारती प्रकाशन, इलाहाबाद।
52. शर्मा, रमेश चन्द्र (1999): *हिंदी-अंग्रेजी व्याकरणिक संरचना*, अशर्मा प्रकाशन, नई दिल्ली।
53. श्रीवास्तव, रवीन्द्रनाथ (2010): *अनुप्रयुक्त भाषाविज्ञान सिद्धांत एवं प्रयोग* राधाकृष्ण प्रकाशन नई दिल्ली।
54. श्रीवास्तव, रवीन्द्रनाथ; गोस्वामी, कृष्णकुमार (2010): *अनुवाद सिद्धांत और समस्याएँ*, आलेख प्रकाशन, दिल्ली।

55. श्रीवास्तव, रवीन्द्रनाथ; श्रीवास्तव, बीना (1997): *भाषाविज्ञान: सैद्धांतिक चिंतन*, राधाकृष्ण प्रकाशन, नई दिल्ली।
56. सक्सेना, बाबूराम (1995): सामान्य भाषा विज्ञान, हिन्दी साहित्य सम्मेलन, इलाहाबाद।
57. सक्सेना, रामप्रकाश (1979): भारतीय भाषाओं में लिप्यंतरण, नागपूर विद्यापीठ, नागपूर।
58. सक्सेना, रामप्रकाश (2004): *लिप्यंतरण : सिद्धांत एवं प्रयोग*, वैभव प्रकाशन, छत्तीसगढ़।
59. सहगल मनमोहन (2008) : हिन्दी शोध तंत्र की रूपरेखा, पंचशील प्रकाशन, जयपूर।
60. सहगल, नगीन चंद (1993): काव्यानुवाद सिद्धान्त और समस्याएँ, हिंदी माध्यम कार्यान्वयन निदेशालय, दिल्ली विश्वविद्यालय, नई दिल्ली।
61. सहाय, चतुर्भुज (2007): *हिंदी पदविज्ञान*, कुमार प्रकाशन, आगरा।
62. सहाय, चतुर्भुज, पिल्ले टी.के.नारायण (संपा.), (1992) : *हिंदी क्रिया, काल, पक्ष एवं वृत्ति*, केन्द्रीय हिन्दी संस्थान, आगरा ।
63. सिंह, रामगोपाल (2009): *अनुवाद विज्ञान स्वरूप और समस्याएँ*, साहित्य संस्थान, गाजियाबाद।
64. सिंह, सूरजभान (2000): *हिंदी का वाक्यात्मक व्याकरण*, साहित्य सहकार, नई दिल्ली।
65. सिंह, सूरजभान (2003): *अंग्रेजी-हिंदी अनुवाद व्याकरण*, प्रभात प्रकाशन, नई दिल्ली।
66. सिंह, हरिवंश (2003) : *कारक व्याकरण की दृष्टि से हिन्दी परसर्ग विचार*, केन्द्रीय हिन्दी संस्थान, आगरा।

67. सिन्हा, लक्ष्मण प्रसाद (1985): *हिन्दी भाषा का रूपिमीय विश्लेषण*, अंशुकमल प्रकाशन पटना।
68. सिंह, दिलीप (2009): *हिंदी भाषा चिंतन*, वाणी प्रकाशन नई दिल्ली।
69. सिंह, सूरजभान (2000): *हिंदी का वाक्यात्मक व्याकरण*, साहित्य सहकार, नई दिल्ली।
70. सिंह, सूरजभान (2008): *हिंदी भाषा संदर्भ और संरचना*, साहित्य साकार, नई दिल्ली।

### **English Books**

71. A Mapping Applied Linguistics (2011): A guide for students & practitioners edited by Christopher J. hall, Patrick H. smith and Rachel Wicaksono, by Routledge, Oxon, USA. & Canada.
72. Abbi, Anvita (1980): *Semantic Grammar of Hindi: A Study in Reduplication*, Bahri Publications, New Delhi.
73. Arnold Douglas, Balkan Lorna, Meijer Siety, Humphreys R. Lee, Sadler Louisa (1994): *Machine Translation: an Introductory Guide*, Blackwell Publishers, London UK.
74. Bassett, Susan (2002): *Corpora in Applied Linguistics*. London, Cambridge University, London.
75. Bharati, Akshar, Chaitanya, Vineet and Sangal, Rajeev (2000): *Natural Language Processing*, Prentice -Hall of India, New Delhi.
76. Bharti Akshar, Chaitanya Vineet, Sangal Rajeev. (2000): *Natural Language Processing: a Paninian Perspective*, Prentice -Hall of India, New Delhi.
77. Bloomfield, L. (1933): *Language*, Holt, New York.

78. Catford, J.C. (1965): *A Linguistic Theory of Translation*, Oxford University Press, London.
79. Central Hindi Directorate (1972): *A Basic Grammar of Modern Hindi*, Ministry of Education and Social welfare, Govt. of India, New Delhi.
80. Chomsky, N. (1957): *Syntactic Structure*, The Hague Mouton.
81. Chomsky, N. (1965): *Aspect of the theory of Syntax*, Cambridge Mass, MIT Press.
82. Corder, P. (1971): *Introducing Applied Linguistics*, Penguin, London.
83. Dorr, Bonnie (1993): *Machine Translation a view from the lexicon*, The MIT Press, USA.
84. Dorr, Bonnie J. (1993): *Machine Translation: A View from the Lexicon*, The MIT Press, Cambridge.
85. Fillmore, Charles J. (1968): *The Case for Case*, In Bach and Harms (Ed.): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston.
86. Hook, Peter Edwin (1974): *The Compound Verb in Hindi*, The University of Michigan, Ann Arbor.
87. Hutchins, John W., Somers, Harold L. (1992): *An Introduction to Machine Translation*, Academic Press, London.
88. Hutchins, W.J. (1982): *The Evolution of Machine Translation Systems in V. Lawson (ed.), Practical Experience of Machine Translation*. North-Holland, Amsterdam.
89. Hutchins, W.J. (1986): *Machine Translation: Past, Presence, Future*. Ellis Horwood Series in Computers and their Applications. Chichester. New York.
90. Jacobsen, Bent (1977): *Transformational Generative Grammar: an introductory survey its genesis and development*, North-Holland publishing company, Amsterdam. Oxford Press, New York.

91. John S. White. (2003): How to evaluate machine translation. Computers and Translation, Harold Somers (Ed.). John Benjamins Publishing Company.
92. Kachru, Yamuna (1966): An Introduction to Hindi Syntax, The University of Illinois, Urbana.
93. Kachru, Yamuna (1980): Aspects of Hindi Grammar, Manohar Publications, New Delhi.
94. Kachru, Yamuna (2006): Hindi, John Benjamin's publishing Company, Amsterdam/Philadelphia.
95. Karunkaran K. & Jaykumar M. (ed.), (1988): Translation As Synthesis, Bahari Publications, New Delhi.
96. Koul, Omkar N. (2008): Modern Hindi Grammar, Dunwoody Press, USA.
97. Kothari, C.R. (2004): Research Methodology: Methods and Techniques (2<sup>nd</sup> edi.), New Age International Publishers, New Delhi.
98. Masterman, M. (1979): The Essential Skills to Be Acquired for Machine Translation, in B.M. Snell (ed.), Translating and the computer, North-Holland, Amsterdam.
99. McGregor, R. S. (1995): Outline of Hindi Grammar, Oxford University Press, Oxford. (3rd edition).
100. Mona Baker (ed.), (1998): The Routledge Encyclopaedia of Translation Studies: Routledge, London/New York.
101. Nida and Taber, Charles, R. (1969): The theory and practice of Translation, E.J. Brill, Leiden.
102. Nirenburg, Sergei (1987): Machine Translation Theoretical and Methodological issues, Cambridge University Press, London.

103. Sharma, Aryendra (1994): A Basic Grammar of Modern Hindi, Central Hindi Directorate, New Delhi.
104. Sinha, K.Binod (1986): Contrastive Analysis of English and Hindi Nominal Phrase, Bahari Publications, New Delhi.
105. Slocum, Jonathan (ed.), (1988): Machine Translation Systems, Cambridge University Press, New York.
106. Subbarao, K.V. (1984): *Complementation in Hindi Syntax*. Academic Publications, New Delhi.
107. Syal, Pushpinder and Jindal, D.V. (2003): An Introduction to Linguistics Language, Grammar and Semantics, Ashok K. Ghosh, Pasafic -Hall of India, New Delhi.
108. Thomson A.J., Martinet A.V. (edi.), (1987): A Practical English Grammar, Oxford University Press, New Delhi.
109. Verma, M.K. (1971): The Structure of Noun Phrase in English and Hindi, Motilal Banarsidas, Delhi.
110. Wilks, Yorick (2009) Machine Translation: Its Scope and Limits, UK.
111. Yule, George (1996): The study of Language, Cambridge University Press, United Kingdom.
112. Yves Gambier, Luc van Doorslaer (ed.), (2010): Handbook of Translation Studies, John Benjamins Publishing.

### **शोध पत्र (Research Papers)**

113. Alina Secar ˘a. (2005): Translation evaluation – a state of the art survey. In Proceedings of the eCoLoRe/MeLLANGE Workshop, Leeds, UK. pages 39-44.
114. Ananthakrishnan, R., Bhattacharyya, P., Sasikumar, M., Shah, & Ritesh M. (2007). "Some Issues in Automatic Evaluation of English-Hindi MT:

- More Blues for BLEU". In Proceedings of 5th International Conference on Natural Language Processing, Macmillan Publishers, India.
115. Ankush Gupta, Sriram Venkatapathy and Rajeev Sangal (2010): METEOR-Hindi: Automatic MT Evaluation Metric for Hindi as a Target Language, Proceedings of ICON-2010:8th International conference on Natural language processing, Macmillan Publishers, India.
116. Arnold, D., and R. L. Johnson, "Robust Processing in Machine translation", presented at the International Conference on the methodology and Techniques of Machine Translation, Cranfield Institute of Technology, England, 13-15 February, 1984.
117. Atul Kr Ojha, Akanskha Bansal, Sumedh Hadke, Girish Nath Jha, Evaluation of Hindi-English MT Systems, proceedings of the "2nd Workshop on Indian Language Data Resource and Evaluation" held under LREC at Reykjavik, Iceland, 26-31 May 2014.
118. Avinesh P.V.S., Transfer Grammar Engine and Automatic Learning of Reorder Rules in Machine Translation, M.sc. Computer Science Thesis, September 2010, pp.14-15.
119. Banerjee, S., Lavie, A. (2005): METEOR: An automatic metric for MT Evaluation with improved correlation with human judgments. In: Proceedings Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and Summarization of Annual Meetings of Association of Computational Linguistics, pp. 65-72.
120. Banik, D., Bhattacharyya, P., Ekbal, A., & Sen, S. (2016): IITP English-Hindi Machine Translation System at WAT 2016. Proceedings of the 3rd Workshop on Asian Translation, Osaka, Japan, December 11-17 2016. Pp. 216-222.
121. Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., Schroeder, J. (2007): (Meta) Evaluation of machine translation. In: Proceedings of the Second Workshop on Statistical Machine Translation, pp. 136-158.

122. Carbonell, J.G., and M. Tomita (1985): New Approaches to Machine translation, in S.Nirenburg (ed.), *Proceedings of the conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*.
123. Costa-Jussà, M.R., Farrús, M., & Mariño, J.B. (2010): Linguistic-based evaluation criteria to identify statistical machine translation errors. In Proceedings of EAMT, Saint Raphaël, France. pp. 52-57.
124. D. Gupta and N. Chatterjee. (2003): Identification of Divergence for English to Hindi EBMT. In *Proceedings of MT SUMMIT IX*, New Orleans, Louisiana, USA.
125. Dabbadie, M., Hartley, A., King, M., Miller, K., Hadi, W.M.E., Popescu-Belis, A., Reeder, F., Vanni, M. (2002): A Hands-On Study of Reliability and Coherence of Evaluation Metrics. In: Handbook of LREC 2002 Workshop Machine Translation Evaluation: Human Evaluation Meet Automated Metrics.
126. Dave Shachi, Parikh Jignashu, Bhattacharyya Pushpak, (2001): *Interlingua-Based English-Hindi Machine Translation and Language Divergence*, Source: Machine Translation, Vol. 16, No. 4, pp. 251-304 Published by: Springer
127. David Vilar, Jia Xu, Luis Fernando D'Haro, and Hermann Ney (2006): Error analysis of machine translation output. In Proceedings of LREC, Genoa, Italy. pp.697-702.
128. Debbie Elliott, Anthony Hartley, and Eric Atwell (2004): A fluency error categorization scheme to guide auto-mated machine translation evaluation. In Proceedings of AMTA, Washington, DC, USA. pages 64-73.
129. Denkowski M. and Lavie A. (2011), Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems, Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation.

130. Doddington, G. (2002): Automatic evaluation of machine translation quality using n-gram co occurrence statistics. Proceedings of the Human Language Technology Conference (HLT), San Diego, CA, pp. 128-32.
131. Durgesh Rao. (2001): Machine Translation in India: A Brief Survey. In "Proceedings of SCALLA 2001 Conference, Bangalore, India.
132. Dwivedi, Sanjay Kumar & Sukhadeve Pramod Premdas (2010): *Machine Translation System in Indian Perspectives*, IJCA Journal of Computer Science 6 (10): 1111-1116, ISSN 1549-3636 © 2010 Science Publications.
133. English-Hindi machine translation. Symposium on Modeling and Shallow Parsing of Indian Languages (MSPIL'06). IIT Bombay, India.
134. Farrús Cabeceran, M., Ruiz Costa-Jussà, M., Mariño Acebal, J. B., & Rodriguez Fonollosa, J. A. (2010): Linguistic-based evaluation criteria to identify statistical machine translation errors. In 14th Annual Conference of the European Association for Machine Translation, pp. 167-183.
135. Frost & Sullivan, Local Language Information Technology Market in India for MAIT COIL Tech, TDIL, Department of IT, Ministry of Communications and Information Technology, 2003.
136. Johnson, R.L., and M.A. Rosner, (1987): Machine Translation and Software Tools, in M. King (ed.), *Machine Translation Today: The State of the Art [Proceeding of the Third Lugano Tutorial , Lugano, Switzerland, 2-7 April 1984]* Edinburgh University Press, pp.154-167.
137. Jos ´e A. R. Fonollosa. (2010): Linguistic-based evaluation criteria to identify statistical machine translation errors. In Proceedings of EAMT, Saint Rapha  el, France. Pp. 52-57.
138. Joshi N., Darbari H., Mathur I. (2012): Human and Automatic Evaluation of English to Hindi Machine Translation Systems. In: Wyld

- D., Zizka J., Nagamalai D. (eds) *Advances in Computer Science, Engineering & Applications. Advances in Intelligent and Soft Computing*, vol 166. Springer, Berlin, Heidelberg, pp.423-432.
139. Joshi, N., Darbari, H., & Mathur, I. (2012): Human and Automatic Evaluation of English to Hindi Machine Translation Systems, In *Advances in Computer Science, Engineering & Applications*, pp. 423-432. Springer, Berlin Heidelberg.
140. Joshi, N., Mathur, I., Mathur, S. (2011): Translation Memory for Indian Languages: An Aid for Human Translators. In: *Proceedings of 2nd International Conference and Workshop in Emerging Trends in Technology*.
141. Kalyani Aditi ,Kumud Hemant, Singh Shashi Pal, Kumar Ajai, Darbari Hemant, Evaluation and Ranking of Machine Translated Output in Hindi Language using Precision and Recall Oriented Metrics, *International Journal of Advanced Computer Research*, March- 2014, vol.4(1) Issue-14, pp.54-55.
142. Kittredge, R., L. Bourbeau, and P. Isabelle, "Design and Implementation of an English-French Transfer Grammar", *Proceeding of the 6<sup>th</sup> ICCL [COLING]*, Ottawa, Canada, 1976.
143. Knowles, F.E., (1979): Error Analysis of SYSTRAN Output: A Suggested Criterion for the 'Internal' Evaluation of Translation Quality and a Possible Corrective for System Design", In B.M. Snell (ed.), *Translating and the computer*. North-Holland, Amsterdam pp. 109-133.
144. Lavie, A., Agarwal, A. (2007): METEOR: an automatic metric for evaluation with high levels of correlation with human judgments. In: *Workshop on Statistical Machine Translation at the 45th Annual Meeting of the Association of Computational Linguistics*.

