"हिंदी-मराठी मशीनी अनुवाद में शाब्दिक अस्पष्टता (समान उच्चारण वाले शब्दों के संदर्भ में)"

"LEXICAL AMBIGUITY IN HINDI-MARATHI MACHINE TRANSLATION (IN THE CONTEXT OF HOMONYMS)"

(एम.फिल्.उपाधि हेतु प्रस्तुत लघु शोध-प्रबंध)

शोधार्थी

कांबले प्रकाश अभिमन्यु

शोध-निर्देशक
प्रो.चमन लाल

शोध सह-निर्देशक
डॉ.गिरीश नाथ झा

भारतीय भाषा केंद्र

भाषा, साहित्य एवं संस्कृति अध्ययन संस्थान

जवाहरलाल नेहरू विश्वविद्यालय

नई दिल्ली - 110067

2009

मेरे परिवार

और

मेरे जीवन में आये

सभी गुरुजनों को

सस्नेह . . .

जवाहरलाल नेहरू विश्वविद्यालय

JAWAHARLAL NEHRU UNIVERSITY

Centre of Indian Languages,

School of Language, Literature & Culture Studies

New Delhi – 110067, INDIA

Centre of Indian Languages

Date - 16th July, 2009

DECLARATION

I declare that the material in this dissertation entitled in Hindi "हिंदी-मराठी मशीनी अनुवाद में शाब्दिक अस्पष्टता (समान उच्चारण वाले शब्दों के संदर्भ में)" " Hindi Transcription "HINDI-MARATHI MASHINI ANUVAAD MEIN SHAABDIK ASPASHTATA(SAMAAN UCHHARAN VAALE SHABDON KE SANDARBHA MEIN)" in English "LEXICAL AMBIGUITY IN HINDI-MARATHI MACHINE TRANSLATION (IN THE CONTEXT OF HOMONYMS)" submitted by me is an original research work and has not been previously submitted for any other degree in this or any other university /institution.

Kamble Prakash Abhimannu

RESEARCH SCHOLAR

PROF.CHAMAN LAL	Dr. GIRISH NATH JHA	PROF.CHAMAN LAL
SUPERVISOR	CO-SUPERVISOR	CHAIRPERSON
CIL / SLL&CS	Special Centre for Sanskrit Studies,	CIL / SLL&CS
Jawaharlal Nehru University	Jawaharlal Nehru University	Jawaharlal Nehru University
New Delhi – 110067	New Delhi – 110067	New Delhi – 110067

Dr. GIRISH NATH JHA
Associate Professor
Special Centre for Sanskrit Studies
Jawaharlal Nehru University
New Delhi - 110067

GRAM:JAYENU TEL.:(91-001) 6107676, 6167557TELEX:031-73167 JNU IN FAX:91-011-6865886

आभार

महात्मा गांधी अंतर्राष्ट्रीय हिंदी विश्वविद्यालय से जे.एन.यू.में इसी उमंग और उल्लास के साथ आया था कि जे.एन.यू.में आकर "अनुवाद प्रौद्योगिकी(Translation Technology)" में किए गए एम.ए.के आधार पर ही एम.फिल्. में भी मशीनी अनुवाद पर ही लघु शोध-प्रबंध पर काम करु। भारतीय भाषा केंद्र, जे.एन.यू. में मशीनी अनुवाद के लिए विशेष प्रोफेसर या व्याख्याता न होने के बावजूद प्रो.चमन लाल सर ने मुझे डॉ.गिरीश नाथ झा सर के साथ जुडकर काम करने का मौका दिया। उसी दिन से मेरी उमंगों ने उड़ान भरी और मैंने इस अवसर को न खोते हुए शोध कार्य आरंभ कर दिया। इस विषय के अर्थी महत्व को ध्यान में रखकर एवं मशीनी अनुवाद में मेरी रुचि को देखते हुए प्रो.चमन लालजी ने मुझे इस कार्य की ओर प्रेरित किया और मशीनी अनुवाद से संबंधित-"हिंदी-मराठी मशीनी अनुवाद में शाब्दिक अस्पष्टता(समान उच्चारण वाले शब्दों के संदर्भ में)" इस विषय पर लघु शोध-प्रबंध पर काम करने की अनुमति दे कर मेरी शोध रुचि को आगे बढ़ाने का महत्वपूर्ण काम किया।

इस लघु शोध-प्रबंध को पूर्ण करने का सर्वाधिक श्रेय जाता है इस शोधकार्य के सह–निर्देशक डॉ.गिरीश नाथ झा सर को, सर की सहायता के बगैर मैं यह कल्पना भी नहीं कर सकता था कि मेरा लघु शोध-प्रबंध मशीनी अनुवाद के क्षेत्र में पूरा हो पायेगा। जब साक्षात्कार के दौरान मैं जे.एन.यू. आया उसी समय से सर ने मुझे जे.एन.यू. में प्रवेश लेने के लिए मेरा हौसला बढ़ाया और मशीनी अनुवाद के क्षेत्र में कार्य करने के लिए प्रेरित किया। बगैर संगणक के यह काम पूर्ण नहीं हो पाएगा और मेरे पास कंप्यूटर न होने की सूचना मैंने सर को दी। सर ने उसी समय मुझे विशिष्ट संस्कृत अध्ययन केंद्र की लैब में काम करने की अनुमति दी। जिसका उपयोग मैं किसी भी समय जाकर रात एक-दो बजे तक भी कर सका। जिसका मैं विशेष आभार प्रकट करता हूँ। इसी अंतराल में मेरे संपर्क में आये उन सभी विशिष्ट संस्कृत केंद्र के मित्रों का भी आभार प्रकट करना चाहूंगा, जिन्होंने मेरे कार्य पर हमेशा एक मित्र, स्नेही, आलोचक एवं मार्गदर्शक के रुप में चर्चा कर एक पैनी दृष्टि रखी। जिनमें दिवाकर मणि, दिवाकर मिश्र, मंजी बद्रा, सुरजीतसर आदि का प्रमुख रुप से नाम लेना चाहूंगा।

यह कार्य पूर्ण रुप से नवीन होने के कारण इसमें केवल दो ही आधार ग्रंथों की विशेष रुप से सहायता मिली है। 1 'हिंदी में अनेकार्थकता का अनुशीलन - डॉ.त्रिभुवन ओझा" एवं 2. "**Learning Form-Meaning Mappings in Presence of Homonymy: a linguistically motivated model of learning**

inflection – by Katya Pertsova" इन दो किताबों के सैद्धांतिक पहलुओं से परिचित होने से शोधकार्य को एक दिशा मिली जिसके लिए मैं इनका आभार व्यक्त करता हूँ। यह कार्य केवल सैद्धांतिक पहलूओं तक ही सीमित नहीं था इसके लिए अनेकार्थक शब्दों को संकलित कर इन शब्दों का मराठी अनुवाद भी करना आवश्यक था। इस कार्य के लिए मुझे आई.आई.टी.मुंबई द्वारा बनाए गए हिंदी एवं मराठी वर्डनेट की सामग्री संकलन एवं विषय विवेचन में सहायता मिली। जिसके कारण मैं समय की बचत कर शोध के अन्य कार्यों पर ध्यान दे सका।

इकबाल, विश्वेश, रजनीश, प्रियंका, माधव गोपाल ने इस शोध कार्य के टाईप होने से प्रूफ़ एवं डेटा बेस को आकार देने में लगातार मदद की। प्रदीप, नीतेश, सौम्या ने यूरोपिय भाषाओं में अनेकार्थक शब्दों के उदाहरणों को स्पष्ट करने लिए मदद की। इस शोध की सामग्री संकलन में साहित्य अकादमी पुस्तकालय से हिंदी-मराठी शब्द कोशों के साथ शोध संबंधित पुस्तकों की सहायता मिली। केंद्रीय हिंदी संस्थान, दिल्ली(पुस्तकालय), विशिष्ट संस्कृत अध्ययन केंद्र(पुस्तकालय), जे.एन.यू., केंद्रीय पुस्तकालय जवाहरलाल नेहरु विश्वविद्यायल का महत्वपूर्ण सहयोग मिला। जिनके कर्मचारियों ने मुझे कई ऐसी पुस्तकें उपलब्ध करवाई जो इनके पास नहीं थीं। इस महत्वपूर्ण सहायता के लिए मैं इन सभी का आभार व्यक्त करता हूँ।

कोई भी कार्य मित्रता एवं स्नेह के बगैर पूरा नहीं हो सकता। घर से मिला माता-पिता का स्नेह और फोन पर हुए भैया-भाभी के मार्गदर्शन ने और जीजा-जीजी के बातचीत से शोध कार्य के प्रति लगन रखने की भावना ने मुझे हमेशा प्रेरित किया। वर्गमित्रों में नसीम, राजीव, राजेश, तृप्ती, हाफिज आदि ने शोध विषय पर निरंतर चर्चा की तो प्रमथेश ने शोध में मराठी के होने की हर बार याद दिलाई। जिसने शोध में रुचि रखने का काम किया। उन सभी नए और पुराने सीनियर जिनमें विवेक सर, रणजीत सर, हरिओम भैया, अखिलेशजी, अंजनी सर, अशोक सर, का भी आभारी हूँ जिन्होंने मशीनी अनुवाद एवं अनुवाद के क्षेत्र में सफलता पूर्वक कार्य कर मुझे भी यह कार्य करने की प्रेरणा दी।

इस अनुसंधान में एक महत्वपूर्ण काम यह भी था कि अनुवाद किए गए हिंदी से मराठी के शब्दों को विश्लेषित करना एवं उनकी जाँच करना। इस कार्य में प्रमुख भूमिका निभाई अर्चना बलविर(व्याख्याता,केंद्रीय हिंदी संस्थान,दिल्ली), रमेश गायकवाड(पी.एचडी), सीमा जाधव(पी.एचडी), अर्चना थूल(व्याख्याता) आदि का प्रमुख रुप से आभार प्रकट करता हूँ। जिनके बगैर यह कार्य संपूर्ण नहीं हो पाता।

मैं इस अनुसंधान कार्य में कितना सफल हो पाया हूँ यह परीक्षक एवं अध्येता ही निश्चित कर पायेंगे लेकिन मेरे इस कार्य से यदि भविष्य में मशीनी अनुवाद में कार्य कर रहें, विशेष रुप से अनेकार्थक शब्दों पर कार्य करने वाले अभ्यासकों को नाममात्र भी सहायता होगी तो मैं इस कार्य को सार्थक मानूंगा। मैं यह भी आशा करता हूँ कि इस कार्य को संपूर्ण रूप देकर न केवल हिंदी से मराठी बल्कि इसी आधार पर अन्य भारतीय भाषाओं में भी अनेकार्थक शब्दों के अनुवाद कार्य के लिए इस शोध कार्य का उपयोग हो।

इस लघु-शोधप्रबंध का कार्य प्रो.चमन लालजी के सान्निध्य में पूर्ण हुआ है और इस लघु शोध-प्रबंध को आकार देने में डॉ.गिरीश नाथ झा सर का स्नेहिल एवं मनोयोग पूर्ण निर्देशन रहा है। उनके प्रति मैं हार्दिक श्रद्धा और आदर प्रस्तुत करता हूँ।

कांबले प्रकाश अभिमन्यु

कक्ष संख्या 28, पेरियार छात्रावास,

जवाहरलाल नेहरु विश्वविद्यलय,

नई दिल्ली -110067

अनुक्रम

आभार ..... vii-viivii

संक्षिप्त अक्षर / टेबल / चित्र पृष्ठ क्र. ..... vii-viii

विषय प्रवेश ..... 01 -03

पहला अध्याय: मशीनी अनुवाद की समस्याएँ ..... 04 - 22

पहले अध्याय का सार 05-05

1.मशीनी अनुवाद का परिचय 06-07

1.1 मशीनी अनुवाद की प्रक्रिया और संक्षिप्त इतिहास 07-11
1.2 भारत में मशीनी अनुवाद की आवश्यकता 11-11
1.3 म.अनु.में स्रोत भाषा विश्लेषण (The Source Language Analysis) की आवश्यकता। 11-11

2.मशीनी अनुवाद की समस्याएँ। 11-22

2.1. अनुवाद पूर्व समस्याएँ (pre -Translation Problems) 12–13
2.2. अनुवाद की समस्याएँ (Translation Problems) 12–15
2.3. अनुवादोत्तर समस्याएँ (Post-Translation Problems) 15–16
2.4. अनुवादक की समस्याएँ (Translator/User problems) 16-16

3. वैश्विक स्तर पर मशीनी अनुवाद की समस्याएँ 16-18

4. भारत में मशीनी अनुवाद की समस्याएँ 18-22

दूसरा अध्याय: समान उच्चारण वाले शब्दों से अनुवाद एवं मशीनी अनुवाद में

होने वाली समस्याएँ 23–61

दूसरे अध्याय का सार ..... 24 - 24

1. समान उच्चारण वाले शब्द का परिचय और परिभाषा..... 25 - 31

2. अन्य भाषाओं में मिलने वाले समान उच्चारण वाले शब्दों के उदाहरण 31- 36

3. समान उच्चारण वाले शब्दों के प्रकार एवं व्याकरणिक सूत्र ..... 36 - 40

4. मानव द्वारा होने वाले अनुवाद एवं मशीनी अनुवाद में समान उच्चारण वाले शब्दों की समस्याएँ ..... 40 - 61

4.1 मानव द्वारा होने वाले अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं के उदाहरण 42-49

4.2 मशीन द्वारा होने वाले अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं के उदाहरण 49-61

तीसरा अध्याय : हिंदी-मराठी मशीनी अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं का भाषिक सूत्र 62-120

तीसरे अध्याय का सार 63-63

1. हिंदी में समान उच्चारण वाले शब्द का प्रचलन और समान उच्चारण वाले शब्द का विभाजन 64-74

2. समान लिपि, समान उच्चारण, भिन्नार्थक शब्द 75-92

3. भिन्न लिपि, समान उच्चारण, भिन्नार्थक शब्द 93-104

4. अंग्रेजी से हिंदी में आये अनेकार्थक शब्द 105-120

चौथा अध्याय : समान उच्चारण वाले शब्दों की समस्याओं का

संगणकीय निराकरण 121-140

चौथे अध्याय का सार 122-122

1. समान उच्चारण वाले शब्द की अन्य विशेषतायें 123-128

2. संगणक कलनविधि 128-135

3. उपयोग 135-140

1. भारतीय भाषाओं के शब्दकोशीय संसाधनों के उपयोग के लिए। 135- 138

2. भारतीय भाषाओं से भारतीय भाषाओं के मशीनी अनुवाद के उपयोग के लिए।... 139-140

उपसंहार 141-148

ग्रंथानुक्रमणिका 149-156

1.आधार ग्रंथ 150-150

2.संदर्भ ग्रंथ, लेख, कोश एवं पत्रिका 151-155

3.इंटरनेट वेब संदर्भ 155-156

संक्षिप्त अक्षर / टेबल / चित्र

संक्षिप्त अक्षर

No	Abbreviations	Description
1	ALPAC	Automatic Language Processing Advisory Committee
2	COLING	Computational Linguistics
3	C-DAC	Centre for Development in Advance Computing
4	IIT	Indian Institute of Technology
5	JSP	Java Server Pages
6	NLP	Natural Language Processing
7	PL	Plural
8	OBJ	Object
9	POS	part-of-speech tagging (POS tagging or POST)
10	SG	Singular
11	STT	Speech to Text
12	TDIL	Technology Development for Indian Language
13	TTS	Text To Speech
14	VR	Verb Root
15	R&D	Research and Development
16	AI	Artificial Intelligence
17	AUX	Auxiliary
18	PREF	Prefix
19	SG	Singular
20	VR	Verb Root
21	SUFF	Suffix
22	SUBJ	Subject
23	OP	Output
24	FL	formed language
25	LTs	Language Tools
26	MT	Machine Translation
		हिंदी संक्षिप्ताक्षर
27	म.अनु.	मशीनी अनुवाद
28	NP	संज्ञा
29	ProNP	सर्वनाम
30	ADJ	विशेषण
31	VP	क्रिया
32	NP	निपात
33	ADV	क्रिया-विशेषण