145. Lehrberger, J., Bourbeau, L. (1988): Machine Translation: Linguistic Characteristics of MT Systems and General Methodology of Evaluation. John Benjamin Publishers.
146. Leusch, G., Ueffing, N., Ney, H. (2006): CDER: Efficient MT Evaluation Using Block Movements. In: Proceedings of 11th Conference of the European Chapter of the Association for Computational Linguistics.
147. Machine learning system architecture – Microsoft Translator, a Case Study: Vishal chowdhary Strata 2014. <https://www.slideshare.net/talktovishal/vishal-chowdhary-strata2014> (Retrieved on 18.05.2017)
148. Maja Popović and Aljoscha Burchardt. (2011): From human to automatic error Classification for machine translation output. In Proceedings of EAMT, pages 265–272, Leuven, Belgium.
149. Maja Popovic, Adriana de Gisper, Deepa Gupta, Patrik Lambert, Hermann Ney, Jose Mariño, and Rafael Banchs. (2006): Morpho-syntactic information for automatic error analysis of statistical machine translation output. In Proceedings of WMT, New York City, New York, USA. pages 1–6.
150. Mala Christopher, (2009): *Implementation of Transfer Grammar in Telugu-Hindi Machine Translation System*, Proceedings of SCONLI-2009: 3rd Students Conference of Linguistics in India.
151. Mary Flanagan. (1994): Error Classification for MT evaluation. In Proceedings of AMTA, Columbia, Maryland, USA. pages 65–72.
152. Miller, G.A., Beebe-Center, J.G. (1956): Some Psychological Methods for Evaluating the Quality of Translation. Mechanical Translation, vol.3, pp. 73-80.
153. Malik Pooja, Gupta Abhilasha, Baghel Anurag (2013): Key Issues in Machine Translation Evaluation of English-Indian Languages,

- International Journal of Engineering Research & Technology (IJERT),  
October - 2013, Vol. 2 Issue 10, pp.3244-3245.
154. Malik Pooja, Baghel Anurag Singh, An Improvement in BLEU Metric for English-Hindi Machine Translation Evaluation, International Conference on Computing, Communication and Automation (ICCCA2016), pp.331-332.
155. Murthy. K, (2002): MAT: A Machine Assisted Translation System”, In Proceedings of Symposium on Translation Support System (STRANS-2002), IIT Kanpur. pp. 134-139.
156. Nakamura, J., Tsujii J., and Nagao M., “Solutions for Problems of MT Parser. Methods Used in Mu-Machine Translation Project” *Proceedings of the 11<sup>th</sup> ICCL [COLING 86]*, Bonn, West Germany, 25-29 August 1986, pp.133-135.
157. Naskar S. and Bandyopadhyay Sivaji. (2005): Use of Machine Translation in India: Current Status. *In proceedings of MT SUMMIT X*, September 13-15, Phuket, Thailand. pp. 465-470.
158. Papineni, K., Roukos, S., Ward, T., Zhu, W.-J. (2001): Bleu: a method for automatic evaluation of machine translation, RC22176. Technical Report, IBM T.J. Watson Research Center
159. Parameswari K, Sreenivasulu N.V., Uma Maheshwar Rao G & Christopher M, (2012): Development of Telugu-Tamil Bidirectional Machine Translation System: A special focus on case divergence, *in proceedings of 11th International Tamil Internet conference*, pp 180-191.
160. Pfafflin S.M. (1956), Evaluation of Machine Translations by Reading Comprehension Tests and Subjective Judgments, *Mechanical Translation and Computational Linguistics*, vol. 8, pp 2-8.
161. Popović, M., & Burchardt, A. (2011, 30-31 May). From human to automatic error classification for machine translation output. Paper

- presented at the EAMT 11: 15th International Conference of the European Association for Machine Translation, Leuven, Belgium, PP. 265-272.
162. R Ananthakrishnan, M Kavitha, J Hegde Jayprasad, Chandra Shekhar, Ritesh Shah, Sawani Bade, M Sasikumar, (2006) : *MaTra: A Practical Approach to Fully-Automatic Indicative EnglishHindi Machine Translation*, In the proceedings of MSPIL-06.
  163. R.M.K. Sinha and A. Jain. (2003): *AnglaHindi: An English to Hindi Machine-Aided Translation System*. In Proceedings of MT SUMMIT IX, New Orleans, Louisiana, USA.
  164. R.M.K. Sinha and Anil Thakur (2005): *Translation Divergence in English-Hindi MT*, In the Proceeding of EAMT Xth Annual Conference, Budapest, Hungary, 30-31 May.
  165. R.M.K. Sinha. (2005): *Integrating CAT and MT in Anglabharati-II Architecture*. In "Proceedings of EAMT 2005", Budapest, Hungary.
  166. Ramnathan, A., Rao, D. (2003): *A Lightweight Stemmer for Hindi*. In: Proceedings of Workshop on Computational Linguistics for South Asian Languages, 10th Conference of the European Chapter of Association of Computational Linguistics, pp. 42-48.
  167. Rao D. (2001). *Machine Translation in India: A Brief Survey*. In *proceedings of SCALLA2001 Conference*, November 21 -23, NCST, Bangalore, India. [Internet Source:
  168. **REPORT:** ALPAC. *Languages and Machines: computers in translation and Linguistics*. (1966) Report of the Authentic Language Processing Advisory Committee, Division of Behavior Sciences, National Academy of Sciences, National Research council Publication 1416, Washington, D.C.

169. Root, R., (1985) : A Two-way Approach to Structural Transfer in MT, *Proceeding of the Second Conference of the European Chapter of the ACL*, University of Geneva, Switzerland, 28-29 March 1985, pp.70-72.
170. Sampark: Machine Translation System among Indian languages (2009) [http://tdildc.in/index.php?option=com\\_vertical&parentid=74](http://tdildc.in/index.php?option=com_vertical&parentid=74) , <http://sampark.iiit.ac.in/> (Retrieved on 29.12.2014)
171. Sara Stymne. (2011): Blast: A tool for error analysis of machine translation output. In Proceedings of ACL, demonstration session, Portland, Oregon, USA.
172. Shachi Dave, Jignashu Parikh and Pushpak Bhattacharyya (2001): Interlingua-based English-Hindi Machine Translation and Language Divergence. *Journal of Machine Translation*, 16 (4): 251-304.
173. Sinha, R. M. K., Jain, R. and Jain A. (2001): *Translation from English to Indian languages: ANGLABHARTI Approach*, In proceedings of Symposium on Translation Support System STRANS 2001. February 15-17, IIT Kanpur, India.
174. Sinha, R.M.K. & Anil Thakur et.al. (2005): '*Divergence Patterns in Machine Translation between Hindi and English*' Proceedings of MT Summit X, Thailand, 12-16 September.
175. Sitender, Bawa Seema (2012) : *Survey of Indian Machine Translation Systems*, JCST Vol.3, Issue 1, Jan-March 2012, ISSN:0976-8491(online)
176. Sivaji Bandyopadhyay (2000): State and Role of Machine Translation in India. *Machine Translation Review*, 11: pp. 25-27.
177. Sivaji Bandyopadhyay (2002): Teaching MT – An Indian Perspective. In "Proceedings of the 6<sup>th</sup> EAMT Workshop on Teaching Machine Translation", Manchester, UK, pp. 13-22.
178. Sivaji Bandyopadhyay (2004) Use of Machine Translation in India. *AAMT Journal*, 36: 25-31.

179. Slocum, Joshua, (1984): Machine Translation: it's History, Current Status, and Future Prospects, *Proceedings of the 10<sup>th</sup> ICCL [COLING 84] and the 22d Annual Meeting of the ACL*, Stanford University California, 2-6 July 1984, pp. 546-561.
180. Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J. (2006): A Study of Translation Edit Rate with Targeted Human Annotation. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA)*, pp. 223-231.
181. Sudip Naskar & Shivaji Bandyopadhyay, (2005): Use of Machine Translation in India: Current status, *AAMT Journal*, pp. 25-31.
182. Tomer N., Sinha D. and Rai P.K. (2012): "Evaluating Machine Translation Evaluation's F-Measure Metric for English to Hindi Language Machine Translation", *International Journal of Academy Research Computer Engineering and Technology-IJAR CET*, 1(7), 151-156.
183. White, J.S. (1986): What Should Machine Translation Be?, *Proceedings of the 24<sup>th</sup> Annual Meeting of the ACL*, Columbia University, New York, July 1986, pp.267.
184. Yadav, Vineet, (ed.) "Inducing Transfer Grammar from Word Aligned Corpus", *Proceedings of ICON-2009: 7th International Conference on Natural Language Processing*, Macmillan Publishers, India.

### **शोध प्रबंध और लघु-शोध प्रबंध (Thesis and Dissertation)**

185. Asamidinova Aainura (2007): Knowledge Base for Russian-English Machine Translation Divergences, (Ph.D Thesis) JNU, Delhi.
186. Avinesh, P.V.S. (2010): *Transfer Grammar Engine and Automatic Learning of Reorder Rules in Machine Translation*, M.Sc. (Thesis), LTRC-IIIT, Hyderabad.

187. Banerjee, Esha (2011): *Intra-phrasal agreement within the Hindi Noun Phrase: Grammar Checker Approach*, M.Phil (dissertation), JNU, New Delhi.
188. Choudhary, Narayan Kumar (2006): *Developing a Computational Framework for the verb morphology of Great Andamanese*, M.Phil (dissertation), JNU, New Delhi.
189. Goyal, Vishal (2010): *Development of a Hindi to Punjabi Machine Translation System* PhD (Thesis), Department of Computer Science, Punjabi University, Patiala.
190. Greenfield, C.C, and D. Serain. (1977): *Machine-Aided translation: From Terminology Banks to Interactive Translation Systems*. Department of Computer Science. Carnegie-Mellon University, Pittsburgh, Pennsylvania, and Institute de Recherché en Informatique et. Automatique, Paris, France, Available from EDRS.
191. Gupta, Deepa (2005) : *Contributions to English to Hindi Machine Translation Using Example Based Approach*, PhD (Thesis), Department of Mathematics, IIT, Delhi.
192. Jha Girisha Nath (1996): *Designing a Lexical Semantic Component for English - Hindi Machine Translation in the domain of Share-Market*, PhD. Thesis, Centre of Linguistics & English School of Languages, JNU, New Delhi.
193. Mathur, Prashant (2011): *Automatic Translation of Noun Compounds from English to Hindi*, M.Sc. (Thesis), LTRC-IIIT, Hyderabad.
194. Mohamed Azza Abd Ei- Moniem, (2000): *Machine Translation of Noun Phrases: From English to Arabic*, M.Sc. (Computer Science Thesis), Faculty of Engineering, Cairo University, Giza, Egypt.
195. Niranjana keshav, (2013): *Natural Language Interface Database*, Lap Lambert Academic Publishing, Germany.

196. Pathak, Kumar Nripendra (2011): *challenges in Sanskrit-Hindi Noun Phrase Mapping*, M.Phil (dissertation), JNU, New Delhi.
197. Sachdeva, Kunal, (2016): Hindi to English Machine Translation, ( M.Sc Thesis), International Institute of Information Technology, Hyderabad.
198. Sharma, Nakul, (2011): English to Hindi Statistical Machine Translation system, ME (Thesis), Computer Science and Engineering Department, Thapar University, Patiala.
199. कांबले, प्रकाश अभिमन्यु (2009) : *हिंदी-मराठी मशीनी अनुवाद में शाब्दिक अस्पष्टता (समान उच्चारण वाले शब्दों के संदर्भ में)* एम.फिल. (लघु-शोध प्रबंध), ज.ने.वि; नई दिल्ली.
200. दीक्षित, नेहा (2014): वाक्यात्मक परिप्रेक्ष्य में हिंदी क्रिया वर्ग अभिज्ञानक, (पीएच. डी.शोध-प्रबंध), प्रौद्योगिकी अध्ययन केंद्र, महात्मा, गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा।
201. पाटिल, मिलिंद बापुराव, (2013): मशीनी अनुवाद के लिए रूपात्मक सर्जक, (पीएच.डी.शोध-प्रबंध), अनुवाद अध्ययन विभाग, महात्मा, गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा।

### **पत्रिकाएँ (Magazines)**

202. Vishwbharat, A Ministry of Information Technology Publication on Language Technology in India, Issues, May,2001., January, 2003; October, 2003; Jan-April-2004,
203. अनुसृजन द्विभाषी ई-पत्रिका अंक-02, जुलाई-सितंबर (2015) : श्रीवास्तव विशेष, प्राकृतिक भाषा संसाधन में बहुशब्दीय अभिव्यक्तियों के अनुवाद की समस्याएँ, अनुवाद अध्ययन विभाग, महात्मा, गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा।