टेबल

Name	Page NO.
1.Table NO.1: Marathi-Hindi MT is one of the major language pair in this project.	3
2.Table No. 2. types of homonyms	36
3. Table No. 3 - अनेकार्थक शब्दों के शब्द-भेद.....	71
4. Table No. 4. अविकारी शब्दों के शब्द-भेद.....	73
5. Table No. 5 एक भाषा में मिलने वाले अर्थ निर्देशक शब्द	124
6. Table No. 6 व्यक्ति समन्वयात्मक अर्थ निर्देशक शब्द.....	125
7. प्राकृतिक वर्ग के अनुसार मिलने वाले उदाहरण एवं शून्य अनेकार्थक शब्द	127

चित्र

1. Graph No.1 process of MT Graph No.1 10
2. Graph No.2 Error analysis, Frameworks – example 54
3. Graph No.3 process of Homonymy word Translation Tool(Hindi to Marathi) 129
4. Graph No.4. CLIR Schematic for input being Marathi 137
5. Graph No.5 Hindi-Marathi Homonym Marker 133

विषय प्रवेश

आधुनिकीकरण के युग में भारतीय विद्वानों को यह कभी नहीं भूलना चाहिए की भारत एक बहुभाषिक देश है। बहुभाषिकता की समस्या को अनुवाद के माध्यम से दूर किया गया लेकिन सूचना प्रौद्योगिकी(Information Technology) के युग में भाषा का मुकाबला मशीन से है। वर्तमान युग में वैश्विक स्तर पर मशीनी अनुवाद के क्षेत्र में कुछ प्रमुख समस्याओं को निम्न रुप से देख सकते हैं, जिनका निराकरण अभी तक नहीं हो पाया है। समान उच्चारण वाले शब्द (Homonymy), वाक्यगत द्विअर्थकता(Syntactical Ambiguity), संदर्भ परक द्विअर्थकता (Referential Ambiguity), अस्पष्ट पद (Fuzzy), संकेत, कहावतें, मुहावरे (Metaphors and Symbols), विकसित नए शब्द (New Vocabulary Development) आदि। इन समस्याओं में समान उच्चारण वाले शब्द (Homonymy) की समस्या एक प्रमुख समस्या है, जिसपर वैश्विक स्तर पर विचार किया जा रहा है। भारतीय भाषाओं में भी इस विषय पर अनुसंधान की आवश्यकता है। "Homonyms" शब्द ग्रीक के Homo + onyms शब्दों के पूर्वसर्ग और परसर्ग से बना है। जिसका हिंदी अर्थ "समान उच्चारण वाले शब्द", "समान उच्चारण वाले भिन्नार्थक शब्द" या "अनेकार्थक" शब्द है। छोटे बच्चे, सामान्य पाठक, श्रोता, मानव द्वारा किये जा रहे अनुवाद (Human Translation), आशु अनुवाद(interpretion), नई भाषा सीखते समय(Language Acquisition), और कृत्रिम बुद्धि निर्माण-कर्ता विशेषज्ञों (Artificial Intelligence specialists) को समान उच्चारण वाले शब्दों (Homonymy) के कारण अर्थ के अस्पष्टता की समस्या उत्पन्न होती है। मानव अपनी कुशाग्र बुद्धि, पूर्व संदेशों की सहायता एवं शारीरिक गतिविधियों को देखते हुए संदेशों का अर्थ समझता है। यह प्रक्रिया पूर्ण रुप से मानव मस्तिष्क पर आधारित है, जो मानव मस्तिष्क में होने वाले विशिष्ट कोश, वैश्विक ज्ञान एवं भाषिक ज्ञान की सहायता से पूर्ण होती है। इस समस्या के निराकरण के लिए मशीन को मशीन की भाषा से अवगत कराया जाता है लेकिन यह भाषिक साधन अभी पूर्ण रुप से विकसित नहीं हुए हैं। मशीनी अनुवाद की इन समस्याओं में समान उच्चारण वाले शब्दों (Homonyms) की समस्या गंभीर रुप से सामने आई है। समान उच्चारण वाले शब्दों के निराकरण की समस्या विश्व के अधिकतम मशीनी अनुवाद यंत्रों में है। मशीनी अनुवाद में निर्माण होने वाली इस वैश्विक स्तर की समस्या पर हिंदी-मराठी मशीनी अनुवाद में यह पहला अनुसंधान है।

इस अनुसंधान के प्रथम अध्याय में मशीनी अनुवाद की समस्याओं के संबंध में चर्चा की गई है। इसी के साथ मशीनी अनुवाद का संक्षिप्त इतिहास और प्राकृतिक भाषा संसाधन(एनएलपी) की भी चर्चा की गई है। भारतीय एवं पाश्चात्य भाषाओं के मशीनी अनुवाद यंत्रों से तुलना करने के उद्देश्य से मशीनी अनुवाद की समस्याओं को वैश्विक स्तर के म.अनु. की समस्याओं एवं भारतीय भाषाओं के म.अनु. की समस्याओं में विभाजित किया गया है।

द्वितीय अध्याय में अनुसंधान के प्रमुख विषय समान उच्चारण वाले शब्द, समान उच्चारण वाले शब्दों के प्रकार और समान उच्चारण वाले शब्दों के सामान्य व्याकरणिक सूत्रों पर चर्चा की गई है। जिसमें समान उच्चारण वाले शब्दों के प्रकारों की समस्याओं को मानव द्वारा किये जाने वाले अनुवाद एवं मशीनी अनुवाद में किस प्रकार समस्याऐं उत्पन्न होती है, और मनुष्य इन समस्याओं का किस प्रकार समाधान निकालता है। इस पर विचार किया गया है।

तृतीय अध्याय में समान उच्चारण वाले शब्दों का विभाजन चार भागों में किया गया है। 1. हिंदी में समान उच्चारण वाले शब्द का प्रचलन और समान उच्चारण वाले शब्द का विभाजन 2.समान लिपि, समान उच्चारण, भिन्नार्थक शब्द 3. भिन्न लिपि, समान उच्चारण, भिन्नार्थक शब्द 4. अंग्रेजी से हिंदी में आये समान उच्चारण वाले शब्द। इन विभाजित हिंदी-मराठी शब्दों में संकलन किया गया है। संकलित शब्दों पर विस्तृत चर्चा कर हिंदी से मराठी में उचित अर्थों को स्पष्ट किया गया है। समान उच्चारण वाले (Homonymy) शब्दों को व्याकरणिक कोटियों की सहायता से स्पष्ट किया गया है, जिससे शब्द का अर्थ स्पष्ट हो सके। यह पाठ हिंदी-मराठी मशीनी अनुवाद में प्राकृतिक भाषा संसाधन के भाषिक संसाधनों में समान उच्चारण वाले शब्दों की अस्पष्टता को दूर करने के लिए भाषा संसाधन(Language Tool) निर्माण में लघु शब्दकोश के रुप में महत्वपूर्ण योगदान देगा।

चतुर्थ अध्याय मुख्य रुप से 'कंम्प्यूटर भाषा प्रोग्रामिंग' के लिए रखा गया है जिसमें समान उच्चारण वाले शब्दों पर किये गये अनुसंधान की सहायता से मशीनी अनुवाद करते समय यह कार्य किस प्रकार सहायता करेगा इसकी प्रोग्रामिंग की गई है। जिसके निर्देशानुसार समान उच्चारण वाले (Homonymy) शब्दों के विशिष्ट सहायक लघु शब्दकोश की सहायता से मशीन कार्य करेगी। इसी के साथ कुछ व्याकरणिक नियमों की सहायता से भी कलनविधि संपन्न की गई है। अनुसंधान का अंतिम चरण अनुसंधान के उपयोग को स्पष्ट करता है। जिससे किया गया कार्य किसी भी प्रकार केवल सिद्धांतों तक ही सीमित न रह पाये इसे स्पष्ट करता है। उपयोगिता को स्पष्ट करने के लिए हिंदी-मराठी शब्द-जाल एवं मशीनी अनुवाद में अनुसंधान का अधिक से अधिक किस प्रकार उपयोग

होगा इस पर चर्चा की गई है। यह अनुसंधान कार्य चार अध्यायों में लिखा गया है जिसे अनुसंधान की आवश्यकता के अनुसार विभाजित किया गया है।

२००६ में भारत सरकार ने मशीनी अनुवाद के संबंध में काफी सकारात्मक रवैया अपनाते हुए भारत की अधिकतर प्रमुख भाषाओं में मशीनी अनुवाद यंत्र बनाने की परियोजना बनाई है। जिसके निर्माण से मशीनी अनुवाद जगत में वैश्विक स्तर पर एक नया इतिहास बन जायेगा लेकिन यह कार्य काफी कठिन है। भारतीय भाषाओं से भारतीय भाषाओं में मशीनी अनुवाद यंत्र निर्माण की यह परियोजना केवल मशीनी अनुवाद के लिए ही नहीं बल्कि संपूर्ण अनुवाद क्षेत्र के लिए काफी सकारात्मक है। जिसका विवरण टेबल नंबर एक में देख सकते है।

Name of the Project	Objectives
Printed Text OCR	• Development of OCR system for printed Indian scripts. • Scripts: Bengali, Devanagari, Malayalam, Gujarati, Telugu, Tamil , Oriya, Tibetan , Nepali, Gurumukhi
On-line Handwriting Recognition system	• Development of On-line Handwriting Recognition System : • Scripts: Bengali, Devanagari, Tamil , Telugu , Kannada and Malayalm Scripts
Cross-lingual Information Access (CLIA)	• Development of web-based cross-lingual information Access • The query language –Retrieval language pairs are: Hindi -Hindi , English Bengali Bengali , Hindi, English Marathi- Marathi , Hindi, English Punjabi- Punjabi , Hindi, English Tamil- Tamil, Hindi, English Telugu- Telugu , Hindi, English
Indian Language to Indian Language Machine Translation System (IL-IL MT)	• Development of Machine Translation System from one Indian Language to another Indian language. • The language pairs are: – Tamil-Hindi ; Hindi-Tamil – Telugu-Hindi; Hindi-Telugu – Marathi-Hindi; Hindi-Marathi – Bengali-Hindi; Hindi-Bengali – Tamil-Telugu; Telugu-Tamil – Urdu-Hindi; Hindi-Urdu – Kannada-Hindi; Hindi-Kannada – Punjabi-Hindi, Hindi-Punjabi – Malayalam-Tamil ; Tamil-Malayalam
English to Indian Language Machine Translation system (E-II, MT)	• Development of Machine Translation System from English to Indian languages. • The language pairs are: – English-Hindi – English-Bengali – English-Marathi – English-Oriya – English-Urdu – English-Tamil.
English to Indian Language Machine Translation system with Angla-Bharti Technology	• Development of Machine Translation System from English to Indian languages. • The language pairs are: – English-Urdu – English-Punjabi – English-Bengali – English-Malayalam •Angla-Bharti-II Technology will be used in this project for the development of MT system for the specific language pairs

Tabl. NO.1[1]: Marathi-Hindi MT is one of the major language pair in this project.

[1] paradigm Shift Of Language Technology initiatives Under Programme - विश्वभारत@TDIL-page 4

पहला अध्याय

मशीनी अनुवाद की समस्याएँ

पहले अध्याय का सार

1. मशीनी अनुवाद का परिचय
   - 1.1 मशीनी अनुवाद की प्रक्रिया और संक्षिप्त इतिहास
   - 1.2 भारत में मशीनी अनुवाद की आवश्यकता
   - 1.3 म.अनु.में स्रोत भाषा विश्लेषण (The Source Language Analysis) की आवश्यकता।
2. मशीनी अनुवाद की समस्याएँ।
   - 2.1. अनुवाद पूर्व समस्याएँ (pre -Translation Problems)
   - 2.2. अनुवाद की समस्याएँ (Translation Problems)
   - 2.3. अनुवादोत्तर समस्याएँ (Post-Translation Problems)
   - 2.4. अनुवादक की समस्याएँ (Translator/User problems)
3. वैश्विक स्तर पर मशीनी अनुवाद की समस्याएँ
4. भारत में मशीनी अनुवाद की समस्याएँ

पहले अध्याय का सार

इस शोध-प्रबंध के पहले अध्याय में मशीनी अनुवाद की प्रमुख समस्याओं के साथ मशीनी अनुवाद की सामान्य जानकारी, मशीनी अनुवाद का संक्षिप्त इतिहास, मशीनी अनुवाद की प्रक्रिया एवं प्राकृतिक भाषा संसाधन (NLP) के संबंध में चर्चा की गई है। मशीनी अनुवाद की समस्याओं को अधिक संशोधित करने के उद्देश्य से भारतीय भाषाओं के मशीनी अनुवाद में आने वाली समस्याओं को पाश्चात्य भाषाओं के मशीनी अनुवाद की समस्याओं के साथ तुलना करने के उद्देश्य से चार भागों में विभाजित किया है। 1) मशीनी अनुवाद का परिचय 2) मशीनी अनुवाद की समस्याएँ। 3) वैश्विक स्तर पर मशीनी अनुवाद की समस्याऐं 4) भारतीय भाषाओं में मशीनी अनुवाद की समस्याऐं। इसी के साथ म.अनु. की समस्याओं को मानव अनुवाद (Human Translation) एवं मशीनी अनुवाद(Machine Translation) के साथ जोड़कर मशीनी अनुवाद की समस्याओं को चार भागों में स्पष्ट किया गया है। जिसके कारण म.अनु की समस्याओं पर विस्तृत चर्चा हो पाई है। साथ ही अनुसंधान का उद्देश्य और अनुसंधान की संरचना पर भी पाठ के अंत में सामान्य चर्चा की गई है जिससे इस शोध-प्रबंध को आगे बढ़ाने में सहायता मिली।