204. ओम विकास, 'तकनीकी लेखन का माध्यम हिंदी', अनुवाद त्रैमासिक, जन-मार्च 1984, पृ.60-65.
205. ओम विकास, 'हिंदी का वर्तमान और भविष्य की दृष्टि', गर्भनाल, (www.garbhanal.com) जनवरी, 2010, अंक 38, पृ. 4-6
206. गवेषणा, अंक-89, जनवरी-मार्च, (2008) : *भाषा शिक्षण पर विशेष*, केंद्रीय हिन्दी संस्थान, आगरा।
207. गवेषणा, अंक-90, अप्रैल-जून (2008) : *भाषाविज्ञान आज*, केंद्रीय हिन्दी संस्थान, आगरा।
208. गवेषणा, अंक-93, जनवरी-मार्च (2009) : *भारत में कोशविज्ञान पर विशेष*, केंद्रीय हिन्दी संस्थान, आगरा।
209. जिंदे, सुधीर, (2015), मशीनी अनुवाद और संदिग्धार्थकता, अनुसृजन द्विभाषी ई-पत्रिका, प्रवेशांक, अप्रैल-जून, अनुवाद अध्ययन विभाग, महात्मा, गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा।
210. पाण्डे, महेन्द्रकुमार सी. अनुक्रमणिक व्याकरण: एक संगणकीय पक्ष, बहुवचन अंक - 18, महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, प्रकाशन, वर्धा।
211. श्रीवास्तव, विशेष कुमार, सिन्हा स्वर्णलता, (2016): मशीनी अनुवाद और समस्याएँ, अनुसृजन द्विभाषी ई-पत्रिका, अंक-04, जनवरी-मार्च, अनुवाद अध्ययन विभाग, महात्मा, गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा। [ ISSN2454-7131(Online)]
212. गुप्ता नीता,(2004) : कंप्यूटर अनुवाद विशेषांक, अनुवाद अंक -19 भारतीय अनुवाद परिषद, नई दिल्ली।

213. त्रिपाठी, अरिमर्दन कुमार, सिंह दिलीप कुमार, (2007) : भाषा प्रौद्योगिकी :स्वरूप एवं संभावना, गवेषणा, केंद्रीय हिन्दी संस्थान, आगरा।

### **शब्दकोश (Dictionary)**

214. Bulcke, Father K. (2002): *An English- Hindi Dictionary*, S. Chand & Company Ltd., New Delhi.
215. Crystal, D. (2003): *A Dictionary of Linguistics and Phonetics*, Blackwell Publishing, USA.
216. भाषाविज्ञान शब्दावली (1992) : वैज्ञानिक तथा तकनीकी शब्दावली आयोग, नई दिल्ली।

### **विश्वकोश (Encyclopaedia)**

217. David, Crystal (1987): *The Cambridge Encyclopaedia of Language*, Cambridge University Press, New Work.
218. Mona Baker and Gabriela Saldanha (2009): *Rutledge Encyclopaedia of Translation studies* edited book, Rutledge Publication, Abingdon, London. Second edition.

### **Internet Website Links**

<http://anusaaraka.iiit.ac.in/>

<http://babelfish.yahoo.com/translate.txt>

<http://corpus.byu.edu/corpora.asp>

[http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus\\_ling/content/history.html](http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/history.html)

<http://dict.hinkhoj.com/>

[http://en.wikipedia.org/wiki/Google\\_Translate](http://en.wikipedia.org/wiki/Google_Translate)

[http://en.wikipedia.org/wiki/Yahoo!\\_Babel\\_Fish](http://en.wikipedia.org/wiki/Yahoo!_Babel_Fish)

<http://hindi-english.org/index.php?input=SYSTRAN&trans=Translate&direction=AU>  
<http://ltrc.iiit.ac.in>  
[http://prayaslt.blogspot.in/2009/08/blog-post\\_9166.html](http://prayaslt.blogspot.in/2009/08/blog-post_9166.html)  
<http://pune.cdac.in>  
<http://sampark.iiit.ac.in>  
<http://shabdkosh.raftaar.in>  
<http://shakti.iiit.ac.in/>  
<http://shubh-yatra.in/>  
<http://tdil-dc.in>  
[http://tdil-dc.in/index.php?option=com\\_vertical&parentid=74](http://tdil-dc.in/index.php?option=com_vertical&parentid=74)  
<http://translate.google.com/>  
<https://www.bing.com/translator>  
[http://www.au-kbc.org/research\\_areas/nlp/projects/anusaaraka.html](http://www.au-kbc.org/research_areas/nlp/projects/anusaaraka.html)  
<http://www.cdacindia.com/html/about/success/mantra.aspx>  
<http://www.cfilt.iitb.ac.in/>  
<http://www.cse.iitk.ac.in/users/rmk/>  
<http://www.drdo.gov.in/drdo/English/index.jsp?pg=public.jsp>  
[http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus\\_ling/contentl](http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/contentl)  
[http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus\\_ling/content/corpora/list/index2.html](http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/index2.html)  
<http://www.free-online-ocr.com/>  
<http://www.hd.uib.no/icame/kolhapur/kolman.htm#intro>  
<http://www.languageinindia.com>  
<http://www.languagetechologies.ac.in/lerc/mat/mat.htm>  
<http://www.mantra-rajbhasha.cdac.in/mantrarajbhasha/>  
<http://www.microsofttranslator.com/>

<http://www.ncst.ernet.in/matra/>  
<http://www.rajbhasha.nic.in/hi/annual-programme-and-reports>  
<http://www.shabdkosh.com/>  
<http://www.shabdkosh.com/hi/translate?e=indian&l=hi>  
<http://www.systransoft.com>  
<http://www.translatorsbase.com/search.aspx?s=34&t=58&c=99>  
<https://convertio.co/ocr/>  
[https://en.wikipedia.org/wiki/List\\_of\\_magazines\\_in\\_Hindi](https://en.wikipedia.org/wiki/List_of_magazines_in_Hindi)  
<https://www.forwardpress.in/about/>  
<https://www.hindlish.com/search.aspx?q=magazine&p=2>  
<http://anusaaraka.iiit.ac.in/>  
<http://ltrc.iiit.ac.in>  
<http://www.cdacindia.com/html/about/success/mantra.aspx>  
<http://pune.cdac.in>  
<http://202.141.152.9/matra/>  
<http://www.ncst.ernet.in/matra/>  
<http://shakti.iiit.ac.in/>  
<http://www.languagetechnologies.ac.in/lerc/mat/mat.htm>  
<http://www.cfilt.iitb.ac.in/>  
[http://www.au-kbc.org/research\\_areas/nlp/projects/anusaaraka.html](http://www.au-kbc.org/research_areas/nlp/projects/anusaaraka.html)  
<http://sampark.iiit.ac.in>

## परिशिष्ट

✚ गूगल अनुवादक आउटपुट स्कोर

✚ माइक्रोसॉफ्ट अनुवादक आउटपुट स्कोर

✚ मंत्र-राजभाषा आउटपुट स्कोर

✚ आंग्ल-हिंदी आउटपुट स्कोर

## परिशिष्ट (APPENDIX)

### गूगल अनुवादक आउटपुट स्कोर

S.No.	English Sentences	Google Translate	Domain	User 1 Score	User 2 Score	User 3 Score	User 4 Score	User 5 Score
1	Ruins of archaeological importance are found in Chandragiri and Chingitarai.	पुरातात्विक महत्व के अवशेष चंद्रगिरी और Chingitarai में पाए जाते हैं।	Tourism	3	3	3	3	4
2	Globalization-induced modernism has spread to all parts of the country.	वैश्वीकरण प्रेरित आधुनिकता देश के सभी भागों में फैल गया है।	Education	4	4	3	4	4
3	Instead of the boy, it is the girl who drives tractors and rides large motorcycles.	लड़के के बजाय, यह लड़की है जो ट्रैक्टर ड्राइव और बड़ी मोटरसाइकिल की सवारी है।	Film	3	2	2	2	4
4	She wants control over her life. She decides whom to love and marry.	वह अपने जीवन पर नियंत्रण चाहता है। वह जिसे प्यार करते हैं और शादी करने का फैसला किया।	Film	2	3	2	3	3
5	The government reduced all the prices.	सरकार सभी की कीमतों में कमी कर दी।	Economy	3	3	3	3	2
6	The natural place for sandal trees in Kerala is Marayur.	केरल में चंदन के पेड़ के लिए प्राकृतिक जगह Marayur है।	Tourism	3	4	2	3	4

7	Manjule underlines this horrendous truth in all its ugliness.	Manjule अपने सभी बदसूरती में इस भयावह सच्चाई यह रेखांकित करता है।	Film	2	1	2	1	3
8	Everyone waited for the customers.	हर कोई ग्राहकों के लिए इंतजार कर रहे थे।	Economy	3	2	3	2	3
9	Munnar is a hilly area situated at a height of 1600 meters above the sea level.	मुन्नार समुद्र तल से 1600 मीटर की ऊंचाई पर स्थित एक पहाड़ी इलाका है।	Tourism	4	3	4	4	4
10	The film is a warning to the gutsy couples opting for inter-caste marriages that they are not safe even in the metropolises.	फिल्म अंतर्जातीय विवाह के लिए चयन है कि वे महानगरों में भी सुरक्षित नहीं हैं साहसी जोड़ों के लिए एक चेतावनी है।	Film	2	2	2	2	3
11	Indian modernity is never very far from hatred and enmity.	भारतीय आधुनिकता नफरत और शत्रुता से कभी नहीं है बहुत दूर।	literature	2	3	3	2	2
12	Therefore these caves also have religious importance.	इसलिए इन गुफाओं भी धार्मिक महत्व है।	Tourism	3	3	2	3	3
13	Munnar was one of the important summer residences of India.	मुन्नार भारत की महत्वपूर्ण गर्मियों निवासों में से एक था।	Tourism	2	3	2	2	4
14	Kailash cave is at a height of 50 meter and is 2500 meters long.	कैलाश गुफा 50 मीटर की ऊंचाई पर है और 2500 मीटर लंबा है।	Tourism	4	2	4	1	3
15	It was the owners of the British gardens who got the tea garden planted in Munnar.	यह ब्रिटिश बागानों जो चाय बागान मुन्नार में लगाए	Tourism	2	2	2	0	3

		मिला के मालिकों था।						
16	Now Shah Jahan and MumtaZ mahal graves are in the Taj mahal.	अब शाहजहां और मुमताज महल कब्र ताजमहल में कर रहे हैं।	Tourism	2	3	2	2	3
17	This foreigner has also come to see Agra.	यह विदेशी भी आगरा देखने के लिए आ गया है।	Tourism	3	1	3	3	3
18	In comparison with the traditional surgical operation this method has many advantages.	पारंपरिक शल्यक्रिया के साथ तुलना में इस विधि के कई फायदे हैं।	Health	3	2	4	4	4
19	It is here that the world famous flower called Neelkurinji blooms.	ऐसा नहीं है कि विश्व प्रसिद्ध फूल खिलता Neelkurinji कहा जाता है यहां।	Tourism	2	4	2	0	2
20	One can also get to see the glimpse of Tamilnadu from the top of these hills.	एक भी इन पहाड़ियों के ऊपर से तमिलनाडु की झलक देखने को मिल सकता है।	Tourism	3	3	2	3	2
21	Perhaps the prime minister will go to Egypt.	शायद प्रधानमंत्री मिस्र के लिए जाना जाएगा।	IR	3	1	3	1	2
22	The first patient of AIDS in India was found in 1986.	भारत में एड्स का पहला मरीज 1986 में मिला था।	Health	4	4	4	4	4
23	These flowers are found in about 60 shades of blue colour.	इन फूलों को नीले रंग के बारे में 60 रंगों में पाए जाते हैं।	Miscellaneous	2	2	2	2	3
24	When it blooms it appears as if whole nature has bathed in blue colour.	जब यह खिलता ऐसा लगता है जैसे सारी प्रकृति नीले रंग में	Miscellaneous	2	3	2	4	3

		नहाया गया है।						
25	The teacher told me the story speaking slowly.	शिक्षक ने मुझे कहानी सुनाई धीरे से बोल रहा हूँ।	Educational	2	2	2	1	2
26	His mind was relieved when he saw his brother.	जब वह अपने भाई को देखा उसका मन राहत मिली थी।	General	3	2	2	2	2
27	Will you go to india via Japan or via Europe?	आप जापान के माध्यम से या यूरोप के माध्यम से भारत के लिए जाना जाएगा?	IR	2	3	2	2	1
28	While fighting, the girl threw out all her friends.	लड़ते हुए, लड़की को उसके सभी दोस्तों से बाहर फेंक दिया।	Society	1	4	1	1	1
29	I saw the cobbler repair the shoes.	मैं मोची जूते की मरम्मत देखा।	General	1	3	1	1	2
30	The five year plan in progress these days.	प्रगति में पांच साल की योजना इन दिनों।	Economy	1	2	2	1	1
31	They call it the pink city. Because many of its buildings are made of pink stone.	वे इसे गुलाबी शहर कहते हैं। क्योंकि अपने भवनों के कई गुलाबी पत्थर के बने होते हैं।	Art and culture	3	0	2	2	3
32	He uses the tools of the mainstream cinema but tells his tale.	उन्होंने मुख्यधारा के सिनेमा के उपकरणों का उपयोग करता है, लेकिन उसकी कहानी कहता है।	Film	2	1	2	2	2

33	From here the old bazaar begins.	यहाँ से पुरानी बाजार शुरू होता है।	Tourism	4	4	3	4	3
34	These people are selling cloth on the side of the street.	इन लोगों को सड़क के किनारे पर कपड़े बेच रहे हैं।	General	3	3	2	3	2
35	They come to sell their vegetables every day.	वे हर दिन उनकी सब्जियों को बेचने के लिए आते हैं।	General	4	2	3	3	3
36	Many foreigners come to Jaipur.	कई विदेशियों जयपुर के लिए आते हैं।	Tourism	3	1	2	3	3
37	That fat boy will eat up all the rasagullas.	यही कारण है कि वसा लड़का सभी rasagullas खा जाएगा।	Food	2	4	2	0	2
38	The exhibition of different levels of the production of tea leaf is very attractive.	चाय पत्ती का उत्पादन के विभिन्न स्तरों की प्रदर्शनी बहुत आकर्षक है।	Food	1	3	3	3	2
39	The blisters of the mouth does not cure and medicines do not affect.	मुँह के छाले ठीक नहीं हैं और दवाओं को प्रभावित नहीं करते।	Health	1	2	2	2	2
40	The school is the focal point of almost all his films.	स्कूल के लगभग सभी अपनी फिल्मों का केन्द्र बिन्दु है।	Film	2	3	2	1	2
41	The students wash clothes downstairs in the kitchen.	छात्रों को रसोई घर में नीचे कपड़े धो लें।	Education	2	4	2	1	2
42	The hard wind is tearing all the flags.	मुश्किल हवा सभी झंडे फाड़ है।	Geography	1	3	1	2	2

43	Where the world bares its secrets.	जहां दुनिया उसके रहस्य bares।	Miscellaneous	1	2	2	1	2
44	The young lead characters of both his films write poetry.	दोनों अपनी फिल्मों के युवा नेतृत्व पात्रों कविता लिखना।	Film	2	3	2	0	2
45	The lady helped the poor man a lot.	महिला गरीब आदमी एक बहुत मदद की।	Society	1	4	4	0	2
46	Indians give their guests tea to drink.	भारतीयों को अपने मेहमानों को चाय पीने के लिए दे।	Food	1	3	2	1	2
47	The customer will have the fruit seller give the girls oranges to eat.	ग्राहक फल विक्रेता खाने के लिए लड़कियों संतरा देना होगा।	Food	1	3	2	1	2
48	The film takes you back to your childhood and adolescence.	फिल्म अपने बचपन और किशोरावस्था के लिए तुम वापस लेता है।	Film	2	2	3	1	2
49	It touches your heart at its most tender part.	यह अपने सबसे निविदा भाग में अपने दिल को छू लेती है।	General	2	4	2	2	1
50	Fatehpur Sikri is 25 miles from Agra.	फतेहपुर सीकरी आगरा से 25 मील की दूरी पर है।	Tourism	4	3	4	4	2
51	To present the narrative of the worldly wisdom, some amount of detailing is necessary.	सांसारिक ज्ञान की कथा को पेश करने के लिए, का ब्यौरा के कुछ राशि आवश्यक है।	Philosophy	3	2	2	2	3

52	The guide is telling the foreigner about Akbar and the buildings in Sikri.	यह गाइड विदेशी को अकबर के बारे में बता रहा है और सिकरी में इमारतों।	Tourism	1	3	3	2	3
53	I am about to speak to the teacher.	मैं के बारे में शिक्षक के लिए बात करने के लिए कर रहा हूँ।	Education	1	2	2	0	2
54	There is a beautiful park also just near the Church.	एक सुंदर पार्क भी सिर्फ चर्च के पास नहीं है।	Tourism	0	3	2	0	3
55	The combined vaccine is made for children and adults.	संयुक्त टीका बच्चों और वयस्कों के लिए किया जाता है।	Health	2	3	3	3	2
56	Sita will be at home at 8 o'clock.	सीता 8 बजे घर पर किया जाएगा।	Society	1	3	1	2	2
57	There is a shortage of water in the desert.	रेगिस्तान में पानी की कमी नहीं है।	Tourism	2	2	1	0	2
58	The washermen who work in the Ashoka Hotel get a lot of money.	धोबी जो अशोक होटल में काम पैसे की एक बहुत कुछ मिलता है।	social issues	1	2	1	0	2
59	Lucknow is the capital of Uttar Pradesh.	लखनऊ उत्तर प्रदेश की राजधानी है।	Tourism	4	2	4	4	4
60	These temples are protected monuments by the Archaeology Department.	इन मंदिरों में पुरातत्व विभाग द्वारा स्मारकों को संरक्षित कर रहे हैं।	Art and culture	3	1	1	2	3
61	The family welfare programme was run at the national level in the year of 1952.	परिवार कल्याण कार्यक्रम 1952 के वर्ष में राष्ट्रीय स्तर पर चला गया था।	Health	2	2	3	4	3

62	Therefore the concept of Reproductive and Child Health Programme was born.	इसलिए प्रजनन एवं बाल स्वास्थ्य कार्यक्रम की अवधारणा का जन्म हुआ।	Health	3	2	4	4	3
63	Do you know anyone among these people?	आप इन लोगों के बीच किसी को जानते हैं?	Miscellaneous	4	3	3	4	3
64	All the people stand up together and pray.	सभी लोगों को एक साथ खड़े हो जाओ और प्रार्थना करते हैं।	Religion	2	3	1	2	2
65	Archi is the daughter of a landlord and politician as well as she is the sister of a musclemans brother.	Archi एक मकान मालिक और राजनीतिज्ञ की बेटी के रूप में अच्छी तरह से वह एक बाहुबली भाई की बहन के रूप में है।	Community	1	2	2	2	3
66	Reproductive and Child Health Programme is being run since the year of 1997 in India.	प्रजनन एवं बाल स्वास्थ्य कार्यक्रम भारत में वर्ष 1997 के बाद से चलाया जा रहा है।	Health	3	3	4	3	3
67	For them all these children have come.	उनके लिए इन सभी बच्चों के लिए आए हैं।	Miscellaneous	2	2	3	1	2
68	<i>Pink</i> released on Friday, 16 September 2016 in thousands of screens across the country.	गुलाबी 16 सितंबर 2016 और अधिक पढ़ें देश भर में स्क्रीन के हजारों की संख्या में शुक्रवार को जारी की।	Film	3	3	2	1	3

69	He is incapable of speaking foreign languages.	उन्होंने कहा कि विदेशी भाषाओं के बोलने के काबिल नहीं है।	language and teaching	3	2	2	1	2
70	Every newspaper reviewer has given it a rating of over 4 stars.	हर अखबार समीक्षक इसे 4 से अधिक सितारों की रेटिंग दी है।	News	3	2	2	3	2
71	Gowariker is the country's newest "historian and archaeologist".	गोवारीकर देश के सबसे नए "इतिहासकार और पुरातत्ववेत्ता" है।	History	3	2	3	4	2
72	I can not make this kind of food.	मैं भोजन के इस तरह नहीं कर सकता।	Food	2	4	1	1	2
73	Many people regularly used to come and go here.	कई लोगों को नियमित रूप से आते हैं और यहां से जाने के लिए इस्तेमाल किया।	General	2	3	1	1	2
74	Adopt nutritious food and healthy lifestyle.	पौष्टिक भोजन और स्वस्थ जीवन शैली को अपनाने।	Food	2	3	2	2	2
75	Among the districts this rate is the least in Paudi while the most in Haridwar.	जिलों में इस दर पौड़ी में कम से कम है, जबकि हरिद्वार में सबसे अधिक है।	Tourism	2	3	2	3	3
76	This film seems to be a clumsy attempt at discrediting the widely held perception that Mohenjo-Daro was destroyed in an Aryan invasion.	इस फिल्म में व्यापक रूप से आयोजित की धारणा है कि मोहन जोदड़ो एक आर्य आक्रमण में नष्ट हो गया था बदनाम करने पर एक अनाड़ी प्रयास हो रहा है।	Film	1	0	2	2	2

77	In 1999, it was claimed that there were horses in Mohenjo-Daro and that later Vedic Sanskrit was spoken there.	1999 में, यह दावा किया गया था वहाँ मोहन जोदड़ो में और है कि बाद में वैदिक संस्कृत वहाँ कहा गया था घोड़ों थे।	Art and culture	1	3	2	1	2
78	Perhaps he regularly gets his money on Friday.	शायद वह नियमित रूप से शुक्रवार को उसका पैसा हो जाता है।	Miscellaneous	0	2	2	0	2
79	Infected persons live a normal life for 10 to 15 years also.	संक्रमित व्यक्तियों को भी 10 से 15 साल के लिए एक सामान्य जीवन जीते हैं।	Health	1	1	2	3	2
80	It is the desire of son that will affect the future reproductive level of the state.	यह बेटे की इच्छा है कि राज्य के भविष्य के प्रजनन के स्तर को प्रभावित करेगा।	Health	2	3	1	3	3
81	Generally, birth rate is lower the hilly areas than in the plains areas.	आम तौर पर, जन्म दर पहाड़ी क्षेत्रों को कम मैदानों क्षेत्रों की तुलना में है।	Health	1	3	2	3	3
82	The main reason of blindness is cataract which is about 55 percent of all blindness.	अंधापन का मुख्य कारण मोतियाबिंद जो सभी अंधापन के बारे में 55 प्रतिशत है।	Health	3	2	2	4	3
83	He falls in love with Chani (Puja Hegde), a dancer in Mohenjo-Daro.	उन्होंने Chani (पूजा हेगड़े), मोहन जोदड़ो में एक नर्तकी के साथ प्यार में गिर जाता है।	Film	3	3	2	1	1
84	This Vishnu Temple of Narayanpal is a protected monument by the Archaeological Department .	Narayanpal की यह विष्णु मंदिर पुरातत्व विभाग द्वारा संरक्षित स्मारक है।	Tourism	3	4	3	3	2

85	The archaeologists, anthropologists, linguists and historians have yet to arrive at any definite conclusion.	पुरातत्वविदों, मानवविज्ञानी, भाषाविद् और इतिहासकारों अभी तक कोई निश्चित निष्कर्ष पर पहुंचने के लिए लोगों की है।	History	2	1	1	2	2
86	It took the man two weeks to go to India.	यह आदमी दो सप्ताह के भारत जाने के लिए ले लिया।	Tourism	1	3	1	1	2
87	If I had studied a lot, I wouldn't have failed the exam.	अगर मैं एक बहुत अध्ययन किया था, मैं परीक्षा में असफल नहीं होता।	Education	3	2	2	3	2
88	The man wanted to eat, but all the restaurants were closed.	आदमी को खाने के लिए चाहता था, लेकिन सभी रेस्तरां बंद कर दिया गया।	General	2	1	2	2	2
89	Danteshwari Devi is the symbol of the residents of Bastar.	दंतेश्वरी देवी बस्तर के निवासियों का प्रतीक है।	Tourism	3	3	4	4	3
90	These statues are also important from the archaeological point of view.	इन प्रतिमाओं को भी देखने की पुरातात्विक दृष्टि से महत्वपूर्ण हैं।	Art and culture	3	3	2	3	2
91	And that is why, the Aryans – who are influential in the country's present power structure.	और यही कारण है, आर्यों - जो देश के वर्तमान सत्ता संरचना में प्रभावशाली रहे हैं।	History	3	2	2	4	3
92	Whether the Aryans were aliens or indigenous inhabitants has yet to be ascertained.	चाहे आर्यों एलियंस या स्वदेशी निवासी थे अभी तक पता नहीं चल पाया गया है।	History	3	4	2	3	2

93	The kind bio-diversity seen in Bastar is not found anywhere else.	तरह जैव-विविधता बस्तर में देखा कहीं और नहीं पाया जाता है।	Environment	3	3	1	3	2
94	Comprehensive construction work and repairing/renovation of health units	व्यापक निर्माण कार्य और स्वास्थ्य इकाइयों की मरम्मत / नवीकरण	Health	3	3	3	4	3
95	The organizations of Tribals and indigenous inhabitants started asking questions.	आदिवासी और स्वदेशी निवासियों के संगठनों सवाल पूछ शुरू कर दिया।	Society	1	3	2	3	3
96	Further, he scored the lowest marks among all the qualified candidates in General Knowledge.	इसके अलावा, वह सामान्य ज्ञान में सभी योग्य उम्मीदवारों के बीच सबसे कम अंक हासिल।	Education	2	3	2	3	3
97	Das was the lowest-ranked candidate eligible for appointment to the IAS.	दास सबसे कम रैंकिंग वाले उम्मीदवार आईएएस पर नियुक्ति का पात्र था।	Education	2	2	1	3	2
98	But the above mentioned programme is being run in the state with the help of women health workers.	लेकिन जैसा कि ऊपर उल्लेख कार्यक्रम महिलाओं के स्वास्थ्य कार्यकर्ताओं की मदद से राज्य में चलाया जा रहा है।	Health	3	2	2	3	3
99	The scholars the world over rejected this theory a decade ago.	विद्वानों दुनिया में एक दशक पहले इस सिद्धांत को खारिज कर दिया पर।	Philosophy	2	3	1	3	2
100	We have come to know about it only through excavations.	हम केवल खुदाई के माध्यम से इसके बारे में पता चल गया है।	ARCHEOLOGY	3	2	2	2	3

## माइक्रोसॉफ्ट अनुवादक आउटपुट स्कोर

S.No.	English Sentences	Microsoft Translator	Domain	User 1 Score	User 2 Score	User 3 Score	User 4 Score	User 5 Score
1	Ruins of archaeological importance are found in Chandragiri and Chingitarai.	पुरातात्विक महत्व के अवशेष चंद्रगिरी और Chingitarai में पाए जाते हैं।	Tourism	3	3	3	3	3
2	Globalization-induced modernism has spread to all parts of the country.	आधुनिकतावाद भूमंडलीकरण प्रेरित करने के लिए देश के सभी भागों में फैल गया है।	Education	3	3	2	3	2
3	Instead of the boy, it is the girl who drives tractors and rides large motorcycles.	लड़के के बजाय, यह लड़की है जो ट्रैक्टर ड्राइव और बड़ी मोटरसाइकिल की सवारी है।	Film	2	3	2	3	2
4	She wants control over her life. She decides whom to love and marry.	वह उसके जीवन पर नियंत्रण करना चाहता है। वह जिनके प्यार और शादी का फैसला किया।	Film	2	2	2	1	2
5	The government reduced all the prices.	सरकार सभी मूल्य कम है।	Economy	1	2	1	1	1
6	The natural place for sandal trees in Kerala is Marayur.	केरल में चप्पल पेड़ के लिए प्राकृतिक जगह Marayur है।	Tourism	0	2	1	0	1
7	Manjule underlines this horrendous truth in all its ugliness.	Manjule यह भयावह सच में अपने सभी कुरूपता रेखांकित करता है।	Film	3	1	2	2	2

8	Everyone waited for the customers.	हर कोई ग्राहकों के लिए इंतजार कर रहे थे।	Economy	2	2	2	4	4
9	Munnar is a hilly area situated at a height of 1600 meters above the sea level.	मुन्नार 1600 मीटर की दूरी पर समुद्र स्तर से ऊपर की ऊंचाई पर स्थित एक पहाड़ी क्षेत्र है।	Tourism	3	3	2	3	3
10	The film is a warning to the gutsy couples opting for inter-caste marriages that they are not safe even in the metropolises.	फिल्म gutsy युगल अंतर जाति विवाह कि वे महानगर में भी सुरक्षित नहीं हैं के लिए ऑप्ट करने के लिए एक चेतावनी है।	Film	2	2	2	2	2
11	Indian modernity is never very far from hatred and enmity.	भारतीय आधुनिकता कभी बहुत घृणा और शत्रुता से दूर है।	literature	0	1	2	0	2
12	Therefore these caves also have religious importance.	इसलिए ये गुफाएं भी धार्मिक महत्व है।	Tourism	2	1	2	1	2
13	Munnar was one of the important summer residences of India.	मुन्नार भारत के महत्वपूर्ण गर्मी निवासों में से एक था।	Tourism	3	3	3	3	4
14	Kailash cave is at a height of 50 meter and is 2500 meters long.	कैलाश गुफा 50 मीटर की ऊंचाई पर है और 2500 मीटर लंबी है।	Tourism	4	2	3	4	4
15	It was the owners of the British gardens who got the tea garden planted in Munnar.	यह ब्रिटिश बागानों के मालिकों जो चाय उद्यान मुन्नार में लगाया गया था।	Tourism	3	3	2	3	2
16	Now Shah Jahan and MumtaZ mahal graves are in the Taj mahal.	अब शाहजहां और मुमताज महल कब्र ताजमहल में कर रहे हैं।	Tourism	2	2	2	2	2
17	This foreigner has also come to see Agra.	इस विदेशी भी आगरा को देखने के लिए आ गया है।	Tourism	3	2	2	2	4