1.मशीनी अनुवाद का परिचय

अनुवाद के कई प्रकारों में से "मशीनी अनुवाद" भी अनुवाद का एक प्रकार है। मशीनी अनुवाद में मानव को केवल सहायक के रुप में उपयोग किया जाता है और बाकी काम कंप्यूटर के माध्यम से पूर्ण किया जाता है। जिसे एक विशिष्ट संगणक प्रोग्राम के द्वारा संचालित किया जाता है। लेकिन अनुवाद केवल शब्दों का खेल नहीं है। अनुवाद के स्त्रोत पाठ में संस्कृति, भाव और संवेदनाओं का गहरा प्रभाव होता हैं। दो भाषाओं के शब्दों के उलटफेर में संगणक मानव बुद्धि से आगे निकल गया लेकिन जब बात संस्कृति, संवेदना और भावनाओं के भाषांतरण की आई तो संगणक पुन: मानव की शरण में आ गया। इस जटिलता को दूर करने में संगणक इस कदर फँसा हुआ है कि अब बिना मानव सहायता के संगणक से अनुवाद कर पाना मुश्किल हो रहा है। पूर्णत: सफल म.अनु. का निर्माण न होना ही इसका सबसे बड़ा उदाहरण है। फिर भी इस काम में मशीनी अनुवाद निर्माणकर्त्ताओं ने अभी हार नहीं मानी है। भारतीय अनुवाद यंत्र भी इन समस्याओं के बराबर शिकार होते हुए नजर आ रहे हैं। इसलिए अनुवाद यंत्र निर्माण करने से पूर्व अनुवाद की समस्याओं के विषय में जान लेना आवश्यक हो जाता है। मशीनी अनुवाद की संकल्पना को व्यवस्थित रुप से देखने के लिए म.अनु. की परिभाषाओं को निम्न रुप से देख सकते हैं।

प्रो.सूरजभान सिंह के अनुसार मशीनी अनुवाद की सामान्य परिभाषा यह है "अनुवाद की ऐसी प्रक्रिया जिसमें संगणक प्रणाली(system) के जरिए एक भाषा से दूसरी भाषा में अपने आप अनुवाद हो, इस प्रक्रिया में अनुवाद की जाने वाली सामग्री(Text) को आगत शब्द(Input) के रुप में देते है। संगणक की भीतरी प्रणाली जिसमें दोनों भाषाओं के शब्दों, मुहावरो और व्याकरणिक नियमों का ज्ञान संचित रहता है, मशीन अपने आप उस सामग्री का दूसरी भाषा में अनुवाद करती है और कुछ ही क्षणों में निर्गत पाठ (output) के रुप में अनुदित सामग्री प्राप्त हो जाती है।"[2]

डॉ. दीपा गुप्ता के अनुसार म.अनु. की परिभाषा कुछ इस प्रकार है "मशीनी अनुवाद एक ऐसी प्रक्रिया है जो पाठ के इकाईयों को एक भाषा(स्रोत भाषा) से दूसरी भाषा(लक्ष्य भाषा) में कंप्यूटर के माध्यम से अनूदित करती है।"Machine Translation (MT) is the process of translating text units of

---

[2] प्रो. सिंह, सुरजभान, अप्रैल-जून 2004, *कम्पूटर अनुवाद, सूचना प्रौद्योगिकी, हिंदी और अनुवाद, अनुवाद पत्रिका*(कंप्यूटर विषेशांक-२)

one language (source language) into a second language (target language) by using computers."[3]

उपर्युक्त परिभाषाओं से स्पष्ट होता है कि अंशत: मानव सहायता के साथ (यंत्र) कंप्यूटर का उपयोग कर एक प्राकृतिक मानव भाषा का दूसरी प्राकृतिक मानव भाषा में अनुवाद करना ही मशीनी अनुवाद \यंत्रानुवाद है। इसी के आधार पर म.अनु. में मशीन की सहभागिता के आधार पर म.अनु. को तीन भागों में विभाजित किया गया हैं।

1) पूर्णत: मशीनी अनुवाद (Fully Machine Translation)

2) मानव साधित मशीनी अनुवाद (Human Aided Machine Translation)

3) मशीन साधित मानव अनुवाद (Machine Aided Human Translation)[4]

1.1 मशीनी अनुवाद की प्रक्रिया और संक्षिप्त इतिहास

विश्वभर के मशीनी अनुवाद के इतिहास पर विचार किया जाए तो मशीनी अनुवाद का इतिहास लगभग पचास वर्ष पुराना है। सबसे पहले "मशीनी अनुवाद यंत्र" निर्माण करने का प्रयास 1956 में जार्ज टाऊन विश्वविद्यालय वाशिंगटन में शुरु हुआ था। जिस पर अल्पेक रिपोर्ट (ALPAC Report,1966) आने के बाद कुछ वर्ष म.अनु. पर हो रहा कार्य काफी धीमा हो गया था। 1980 के दशक के बाद पुन: म.अनु. का कार्य काफी तेज गति से आगे बढ़ा और आज यह स्थिति है कि विश्व में कई मशीनी अनुवाद यंत्र स्वस्थ रुप से अनुवाद कर रहें हैं। इस बीच कई अनुवाद यंत्र ऐसे भी रहे जो केवल अनुवाद यंत्र निर्माण के अभ्यास के तौर पर निर्माण किए गए थे। इनमें से कुछ म.अनु. यंत्र आज भी काम कर रहे हैं। मशीनी अनुवाद यंत्रों के कुछ नाम इस प्रकार हैं। "GAT, SYSTRAN, LOGOS, METEL, TAUM-METEO, EUROTRA, ATLAS-I, ALTAS-II, TAURUS, PIUOT, ALPA, LOGOS, CULT, TITUS, ARINC-78, METAL, MU, MT.S-NCST आदि।"[5] भारत में मशीनी अनुवाद की प्रथम शुरुआत 1983 से मानी जाती है। जो दक्षिण भारत के तमिल विश्वविद्यालय में रशियन-तमिल TUMTS अनुवाद यंत्र के लिए शुरु हुआ। (In India research on MT began in 1983 at the Tamil University in South India. The TUMTS system is a small-scale 'direct

---

[3] Gupta, Deepa. 2005. *Contributions to English to Hindi Machine Translation Using Example-Based Approach. (Ph.D. Thesis)*Department Of Mathematics, IIT, Delhi, India

[4] **Punjabi Text Generation using Interlingua approach in Machine Translation (**A thesis) by **-SACHIN KALRA, page No.-10**

[5] ओम, डॉ.विकास, (संपादक डॉ.नगेंद) 1995, *अनुवाद विज्ञान सिद्धांत और अनुप्रयोग मशीनी अनुवाद की समस्याएँ,* हिंदी कार्यान्वय निदेशालय,दिल्ली विश्वविज्ञाय,पृष्ठ सं.३४२

translation' system specifically designed for Russian as SL and Tamil as TL and running on a small microcomputer.[6]) लेकिन इस यंत्र की क्षमता काफी कम थी। इस यंत्र का निर्गत पाठ(Output) भी अधिक उच्च स्तर का नहीं था। इसलिए भारत में मशीनी अनुवाद की विधिवत शुरुआत आई.आई.टी कानपुर में शुरु हुए "अक्षर भारती ग्रुप" के "आंग्लभारती मशीनी अनुवाद यंत्र" से मानी जाती है। इसके उपरांत आज भारत में कई मशीनी अनुवाद यंत्रों पर कार्य किया जा रहा हैं। जिसमें (मॅट)MAT, (अनुवादक)ANUVADAK, (मंत्रा)MANTRA, (शिव)SIVA, (मंत्र)MANTR, (अनुवाद-यन्त्र)ANUVADAYAN- TR (CIILM) (अंग्रेजी-हिंदी) (अनुसारका)ANUSARAKA (भारतीय भाषाओं के लिए), (शक्ति) Shakti (अंग्रेजी-मराठी, हिंदी, तेलगु के लिए) है। भारत सरकार ने भारतीय भाषाओं से भारतीय भाषाओं के लिए मशीनी अनुवाद यंत्र की एक नई महत्वपूर्ण योजना आरंभ की है। जिसमें "Indian Language to Indian Language Machine Translation system" निर्मित किए जा रहे हैं। जिसका विवरण टेबल 1 में दिया गया है।

अनुवाद प्रक्रिया अपने-आप में एक अति कठिन कार्य है। मशीनी अनुवाद की प्रक्रिया पर विचार किया जाए तो म.अनु. की प्रक्रिया मानव द्वारा किए जाने वाले अनुवाद(Human Translation) प्रक्रिया के समान होती है। लेकिन म.अनु. की प्रक्रिया में मानव को केवल एक सहायक के रुप में ही कार्य करना पड़ता है। अनुवाद की प्रक्रिया में उसका कोई कार्य नहीं होता। इसीलिए म.अनु. की प्रक्रिया अधिक जटिल बन जाती है। जिसमें अनुवाद के शब्द चयन से लेकर वाक्य व्याकरणिक संरचना और अर्थ चयन का भी काम मशीन को ही करना पड़ता हैं। मशीन भावनाशून्य और सभ्यता-संस्कृति से अनभिज्ञ होने के कारण मशीन 'अनुवाद' प्रक्रिया को सही रुप से अंजाम नहीं दे पाती है। इस प्रक्रिया में पाठ(Text) कई भाषा संसाधनों(LTs) से होकर जाता है। मशीनी अनुवाद की संपूर्ण प्रक्रिया संक्षिप्त रुप में निम्न से देख सकते है।

मशीनी अनुवाद की प्रक्रिया : - मशीनी अनुवाद के लिए सबसे प्रमुख कंप्यूटर में "अभिष्ट भाषिक संरचना का यह व्युत्पादन संगणक के केंद्रीय संसाधक इकाई के माध्यम से होता है। संगणक की इस केंद्रीय संसाधक इकाई में एक नियंत्रक इकाई, एक मुख्य स्मृति पटल और एक गणितीय इकाई होती है। जिसके आदेशों के आधार पर संगणक काम करता है। मशीनी अनुवाद की प्रक्रिया भी इसी के

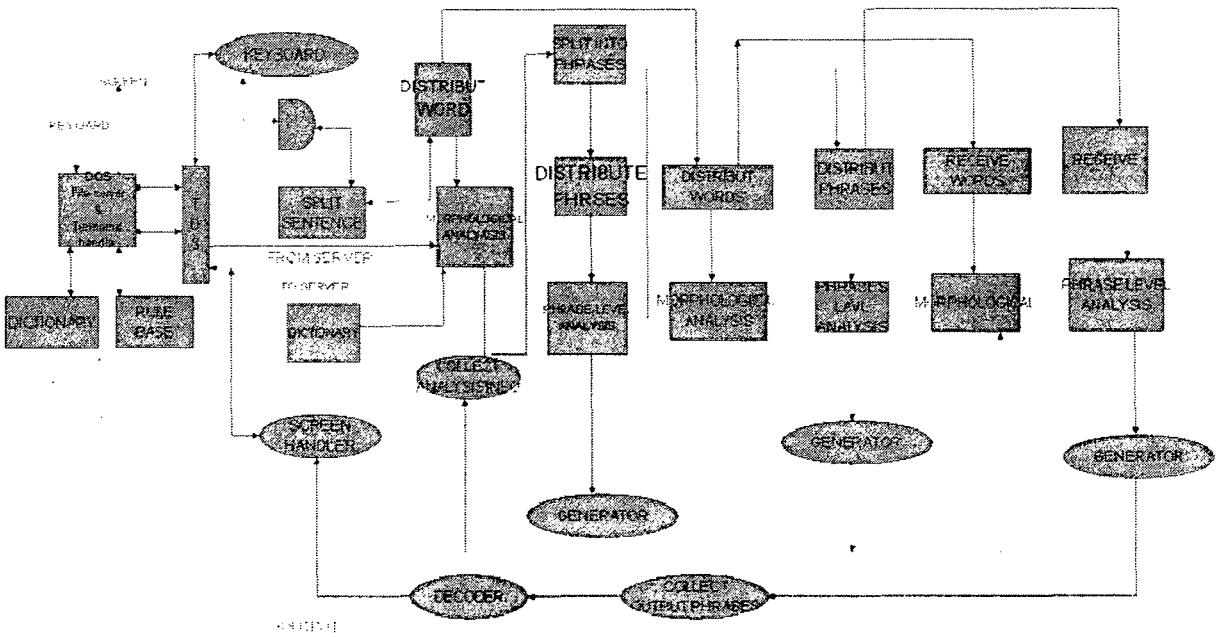
---

[6] Hutchins, W.John. 1988. *Recent developments in Machine Translation a review of the last five years*,University of East Anglia.Norwich

आधार पर आगे बढ़ती है। सबसे पहले पाठ निवेषन का कार्य होता है। प्रत्येक मशीनी अनुवाद यंत्र की यह प्रक्रिया दो रुपों में होती है। 1.लक्ष्य पाठ अनुवाद यंत्र में दिए गए बॉक्स में टाईप कर दिया जाता है। 2.या लक्ष्य पाठ की फाईल को ब्राऊस कर। इस प्रक्रिया के उपरांत ही अनुवाद की प्रक्रिया आरंभ होती है।

सामान्यत: भारतीय मशीनी अनुवाद में अनुसारका मशीनी अनुवाद की प्रक्रिया निम्ना नुसार होती है। 1.स्रोत भाषा पाठ(SOURCE LANGUAGE TEXT) 2.धातु(ROOT WORD) मशीन को दिए गए स्रोत पाठ के शब्दों में धातुओं को देखा जाता है। जिससे उचित अर्थ मिल सके। इसके उपरांत दिए गए पाठ में व्याकरणिक कार्य होता है जो सभी अनुवाद यंत्रों में समान नहीं होते। व्याकरणिक कार्य अधिकतर अनुवाद यंत्र निर्माण कर्ता ही निश्चित करते हैं कि मशीनी अनुवाद यंत्र किस व्याकरण के आधार पर कार्य करेगी। जैसे:- 3. वृक्ष संवादिता व्याकरण(POS TAG) इसके आधार से शब्दों को चिन्हित किया जाता है और आगे कार्य भी चलता है। 4.शब्द चिन्हक CHANKAR MARKING 5.पदसूत्र (PADSUTRA) इस प्रक्रिया में शब्दों को फिर से संघटित किया जाता है। 6.शब्द साधक प्रक्रिया (WORD GRUPING) 7.कथ्य विसंदिग्धिकरण (SENSE DISAMBIGUATION) विसंदिग्धिकरण की प्रक्रिया सबसे महत्वपूर्ण प्रक्रिया होती है। जिससे पाठ में निर्माण होने वाली संदिग्धता दूर की जाती है। पाठ में कई जगह पूर्वसर्ग एवं परसर्ग लगने से कई समस्याएँ निर्माण होती है जिसका निराकरण इस प्रक्रिया के माध्यम से दूर किया जाता है। 8.पूर्वसर्ग गतिविधियाँ(PREPOSITION MOVEMENTS) इस प्रक्रिया में स्रोत पाठ का लक्ष्य पाठ में रुपांतरण करने के लिए शब्दकोश से लिए गए शब्दों का प्रजनन किया जाता है। 9.लक्ष्य भाषा पाठ प्रजनन (TARGET LANGUAGE GENERATION) में कई समस्याएँ निर्माण होती है। शब्द कोश में शब्द का अर्थ या वाक्य पद नहीं मिल पाते हैं ऐसे समय भाषा व्यवहार संग्रह की मदद ली जाती है। 10.भाषा व्यवहार संग्रह(CORPARA) यह एक भाषिक साधन है जिसका कार्य पाठ में आए रुपों को विश्लेषित करना एवं रुप संरचना का विश्लेषण करना है। 11.रूपवैज्ञानिक विश्लेषण MORPHOLOGICAL ANYLISAR 12.विश्लेषक(PARASAR):- दिए गए पाठ का निर्माण एवं पाठ को प्रदर्शित करने के लिए तैयार करने का काम करता है। जिसे हम आगत पाठ भी कहते हैं। 13.सर्जक(GENARATER) आगत पाठ मशीनी अनुवाद का सबसे अंतिम चरण होता है। जिसमें सभी व्याकरणिक, कोशिय, अल्गोरिथमिक एवं भाषिक संसाधनों की प्रक्रिया के अंत में जो पाठ निकल कर