18	In comparison with the traditional surgical operation this method has many advantages.	पारंपरिक शल्य चिकित्सा ऑपरेशन के साथ तुलना में इस विधि के कई फायदे हैं।	Health	2	1	2	3	4
19	It is here that the world famous flower called Neelkurinji blooms.	यह यहाँ कि विश्व प्रसिद्ध फूल खिलता है Neelkurinji कहा जाता है।	Tourism	3	2	2	3	3
20	One can also get to see the glimpse of Tamilnadu from the top of these hills.	एक भी इन पहाड़ियों के ऊपर से तमिलनाडु की झलक देखने के लिए प्राप्त कर सकते हैं।	Tourism	2	3	2	3	2
21	Perhaps the prime minister will go to Egypt.	शायद प्रधानमंत्री ने मिस्र के लिए जाना होगा।	IR	2	2	2	3	2
22	The first patient of AIDS in India was found in 1986.	1986 में भारत में एड्स का पहला मरीज पाया गया था।	Health	3	4	3	4	4
23	These flowers are found in about 60 shades of blue colour.	ये फूल नीले रंग के बारे में 60 रंगों में पाए जाते हैं।	Miscellaneous	3	1	2	3	2
24	When it blooms it appears as if whole nature has bathed in blue colour.	जब यह खिलता है यह प्रतीत होता है मानो पूरी प्रकृति नीले रंग में नहाया है।	Miscellaneous	3	3	3	3	4
25	The teacher told me the story speaking slowly.	शिक्षक मुझे धीरे धीरे बोल कहानी कहा था।	Education	3	2	2	3	2
26	His mind was relieved when he saw his brother.	उनके मन में राहत मिली जब उन्होंने अपने भाई को देखा था।	General	2	1	2	3	3
27	Will you go to india via Japan or via Europe?	आप भारत को जापान के माध्यम से या यूरोप के माध्यम से जाना होगा?	IR	2	3	2	3	2

28	While fighting, the girl threw out all her friends.	लड़ते हुए, लड़की अपने सभी दोस्तों के बाहर फेंक दिया।	society	0	3	2	2	2
29	I saw the cobbler repair the shoes.	मैं मोची जूते की मरम्मत को देखा।	General	2	2	2	3	2
30	The five year plan in progress these days.	पंचवर्षीय योजना प्रगति पर है इन दिनों।	Economy	3	4	3	3	4
31	They call it the pink city. Because many of its buildings are made of pink stone.	वे इसे गुलाबी शहर कहते हैं। क्यों कि कई अपने भवनों के गुलाबी पत्थर के बने हैं।	art and culture	3	2	3	3	3
32	He uses the tools of the mainstream cinema but tells his tale.	वह मुख्यधारा के सिनेमा के उपकरण का उपयोग करता है, लेकिन उसकी कहानी बताता है।	Film	2	3	3	2	2
33	From here the old bazaar begins.	यहाँ से पुरानी बाजार शुरू होता है।	Tourism	3	3	4	3	4
34	These people are selling cloth on the side of the street.	इन लोगों को सड़क के किनारे पर कपड़ा बेच रहे हैं।	General	1	1	2	2	3
35	They come to sell their vegetables every day.	वे हर दिन अपनी सब्जियां बेचने आते हैं।	General	3	3	4	2	4
36	Many foreigners come to Jaipur.	कई विदेशियों के लिए जयपुर आते हैं।	Tourism	2	2	3	3	2
37	That fat boy will eat up all the rasagullas.	कि फैट लड़के को सभी rasagullas खा जाएगा।	Food	1	4	1	0	2

38	The exhibition of different levels of the production of tea leaf is very attractive.	चाय पत्ती के उत्पादन के विभिन्न स्तर की प्रदर्शनी बहुत ही आकर्षक है।	Food	3	3	4	2	4
39	The blisters of the mouth does not cure and medicines do not affect.	मुंह के छाले का इलाज नहीं और दवाओं को प्रभावित नहीं करते।	Health	2	1	2	3	2
40	The school is the focal point of almost all his films.	स्कूल उनकी लगभग सभी फिल्मों का केन्द्र बिन्दु है।	Film	3	3	4	3	4
41	The students wash clothes downstairs in the kitchen.	छात्र नीचे रसोई घर में कपड़े धो लो।	Education	2	3	2	2	3
42	The hard wind is tearing all the flags.	मुश्किल हवा सभी झंडे फाड़ है।	Geography	1	2	2	2	2
43	Where the world bares its secrets.	जहां दुनिया उसके रहस्य bares.	Miscellaneous	1	3	1	0	1
44	The young lead characters of both his films write poetry.	दोनों अपनी फिल्मों के युवा नेतृत्व वर्ण कविता लिखना।	Film	2	2	1	1	1
45	The lady helped the poor man a lot.	महिला गरीब आदमी एक बहुत मदद की।	society	2	1	1	1	2
46	Indians give their guests tea to drink.	भारतीय अपने मेहमानों चाय पीने के लिए दे।	Food	1	3	2	1	2
47	The customer will have the fruit seller give the girls oranges to eat.	ग्राहक फल विक्रेता लड़कियों संतरे खाने के लिए देना होगा।	Food	2	2	1	2	1

48	The film takes you back to your childhood and adolescence.	फिल्म आप वापस अपने बचपन और किशोरावस्था के लिए लेता है।	Film	2	4	1	1	1
49	It touches your heart at its most tender part.	यह अपने सबसे कोमल भाग पर अपना दिल छू लेती है।	General	2	3	2	2	3
50	Fatehpur Sikri is 25 miles from Agra.	फतेहपुर सीकरी आगरा से 25 मील की दूरी पर है।	Tourism	3	2	4	2	4
51	To present the narrative of the worldly wisdom, some amount of detailing is necessary.	सांसारिक ज्ञान की कथा प्रस्तुत करने के लिए, का विवरण की कुछ मात्रा आवश्यक है।	Philosophy	2	1	3	2	2
52	The guide is telling the foreigner about Akbar and the buildings in Sikri.	गाइड foregner अकबर और सीकरी में भवनों के बारे में बता रहा है।	Tourism	1	2	1	3	3
53	I am about to speak to the teacher.	मैं शिक्षक के लिए के बारे में बात करने के लिए कर रहा हूँ।	Education	2	3	1	4	2
54	There is a beautiful park also just near the Church.	वहाँ एक सुंदर पार्क है अभी भी चर्च के पास।	Tourism	3	2	2	2	4
55	The combined vaccine is made for children and adults.	संयुक्त टीके बच्चों और वयस्कों के लिए बनाया है।	Health	3	1	4	2	4
56	Sita will be at home at 8 o clock.	सीता घर पर 8 ओ घड़ी में हो जाएगा।	society	0	3	2	4	2
57	There is a shortage of water in the desert.	रेगिस्तान में पानी की कमी है।	Tourism	4	4	4	4	4

58	The washermen who work in the Ashoka Hotel get a lot of money.	अशोक होटल में काम करने वाले washermen बहुत सारा पैसा मिलता है।	social issues	3	3	3	3	3
59	Lucknow is the capital of Uttar Pradesh.	लखनऊ उत्तर प्रदेश की राजधानी है।	Tourism	3	2	4	3	4
60	These temples are protected monuments by the Archaeology Department.	इन मंदिरों में पुरातत्व विभाग द्वारा संरक्षित स्मारक हैं।	art and culture	3	4	4	3	2
61	The family welfare programme was run at the national level in the year of 1952.	परिवार कल्याण कार्यक्रम 1952 के वर्ष में राष्ट्रीय स्तर पर चल रही थी।	Health	3	3	3	3	4
62	Therefore the concept of Reproductive and Child Health Programme was born.	इसलिए प्रजनन और बाल स्वास्थ्य कार्यक्रम की अवधारणा का जन्म हुआ।	Health	2	2	4	2	4
63	Do you know anyone among these people?	आप इन लोगों के बीच किसी को भी पता है?	Miscellaneous	3	2	2	3	2
64	All the people stand up together and pray.	सभी लोग एक साथ खड़े हो जाओ और प्रार्थना करता हूँ।	Religion	2	1	3	3	2
65	Archi is the daughter of a landlord and politician as well as she is the sister of a muscleman brother.	Archi एक मकान मालिक की बेटी है और राजनीतिज्ञ के रूप में अच्छी तरह के रूप में वह एक muscleman भाई की बहन है।	Community	2	3	1	3	2
66	Reproductive and Child Health Programme is being run since the year of 1997 in India.	प्रजनन और बाल स्वास्थ्य कार्यक्रम चलाया जा रहा है वर्ष 1997 के बाद से भारत में।	Health	3	3	3	3	4

67	For them all these children have come.	उनके लिए ये सभी बच्चे आ गए।	Miscellaneous	2	3	4	2	4
68	<i>Pink</i> released on Friday, 16 September 2016 in thousands of screens across the country.	गुलाबी स्क्रीन के हजारों में शुक्रवार, 16 सितम्बर 2016 पर देश भर में जारी किया गया।	Film	2	3	2	2	2
69	He is incapable of speaking foreign languages.	उन्होंने विदेशी भाषाओं के बोलने में असमर्थ है।	language and teaching	3	3	1	2	3
70	Every newspaper reviewer has given it a rating of over 4 stars.	हर अखबार समीक्षक यह एक से अधिक 4 सितारों की रेटिंग दी है।	News	4	2	1	2	2
71	Gowariker is the country's newest "historian and archaeologist".	गोवारीकर देश के नवीनतम "इतिहासकार और पुरातत्वविद्" हैं।	History	2	1	2	2	4
72	I cannot make this kind of food.	मैं इस तरह का भोजन नहीं कर सकते हैं।	Food	2	4	3	3	2
73	Many people regularly used to come and go here.	कई लोग नियमित रूप से आते हैं और यहाँ जाने के लिए इस्तेमाल किया।	General	3	3	2	3	2
74	Adopt nutritious food and healthy lifestyle.	पौष्टिक भोजन और स्वस्थ जीवन शैली को अपनाने।	Food	3	3	3	4	3
75	Among the districts this rate is the least in Paudi while the most in Haridwar.	जिले के बीच यह दर जबकि हरिद्वार में सबसे अधिक पौड़ी में कम से कम है।	Tourism	2	2	2	2	2
76	This film seems to be a clumsy attempt at discrediting the widely held perception that Mohenjo-Daro was	इस फिल्म पर व्यापक रूप से आयोजित धारणा है कि मोहन जोदड़ो एक आर्यन आक्रमण में नष्ट	Film	1	2	1	3	2

	destroyed in an Aryan invasion.	हो गया discrediting एक अनाड़ी का प्रयास किया जा रहा है।						
77	In 1999, it was claimed that there were horses in Mohenjo-Daro and that later Vedic Sanskrit was spoken there.	1999 में, यह दावा किया था कि मोहन जोदड़ो में घोड़े थे और बाद में वैदिक संस्कृत वहाँबाट की थी।	art and culture	2	3	2	3	2
78	Perhaps he regularly gets his money on Friday.	शायद वह नियमित रूप से शुक्रवार को अपने पैसा मिलता है।	Miscellaneous	3	2	3	3	2
79	Infected persons live a normal life for 10 to 15 years also.	संक्रमित व्यक्ति सामान्य जीवन भी 10 से 15 साल के लिए रहते हैं।	Health	3	3	2	2	4
80	It is the desire of son that will affect the future reproductive level of the state.	यह राज्य का भविष्य प्रजनन स्तर को प्रभावित करेगा बेटे की इच्छा है।	Health	2	2	3	3	2
81	Generally, birth rate is lower the hilly areas than in the plains areas.	आम तौर पर, जन्म दर कम है से पहाड़ी क्षेत्रों को मैदानी क्षेत्रों में।	Health	2	1	3	2	2
82	The main reason of blindness is cataract which is about 55 percent of all blindness.	सभी अंधापन के बारे में 55 प्रतिशत है जो मोतियाबिंद दृष्टिहीनता का मुख्य कारण है।	Health	2	3	3	3	2
83	He falls in love with Chani (Puja Hegde), a dancer in Mohenjo-Daro.	वह Chani (पूजा हेगड़े), मोहन जोदड़ो में एक नर्तकी के साथ प्यार में गिर जाता है।	Film	2	3	2	2	3
84	This Vishnu Temple of Narayanpal is a protected monument by the Archaeological Department.	यह विष्णु मंदिर Narayanpal के पुरातत्व विभाग द्वारा संरक्षित स्मारक है।	Tourism	3	3	4	2	2

85	The archaeologists, anthropologists, linguists and historians have yet to arrive at any definite conclusion.	पुरातत्वविदों, मानवविज्ञानी, भाषाविद् और इतिहासकार अभी तक किसी भी निश्चित निष्कर्ष पर पहुंचे हैं।	History	2	2	3	2	3
86	It took the man two weeks to go to India.	यह आदमी दो सप्ताह के लिए भारत जाने के लिए ले लिया।	Tourism	2	3	1	2	2
87	If I had studied a lot, I wouldn't have failed the exam.	मैं एक बहुत का अध्ययन किया था, तो मैं परीक्षा में नाकाम रहे हैं नहीं होता।	Education	1	2	1	2	1
88	The man wanted to eat, but all the restaurants were closed.	खाने के लिए आदमी चाहता था, लेकिन सभी रेस्तरां बंद किए गए थे।	General	2	1	1	2	2
89	Danteshwari Devi is the symbol of the residents of Bastar.	देवी दंतेश्वरी बस्तर के निवासियों का प्रतीक है।	Tourism	2	3	2	2	4
90	These statues are also important from the archaeological point of view.	ये मूर्तियां भी पुरातात्विक दृष्टि से महत्वपूर्ण हैं।	art and culture	3	2	3	3	4
91	And that is why, the Aryans - who are influential in the country's present power structure.	और यही क्यों, देना आर्यों के- जो देश के वर्तमान सत्ता संरचना में प्रभावशाली रहे हैं।	History	2	1	1	2	1
92	Whether the Aryans were aliens or indigenous inhabitants has yet to be ascertained.	थे कि क्या देना आर्यों विदेशी या स्वदेशी निवासियों अभी तक पता लगाया जा करने के लिए है।	History	1	4	1	2	2
93	The kind bio-diversity seen in Bastar is not found anywhere else.	तरह जैव-बस्तर में देखा विविधता कहीं भी पाया नहीं है।	Environment	2	3	1	2	2
94	Comprehensive construction work and repairing/renovatio	व्यापक निर्माण कार्य और स्वास्थ्य इकाइयों की मरम्मत/नवीकरण	Health	2	2	4	2	4

	n of health units							
95	The organizations of Tribals and indigenous inhabitants started asking questions.	आदिवासियों और स्थानीय निवासियों के संगठन सवाल पूछने शुरू कर दिया।	society	2	1	3	3	2
96	Further, he scored the lowest marks among all the qualified candidates in General Knowledge.	इसके अलावा, वह सामान्य ज्ञान में सभी योग्य उम्मीदवारों के बीच सबसे कम अंक रन बनाए।	Education	2	3	2	3	3
97	Das was the lowest-ranked candidate eligible for appointment to the IAS.	दास ने भारतीय प्रशासनिक सेवा के लिए नियुक्ति के लिए पात्र सबसे कम स्थान पर उम्मीदवार थे।	Education	3	4	2	3	2
98	But the above mentioned programme is being run in the state with the help of women health workers.	लेकिन महिला स्वास्थ्य कार्यकर्ताओं की मदद से राज्य में उपरोक्त कार्यक्रम चलाया जा रहा है।	Health	3	3	3	3	4
99	The scholars the world over rejected this theory a decade ago.	विद्वान दुनिया भर में एक दशक पहले इस सिद्धांत को अस्वीकार कर दिया।	Philosophy	1	2	3	3	2
100	We have come to know about it only through excavations.	हम केवल के माध्यम से खुदाई के बारे में पता करने के लिए आए हैं।	ARCHEOLOGY	2	2	2	1	3

## मंत्र-राजभाषा आउटपुट स्कोर

S.No.	English Sentences	MANTRA-Rajbhasha	Domain	User 1 Score	User 2 Score	User 3 Score	User 4 Score	User 5 Score
1	Ruins of archaeological importance are found in Chandragiri and Chingitarai.	खंडहर पुरातात्विक महत्व की चंद्रगिरि और चिंगिटरे प्रबन्ध किया जाता है	Tourism	1	2	2	2	2
2	Globalization-induced modernism has spread to all parts of the country.	आधुनिकतावाद भूमंडलीकरण प्रवृत्त देश की सभी भागों के लिए फैल जाया है	Education	2	2	2	1	2
3	Instead of the boy, it is the girl who drives tractors and rides large motorcycles.	लड़का, यह के बजाय लड़की जो ट्रैक्टर और सैर बड़ा मोटर साइकिल बाध्य करता है	Film	1	1	1	1	1
4	She wants control over her life. She decides whom to love and marry.	वह उनका जीवन से ज्यादा नियंत्रण को चाहता है, वह जिसको प्यार करने के लिए सुनिश्चित करता है	Film	2	1	1	1	1
5	The government reduced all the prices.	सरकार ने सभी कीमतों को कम हुआ था	Economy	3	2	2	2	2
6	The natural place for sandal trees in Kerala is Marayur.	चप्पल पेड़ों के लिए प्राकृतिक जगह केरल में मारयुर है	Tourism	2	3	1	3	1
7	Manjule underlines this horrendous truth in all its ugliness.	अमांजुल सभी अपना कुरूपता में यह भयानक सच्चाई को नीचे रेखा खींचता है	Film	2	2	1	2	1

8	Everyone waited for the customers.	सभी ने ग्राहकों के लिए प्रतीक्षा किया	Economy	3	2	2	3	2
9	Munnar is a hilly area situated at a height of 1600 meters above the sea level.	मुंनर समुद्र तल ऊपर 1600 मीटर की ऊंचाई में स्थित पहाड़ी क्षेत्र है	Tourism	3	2	3	3	3
10	The film is a warning to the gutsy couples opting for inter-caste marriages that they are not safe even in the metropolises.	फिल्म इंटर जाति शादीयों के लिए विकल्प प्रस्तुत करके गुथी दम्पतियों के लिए चेतावनी है कि वे यहाँ तक कि राजधानीयों में सुरक्षित हैं नहीं	Film	3	1	1	1	1
11	Indian modernity is never very far from hatred and enmity.	भारतीय आधुनिकता कभी नहीं अत्यधिक दूरी पर नफरत और वैर से है	literature	2	2	1	2	1
12	Therefore these caves also have religious importance.	इसलिए ये गुफाएं धार्मिक महत्व भी	Tourism	3	1	1	1	1
13	Munnar was one of the important summer residences of India.	मुंनर भारत की एक महत्वपूर्ण ग्रीष्म ऋतु निवास स्थानों था	Tourism	1	2	2	2	2
14	Kailash cave is at a height of 50 meter and is 2500 meters long.	केलाश गुफा 50 मीटर की ऊंचाई में है और 2500 मीटर लंबी है	Tourism	2	2	3	3	3
15	It was the owners of the British gardens who got the tea garden planted in Munnar.	यह ब्रिटेन का उपवन की मालिकों था जो चाय बागान मुंनर में रोप को प्राप्त किया	Tourism	1	2	1	2	1

16	Now Shah Jahan and MumtaZ mahal graves are in the Taj mahal.	अब शाह जहन और मुमताज माहल गंभीर ताज माहल में हैं	Tourism	1	2	1	1	1
17	This foreigner has also come to see Agra.	यह विदेशी आगरा देखने के लिए भी आया है	Tourism	2	3	2	3	2
18	In comparison with the traditional surgical operation this method has many advantages.	तुलना में पारम्परिक चिकित्सीय परिचालन यह विधि सहित कई लाभ	Health	1	1	1	1	1
19	It is here that the world famous flower called Neelkurinji blooms.	यह यहाँ कि नील्कुरिंजि फूलों कह विश्व मशहूर फूल है	Tourism	1	2	2	1	2
20	One can also get to see the glimpse of Tamilnadu from the top of these hills.	कोई भी पहाड़ियाँ इनमें से शीर्ष से तमिलनाडु की झलक देखने के लिए भी कर सकता है	Tourism	2	2	1	1	1
21	Perhaps the prime minister will go to Egypt.	सम्भवतः प्रधानमंत्री इजिप्ट के लिए जाएँगे	IR	3	2	4	3	4
22	The first patient of AIDS in India was found in 1986.	एड्स की पहला मरीज भारत में 1986 प्रबन्ध किया गया था	Health	2	2	3	1	3
23	These flowers are found in about 60 shades of blue colour.	ये फूलों को नीला रंग की 60 छह ाया के बारे में प्रबन्ध किया जाता है	Miscellaneous	2	1	2	2	2
24	When it blooms it appears as if whole nature has bathed in blue colour.	जब यह इसको फूलों करता है उपस्थित है जैसे सभी प्रकृति ने नीला वर्ण में बदल कर दिया है	Miscellaneous	1	1	1	1	1

25	The teacher told me the story speaking slowly.	शिक्षक धीरे धीरे सदस्य इंजी. वृत्तान्त बोल बताया	Educ ation	2	1	1	1	1
26	His mind was relieved when he saw his brother.	उनका ध्यान को कार्यमुक्त किया जा किया गया था जब वह उनका भाई को देखता है	Gene ral	2	1	2	2	2
27	Will you go to india via Japan or via Europe?	क्या आप जापान या यूरोप द्वारा द्वारा भारत के लिए जाएँगे	IR	1	2	2	1	2
28	While fighting, the girl threw out all her friends.	जबकि युद्ध, लड़की सभी उनका दोस्त को फेंक बाहर करते हैं	socie ty	1	2	1	1	1
29	I saw the cobbler repair the shoes.	मैं मोची देखता है जूते मरम्मत करते हैं	Gene ral	2	2	1	1	2
30	The five year plan in progress these days.	प्रगति में पंचवर्षीय योजना आजकल	Econ omy	1	2	2	1	1
31	They call it the pink city. Because many of its buildings are made of pink stone.	जिले यह दर के बीच हरिद्वार में व्यापक स्तर पर जबकि पॉडि में न्यूनतम है	art and cultu re	2	2	1	2	1
32	He uses the tools of the mainstream cinema but tells his tale.	वह मुख्यधारा चलचित्र की टूल्स को उपयोग करता है किंतु उनका कथा को बताता है	Film	2	2	1	2	2
33	From here the old bazaar begins.	यहाँ, पुरानी बज़र से प्रारंभ करता है	Tour ism	1	2	2	2	1

34	These people are selling cloth on the side of the street.	ये आम आदमी सड़क की तरफ पर वस्त्र विक्रय कर रहे हैं	General	2	4	1	2	2
35	They come to sell their vegetables every day.	वे उनके सब्जियों प्रत्येक दिन बेचने के लिए आते हैं	General	1	3	2	2	2
36	Many foreigners come to Jaipur.	कई विदेशी जयपुर के लिए आते हैं	Tourism	3	4	2	2	3
37	That fat boy will eat up all the rasagullas.	वह चर्बी लड़का सभी रसगुल्लास को खा यू.पी. करेंगे	Food	1	2	3	1	1
38	The exhibition of different levels of the production of tea leaf is very attractive.	चाय पत्ती की प्रस्तुति की विभिन्न स्तर की प्रदर्शनी अत्यधिक आकर्षक है	Food	1	3	4	2	4
39	The blisters of the mouth do not cure and medicines do not affect.	फफोलों मुँह की उपाय नहीं करता है और औषधियों को प्रभावित नहीं करता है	Health	1	2	2	2	2
40	The school is the focal point of almost all his films.	विद्यालय के विषय में उनका फिल्म की केन्द्र बिन्दु है	Film	2	2	2	2	1
41	The students wash clothes downstairs in the kitchen.	विद्यार्थियों कपड़े दोस्तेर्स रसोई में को धो लेते हैं	Education	2	1	2	2	1
42	The hard wind is tearing all the flags.	कठिन हवा सभी झंडे को खरोच रहा है	Geography	2	2	1	1	2

43	Where the world bares its secrets.	जहाँ विश्व अपना सीक्रेट को बंद करता है	Miscellaneous	2	3	1	2	2
44	The young lead characters of both his films write poetry.	दोनों उनका फिल्म की जवान अग्रणी वर्णों कविता को लिखता है	Film	1	2	2	2	3
45	The lady helped the poor man a lot.	महिला ने कमजोर पुरुष काफी को सहायता किया	society	2	1	3	2	3
46	Indians give their guests tea to drink.	भारतीयों पीने के लिए उनके अतिथियाँ चाय प्रदान करते हैं	Food	2	2	3	1	1
47	The customer will have the fruit seller give the girls oranges to eat.	ग्राहक फल विक्रेता वसीयत खाने के लिए लड़कियाँ संतराएँ प्रदान करते हैं	Food	1	2	1	1	2
48	The film takes you back to your childhood and adolescence.	फिल्म पुनः आपके बचपन और किशोर के लिए आप प्राप्त करता है	Film	2	2	2	2	2
49	It touches your heart at its most tender part.	यह अपना व्यापक स्तर पर निविदा भाग में आपके हृदय को स्पर्श करता है	General	1	1	2	1	4
50	Fatehpur Sikri is 25 miles from Agra.	फ़तेहपुर सीकरी आगरा से 25 मील है	Tourism	3	2	4	4	1
51	To present the narrative of the worldly wisdom, some amount of detailing is necessary.	सांसारिक बुद्धिमानों की कहानी उपस्थित करने के लिए, वर्णन करना की कुछ राशि आवश्यक है	Philosophy	2	1	1	1	2

52	The guide is telling the foreigner about Akbar and the buildings in Sikri.	गाइड के रूप में सीकरी में अकबर और भवनों लगभग फोरेग्नेर बता रहा है	Tourism	2	1	2	2	2
53	I am about to speak to the teacher.	मैं शिक्षक के लिए बोलने के लिए लगभग हूँ	Education	2	2	2	2	2
54	There is a beautiful park also just near the Church.	_ भी अब गिरजाघर के पास सुंदर पार्क है	Tourism	1	2	2	1	2
55	The combined vaccine is made for children and adults.	संयुक्त रूप टीका को बच्चे और युवक के लिए किया जाता है	Health	2	2	2	2	1
56	Sita will be at home at 8 o clock.	सीता 8 ओ घड़ी में अनौपचारिक पार्टी होगा	society	1	1	1	1	1
57	There is a shortage of water in the desert.	_ योग्यता में जल की कमी है	Tourism	1	1	1	1	2
58	The washermen who work in the Ashoka Hotel get a lot of money.	वाशरमेन जो अशोक होटल में काम करते हैं बहुत अधिक मुद्रा करते हैं	social issues	1	2	2	2	2
59	Lucknow is the capital of Uttar Pradesh.	लखनऊ उत्तर प्रदेश की पूंजी है	Tourism	2	3	2	3	2
60	These temples are protected monuments by the Archaeology Department.	ये मंदिरों को पुरातत्व विभाग द्वारा स्मारकों रक्षा किया जाता है	art and culture	2	2	2	2	3

61	The family welfare programme was run at the national level in the year of 1952.	परिवार कल्याण कार्यक्रम 1952 की वर्ष में राष्ट्रीय स्तर में संचालित किया गया था	Health	2	2	3	4	2
62	Therefore the concept of Reproductive and Child Health Programme was born.	इसलिए संकल्पना जननीय और चाइल्ड हैल्थ कार्यक्रम की वहन किया गया था	Health	1	2	2	2	1
63	Do you know anyone among these people?	आप करें ये आम आदमी के बीच कोई व्यक्ति जानता है	Miscellaneous	2	1	1	1	4
64	All the people stand up together and pray.	सभी आम आदमी एक साथ खड़े होते हैं और प्रार्थना करते हैं	Religion	1	3	4	4	2
65	Archi is the daughter of a landlord and politician as well as she is the sister of a muscleman brother.	अर्चि ज़मीनदार और राजनीतिज्ञ की पुत्री है और वह मुस्कलेमान भाई की बहन है	Community	2	3	2	3	2
66	Reproductive and Child Health Programme is being run since the year of 1997 in India.	जननीय और चाइल्ड हैल्थ कार्यक्रम को भारत में 1997 की वर्ष चूँकि संचालित किया जा रहा है	Health	3	3	2	3	1
67	For them all these children have come.	उन्हें सभी के लिए, बच्चे ने आया है	Miscellaneous	1	2	1	2	1
68	<i>Pink</i> released on Friday, 16 September 2016 in thousands of screens across the country.	शुक्रवार, 16 सितंबर 2016 पर पिंकरेलीसेड हजारों में परदे की देश भर में	Film	1	1	1	1	2