आता है उसे हम आगत पाठ कहते है। 14.निर्गत पाठ (Output)। इस प्रकार मशीनी अनुवाद की सामान्य प्रक्रिया चलती रहती है। इन साधनों के अतिरिक्त अन्य कई साधन भी इस प्रक्रिया में छोटे स्तरों पर कार्य करते हैं। इसलिए यह कार्य भी महत्वपूर्ण होता है।



(Graph No.1[7] ref. by International Machine Translation)

म.अनु. के सभी साधनों (Tools) का शब्द संचयन(Lexical Data) से अति निकटतम संबंध होता है। जिसके माध्यम से किसी भी पाठ को स्रोत भाषा से लक्ष्य भाषा तक पहुँचाने में सहायता प्राप्त की जाती है। इस प्रक्रिया में होने वाली एक गलती संपूर्ण पाठ के अंदर समस्या निर्माण करती है। ऐसी समस्याओं को मशीन स्वयं संशोधित नहीं कर पाती है। भारत में कई म.अनु. यंत्र बनाये जा रहें हैं जैसे- सी-डॅक नोएडा द्वारा "आँग्ल भारती-मॅट", आई.आई.टी.हैदराबाद "अनुसारका", सी-डॉक पुणे द्वारा "मंत्र", आई.आई.टी.मुम्बई द्वारा "मात्रा", इन्फ़ोसेट द्वारा "अनुवादक" आदि, परंतु इनमें से किसी में भी अनुवाद की शुद्धता की दर शत-प्रतिशत नहीं है, यह भी इसकी एक प्रमुख समस्या है।

[7] S.Raman and Reday , N.Rama Krishna, Vol.No.7.jan-dec.1995, *Parallel MTS for Indian Languages.* Madras, IJT

1.2 भारत में मशीनी अनुवाद की आवश्यकता :- भारत एक बहुभाषिक देश है। जहाँ कई भाषाऐं एक साथ बोली और समझी जाती हैं। विशेषकर उस स्थिति में जहाँ 16 से अधिक प्रशासनिक भाषाएं हों वहाँ मशीनी अनुवाद यंत्र विकसित करना कम से कम कार्यालीय भाषाओं के लिए बहुत ही उपयोगी सिद्ध होगा। इसी के साथ भारतीय भाषाओं की सीमाओं को तोड़ने के लिए भी मशीनी अनुवाद की आवश्यकता है। जिससे प्राशासनिक कार्य अधिक गति से पूर्ण हो सके।

1.3 म.अनु. में स्रोत भाषा विश्लेषण (The Source Language Analysis) **की आवश्यकता।**

किसी भी प्रोग्राम(यंत्र) को बनाने के लिए कई भाषाई संसाधनों की आवश्यकता होती है जिनकी सहायता के बिना प्रोग्राम का कार्य पूर्ण नहीं हो सकता। मशीनी अनुवाद के प्रोग्राम को बनाने के लिए प्रा.भा.सं. (NLP) के कई भाषाई संसाधनों का उपयोग करना पड़ता है। यह भाषाई संसाधन निम्न रुप में देखे जा सकते हैं। (A) वाक(Recognition) (B) बोधन प्रक्रिया(Understanding), (C) जनन (Generation) (D) प्राकृतिक भाषा संसाधन की प्रमुख रुपावली (प्रा.भा.सं. Standard Paradigm) (E) प्रोक्ति विश्लेषक (Discourse Analyser) (F) अर्थ विश्लेषक (Semantic Analyser) (G) रुप विश्लेषक (Morphological Analyser) (H) स्रोत भाषा विश्लेषक (The Source Language Analyser), (I) वाक्य विश्लेषक (Syntactic Analyser) (J) लक्ष्य भाषा प्रजनन (Target Language Generation Content Delimitation) (K) अन्वादेशक(Anaphora) (L) वाक्यगत चयन (Syntactic Selection) (M) पाठगत संरचना (Text Structuring) (N) संदर्भगत आदेश (Constituent Ordering) (O) प्रतिफलन (Realization) (P) शब्दकोश चयन (Lexical Selection) इन भाषिक संसाधनों का उपयोग स्रोत भाषा विश्लेशण के लिए किया जाता है। इन सभी भाषिक संसाधनों के साथ अन्य सहायक संसाधनों की भी आवश्यकता होती है। जिसका निर्माण प्रा.भा.सं. में किए जाने वाले भाषिक अनुसंधानों के बाद हो पाता है, इसलिए म.अनु. के निर्माण में प्रा.भा.सं. की सबसे अधिक आवश्यकता है।

2. मशीनी अनुवाद की समस्याएँ।

अनुवाद मूल रुप से एक बौद्धिक प्रक्रिया है। जिसकी क्षमता केवल मनुष्य के पास है अत: कोई भी संगणक पूरी तरह मानव बुद्धि की जगह नहीं ले सकता। "किसी एक भाषा से दूसरी भाषा में अनुवाद की प्रक्रिया मात्र शब्दों का अनुवाद नहीं है बल्कि यह एक विचार प्रक्रिया है, वाक्य, शब्द और अर्थ के

स्तर पर प्राकृतिक भाषाओं में व्याप्त संदिग्धता को समझ कर जब तक रीतिबद्ध नहीं कर लिया जाता तब तक मशीनी अनुवाद की कल्पना भी नहीं की जा सकती।"[8] मशीनी अनुवाद में एक ओर प्राकृतिक बुद्धि और कृत्रिम बुद्धि की प्रक्रियाओं के मूल तत्वों का समावेश होता है दूसरी ओर संप्रेषण के सिद्धांतों, व्याकरण के नियमों और तर्क विज्ञान (logic science) के सिद्धांतों का भी समावेश होता है। मशीनी अनुवाद की समस्याओं को प्रमुख रुप से चार भागों में विभाजित किया जा सकता हैं।

2.1. अनुवाद पूर्व समस्याएँ (pre -Translation Problems)

2.2. अनुवाद की समस्याएँ (Translation Problems)

2.3. अनुवादोत्तर समस्याएँ (Post-Translation Problems)

2.4. अनुवादक की समस्याएँ (Translator/User problems)

मशीनी अनुवाद की समस्याएँ मशीनी अनुवाद में होने वाले घटकों के अकार्यक्षम व्यवहार के कारण उत्पन्न होती हैं। ये समस्यायें कई बार मशीन में होने वाले प्रोग्राम के कारण उत्पन्न होती हैं, तो कई बार मानव निर्मित होती हैं जिन्हें हम निम्न रुप से देख सकते हैं।

2.1. अनुवाद पूर्व समस्याएँ

यह समस्याएँ मशीनी अनुवाद का कार्य करने से पहले आने वाली समस्याएँ होती हैं। जैसे- भारत एक बहुभाषिक देश होने के बावजूद भारत की अधिकांश जनसंख्या निम्न आय रेखा में रहती है जिसमें कई लोग अनुवाद को अपना व्यवसाय चुनते हैं। ऐसे लोगो के लिए म.अनु. को आर्थिक रुप से उपयोग में लाना कठिन है। जैसे :- "MT is a very expensive endeavor, both in terms of the software development effort required and in terms of the linguistic resources which need to be assembled.[9] दूसरी समस्या कंप्यूटर की सीमित उपलब्धता, उच्च कोटि के या तेज गति से कार्य न करने वाले संगणक हैं। इस परिस्थिति में उपयोग कर्ता/कर्मचारी को संगणक की जानकारी न होना

---

[8] मल्होत्रा, विजय कुमार 2007, *कंप्यूटर के भाषिक अनुप्रयोग*, वाणी प्रकाशन, 21-ए दरियागंज, नयी दिल्ली, पृष्ठ संख्या-73

[9] Gey, Fredric C., 2002, *Prospects for Machine Translation of the Tamil Language*, University of California, Berkeley

या अनुवाद यंत्रों के संबंध में संपूर्ण जानकारी का अभाव भी एक अनुवाद पूर्व समस्या ही कही जा सकती है।

2.2 अनुवाद की समस्याएँ अनुवाद की समस्या प्रमुख रुप से अनुवाद की प्रक्रिया से जुड़ी हुई है। जिसे हम पुन: तीन भागों में विभाजित कर सकते हैं।

2.2.1 भाषा वैज्ञानिक समस्याएँ [10] (Language problems)

2.2.2 व्याकरणिक समस्याएँ (Grammar problems)

2.2.3 भाषिक संसाधनों के उपयोग और निर्माण की समस्याएँ (for use and development of language tool)

2.1 भाषा वैज्ञानिक समस्याएँ :- मशीनी अनुवाद में भाषा वैज्ञानिक समस्याओं (Linguistic problems In म.अनु.) को Harold Somers ने निम्न रुप से विभाजित किया है जिनमें से "भाषा विज्ञान की अधिकतर समस्याओं पर विस्तृत रुप से विचार होना चाहिए।

2.1.1 समस्या विश्लेषक (प्राकृतिक भाषा संसाधन के लिए) (problems Analysis)

(Apply to all NLP) (क) शाब्दिक(Lexical) (ख) वाक्यगत(Syntactic), (ग) अर्थगत(Semantic)

2.1.2. म.अनु.में व्यतिरेकी समस्याएँ. (Contrastive problems In MT)

(क) शब्दगत (Lexical), (ख) संरचनागत (Structural), (ग) संदर्भपरक (Representational), (घ) शैली और प्रोक्तिपरक(problems in Style/Register)

(ड) अर्थगत भिन्नता (conceptual differences), (ढ) शाब्दिक अस्पष्टता (Lexical gaps)

1. Staructural divergence linked to lexical differences

2. Structural divergence linked to grammatical differences

3. Level shift: - Similar grammatical meanings conveyed by Different devices

2.1.3 शाब्दिक विश्लेषण(Lexical Analysis)

---

[10] Harold Somers, summer School 8-9 December 2003, *Machine Translation-1- , -2, CCL, UMIST,* Manchester M60, IQD England, Australian Language Technology, Australia

2.1.3.1 विखंडिकरण प्रक्रिया (Segmentation)

2.1.1.1 रुप विज्ञान (In Morphology)

1. प्रकार्यात्मक रुपविज्ञान(Functional Morphology)

2. व्युत्पादक रुपविज्ञान (Derivational Morphology)

3. द्विअर्थी रुपविज्ञान(Ambiguous Morphology)

2.1.3.3 अपरिचित शब्द (Unknown words):- Misspelled, spelling, not in Dictionary, because it a regular derivation, proper name, Compound

2.1.3.4 शाब्दिक अस्पष्टता (Lexical Ambiguity) :-

"(1) संरचनात्मक (Structural) बनाम (असंरचनात्मक) Unstructured

(2) (वास्तविक)Real बनाम (संयोगी)Accidental

(3) (सीमित)Local बनाम वैश्विक(Global)

(4) विश्लेषित Analytical (ड) वैश्विक अस्पष्टता Global Ambiguity

(5) सतही(Shallow) बनाम गहन अस्पष्टता (Deep Ambiguity) : -

अन्वादेशक Anaphora, (उदय)Raising, (कारक)Case-

रुपरेखा आधारित अस्पष्टता(Frame ambiguity), (परिणाम सूचक)

Quantifier and कर्ता आधारित (operator scope)"[11]

2.1.3.5 अस्पष्ट कोटियाँ (Category Ambiguities):-

1) समान उच्चारण वाले शब्द (Homonymy):- (समध्वन्यात्मक शब्द) Homophones, Homographs (समान लिपि वाले शब्द) (proper names : - many proper names form Homonyms with meaningful words)

2) समोच्चारित शब्द :- (Polysemy) ये भाषिक समस्यायें मानव द्वारा किये जाने वाले

[11] Harold Somers, summer School 8-9 December 2003, *Machine Translation-1- , -2, CCL, UMIST*, Manchester M60, IQD England, Australian Language Technology, Australia

अनुवाद में भी समान रुप से दृष्टिगत (निर्मित) होती हैं।

2.2 व्याकरणिक समस्याएँ - मशीनी अनुवाद के लिए जिस प्रकार के व्याकरणिक प्रारुप की जरुरत होती है वह सामान्यत: परंपरागत व्याकरणिक प्रारुप से भिन्न होता है। जिससे मशीनी अनुवाद के प्रकारो के अनुसार म.अनु. के उपयोग में व्याकरण के प्रकार अलग-अलग दिखाई देते हैं। जिसमें से कुछ इस प्रकार हैं "FUG (kAY-1984), HPSG(Polard and Sag-1994), LFG(Bresnsn-1982), TAG (Joshi and Schables-1992), Panini Grammar"[12] समस्याओं की इन कोटियों में भाषिक और व्याकरणिक समस्याएँ सबसे अधिक जटिल होती हैं।

2.3 भाषिक संसाधानों के उपयोग और निर्माण की समस्याएँ - किसी एक म.अनु. के निर्माण में लाए गए सभी भाषिक संसाधन (Language Tools (LTs)) दूसरी भाषा के मशीनी अनुवाद यंत्र निर्माण के उपयोग में आयेंगे यह निश्चित रुप से नहीं कहा जा सकता। भाषिक संसाधनों की यह समस्या भाषाओं की असमान संरचना के कारण उत्पन्न होती है। यह निर्माण कार्य कई जगह होता है। पाश्चात्य देशों में यह कार्य काफी तेजी से हो रहा है।

2.3.अनुवादोत्तर समस्याएँ

कुछ ऐसी समस्याएँ भी होती है जो अनुवादोत्तर तो होती हैं लेकिन अप्रत्यक्ष रुप से उनका संबंध अनुवाद से ही होता है जिसे हम उपेक्षित (ignore) नहीं कर सकते। म.अनु. के संदर्भ में प्रयुक्त सभी भाषिक संसाधन; किसी अन्य (दूसरी) भाषा के मशीनी-अनुवाद-यंत्र के निर्माण में प्रयुक्त होंगे यह निश्चित नहीं कहा जा सकता। किसी भी मशीनी अनुवाद के निर्माण में आने वाली समस्याएँ केवल अनुवाद से ही जुड़ी नहीं होती अपितु अन्य बहुत सी समस्यायें भी होती हैं, जैसे -

2.31.1 यांत्रिक (कंप्यूटर) समस्याएँ

2.3.1.2 अभिकलनात्मक भाषाविदों की अनुपस्थिति की समस्या

2.3.1.3 संगणकी कृत भाषा और प्रोग्राम को समझने की समस्या

---

[12] Dr. Jha, Girish N., *Computational Linguistics and Sanskrit Computational Linguistics*, Special Centre for Sanskrit Studies, JNU,N.Delhi-67

मशीन द्वारा मिलने वाला (output(OP)) निर्गत पाठ (Formed language) कृत्रिम भाषा में मिलता है। जिसे मानवीय भाषा में लाकर व्यवस्थित रुप प्रदान करने के लिए मशीनी अनुवाद में कोई भाषिक साधन उपलबद्ध नहीं है।