69	He is incapable of speaking foreign languages.	वह विदेशी भाषा बोलना की अक्षम है	language and teaching	2	2	2	2	1
70	Every newspaper reviewer has given it a rating of over 4 stars.	प्रत्येक समाचारपत्र समीक्षकार ने 4 स्टार की से ज़्यादा यह नियत को प्रदान कर दिया है	News	2	2	1	2	1
71	Gowariker is the country's newest "historian and archaeologist".	गोअरिकेर कोउंत्र्यास नयी आहिस्तोरिअन और अर्चेओलोगिस्टा है	History	1	1	1	1	2
72	I can not make this kind of food.	मैं खाद्य की यह किस्म बना नहीं सकता है	Food	2	4	2	1	2
73	Many people regularly used to come and go here.	बहुत से लोग नियमित रूप से प्रयोग यहाँ आने के लिए और जाते हैं	General	2	2	2	3	3
74	Adopt nutritious food and healthy lifestyle.	पौष्टिक भोजन और स्वस्थ जीवनशैली स्वीकार करें	Food	2	4	3	2	2
75	Among the districts this rate is the least in Paudi while the most in Haridwar.	जिले यह दर के बीच हरिद्वार में व्यापक स्तर पर जबकि पाँडि में न्यूनतम है	Tourism	1	2	2	4	2
76	This film seems to be a clumsy attempt at discrediting the widely held perception that Mohenjo-Daro was destroyed in an Aryan invasion.	यह फिल्म अविश्वास उत्पन्न व्यापक रूप में धारित अनुभव में बेढंगा प्रयत्न होने के लिए प्रतीत है कि मोहेंजो दरो अर्यन चढ़ाई में नष्ट किया गया था	Film	1	1	2	2	2

77	In 1999, it was claimed that there were horses in Mohenjo-Daro and that later Vedic Sanskrit was spoken there.	1999 में, इसको का दावा किया गया था कि वहाँ मोहेंजो दरो में घोड़ों थे और वह बाद में वेदिक संस्कृत को ये बोला गया था	art and culture	2	1	2	1	2
78	Perhaps he regularly gets his money on Friday.	सम्भवतः वह शुक्रवार पर उनका मुद्रा को नियमित रूप से प्राप्त करता है	Miscellaneous	3	2	2	1	2
79	Infected persons live a normal life for 10 to 15 years also.	संक्रमित व्यक्ति भी 15 वर्ष के लिए 10 के लिए सामान्य जीवन रहता है	Health	2	1	2	2	2
80	It is the desire of son that will affect the future reproductive level of the state.	यह पुत्र की इच्छा वह राज्य की भावी जननीय स्तर को प्रभावित करेंगे है	Health	2	1	2	1	2
81	Generally, birth rate is lower the hilly areas than in the plains areas.	आम तौर पर, जन्म दर मैदानों क्षेत्र से में निम्न पर्वतीय क्षेत्र है	Health	2	1	1	2	1
82	The main reason of blindness is cataract which is about 55 percent of all blindness.	मुख्य कारण अन्धापन की सभी अन्धापन की मोतियाबिन्द जो 55 प्रतिशत लगभग है	Health	3	1	1	1	1
83	He falls in love with Chani (Puja Hegde), a dancer in Mohenjo-Daro.	अह मोहेंजो दरो में चनि (पूजा हेगडे), नर्तक सहित प्रेम समाप्त है	Film	2	1	1	1	1
84	This Vishnu Temple of Narayanpal is a protected monument by the Archaeological Department.	नरयंपल की यह विष्णु पुरातत्व विभाग द्वारा रक्षा स्मारक है	Tourism	1	2	2	1	2

85	The archaeologists, anthropologists, linguists and historians have yet to arrive at any definite conclusion.	पुरातत्त्वज्ञों, मानव विज्ञानी, भाषा विज्ञानीयों और इतिहासकारों के पास किसी भी स्पष्ट अंत को अभी तक आयाने के लिए	Histo ry	2	1	1	2	1
86	It took the man two weeks to go to India.	यह भारत के लिए जाने के लिए पुरुष दो सप्ताह ग्रहण किया	Tour ism	2	1	2	1	2
87	If I had studied a lot, I wouldn't have failed the exam.	यदि मैं काफी, मैं वऔल्डनात अध्ययन कर दिया था परीक्षा अनुत्तीर्ण हुआ है	Educ ation	1	1	1	1	1
88	The man wanted to eat, but all the restaurants were closed.	पुरुष ने खाने के लिए चाह हुआ था किंतु सभी भोजनालयों को समाप्त किया गया था	Gene ral	1	2	2	1	2
89	Danteshwari Devi is the symbol of the residents of Bastar.	दंतेश्वरि देवी बस्टर की निवासियों की प्रतीक है	Tour ism	2	3	3	2	3
90	These statues are also important from the archaeological point of view.	ये मूर्तियों पुरातात्विक दृष्टिकोण से भी महत्वपूर्ण हैं	art and cultu re	3	3	3	3	3
91	And that is why, the Aryans - who are influential in the country's present power structure.	और वह है क्यों अर्यस आ जो कोउंत्र्यास उपस्थित शक्ति ढांचा में प्रभावशाली हैं	Histo ry	2	1	1	1	1
92	Whether the Aryans were aliens or indigenous inhabitants has yet to be ascertained.	क्या अर्यस अलीस थे या देशी निवासियों को अभी तक सुनिश्चित करना है	Histo ry	1	3	2	3	2

93	The kind bio-diversity seen in Bastar is not found anywhere else.	किस्म जैव विविधता बस्टर में देख कहीं भी के अतिरिक्त पाया गया नहीं या गया है	Environment	1	1	1	1	1
94	Comprehensive construction work and repairing/renovation of health units	स्वास्थ्य इकाइयों की व्यापक निर्माण काम और मरम्मत/नवीकरण	Health	2	3	1	3	1
95	The organizations of Tribals and indigenous inhabitants started asking questions.	जनजातियाँ और देशी निवासियों की संगठनों ने मांग करना को प्रश्न शुरू किया	society	1	2	2	2	2
96	Further, he scored the lowest marks among all the qualified candidates in General Knowledge.	इसके अतिरिक्त, वह सभी योग्य अभ्यर्थी में सामान्य ज्ञान के बीच सबसे कम अंक बनाया	Education	1	2	1	2	1
97	Das was the lowest-ranked candidate eligible for appointment to the IAS.	दास भा.प्र.से. के लिए नियुक्ति के लिए सबसे कम क्रमानुसार व्यवस्थित अभ्यर्थी पात्र था	Education	2	2	1	2	1
98	But the above mentioned programme is being run in the state with the help of women health workers.	किंतु उपरोक्त कार्यक्रम महिलाओं स्वास्थ्य कर्मचारी की सहायता से झगड़ा राज्य की जा रही है	Health	2	2	1	2	1
99	The scholars the world over rejected this theory a decade ago.	पहले विद्वानों विश्व से ज्यादा ने यह सिद्धांत दशक को अस्वीकार किया	Philosophy	2	1	1	1	1
100	We have come to know about it only through excavations.	हम ने उत्खननों के माध्यम से सिर्फ यह लगभग जानने के लिए आया है	ARCHEOLOGY	3	2	2	2	2

## आंग्ल-हिंदी आउटपुट स्कोर

S.No	English Sentences	AnglaMT	Domain	User 1 Score	User 2 Score	User 3 Score	User 4 Score	User 5 Score
1	Ruins of archaeological importance are found in Chandragiri and Chingitarai.	पुरातात्विक महत्व के खंडहर चान्द्रागिरि और चिन्गितारे में खोजे जाते हैं।	Tourism	2	3	2	3	3
2	Globalization-induced modernism has spread to all parts of the country.	सार्वभौमीकरण - उत्प्रेरित आधुनिकतावाद का देश की पूरे भूमिका को फैलाव है।	Education	2	2	2	3	2
3	Instead of the boy, it is the girl who drives tractors and rides large motorcycles.	लड़का के बजाय यह लड़की है जो ट्रैक्टरों को बाध्य करती है और बड़ी मोटरसाइकिलें पीछे लौटते हैं।	Film	1	4	1	2	1
4	She wants control over her life. She decides whom to love and marry.	वह अपने जीवन पर नियन्त्रण चाहती है। वह निश्चय करती है कि किससे पसंद किया जाय और विवाह करती है।	Film	3	3	3	2	3
5	The government reduced all the prices.	सरकार ने सभी कीमतें कम कीं।	Economy	3	2	3	2	3
6	The natural place for sandal trees in Kerala is Marayur.	चन्दन के पेड़ों के लिये केरल में प्राकृतिक स्थान मारायुर है।	Tourism	2	1	3	3	2
7	Manjule underlines this horrendous truth in all its ugliness.	मान्जुले पूरे इसकी बदसूरती में यह	Film	2	3	2	3	2

		भयंकर सच्चाई रेखांकित करता है।						
8	Everyone waited for the customers.	प्रत्येक व्यक्ति ने खरीदारों के लिये इंतजार किया।	Economy	3	2	3	4	3
9	Munnar is a hilly area situated at a height of 1600 meters above the sea level.	मुन्नर समुद्र तल के ऊपर १६०० मीटरों की एक ऊँचाई पर स्थित एक पर्वतीय क्षेत्र है।	Tourism	1	3	2	3	3
10	The film is a warning to the gutsy couples opting for inter-caste marriages that they are not safe even in the metropolises.	फिल्म साहसी जोड़ों के प्रति एक चेतावनी है जो अंतः - जाति मारिआगेस चुनाव कर रही है जो वे भी मेत्रोपोलिसेस में सुरक्षित नहीं हैं।	Film	2	2	3	3	1
11	Indian modernity is never very far from hatred and enmity.	भारतीय की आधुनिकता घृणा और शत्रुता से कभी बहुत दूर नहीं है।	literature	3	4	3	2	2
12	Therefore these caves also have religious importance.	अतः इन गुफाओं को भी धार्मिक महत्व है।	Tourism	3	3	3	3	1
13	Munnar was one of the important summer residences of India.	मुन्नर भारत की कोई एक महत्वपूर्ण गरमी का निवास था।	Tourism	2	2	3	3	1
14	Kailash cave is at a height of 50 meter and is 2500 meters long.	कैलाश की गुफा ५० मीटर की एक ऊँचाई पर है और क्या २५०० लम्बे समय के लिये मीटर हैं।	Tourism	2	3	2	2	1
15	It was the owners of the British gardens who got the tea garden planted in Munnar.	इसने अंग्रेज के बागों का मालिक थे जो मुन्नर में लगाया हुआ जाय चाय बागान हुआ।	Tourism	2	1	3	2	3

16	Now Shah Jahan and MumtaZ mahal graves are in the Taj mahal.	अब शाहजहाँ और मुम्तज़ महल समाधि हैं में टाज महल।	Tourism	3	2	2	2	1
17	This foreigner has also come to see Agra.	यह विदेशी अग्रा को मिलना भी आया है।	Tourism	2	3	2	2	2
18	In comparison with the traditional surgical operation this method has many advantages.	पारम्परिक शल्य आपरेशन से तुलना में इस पद्धति को अनेक लाभ हैं।	Health	2	2	3	3	1
19	It is here that the world famous flower called Neelkurinji blooms.	यह यहाँ है कि विश्व विख्यात फूल ने नील्कुरिन्जि को फूल बुलाए।	Tourism	0	2	2	2	2
20	One can also get to see the glimpse of Tamilnadu from the top of these hills.	लोग डिब्बा भी इकट्ठा कर देख झलक।	Tourism	0	1	0	1	2
21	Perhaps the prime minister will go to Egypt.	सम्भवतः प्रधान मंत्री मिस्र जायेंगे।	IR	3	3	2	2	2
22	The first patient of AIDS in India was found in 1986.	भारत में ऐआईडीएस का प्रथम रोगी १९८६ में खोजा गया था।	Health	4	3	3	2	3
23	These flowers are found in about 60 shades of blue colour.	इन फूल नीले रंग की लगभग ६० छाया में खोजे जाते हैं।	Miscellaneous	3	2	3	3	1
24	When it blooms it appears as if whole nature has bathed in blue colour.	कब/किस समय/जब/जब कि यह फूल यह प्रकट हो जैसे यदि सम्पूर्ण प्रकृति रख भीगा हुआ में नीला रंग।	Miscellaneous	2	2	2	2	2
25	The teacher told me the story speaking slowly.	शिक्षक कथित मुझे कहानी बोलनेवाला धीरे धीरे।	Education	1	1	2	1	0
26	His mind was relieved when he saw his brother.	उसका दिमाग मुक्त था जब उसने अपने	General	2	3	3	2	1

		भाई को देखा।						
27	Will you go to india via Japan or via Europe?	क्या आप यूरोप से होता हुआ या जापन से होता हुआ भारत जायेंगे ?	IR	3	2	3	2	2
28	While fighting, the girl threw out all her friends.	लड़ते समय , लड़की उसके सब मित्र भर्त्स्येकी।	society	0	1	3	2	1
29	I saw the cobbler repair the shoes.	मैंने मोची की मरम्मत जूते देखीं।	General	0	1	2	2	1
30	The five year plan in progress these days.	पाँच साल प्रगति में ये दिन योजना बनाता है।	Economy	1	3	3	1	1
31	They call it the pink city. Because many of its buildings are made of pink stone.	वे यह गुलाबी शहर बुलाते हैं। क्योंकि अधिकतर इसका भवन हैं समझ गुलाबी पत्थर।	art and culture	2	2	3	2	4
32	He uses the tools of the mainstream cinema but tells his tale.	वह मुख्यधारे सिनेमा के उपकरणों का उपयोग करता है लेकिन अपनी कहानी बताता है।	Film	2	1	2	2	3
33	From here the old bazaar begins.	यहाँ से बड़ा बाज़र प्रारम्भ होता है।	Tourism	2	2	4	3	2
34	These people are selling cloth on the side of the street.	इन लोग सड़क का पक्ष कपड़ा बेच रहे हैं।	General	2	4	3	2	3
35	They come to sell their vegetables every day.	वे अपनी तरकारियाँ प्रत्येक दिन बेचना आते हैं।	General	3	3	3	2	4
36	Many foreigners come to Jaipur.	अनेक विदेशी जयपुर आते हैं।	Tourism	4	2	4	2	2
37	That fat boy will eat up all the rasagullas.	वह मोटा लड़का सभी रासागुलास को खा जायेगा।	Food	4	2	3	3	3

38	The exhibition of different levels of the production of tea leaf is very attractive.	चाय की पत्ती के उत्पादन के भिन्न स्तरों की प्रदर्शनी बहुत आकर्षक है।	Food	2	1	3	2	0
39	The blisters of the mouth does not cure and medicines do not affect.	मुँह के फफोले अच्छा नहीं करते हैं और दवाएं प्रभावित नहीं करती हैं।	Health	2	2	3	3	3
40	The school is the focal point of almost all his films.	विद्यालय लगभग उसकी सब फिल्मों का आकर्षण केन्द्र है।	Film	3	2	3	2	2
41	The students wash clothes downstairs in the kitchen.	विद्यार्थी रसोईघर में कपड़े निचली मंजिल पर धोते हैं।	Educational	3	3	3	3	3
42	The hard wind is tearing all the flags.	कठोर हवा सभी ध्वजों को फाड़ रही है।	Geography	2	2	2	3	3
43	Where the world bares its secrets.	कहाँ दुनिया खोल इसका रहस्य।	Miscellaneous	2	1	2	2	1
44	The young lead characters of both his films write poetry.	उसकी दोनों फिल्मों के युवा का प्रथम स्थान के अक्षर कविता लिखते हैं।	Film	2	2	3	2	1
45	The lady helped the poor man a lot.	महइला ने बेचारे आदमी का बहुत सहायता की।	society	3	3	2	3	3
46	Indians give their guests tea to drink.	भारतीय पीना उनके अतिथि चाय देते हैं।	Food	2	2	3	2	3
47	The customer will have the fruit seller give the girls oranges to eat.	खरीदार वसीयत रख फल विक्रेता दे लड़की संतरा तक खा।	Food	2	1	1	1	1
48	The Film takes you back to your childhood and adolescence.	फिल्म आपकी बचपन और किशोरावस्था को आपको वापस लेती है।	Film	2	3	2	3	3
49	It touches your heart at its most tender part.	यह इसकी अत्यधिक नरम भूमिका पर	General	3	4	2	3	3

		आपका हृदय छूता है।						
50	Fatehpur Sikri is 25 miles from Agra.	फतेहपुर सीकरी आगरा से २५ मील है।	Tourism	4	3	3	3	2
51	To present the narrative of the worldly wisdom, some amount of detailing is necessary.	नियुक्त करने का सांसारिक बुद्धिमत्ता, कुछ धनराशि का आख्यान उपस्थित होने के लिये आवश्यक है।	Philosophy	1	2	2	2	1
52	The guide is telling the foreigner about Akbar and the buildings in Sikri.	गाइड अकबर के बारे में और भवन सिक्री में फोरेगैर बता रहा है।	Tourism	2	1	2	2	2
53	I am about to speak to the teacher.	मैं शिक्षक को बोलने के लिये लगभग हूँ।	Education	2	2	2	1	3
54	There is a beautiful park also just near the Church.	वहाँ गिरिजा घर के पास भी यथोचित एक सुंदर पार्क है।	Tourism	2	2	3	1	3
55	The combined vaccine is made for children and adults.	संयुक्त वैक्सीन है बच्चा और वयस्क।	Health	2	1	2	2	1
56	Sita will be at home at 8 o'clock.	सीता होगा पर घर पर ८ ओ घड़ी।	society	0	2	2	1	2
57	There is a shortage of water in the desert.	वहाँ मरुस्थल में पानी की एक कमी है।	Tourism	3	3	3	3	3
58	The washermen who work in the Ashoka Hotel get a lot of money.	धोबी जो आशोका होटल की कोशिश करते हैं बहुत सा धन मिलते हैं।	social issues	3	2	2	3	1
59	Lucknow is the capital of Uttar Pradesh.	लखनऊ उत्तर प्रदेश का बड़े अक्षर है।	Tourism	0	1	1	2	2
60	These temples are protected monuments by the Archaeology Department.	ये मंदिर अर्चैओलोगी देपाट्मेंन्त के द्वारा संरक्षित स्मारक हैं।	art and culture	3	3	3	2	3

61	The family welfare programme was run at the national level in the year of 1952.	परिवार कल्याण कार्यक्रम था आक्रमण कर राष्ट्र स्तरीय में वर्ष का/की/के १९५२।	Health	3	2	1	1	3
62	Therefore the concept of Reproductive and Child Health Programme was born.	अतः रेप्रोदुक्ताईवे और चिल्ड हील्थ प्रोग्रामे की अवधारणा जन्मजात था।	Health	1	2	2	2	3
63	Do you know anyone among these people?	क्या आप इन लोगों के बीच में किसी को जानते हैं ?	Miscellaneous	3	2	2	4	2
64	All the people stand up together and pray.	सभी लोग साथ साथ खड़ा होते हैं और प्रार्थना करते हैं।	Religion	4	1	2	3	3
65	Archi is the daughter of a landlord and politician as well as she is the sister of a muscleman brother.	अर्चि एक भू स्वामी और राजनीतिज्ञ की पुत्री है उतना अच्छा जितना वह एक मुस्कलेमन भाई की बहन है।	Community	2	2	3	3	2
66	Reproductive and Child Health Programme is being run since the year of 1997 in India.	प्रजननीय और चिल्ड हील्थ प्रोग्रामे है होता हुआ रन से वर्ष का/की/के १९९७ में भारत।	Health	1	2	1	1	2
67	For them all these children have come.	उनके लिये ये सब बच्चे आए हैं।	Miscellaneous	1	2	2	2	3
68	<i>Pink</i> released on Friday, 16 September 2016 in thousands of screens across the country.	शुक्रवार को पिन्क्रेलीसेद , देश के एक ओर से दूसरी ओर तक स्क्रीन में से एक हजार में १६ सितम्बर २०१६।	Film	1	3	2	1	2
69	He is incapable of speaking foreign languages.	वह विदेशी भाषा बोलने में असमर्थ है।	language and teaching	4	2	4	1	1

			g					
70	Every newspaper reviewer has given it a rating of over 4 stars.	प्रत्येक समाचार पत्र ने रेविण्वेर यह समाप्त ४ स्टारों का मूल्यांकन करने दिया है।	News	3	2	2	2	2
71	Gowariker is the country's newest "historian and archaeologist".	गोवारिकेर काउन्ट्रिस सबसे नया इतिहासकार और पुरातत्वविद है।	History	2	1	3	3	1
72	I cannot make this kind of food.	मैं खाना यह नहीं बना सकता हूँ।	Food	3	2	3	3	2
73	Many people regularly used to come and go here.	अनेक लोग नियमित रूप से प्रयोग किया हुआ तक आ और जा यहाँ।	General	1	1	2	2	2
74	Adopt nutritious food and healthy lifestyle.	आप पौष्टिक खाद्य और स्वस्थ जीवन शैली को गोद लीजिये।	Food	1	3	2	3	1
75	Among the districts this rate is the least in Paudi while the most in Haridwar.	बीच में जिला यह दर है न्यूनतम मात्रा में पाउदि समय अधिकतम में हरिद्वार।	Tourism	1	2	1	1	1
76	This film seems to be a clumsy attempt at discrediting the widely held perception that Mohenjo-Daro was destroyed in an Aryan invasion.	यह फिल्म व्यापक रूप से आयोजित अनुभव प्रतिष्ठा गिराने पर एक अशोभनीय प्रयास होना प्रतीत होती है कि मोहेन्जो - दारो एक आर्यन के हमले में नष्ट किया गया था।	Film	1	3	3	3	2
77	In 1999, it was claimed that there were horses in Mohenjo-Daro and	१९९९ में यह दावा किया गया था कि वहाँ मोहेन्जो - दारो में	art and culture	2	2	2	3	2

	that later Vedic Sanskrit was spoken there.	घोड़ा थे और वह बाद का वैदिक सांस्कृत वहाँ बोला गया था।						
78	Perhaps he regularly gets his money on Friday.	सम्भवतः वह नियमित रूप से प्राप्त कर उसका धन पर शुक्रवार।	Miscellaneous	2	3	2	1	1
79	Infected persons live a normal life for 10 to 15 years also.	संक्रमित व्यक्ति १० से १५ तक साल सामान्य जीवन भी जीवित रहते हैं।	Health	3	2	2	2	2
80	It is the desire of son that will affect the future reproductive level of the state.	यह पुत्र की इच्छा है जो भविष्य अवस्था का प्रजननीय स्तर प्रभावित करेगी।	Health	2	2	2	2	1
81	Generally, birth rate is lower the hilly areas than in the plains areas.	सामान्यतया जन्म दर है निम्न पर्वतीय क्षेत्र से में समतल क्षेत्र।	Health	2	2	3	1	1
82	The main reason of blindness is cataract which is about 55 percent of all blindness.	अंधापन का मुख्य कारण मोतिया बिंद है जो पूरे अंधापन से लगभग ५५ प्रतिशत हैं।	Health	2	3	1	2	2
83	He falls in love with Chani (Puja Hegde), a dancer in Mohenjo-Daro.	वह चानि ( पूजा हेगडे ) से प्रेम , मोहेन्जो - दारो में एक नर्तकी में गिरता है।	Film	1	2	3	1	2
84	This Vishnu Temple of Narayanpal is a protected monument by the Archaeological Department.	नारायान्पल का यह विष्णु मंदिर अर्चैओलोगिकल देपाट्मेंन्त के द्वारा एक संरक्षित स्मारक है।	Tourism	2	3	4	4	1
85	The archaeologists, anthropologists, linguists and	पुरातत्वविद मानव विज्ञानी भाषाविद और	History	2	4	2	2	3

	historians have yet to arrive at any definite conclusion.	इतिहासकार रख अभीतक तक पहुँच पर किसी भी निश्चित निष्कर्ष।						
86	It took the man two weeks to go to India.	इसने भारत को जाना आदमी का दो सप्ताह लिया।	Tourism	2	1	2	1	1
87	If I had studied a lot, I wouldn't have failed the exam.	यदि मैं रखा विवेचित बहुत मैं वाउल्दंत रख असफल परीक्षा।	Education	1	3	1	1	2
88	The man wanted to eat, but all the restaurants were closed.	आदमी खाना चाहा , लेकिन सभी रेस्तराँ बन्द हो गये थे।	General	2	2	2	3	2
89	Danteshwari Devi is the symbol of the residents of Bastar.	दान्तेश्वारि देवी बस्तर के निवासियों का प्रतीक है।	Tourism	3	1	2	4	3
90	These statues are also important from the archaeological point of view.	ये प्रतिमाएं पुरातात्विक दृष्टिकोण से भी महत्वपूर्ण हैं।	art and culture	3	1	2	2	3
91	And that is why, the Aryans – who are influential in the country's present power structure.	और वह है क्यों अर्यान्स कौन/जिसने/किसने/जो हैं प्रभावशाली मैं काउन्सिल उपहार शक्ति संरचना।	History	2	3	1	1	3
92	Whether the Aryans were aliens or indigenous inhabitants has yet to be ascertained.	कि अर्यान्स थे विदेशी या स्वदेशीय निवासी रख अभीतक तक हो सिद्ध।	History	1	2	1	1	2
93	The kind bio-diversity seen in Bastar is not found anywhere else.	दयालु जैव - विभिन्नता जो बस्तर देखा जाता है और किसी भी जगह नहीं खोजा जाता है।	Environment	2	1	2	2	1

94	Comprehensive construction work and repairing/renovation of health units	व्यापक निर्माण कार्य और स्वास्थ्य की इकाइयों का मरम्मत होनी/नवीनीकरण।	Health	3	1	2	2	2
95	The organizations of Tribals and indigenous inhabitants started asking questions.	त्रिबाल्स और स्वदेशीय निवासियों के आयोजनों ने पूछनेवाले प्रश्न शुरू किए।	society	2	2	2	2	3
96	Further, he scored the lowest marks among all the qualified candidates in General Knowledge.	आगे बढ़ाइये , उसने जनरल कौल्लेद्गे में सभी योग्य अभ्यर्थियों के बीच में सबसे निम्न मार्क अंक दिया।	Educati on	1	3	3	2	1
97	Das was the lowest-ranked candidate eligible for appointment to the IAS.	दास आईएएस को आबन्ध के लिये योग्य सबसे निम्न - श्रेणीबद्ध अभ्यर्थी था।	Educati on	2	2	3	2	3
98	But the above mentioned programme is being run in the state with the help of women health workers.	लेकिन उपरोक्त उल्लिखित कार्यक्रम महिलाएं का स्वास्थ्य के कार्यकर्ताओं की सहायता के साथ अवस्था में चलाया जा रहा है।	Health	2	3	3	3	1
99	The scholars the world over rejected this theory a decade ago.	विद्वान दुनिया ओवर अस्वीकृत यह सिद्धांत इक दशक पहले।	Philoso phy	1	2	1	1	3
100	We have come to know about it only through excavations.	हमारे पास उत्खनन से अकेले इसके बारे में ज्ञात हुआ हुआ है ।	ARCHE OLOGY	2	1	3	3	2

## प्रकाशित शोध पत्र (Published Research paper)

---

1. हाडके सुमेध खुशालराव, झा गिरीश नाथ (2017): *अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों की मूल्यांकन पद्धतियाँ*, International Journal of Hindi Research, Gupta Publication, New Delhi, Vol.3, Issue 4; July 2017; ISSN : 2455-2232; Page No. 16-19.
2. हाडके सुमेध खुशालराव, *सूचना प्रौद्योगिकी के युग में मशीनी अनुवाद तंत्रों की उपादेयता*, सूचना प्रौद्योगिकी : कल, आज, और कल, संपादक : सुरेश कुमार जिंदल, फूलदीप कुमार, डी.आर.डी.ओ. विशेष प्रकाशन श्रृंखला, रक्षा वैज्ञानिक सूचना तथा प्रलेखन केंद्र (डेसीडॉक), दिल्ली, 2015, ISBN:9788186514689, Page No. 179-188.
3. हाडके सुमेध खुशालराव, *मशीनी अनुवाद के द्वारा कॉर्पोरा आधारित हिंदी-अंग्रेजी अनूदित वाक्यों का त्रुटि विश्लेषण*, समकालिन वैज्ञानिक अनुसंधान, संपादक: सुरेश कुमार जिंदल, फूलदीप कुमार, डी.आर.डी.ओ. विशेष प्रकाशन श्रृंखला, रक्षा वैज्ञानिक सूचना तथा प्रलेखन केंद्र (डेसीडॉक), दिल्ली, 2013, ISBN:9788186514382, Page No. 60-66.
4. Evaluation of Hindi-English MT Systems, Atul Kr Ojha, Akanskha Bansal, Sumedh Hadke, Girish Nath Jha, proceedings of the "2nd Workshop on Indian Language Data Resource and Evaluation" held under LREC at Reykjavik, Iceland, 26-31 May 2014
5. Using the ILCI Annotation Tool for POS Annotation: A Case of Hindi, Ritesh Kumar, Shiv Kaushik, Pinkey Nainwani, Esha Banerjee, Sumedh Hadke, Girish Nath Jha, International Journal of Computational Linguistics and Applications (IJCLA), VOL. 3(2), JUL-DEC 2012, PP. 93-104, guest editor: Yasunari Harada, Editor in-chief: Alexander Gelbukh, Bahri Publications, Delhi, ISSN: 0976-0962.