2.4. अनुवादक की समस्याएँ

प्रत्येक समस्या का प्रभाव अनुवाद पर होता है लेकिन कुछ समस्याएँ मूल रुप से अनुवादक/प्रयोग कर्त्ता की होती हैं। अनुवाद पूर्व या अनुवाद के बाद निर्गत पाठ में कोई समस्या उत्पन्न हो जाने पर अनुवाद यंत्र ठीक करने के लिए म.अनु. का उपयोग कर रहे उपभोक्ता को मशीन में सुधार के लिए मशीनी अनुवाद यंत्र निर्माण करने वाली संस्था से संपर्क करना पड़ेगा। समस्या किसी भी प्रकार की हो प्रयोग कर्त्ता स्वयं अनुवाद यंत्र ठीक नहीं कर सकता। इन समस्याओं के साथ मशीन ही नहीं अनुवादक/उपयोग कर्ता (Translator/user) भी जूझता है। मशीनी अनुवाद में समस्यायें तो बहुत हैं जिन पर एक विस्तृत चर्चा की आवश्यकता है। मशीनी अनुवाद की समस्याएँ भाषाओं के साथ कुछ भाषिक कारणों से बदल जाती हैं, लेकिन इनकी मात्रा कुछ हद तक कम होती है।

पाश्चात्य देशों मे संगणक भारत से कई वर्ष पूर्व आया जिसका उपयोग पाश्चत्य भाषा वैज्ञानिक और संगणक विज्ञानियों ने भली-भाँति समझ कर भाषा विश्लेषण के लिए संगणक का उपयोग किया। यह कार्य भारतीय मशीनी अनुवाद कार्य के शुरु होने से लगभग पच्चीस वर्ष पूर्व हुआ। जिसका परिणाम आज यह हुआ है कि भारत में मशीनी अनुवाद पर कार्य कर रहे विद्वान पाश्चात्य मशीनी अनुवाद यंत्रों को स्रोत यंत्रों के रुप में देखते हैं।

3. वैश्विक स्तर पर मशीनी अनुवाद की समस्याएँ

वैश्विक स्तर पर साहित्य, काव्य, ज्ञान-विज्ञान के क्षेत्र आदि में भी मशीनी अनुवाद की आवश्यकता है परन्तु कई मशीनी अनुवाद विज्ञानियों ने पहले ही स्पष्ट कर दिया है कि साहित्य, कविता जैसे पाठ का अनुवाद मशीन के द्वारा करना संभव नहीं है। यह समस्या वैश्विक स्तर की सबसे बड़ी समस्या है। साथ ही वैश्विक स्तर पर कुछ ऐसी समस्याएँ भी हैं जो विश्व की सभी मशीनी अनुवाद यंत्रो में समान रुप से पाई जाती हैं। इन समस्याओं का निराकरण अब तक नहीं हो पाया है। "शब्दों के पीछे निहित व्यंग्यार्थ, ध्वनि व्यंजना और अस्पष्ट/समान उच्चारण वाले शब्दों को समझने और उन्हें पुनः प्रस्तुत करने में आज कोई भी यंत्र सक्षम नहीं है। किसी भी पाठ में अपूर्णता (incompleteness)

लेखक यह मानकर विवरण छोड देता है कि पाठक इनका अनुमान लगा सकेगा। ऐसे वाक्य मशीन अनुदित नहीं कर सकेगी।"[13] क्योंकि अभी कंप्यूटर में अपूर्ण(incomplete) वाक्यों को समझने की क्षमता विकसित नहीं हुई है।

गूगल ने १८ भाषाओं में अनुवाद करने वाला ऑनलाइन सॉफ्ट्वेअर तैयार किया है। जिसमें भाषाओं की मात्रा (quantity) तो है लेकिन गुणवत्ता(quality) नहीं है। यह अनुवाद यंत्र मात्र एक प्रकार से स्रोत भाषा का लक्ष्य भाषा में शब्दांतरण है। कुछ अनुवादक तो इसे शब्दांतरण भी नहीं मानते महज एक भाषा की शब्द आकृति को दूसरी भाषा की शब्द आकृति के समकक्ष रखने का काम मात्र मानते है। जिसमें एक शब्द के भाव का दूसरे शब्द के भाव के साथ किसी प्रकार का संबंध नहीं होता। किसी भी भाषा में अर्थ शब्दों के संबंधों पर निर्भर करता है जिससे बोले गए या लिखे गए वाक्यों का अर्थ स्पष्ट होता है। इसकी गुणवत्ता गूगल मशीनी अनुवाद यंत्र के पास नहीं है। यह समस्या केवल गूगल में प्रस्तुत म.अनु. की ही नहीं बल्कि वैश्विक स्तर पर उपयोग में लाए जा रहे कई मशीनी अनुवाद यंत्रों की हैं। केवल तेज गति से एक भाषा के शब्दों को दूसरी भाषा के शब्दों की जगह रखना अनुवाद नहीं कहलाता। गूगल की ये समस्याएँ सामान्य पाठकों के लिए भी एक वैश्विक मुद्दा बन गयी हैं।

भाषिक रुप से देखने पर वैश्विक स्तर पर प्रमुख रुप से निम्न समस्याएँ मिलती हैं। 1) अनुवाद में उच्चारण और लिखित क्रिया में पारस्पारिक संबंध 2) खंडित पद और पदांश 3) उच्च और निम्न प्रकार्य (High function and low function term) 4) अस्पष्टता (Vagueness) 5) अन्वादेश (Anaphora) 6) संहिता (juncture) 7) बोलचाल की भाषा- मुहावरे और अशिक्षितों की भाषा/गवांरु बोली 8) कोड-परिवर्तन 9) भाषिक प्रकार्य 10) अनुवाद इकाइयों का विभाजन (Segmentation of unit) आदि। यह भी विचार किया जा रहा है कि स्त्रोत भाषा और लक्ष्य भाषा के बीच एक मध्यस्थ(mediator language) हो वह "आदर्श मध्य भाषा भी ऐसी व्युत्पत्ति परक (Derivational) भाषा हो जिससे समस्त स्रोत भाषाओं की अभिव्यक्ति का अर्थ निरुपित (Represent) किया जा सकें।[14] इसपर भी अबतक किसी ने अनुसंधान परक कार्य नहीं किया है और न ही आदर्श मध्य भाषा

---

[13] ओम, डॉ.विकास, (संपादक डॉ.नगेंद) 1995, *अनुवाद विज्ञान सिद्धांत और अनुप्रयोग मशीनी अनुवाद की समस्याएँ,* हिंदी कार्यान्वय निदेशालय,दिल्ली विश्वविज्ञाय,पृष्ठ सं.340

[14] डॉ. शर्मा, प्रमोद, अनुवाद पत्रिका, अप्रैल-जून 2004, *मशीनी अनुवाद की समस्याएँ और संभवनाऐं,* अनुवाद पत्रिका(कंप्यूटर विषेशांक-२)

का निर्माण हुआ है। प्रमुख रुप से वैश्विक स्तर पर ये समस्याएँ सामने आई हैं। इसी के साथ ही व्याकरण परक एवं भाषिक समस्याएँ भी सामने आई हैं, जो विश्व के सभी अनुवाद यंत्रों में समान रुप से पाई जाती हैं।

4. भारत में मशीनी अनुवाद की समस्याएँ

भारत में संगणक की साक्षरता अभी पूर्ण रुप से नहीं हो पाई है, आज भी कई कार्यालयों में संगणक का उपयोग कम मात्रा में किया जाता है। इसी का नतीजा है कि भारत में सबसे पहले अनुवादक को अनुवाद सॉफ्टवेर की जटिलताओं और संगणक कार्य शैली की कुशलता से ज्ञात कराना पड़ता है। उन्हें यह भी ध्यान रखना होता है कि उपयोग कर्त्ता को विंडोज और "लिनिक्स" संगणक प्रणालियों के दोनों परिवेशों से ज्ञात कराना है। क्योंकि अनुसारका लिनक्स परिवेश में विकसित किया गया है यदि इन्हें विंडोज के परिवेश में भी विकसित किया जाता तो प्रयोक्ता उसका और अधिक लाभ उठा पाते।[15] म.अनु. निर्माण कर्त्ताओं की इस गलती के कारण म.अनु. उपयोग कर्त्ताओं (MT users) को कई कठिनाईयाँ उठानी पडती हैं। अनुसंधान की पद्धतियों पर विचार किया जाए तो "दूसरे देशों में वैयक्तिक अनुसंधान भी होता है भारत में मशीनी अनुवाद के क्षेत्र में वैयक्तिक अनुसंधान नहीं होता इसके फलस्वरुप भारत जैसे देश में जहाँ कुल आबादी के ३% से भी कम लोग अंग्रेजी जानते हो (Sinha and Jain, 2003) वहाँ आवश्यकता है कि अंग्रेजी से भारतीय स्थानीय भाषा के लिए मशीनी अनुवाद यंत्र निर्माण किया जाये।" "Private research also in other countries but not in India. Consequently, in a country like India, where English is understood by less than % of the population (Sinha and Jain, 2003), the need for developing MT systems for translating from English into some native Indian languages is very acute.[16]

भारत एक बहुभाषी देश होने के कारण भारत में मशीनी अनुवाद की संभावनाओं को नकारा नहीं जा सकता। भारत में मशीनी अनुवाद की प्रक्रिया को अधिक गति प्रदान करना इसलिए भी महत्वपूर्ण है कि पाश्चात्य भाषाऐं भाषा प्रौद्योगिकी(Language Technology) के क्षेत्र में तेजी के साथ प्रगति कर रहे हैं। भारतीय भाषाओं को भी प्रौद्योगिकी एवं जागतिक स्तर पर अपनी भाषा की अस्मिता को बनाए रखने के लिए मशीनी अनुवाद में अधिक विकसित होना अत्यावश्यक है। जिससे भारतीय भाषाओं

---

[15] प्रो. दास, ठाकुर,अप्रैल-जून 2004, *मशीनी अनुवाद:भारतीय परिदृश्य अनुवाद पत्रिका* (कंप्यूटर विषेशांक-२)

[16] Gupta, Deepa, 2005, *Contributions to English to Hindi Machine Translation Using Example-Based Approach. (Ph.D. Thesis)*Department Of Mathematics, IIT, Delhi, India

का प्रचार-प्रसार विश्व की अन्य भाषाओं की तरह समान रुप से चलता रहे। जिससें सभी भाषाओं का संतुलन बना रहे। किसी एक भाषा के अधिक विकसित होने या उपयोग में लाने से अन्य भाषाओं के विकास में बाधा तो पहुंचती ही है और कभी-कभी कम संख्या में बोली जाने वाली भाषा समाप्त भी हो जाती है। भारत के मशीनी अनुवाद यंत्रों की स्रोत भाषाओं पर विचार किया जाए तो भारत में अधिकतर मशीनी अनुवाद का कार्य अंग्रेजी से भारतीय भाषाओं में शुरु हुआ। अंग्रेजी भाषा की संरचना भारतीय भाषाओं से भिन्न होने के कारण कई समस्याओं का सामना करना पड़ा। आज भी मशीनी अनुवाद का प्रमुख कार्य अंग्रेजी से भारतीय भाषाओं के मशीनी अनुवाद यंत्रों पर ही होता है। संभवता इस संकल्पना में बदलाव की गुंजाइश दिखाई देती है। कुछ विद्वान भारत की सभी भाषाओं के लिए म.अनु. की संकल्पना करते हैं लेकिन भारत के सभी भाषाओं के लिए अनुवाद यंत्र निर्माण कर पाना आज की परिस्थिति में संभव नहीं है। To our knowledge there is no direct machine translation software development being done on Indian subcontinent languages – languages spoken by nearly twenty percent of the world's people. The reasons are simple, the resources available for traditional machine translation are simply non-existent and the commercial viability of such development seems dim.[17] However, the new field of research into statistical machine translation offers hopes that Tamil translation may be just around the corner, to the pleasure and social benefit of all. अब तक भारतीय उपमहद्वीप की भाषाओं के लिये कोई भी पूर्णत: मशीनी अनुवाद यंत्र निर्माण नहीं हो पाया है जबकि २०% लोग इन भाषाओं को बोलते हैं। इसका एक कारण यह भी है कि पारंपरिक मशीनी अनुवाद यंत्र में उपयोग किए जाने वाले भाषिक संसाधन लगभग अनुपलब्ध है, और ऐसे प्रयोगों की व्यावसायिक उपयोगिता बहुत ही सीमित है। तदोपरांत मशीनी अनुवाद में नए रुप से हो रहे अनुसंधान से आशा की जा सकती है कि तमिल अनुवाद यंत्र सामाजिक रुप से भी महत्वपूर्ण काम करेगा। "Two major problems exist in connection with machine translation and cross-language retrieval of Tamil (and other Indian languages). First is the lack of machine-readable resources for either machine translation or cross-language dictionary lookup. Such dictionaries need to be pared of poetic and Classical terminology and augmented with modern words as found in recent newspaper and Magazine texts. Tamil is a phonetic language. By default each Tamil consonent is followed by an 'a' vowel sound. The vowel sound is modified by the addition of glyphs, such as the curlicue following the 'm' which changes it sound

[17] Gey, Fredric C., 2002. *Prospects for Machine Translation of the Tamil Language*. University of California, Berkeley

from 'ma' to 'mi'. To suppress the vowel sound one places a dot over the consonant. This means that borrowed words from English or other languages will 'sound' similar in Tamil to their native language (Malten 1996)."[18] तमिल मशीनी अनुवाद यंत्र पर विचार किया जाए तो तमिल (और अन्य भारतीय भाषाओं के) मशीनी अनुवाद यंत्र में दो प्रमुख समस्याएँ दिखाई देती है। सबसे पहले कई मशीन साधित भाषाई संसाधन मशीनी अनुवाद के लिए या अंतर-भाषाओं के लिए कोश। कुछ कोशों में काव्यात्मक और शास्त्रीय तकनीक के साथ नए विकसित शब्द हों जो समाचार पत्रों और पत्रिकाओं से लिए जाएँ। तमिल एक स्वनिम आधारित भाषा है, नियमानुसार प्रत्येक तमिल व्यंजन के बाद एक 'अ' स्वर आता है। इस स्वर में विशेष आकृति द्वारा परिवर्तन किया जाता है। जैसे कि 'म' के बाद लगने वाली मात्रा 'मा" ध्वनि को 'मि' में परिवर्तित कर देती है। इस स्वर को विलोपित करने के लिए व्यंजन पर एक बिंदी लगाई जाती है। अर्थात अंग्रेजी या अन्य भाषाओं से लिए गए शब्दों का उच्चारण तमिल में भी उनके मूल उच्चारण के समान ही होगा। इस क्षेत्र की एक समस्या कोशीय-संसाधनों का अभाव है। इस दिशा में एकल कारपोरा की शब्दवर्ग अंकक(टैगिंग) एवं एनोटेशन समानांतर कारपोरा का विकास, पदबंध कोशों का निर्माण, शब्दजाल विकास आदि कार्य किए जाने की आवश्यकता है। तमिल द्रविड़ भाषा परिवार की भाषा होने के बावजूद आर्य भाषा परिवार की भाषाओं के साथ महत्वपूर्ण संबंध रखती है। हिंदी-तमिल एवं तमिल-हिंदी मशीनी अनुवाद यंत्रों पर काम किया जा रहा है इसलिए तमिल मशीनी अनुवाद यंत्र में होने वाली समस्याओं को नजर अंदाज नहीं किया जा सकता। यह कार्य मूलत: भाषापरक कार्य है अत: इसमें भाषा विज्ञानियों, कोश विज्ञानियों, संगणक विज्ञानियों और अनुवाद विज्ञानियों आदि की प्रमुख भूमिका है। "अधिकतर अनुवाद यंत्र सरकारी वित्तपोषण से विकसित किए गए है उनमें भाषा-विशेषज्ञों का सहयोग अधिक परिलक्षित होता है। "यद्यपि इसका दोष भाषा विशेषज्ञों के ही मत्थे मढ़ा जाता है। किंतु वस्तुत: स्थिति यह है कि कंप्यूटर विज्ञानी भाषा विशेषज्ञों का सहयोग लेने से नकारते हैं।"[19] एक दूसरे पर आरोप-प्रत्यारोप से अधिक, सहयोग की आवश्यकता है। इन कारणों से भी भारतीय म.अनु. की दुर्गति दिखाई देती है। भारत में म.अनु. की एक समस्या यह भी है कि कई अंग्रेजी से हिंदी अनुवाद यंत्रों पर एक साथ कई संस्थानों में कार्य हो रहा है और हर कोई अपने यंत्र को अधिक उपयुक्त एवं अधिक प्रगत मान रहा है, जबकि अब तक कोई भी

---

[18] Gey, Fredric C., 2002, *Prospects for Machine Translation of the Tamil Language*, University of California, Berkeley

[19] वहीं - पृष्ठ सं.- ४३

मशीन पूर्ण रुप से अनुवाद के लिए सक्षम नहीं है। हालांकि इन अनुवाद यंत्रों के निर्गत पाठ के प्रतिशत में अंतर जरुर है।

भारतीय भाषाओं के मशीनी अनुवाद यंत्रों में निर्माण होने वाली कुछ मुख्य समस्याओं को निम्न रुप से देखा जा सकता है। 1. भाषिक संरचना 2.वाक्यगत भाषिक अस्पष्टता 3.स्रोत भाषा का अधिक विकसित होना 4.संदर्भ परक अस्पष्टता, 5. मूल पाठ के छोटे एवं लंबे पदों के शून्य अनुवाद की समस्या 6.अलंकारों एवं संकेतों के अनुवाद की समस्या, 7.विकसित नए शब्दों के अनुवाद की समस्या, 8.संज्ञा और सर्वनाम के शून्य अनुवाद की समस्या 9.वाक्य, शब्द एवं प्रोक्ति की अपूर्णता की समस्या 10.शब्दगत संबंध की समस्या। इन समस्याओं का मूल कारण है भाषिक संसाधनों का पूर्ण विकसित न होना। जिसके चलते किसी भी एक भाषिक संसाधन के द्वारा होने वाली गलती के कारण अनुवाद यंत्र में कई समस्याएँ उत्पन्न होती हैं।

बावजूद इतनी सारी समस्याओं के मशीनी अनुवाद पर यह बार-बार आक्षेप लगायाजाता है कि कोई भी मशीनी अनुवाद यंत्र सौ प्रतिशत शुद्धता का दावा नहीं कर सकता। मशीनी अनुवाद की प्रक्रिया सफल करने के लिए म.अनु. की प्रक्रिया को मानव अनुवाद की प्रक्रिया के सोपानों में ढाला जाए। जिससे गलतियों का प्रतिशत कम हो सके। वर्तमान कंप्यूटर प्रणालियां अनेक गलतियाँ करती हैं जो मानव अनुवाद में काफी आसान होती हैं। म.अनु. द्वारा की गई छोटी-छोटी गलतियों का मशीन स्वयं पता नहीं लगा सकती है। जिसके कारण निर्गत पाठ(Output) बगैर वाक्य संरचना या अनुवाद की गलतियाँ ठीक किए बगैर मिलता है, जो एक अशुद्ध पाठ होता है।

इस क्षेत्र की एक समस्या कोशीय संसाधनों का अभाव है। इस दिशा में एकल कारपोरा की टैगिंग एवं एनोटेशन समानांतर कारपोरा का विकास, पदबंध कोशों का निर्माण, वर्डनेट विकास आदि कार्य किए जाने की आवश्यकता है। मशीनी अनुवाद की असफलता का मुख्य कारण यह भी है कि अब तक ऐसी कोई तकनीक उपलब्ध नहीं थी, जिसकी सहायता से अर्थगत और संदर्भगत संदिग्धताओं को स्पष्ट किया जा सके किंतु कृत्रिम बुद्धि (Artificial Intelligence) के अंतर्गत ज्ञान निरुपण(Knowledge Representation) की ऐसी विधियां विकसित की जा रही हैं, जिसकी सहायता से प्राकृतिक भाषाओं

TH-17822

में सन्निहित वाक्यपरक, अर्थपरक और संदर्भपरक प्रक्रियाओं को कंप्यूटर मनुष्य से भी अधिक अच्छी तरह समझ सकेगा।[20] जो म.अनु. को सफल बनाने में मदद करेगा।

बीसवीं सदी के अंत के दशकों में मशीनी अनुवाद ने हमारे दृष्टि पटल पर एक अमिट छाप छोडी है, जिसकी कल्पना ने ही हमें अब तक मशीनी अनुवाद से बांधे रखा है। संगणक की उक्त सीमाओं पर अनुसंधान कर उन्हें धीरे-धीरे दूर किया जा सकता है। इसके लिए अनेक भाषाओं पर एक साथ काम करने के बजाए एक ही भाषा पर कई विद्वानों को काम करना चाहिए। उस भाषा में मिलने वाली प्रमुख समस्याओं को खंडों में बांट कर या दूसरे शब्दों में कहा जाए तो समस्याओं का खंडीकरण कर बड़ी समस्या को छोटी-छोटी समस्याओं में विभाजित कर छोटी समस्याओं का ऐसा समाधान खोजा जाए जो अंत में मिलकर बड़ी समस्या के समाधान के रुप में सामने आए। इस प्रकार एक-एक भाषाओं पर काम कर मशीनी अनुवाद यंत्रों का निर्माण किया जाए तो भाषाओं के साथ ही उस भाषा को बोलने वाले समाज का भी विकास होगा जैसे तमिल मशीनी अनुवाद यंत्र के संबंध के लिए कहा जाता है। "Building translation system From and to Tamil helps the Tamil community all over the world in accessing the information in Tamil." तमिल भाषा में और तमिल भाषा से म.अनु. का विकास कर संपूर्ण तमिल समाज विश्व से सूचना और ज्ञान-अर्जन का कार्य करेगा। यह कार्य मशीनी अनुवाद केवल तमिल के लिए ही नहीं बल्कि भारत की सभी भाषाओं के लिए लागू होगा जिसके लिए म.अनु. की समस्याओं को दूर करना आवश्यक है। अंत में म.अनु. के समाधान और समस्याओं के संबंध में फेडरिक सी.गेए(Fredric C. Gey)म.अनु. के संसाधनों पर चर्चा के उपरांत यही कहते है कि "मशीनी अनुवाद के संसाधनों को और विकसित करने की अत्यावश्यकता है।" "After having discussed the various components of an MT system, and the resources that might be needed to be build for MT "[21]

---

[20] कंप्यूटर के भाषिक अनुप्रयोग – विजय कुमार मल्होत्रा – पृष्ठ संख्या - ७५

[21] Gey, Fredric C., 2002. *Prospects for Machine Translation of the Tamil Language*, University of California, Berkeley

दूसरा अध्याय

समान उच्चारण वाले शब्दों से अनुवाद एवं मशीनी अनुवाद में होने वाली समस्याएँ

दूसरे अध्याय का सार

1. समान उच्चारण वाले शब्द का परिचय और परिभाषा

2. अन्य भाषाओं में मिलने वाले समान उच्चारण वाले शब्दों के उदाहरण

3. समान उच्चारण वाले शब्दों के प्रकार एवं व्याकरणिक सूत्र

4. मानव द्वारा होने वाले अनुवाद एवं मशीनी अनुवाद में समान उच्चारण वाले शब्दों की समस्याएँ

    4.1 मानव द्वारा होने वाले अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं के उदाहरण

    4.2 मशीन द्वारा होने वाले अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं के उदाहरण

दूसरे अध्याय का सार

दूसरे अध्याय में अनुसंधान के प्रमुख विषय समान उच्चारण वाले शब्दों के संदर्भ में विस्तृत चर्चा की गई है। प्रथमत: समान उच्चारण वाले शब्द किस प्रकार निर्माण होते है, इसपर प्रकाश डाला गया है। समान उच्चारण वाले शब्द लगभग सभी भाषाओं में समान रुप से पाए जाते हैं। इन शब्दों का प्रत्येक भाषा में किस प्रकार उपयोग होता है और म.अनु. के निर्माण में होने वाली यह समस्या भाषिक अस्पष्टता की किस कोटि में प्रविष्ट होती है इस पर भी चर्चा की गई है। भाषा में निर्माण होने वाली द्विअर्थकता को म.अनु. के निर्माण में होने वाली समस्या के साथ जोड़ा गया है। समान उच्चारण वाले शब्दों पर विस्तृत चर्चा करने के उद्देश्य से इन शब्दों का विभाजन किया गया है। इसी के साथ समान उच्चारण वाले शब्दों के सामान्य व्याकरणिक सूत्रों पर भी चर्चा की गई है। जिसमें समान उच्चारण वाले शब्दों के प्रकारों के कारण मानव द्वारा किये जाने वाले अनुवाद एवं मशीनी अनुवाद में किस प्रकार समस्याएँ उत्पन्न होती हैं और मानव इन समस्याओं का कैसे समाधान निकालता है, इस पर विचार किया गया है। हिंदी से मराठी में अनुवाद का अब तक कोई मशीनी अनुवाद यंत्र पूर्ण रुप से निर्मित नहीं हो पाया है। इसलिए अंग्रेजी-हिंदी मशीनी अनुवाद यंत्रों में समान उच्चारण वाले शब्दों को देकर मशीनी अनुवाद के निर्माण में होने वाली समस्याओं पर विचार किया गया है।

कुछ अनुवाद यंत्रों में शब्दों के निराकरण के लिए प्रयास किया गया है लेकिन यह काम नाही संपूर्ण है और नाही समाधान परक। मशीनी अनुवाद यंत्रों की संख्या भारतीय परिदृष्य में भी बढ़ी है। कुछ म.अनु. यंत्रों को लेकर भी इस अध्याय में चर्चा की गई है।

1. समान उच्चारण वाले शब्द का परिचय और परिभाषा

हिंदी एक अंतर्राष्ट्रीय भाषा है, हिंदी का भाषिक संपर्क अन्य भारतीय भाषाओं की तुलना में अधिक है। इस उपलब्धि के साथ ही हिंदी ध्वनिमूलक भाषा है। भारतीय एवं पाश्चात्य भाषाओं के शब्दों का उच्चारण हिंदी में ध्वनि के आधार पर करना संभव है। ध्वनिमूलकता की देन से हिंदी में आए कई पाश्चात्य एवं भारतीय परिवार की भाषाओं के शब्द अन्य भाषाओं की तुलना में अधिक पाये जाते हैं। ऐसे में सवाल यह उठता है कि क्या अन्य भाषाओं के शब्द हिंदी में आने से हिंदी के मूल शब्दों(धातुओं) पर इसका कोई असर नहीं होगा? क्या इन संभावनाओं को नहीं नकारा जा सकता कि दो हिंदीतर भाषाओं से आयातित शब्द समान उच्चारण वाले लेकिन भिन्न अर्थ एवं भिन्न सांस्कृति के हों? आज यह स्थिति पैदा हो गई है कि हिंदी में अन्य भाषाओं एवं हिंदी के मूल शब्दों में समान उच्चारण वाले शब्द भारी मात्रा में आ चुके हैं। कई विद्वानों का मानना है कि अन्य भाषाओं के शब्दों को हिंदी में स्थान देने से परहेज नहीं होना चाहिए। इससे हिंदी का विकास ही होगा और हिंदी की जीवंतता बनी रहेगी। लेकिन यह भी समझना आवश्यक है कि अगर हिंदी में कोई शब्द पहले से उपलब्ध हो तो उस शब्द के स्थान पर दूसरी भाषा से समान उच्चारण वाला शब्द हिंदी में न आये। ऐसे शब्दों को हिंदी में स्वीकृत करने से पूर्व उस शब्द को अस्पष्टता से मुक्त किया जाए। जिससे द्विअर्थकता का निर्माण न हो सके।

हिंदी देवनागरी लिपि में लिखी जाती है। "देवनागरी लिपि में वर्तनी की भिन्नता से उच्चारण में भी भिन्नता आ जाती है। जैसे - अविराम-अभिराम; सम-शम; सूर-शूर: कुल-कूल; आसन-आसन्न आदि।"[1] ऐसे समय पर कई बार लेखन की गलतियों से भी समान उच्चारण वाले शब्द उत्पन्न हो जाते हैं। यह गलतियाँ कभी-कभी व्यक्ति जान बूझकर करता है, तो कभी मनोरंजन के लिए या शब्द ज्ञान के दिखावे के लिए जान बूझकर करता है। लेकिन यह कार्य वही व्यक्ति कर सकता है। जो शब्दों के अर्थों को समझने में विद्वान हो, जिसकी शब्दों पर मजबूत पकड़ हो। "भाषा के विकास में एक अर्थ के लिए

---

[1] ओझा, डॉ.त्रिभुवन,1994, *हिंदी में अनेकार्थकता का अनुशीलन,* विश्वविद्यालय प्रकाशन, वाराणसी पृष्ठ सं.-76

एक ध्वनि संकेत का प्रयोग भाषा के विकास का पहला चरण माना जाता रहा है, उसी प्रकार लेखन के विकास में भी सर्व प्रथम एक संकल्पना के लिए एक चित्र का दृश्य प्रतीक के रुप में प्रयोग किया जाता रहा है।"[2] संगणक अक्षरों को चित्रों के रुप में ही समझता है। लेकिन इन चित्रों के आधार पर अर्थों को समझना संगणक के लिए कठिन कार्य है। इस कार्य के लिए संगणक को अन्य चित्रों का सहारा भी दिया जाता है। जैसे एक डेटा बेस में शब्दों के साथ समान अर्थ वाले शब्द, व्याकरणिक कोटि, शब्द संख्या क्रमांक आदि को भी स्पष्ट करना पड़ता है। स्पष्टीकरण के बाद भी जो समस्याएँ आती है, उन्हें दूर करना भाषा प्रौद्योगिकी के लिए सबसे कठिन कार्य है।

अक्षरों को यदि चित्र कहा जाए तो एकाधिक चित्रों से शब्द का निर्माण होता है। जब शब्दों का उपयोग पद के रुप में होता है, तो उन्हें अर्थ प्राप्त होता है और शब्दों के उपयोग के अनुसार शब्दों को व्याकरणिक कोटियाँ भी प्राप्त होती हैं। शब्दों को मिले अर्थों के अनुसार मानव शब्दों का उपयोग समाज में संप्रेषण के लिए करता है। यह प्रक्रिया तब तक समान रुप से चलती है जब तक संप्रेषण में किसी भी प्रकार की रुकावट नहीं आती लेकिन भाषिक संप्रेषण में शब्दों की अस्पष्टता के कारण गतिरोध उत्पन्न हो जाता है।

यह रुकावट एकाधिक भाषा होने से उत्पन्न होती है। मानव समाज में एकाधिक भाषाओं के होने से दो भाषाओं में संबंध स्थापित होते हैं और एक समाज में बोली जाने वाली भाषा के शब्दों का उपयोग दूसरी भाषा में होना आरंभ हो जाता है। एक भाषा से दूसरी भाषा में होने वाले इस संक्रमण के कारण दोनों भाषाओं पर विपरीत एवं लाभदायक दोनों प्रकार का प्रभाव दिखाई देता है। लाभ यह है कि एक विशिष्ट समाज में उपयोग किए जाने वाला विचार(concept), समाज व्यवस्था, संस्कृति, भाषा, शब्द, वेशभूषा किसी भी संदर्भ में दूसरे समाज में निर्माण होने की संभावना बढ़ जाती है।

आयातित शब्दों के कारण विपरीत परिणाम भी होते हैं। शब्दों का यह परिणाम हम सामाजिक दृष्टि से देख सकते हैं। भाषिक दृष्टि से विचार किया जाए तो इस क्षेत्र में दोनों प्रकार के परिणाम सामने आते हैं। प्रथमत: हम आयातित शब्दों के कारण नए शब्दों से

---

[2] प्रो.नारंग, वैष्णा,2006,*समसामायिक भाषा विज्ञान,* यश पब्लिकेशन,दिल्ली,सं.- नं. 87

परिचित होते है और भाषा की शब्द संख्या बढ़ती है। विपरीत दृष्टि से देखा जाए तो पहले से मौजूद शब्दों का लोप हो जाना और भाषा में द्विअर्थकता का निर्माण हो जाना प्रमुख। जैसे - हिंदी में एसे कई शब्द हैं जो अन्य भाषाओं से आये नए शब्दों के अधिक उपयोग के कारण कम उपयोग में आने लगे। जिसके कारण आज पहले से उपलब्ध शब्द लुप्त हो रहे हैं। हिंदी में ऐसे कई शब्द हैं जो पहले से हैं लेकिन अन्य भाषाओं से आये शब्दों के कारण हिंदी में द्विअर्थक हो गए है। जैसे - पास (समीप), पास(उतीर्ण), बटन(घुंडी), बस(नियंत्रण), बस(मोटर लॉरी), बस(केवल) आदि।

लेखन की दृष्टि से इन शब्दों पर विचार किया जाए तो यह शब्द चित्र रुप में समान दिखाई देते हैं। जब यह शब्द वाक्य में प्रयोग किए जाते हैं तो संगणक यह नहीं समझ सकता कि शब्द का उचित अर्थ क्या है? इस कार्य के लिए मशीन को एक भाषाई साधन की आवश्यकता होती है। शब्द की यह अवस्था केवल मशीन को ही नहीं; अपितु मानव को भी दुविधा में डाल देती है। मानव इन समस्याओं को दूर करने के लिए कोश का सहारा लेता है। लेकिन "शब्दों की समस्वनता (Homophony) और समवर्णता (Homography) उनके अर्थों में जो वैशिष्ट्य उत्पन्न करती है, उसे शब्दकोश में प्रविष्टियों के अर्थ-प्रकटन की एक समस्या के रूप में देखा जा सकता है।"[3] समान उच्चारण वाले शब्दों को भारतीय परिदृष्य में पाणिनि के युग से ही देखा जाता रहा है।

इस पर संस्कृत के कई विद्वानों ने कार्य किया है। एच.सी.अनंतनारायन के अनुसार "संस्कृत में 6वीं शताब्दी में पाणिनि ने सर्वप्रथम अपने ग्रंथ अष्टाध्यायी में समान उच्चारण वाले शब्दों पर चर्चा की है।" "in Sanskrit one comes across discussion on homonymy among the suffixes. Páṇini (6th cent.B.C) treats them in his Aṣṭádhyáyi"[4] संकृत में शब्दों पर चर्चा करने वाले अग्रणी व्याकरणकार भर्तृहरि रहें हैं। भारतीय लेखकों

---

[3] भाटिया, डॉ. कैलाशचंद्र, डॉ.युगेश्वर,डॉ. कपूर, बदरीनाथ, संपादक-डॉ.भोलानाथ तिवारी, डॉ.हरदेव बाहरी,1989, *कोश विज्ञान सिद्धांत और प्रयोग,* (आचार्य रामचन्द्र वर्मा जन्मशती ग्रंथ)

[4] Ananthanarayana, H.S. , Jan. 9-11, 1980, *Treatment of homonymy in Páṇini Aṣṭádhyáyi*,Papers presented at the Second International Conference on South Asian Languages and Linguistics, Dept. of Linguistics, Osmania University, Hyderabad, India page No.- 50

में भर्तृहरि(७वीं शताब्दी) ने सर्वप्रथम इस विषय पर व्यवस्थित रुप से विस्तृत चर्चा करने का काम किया। "Among the Indian writers, Bhartrihari(7th cent .A.D.) was probably the first to discuss this question, systematically."[5] भर्तृहरि(७वीं शताब्दी) ने शब्दों पर चर्चा करते हुए ऐसे शब्दों को छह भागों में विभाजित किया है। जिसके आधार पर शब्दों से द्विअर्थकता को नष्ट किया जा सकता है। "उचित" अर्थ को संप्रेषित करने के लिए भर्तृहरि 'वाक्यपर्दम्य' में छह रुपों का उपयोग करते है। वह इस प्रकार है **"वाक्य" syntactic construction, "प्रकरण" context, "अर्थ" goal, "औचित्य" propriety, "दिशा" place, एवं "काल" time.** "Bhartihari lists in his vakyapadtya(2.314) six factors which may be used as helpful cues in correctly interpreting the intended sense. They are: "vakya" syntactic construction. "prakarana" context, "artha" goal, "aucitya" propriety, "desa" place and "kala" time."[6] इस प्रकार भर्तृहरि समान उच्चारण वाले शब्दों के संबंध में अर्थ को स्पष्ट करने की चर्चा करते है। यही संदर्भ मशीनी अनुवाद के निर्माण में होने वाले भिन्नार्थक शब्दों के निराकरण के लिए किया जाए तो कुछ सीमा तक सफलता मिल सकती है। इस सूत्र का उपयोग संपूर्णत: मशीनी अनुवाद में नहीं हो पाया है। इस सूत्र का उपयोग पूर्ण रुप से किया जाए तो इस सूत्र के माध्यम से अस्पष्टता की समस्या को दूर किया जा सकता है। लेकिन मशीनी अनुवाद में उत्पन्न होने वाली समस्या केवल ऐसे शब्दों का निराकरण करना ही नहीं बल्कि ऐसे शब्दों की अस्पष्टता को समझ कर स्पष्ट किए गए शब्दों का मशीन के द्वारा ही अनुवाद करना है। प्राथमिक स्तर पर कठिन होने वाली यह प्रक्रिया मशीनी स्तर पर आकर और भी अधिक जटिल हो जाती है। डॉ.त्रिभुवन ओझा के अनुसार "कुछ अर्से से ऐसे शब्दों के वैज्ञानिक अध्ययन के लिए भाषा विज्ञान की एक नई शाखा 'होमोनिमिक्स' (Homonymics) भी विकसित हो गई है।" जिसके माध्यम से केवल समान उच्चारण वाले शब्दों से अस्पष्टता उत्पन्न होने वाली समस्याओं के निराकरण से संबंधित संशोधन होता है।

समान उच्चारण वाले शब्दों की परिभाषा

---

[5] वहीं पृष्ठ संख्या - 49

[6] वहीं पृष्ठ संख्या - 50

1. द वर्ल्ड बुक इन्सायक्लोपीडिया(The world book encyclopedia):- भाषा की वह स्थिति जिसमें दो या अधिक शब्दों में समान ध्वनि परन्तु भिन्न अर्थ होता है शब्द की लिपि समान हो भी सकती है या नहीं भी। अंग्रेजी भाषा में शब्दों का होना साधारण बात है "Is a language term for two or more words that have the same sound but different meaning. The word may or may not be spelled alike Homonyms are common in the English language."[7]

2. एम.ट्रेसा कॅबर (M.Teresa Cabr's) :- पारंपरिक भाषा विज्ञान के अनुसार शब्दों में पृथकता के संबंध में ध्वनि युक्त अनेकार्थक शब्द (समान ध्वनि वाले शब्द) होते है और वर्तनी मूलक शब्द (समान लिपि वाले शब्द) होते है। समान ध्वनि वाले शब्द वह इकाई होते है जो उच्चारण में समान लेकिन लिपि में भिन्न होते है। "Tradition linguistics draws a distinction within homonymy between phonological Homonymy (Homophony) and Orthographic homonymy (homograph) homophones are unit that are pronounced the same but are spelled differently : - e.g.night/knight, here/hear, bare/bear, whereas homographs are words that have the same spelling but differ in origin."[8]

3. डॉ. दयाल जैन :- "हिंदी में शब्दों के कुछ ऐसे जोड़े मिलते हैं जो उच्चारण और वर्तनी में समान है परन्तु स्त्रोत और अर्थ में भिन्न हैं ऐसे शब्दों को भिन्नार्थक, समध्वन्यात्मक, समध्वनीय शब्द कहा गया है।[9] अंग्रजी में समान उच्चारण वाले शब्द को (होमोनिम्स Homonyms) कहते हैं।

4. एम.टेरीसा कॅबर्स (M.Teresa Cabr's):- के अनुसार समान उच्चारण वाले शब्द वह शब्द होते हैं जो ध्वनि में दूसरे शब्दों से समरुप होते है लेकिन अर्थ में भिन्न होते हैं। जैसे :- अंग्रेजी का एक शब्द "bank" (बैंक) देखा जाए तो यह शब्द कृत्रिम रुप से दूसरे शब्द के साथ भिन्न अर्थ में भी आता है। जैसे:- "bank" नदी का किनारा भी होता है और 'bank' का अर्थ बैंक (पैसे जमा करने की जगह) भी होती है। समान उच्चारण वाले शब्द ऐसे भी दो शब्द होते है जो उच्चारण में समान होते है लेकिन लिपि भिन्न होती है। जैसे:- "bare"

---

[7] The world book encyclopedia volume 9 (H) page 275-276

[8] **Terminology Theory, Methods and Applications** - M.Teresa Cabr's, page No.111

[9] डॉ. जैन,दिन दयाल,*हिंदी शब्द रचना,* पृष्ठ संख्या -28

एवं "bear" समान उच्चारण वाले शब्द समान ध्वनि वाले शब्दों से अलग होते है। समोच्चारित या समान ध्वनि वाले शब्द जैसेः- "mouth" और "the mouth of the river," यह एक ही शब्द होते है लेकिन संदर्भ के अनुसार इन शब्दों का अर्थ बदलता जाता है। A word that sound identical to another word with a different meaning; e.g. "bank", pertaining to the ground boarding a river and also pertaining to a place for keeping money. A Homonym also be word that sound identical to another word that is spelled differently; e.g. "bare" and "bear" Homonym is distinct from polysemy or multiple meaning; e.g. "mouth" as in "the mouth of the river," is a single word with two related meanings.[10]

उपर्युक्त परिभाषाओं से स्पष्ट होता है कि केवल हिंदी में ही नहीं; अपितु अन्य भाषाओं में भी शब्दों का मिलना सामान्य बात है। लिपि भिन्नता के कारण इन समस्याओं का अन्य भाषा की लिपि के अनुसार इन शब्दों की अस्पष्टता पर काम चल रहा है। परिभाषाओं से यह भी स्पष्ट है कि समान उच्चारण वाले शब्दों में शब्द एक या एक से अधिक हो सकते हैं लेकिन इन शब्दों का उच्चारण समान होता है और अर्थ में भिन्नता होती है। किसी भी विद्वान ने समान उचारण वाले शब्द एक ही प्रकार के होते हैं; ऐसा नहीं कहा है। यह शब्द दो या दो से अधिक भी हो सकते हैं। लेकिन एम.टेरीस कॅब्र्स (M.Teresa Cabr's) के अनुसार स.उ.शब्द एक ही होते है जैसे अपनी परिभाषा में उन्होंने "mouth" शब्द का उदाहरण दिया है। इसलिए समान ध्वनि वाले शब्द और समान उच्चारण वाले शब्दों में भिन्नता देखने को मिलती है। समान ध्वनि वाले शब्दों और समान उच्चारण वाले शब्दों के संबंध में अर्थ की दृष्टि से विचार किया जाए तो दोनों शब्दों का अर्थ भिन्न होना अनिवार्य होता है। इस नियम में समानता दिखाई देती है। जिसके आधार पर समान ध्वनि वाले शब्दों को शब्दों के अंतर्गत रखा जाता है। कई विद्वानों ने समान उच्चारण वाले शब्दों को विभाजित करते समय समोचारित शब्दों को समान उच्चारण वाले शब्दों के अंतर्गत ही रखा है। अंग्रेजी भाषा में शब्दों की तुलना में समोच्चारित शब्दों पर अधिक काम हुआ है लेकिन अबतक अधिक सफलता हासिल नहीं हुई है। आज भारतीय भाषाओं में भी मशीनी अनुवाद के क्षेत्र में बड़ी मात्रा में काम हो रहा है जिसके अंतर्गत समान उच्चारण वाले शब्दों की अस्पष्टता के निराकरण पर विचार हो रहा है।

---

[10] Encyclopedia Britannica – vol.No.5.H :- page No - 107

समान उच्चारण वाले शब्द की उत्पत्ति पर विचार किया जाए तो अग्रेंजी में "Homonym" शब्द ग्रीक के 'Homo' + 'onyms' इन शब्दों के पूर्वसर्ग और परसर्ग से "homonyms" शब्द बना है। अंग्रेजी में 'Homo' शब्द का अर्थ है 'समान' और 'onym' 'नाम' इस प्रकार अंग्रेजी में "Homonymy" का निर्माण हुआ।[11] संस्कृत में "Homonymy" के लिए "श्रृतिसम्मत भिन्नार्थक शब्द" शब्द प्रचलित है। हिंदी में अब तक कोई एक शब्द "Homonymy" के लिए नहीं दिया गया है। अंग्रेजी-हिंदी शब्दकोशों में अलग-अलग परिभाषाओं को देखा गया है। जैसे - "समध्वनीय" "समान उच्चारण वाले शब्द" "समान उच्चारणवाले भिन्नार्थक शब्द" "समन लिपि वाले शब्द" **"अनेकार्थक शब्द"** तथा इसके अलावा और भी कई उदाहरण प्राप्त होते हैं। **इस अनुसंधान में अंग्रेजी के "Homonymy" शब्द के लिए हिंदी में "अनेकार्थक शब्द" एवं "समान उच्चारण वाले शब्द" इन दोनों शब्दों का उपयोग किया गया है,** जो आई.आई.टी.मुंबई के अंग्रेजी-हिंदी ऑन लाईन शब्दकोश से लिया गया है। अनेकार्थक शब्दों के संदर्भ में यह स्पष्ट नहीं हो पाया है कि किस भाषा में सबसे पहले इन शब्दों का विकास हुआ। यह भी स्पष्ट नहीं हो पाया है कि किस भाषा में ऐसे शब्दों की अधिक संख्या पायी जाती है। सामान्यत: समान उच्चारण वाले शब्द विश्व की अधिकतर भाषाओं में समान रुप से मिलते हैं।

2.अन्य भाषाओं में मिलने वाले समान उच्चारण वाले शब्दों के उदाहरण

1.हिंदी उदाहरण

अंकुश(गजांकुश)	अंकुश(रोक)
अंचल(प्रदेश)	अंचल(आँचल)
अचल(स्थिर)	अचल(पर्वत)
अदा(चुकता)	अदा(हावभाव)
आदि(इत्यादि)	आदी(लत लगना)
आदि(इत्यादि)	आदि(अदरक)
उत्तर(परवर्ती)	उत्तर(जवाब)

[11] http://efl.htmlplanet.com/polysemy.htm

2.मराठी उदाहरण

दंड(जुर्माना)	दंड (बाहू)
पालक(सब्जी)	पालक(पालन कर्ता)
बस(मोटर बस)	बस(बस करना)
जाते(चक्की)	जाते(जाती हूँ)
वास(गंध)	वास(रहना)
नट(हीरो)	नट (नट-बोल्ट)
गोळी(बंदूक की गोली)	गोळी (डॉक्टरकी गोली)

3.अंग्रेजी उदाहरण

be(होना)	bee (मधुमक्खी)
knot(गुच्छा)	not (नहीं)
dear(प्रिय)	deer (हिरण)
hour(घंटा)	our(हमारा)
wind(समेटना)	wind(हवा)
dove (बतख)	dove(पड़की)
mobail (चपल)	mobile(दूरध्वनी यंत्र,मोबाईल)

4.जर्मन उदाहरण

जर्मन	अंग्रेजी	हिंदी	जर्मन	अंग्रेजी	हिंदी
Morgan	tomorrow	(कल)	morgan	morning	(सुबह)
arm	poor	(गरीब)	Arm	arm	(भुजा)
elf	eleven	(ग्यारह)	Elf	elf	(परी)
Steuer	tax	(कर)	Steuer	wheel	(पहिया)
Tau	dew	(ओस)	Tau	line	(रेखा)
Bis	until	(तक)	Biss	bite	(काटना)
Boot	boat	(नाव)	bot	offered	(प्रस्तावित)

Bund	federation	(संघ)	bunt	colorful	(रंगीन)
Chor	chorus	(गायक दल)	Korps	corps	(निकाय)
das	he	(वह)	dass	that	(क्योंकि)
Heer	army	(सेना)	her	from	(से)
Isst	eats	(खाता है)	ist	is	(है)
Meer	sea	(सागर)	mehr	more	(अधिक)
Nahmen	took	(लिया)	Namen	name	(नाम)
Rad	wheel	(पहिया)	Rat	advice	(सलाह)
seid	are	(हैं)	seit	since	(के बाद)
Tod	death	(निधन)	tot	dead	(बंद)
viel	much	(बहुत)	fiel	fell	(खाल)
wahr	true	(सच)	war	was	(होना)
Waise	orphan	(अनाथ)	Weise	way	(मार्ग)

5.रशियन उदाहरण[12]

रशियन	अंग्रेजी	हिंदी	रशियन	अंग्रेजी	हिंदी
max	stroke	प्रहार	Max	stretch	तानना
cegok	hoursman	घुड़सवार	cegok	passenger	प्रवासी
Beulu	will pouer	इच्छाशक्ति	Beulu	freedov	-
Digx	sprit		Digx	scent	इत्र

6.स्पॅनिश उदाहरण[13]

स्पेनिश	अंग्रेजी	हिंदी	स्पॅनिश	अंग्रेजी	हिंदी
Solo	(alone)	अकेला)	***sólo***	(only)	(मात्र)

[12] Langenslheiots, Russian Dicationary, by – E.Wedel
[13] www.spanicity.com

si	(if)	(यदि)	***sí***	(yes)	(हाँ)
sumo	(supreme)	(सर्वोच्च)	***zumo***	(juice)	(रस)
tasa	(rate)	(दर)	***taza***	(cup)	(प्याला)
ti	(you)	(आप)	***ti*** (note of the musical scale)		(ट)
tu	(your)	(तुम्हारा)	***tú***	(you)	(आप,तुम)
tubo	(pipe)	(नली)	***tuvo***	(to have)	(पास होना)
vino	(wine)	(शराब)	***vino***	(to come)	(आना)

भारतीय एवं पाश्चात्य भाषाओं में अधिकतर जगह शब्द विज्ञान एवं कोश विज्ञान के माध्यम से अनेकार्थक शब्दों को स्पष्ट किया गया है। अंग्रेजी, जर्मन, स्पॅनिश, चीनी, तमिल, इटालिअन, लैटिन, फ्रेंच, तेलुगु, अरेबिक आदि[14] भाषाओं में मशीनी अनुवाद का विकास तेजी से हो रहा है। जिसके चलते इन भाषाओं में अनेकार्थक शब्दों पर मशीनी अनुवाद की दृष्टि से विचार किया जा रहा है। आवश्यकता है कि भारतीय भाषाओं में भी इस विषय पर सभी विषयों की दृष्टि से कार्य किया जाए। सामान्यत: अनेकार्थक शब्दों को विभाजित कर संशोधित किया गया है। प्राय: इन शब्दों की परिभाषाओं की भिन्नता इतनी सूक्ष्म है कि सामान्य व्यक्ति, विद्यार्थी को इन शब्दों की परिभाषा को स्पष्ट रुप से समझने में काफी कठिनाईयों का सामना करना पड़ता है।

देखा जाए तो विश्व की किसी भी भाषा में अनेकार्थक शब्द किस प्रकार निर्माण होते हैं इसपर भी विचार होना आवश्यक है। विश्व में मानव द्वारा बोली जानी वाली जितनी भी विकसित भाषाएँ हैं उन सभी भाषाओं में मानवीय गुण विद्यमान हैं। अनेकार्थक शब्दों की उत्पत्ति के कारणों की ओर देखने पर हमें यह पता चलता है कि भाषा में अनेकार्थकता निर्माण होने के अधिकांश कारण मानव द्वारा अपने आपको अधिक विकसित करने की

[14] Manual of Lexicography, Homonymy – IGUSTA (Lexical meaning) – page No. 75

होड़। अन्य समाजों, जातियों की भाषा के साथ संपर्क स्थापित कर अन्य भाषाओं के शब्दों को अपनी भाषा में सम्मिलित करने की प्रक्रिया, मानव में होने वाले संप्रेषण को अधिक विकसित करने का परिणाम आदि कुछ प्रमुख कारण हैं। पशु, पक्षी और जानवरों की भाषा में अनेकार्थक शब्दों का कोई उदाहरण नहीं मिल पाता। इस पर अभी अनुसंधान नहीं हो पाया है। जिससे यह स्पष्ट रुप से नहीं कहा जा सकता की पशु-पक्षियों की भाषा में यह गुण है या नहीं।

किसी भी भाषा में अनेकार्थक शब्द निर्माण होने के क्या कारण हो सकते है? इस पर विचार किया जाए तो निम्न रुप से कुछ प्रमुख कारण दिखाई देते हैं।

1. मानसिक सुविधा
2. सांकृतिक आवश्यकता
3. पारिभाषिकता
4. अर्थ का स्थानांतरण
5. शब्दों का गुणार्थक प्रयोग
6. शब्दों का आलंकारिक प्रयोग
7. शब्दों का संक्षिप्तिकरण
8. अर्थ का बोलीगत या क्षेत्रीय विकास
9. संरक्षणशीलता
10. विशेषण का संज्ञा के लिए भी उपयोग
11. शब्दों का व्याकरणिक प्रयोग
12. भिन्नार्थकता - भाषा मिश्रण, विभिन्न शब्द-भेदों की समरुपता, पद्यात्मक समरुपता

अनेकार्थक शब्द निर्माण होने के यह कुछ प्रमुख कारण हैं। देवनागरी लिपि ध्वनिमूलक होने के कारण रोमन भाषा में निर्माण होनेवाली समस्याओं से भिन्न प्रकार की समस्यायें हिंदी अनेकार्थक शब्द निर्माण प्रक्रिया में उत्पन्न होती हैं। जैसे - "अनेकार्थक शब्द शब्द-युग्म होते हैं, जिनका उच्चारण प्राय: समान होता है, किंतु उनके अर्थ भिन्न होते हैं और कई जगह वर्तनी समान या भिन्न भी हो सकती है।" हिंदी में कुछ ऐसे शब्द हैं जिनका उच्चारण समान है। वर्तनी भिन्न है, और अर्थ समान है लेकिन इन शब्दों की संख्या कम है। अंग्रेजी में ऐसे शब्दों को

भी अनेकार्थक शब्दों की कोटि में रखा गया हैं। हिंदी में यह प्रयोजन अभी सार्थक रुप नहीं ले पाया है। हिंदी ध्वनिमूलक भाषा होने के कारण इसे हम अपवाद भी समझ सकते हैं। सबसे पहले अनेकार्थक शब्दों को कितने भागों में विभाजित किया जा सकता है, इसपर विचार करना आवश्यक है। जिससे इन शब्दों का अधिक से अधिक संशोधित रुप हमारे सामने आ सके। जिससे हमें आगे बढ़ने में आसानी होगी। सामान्य रुप से अनेकार्थक शब्दों को निम्न भागों में विभाजित करके देखा जा सकता है।

1.HOMOGRAPHS

2.HOMOPHONES

3.HETERONYMS

4.POLYSEMES

5. CAPITONYMS[15]

	Same pronunciation	Different pronunciation
Same spelling	**Homonym**	Homograph
Different spelling	Homophone	Heteronym/ Heterophony

Tabal No. 2.

उच्चारण (Pronunciation), समध्वनि और भिन्नार्थकता के निकषों पर शब्द-युग्मों को हम विभाजित कर सकते हैं। जिससे हम अनेकार्थक शब्दों की असमानता(inequality) को आसानी से समझ सकते हैं। विभाजित प्रकारों की परिभाषायें इस प्रकार हैं।

3. समान उच्चारण वाले शब्दों के प्रकार एवं व्याकरणिक सूत्र

1.समान लिपि वाले शब्द (HOMOGRAPHS) - वह शब्द जो समान वर्तनी के साथ आते हों चाहे उनका उच्चारण समान हो या भिन्न। समान लिपि वाले शब्दों का उच्चारण अधिकतर समान ही होता है। ऐसे समय यह शब्द समान उच्चारण वाले शब्द भी हो सकते हैं। उदा. :- *bark* (the sound of a dog) and *bark* (the skin of a tree). इसी के साथ जब इन शब्दों का उच्चारण भिन्न हो तो ऐसे समय यह शब्द 'एकाधिक अर्थों वाले शब्द' (heteronyms) हो जाते है। जैसे:- *bow* (the front of a ship) and *bow* (a type of knot).

[15] http://en.wikipedia.org/wiki/Homonym

2.समान ध्वनि वाले शब्द (HOMOPHONES) - समान ध्वनि वाले शब्द उन शब्दों को कहते हैं जिनका उच्चारण समान होता है। इन शब्दों का संबंध वर्तनी से नहीं होता, वर्तनी भिन्न भी हो सकती है। कई बार समान ध्वनि वाले भिन्नार्थ शब्द, समान लिपि वाले भिन्नार्थक शब्द भी हो सकते हैं। (जब दो भिन्नार्थक शब्दों की लिपि समान हो। समान ध्वनि वाले भिन्नार्थक शब्द का उदाहरण है :- Homographic examples include *tire* (to become weary) and *tire* (on the wheel of a car). समान उच्चारण वाले लेकिन भिन्न लिपि वाले भिन्नार्थक शब्द(Heterographic) का उदाहण:- *to, too, two,* and *there, their, they're.*

3.एकाधिक अर्थों वाले शब्द (HETERONYMS) - इन शब्दों को समान लिपि वाले भिन्नार्थक शब्दों के उपवर्गों की कोटि में भी रखा जाता है। इन शब्दों की लिपि समान होती है लेकिन इन शब्दों का उच्चारण भिन्न होता है। इसलिए यह शब्द समान लिपि वाले शब्द हो सकते है लेकिन समान उच्चारण वाले शब्द नहीं हो सकते। जैसे कुछ शब्द :- *desert* (to abandon) and *desert* (arid region); *row* (to argue or an argument) and *row* (as in to row a boat or a row of seats). यह शब्द समान उच्चारण वाले शब्दों में संस्थापित होते है। कई बार यह शब्द Heterophony's के रुप में भी पहचाने जाते हैं।

4. समोच्चारीत शब्द_(POLYSEMES) - वह शब्द जिनकी लिपि/वर्तनी समान हो किंतु अर्थ भिन्न होते हुए भी एक दूसरे से संबंधित **हों**। प्राय: अनेकार्थक शब्दों में एवं इन शब्दों में अंतर अति सूक्ष्म एवं विषय परक होता है। अत: निश्चित तौर पर यह नहीं कहा जा सकता कि समोच्चारीत शब्द अनेकार्थक शब्द होते हैं, यह विषय पर निर्भर करता है। जैसे :- Words such as "mouth", meaning either the orifice on one's face, or the opening of a cave or river, are polysemous and may or may not be considered homonyms.

5.लिपि परिवर्तक अनेकार्थक शब्द (CAPITONYMS) - वह शब्द जो समान लिपि में आते हैं लेकिन उनका अर्थ भिन्न होता है। यह शब्द मुख्य रुप से शब्द की वर्तनी के परिवर्तन पर निर्भर होते हैं। जैसे :- *polish* (to make shiny) and *Polish* (from Poland).

अनेकार्थक शब्दों को जिन पाँच भागो में विभाजित करके देखा जाता है उनकी पाचों परिभाषाओं में कुछ समानता पाई जाती हैं। लेकिन इन परिभाषाओं में पायी जाने वाली सूक्ष्म भिन्नता के कारण इन्हें विभाजित करना आवश्यक है। यह भिन्नता इतनी सूक्ष्म है, कि सामान्य व्यक्ति इसे आसानी से समझ नहीं पायेगा। समान उच्चारण वाले शब्द और समान लिपि वाले शब्दों में केवल उच्चारण एवं लिपि का भेद पाया जाता है। अर्थ भिन्न होता है लेकिन उच्चारण समान होता है। एकाधिक अर्थों वाले शब्द (HETERONYMS) विशेषत: रोमन लिपि में अधिक पाये जाते हैं या विशेष रुप से उन भाषाओं में अधिक पाये जाते है जिसकी लिपि ध्वनि मूलक नहीं होती। एकाधिक अर्थों वाले शब्दों में लिपि समान होती है लेकिन उच्चारण भिन्न होता है। इसलिए इन शब्दों में लिपि का महत्व अधिक होता है। जो लिखा जाता है, वही उच्चारण किया जाता है। इसलिए ऐसे शब्द देवनागरी भाषा में नहीं पाये जाते। अर्थ के आधार पर समान ध्वनि वाले शब्दों को भी अनेकार्थक शब्दों के विभाजन में रखा गया है। जिसमें शब्द एक ही होता है लेकिन शब्द का अर्थ संदर्भ के अनुसार बदलता रहता है। अंतिम परिभाषा लिपि परिवर्तक अनेकार्थक शब्द (CAPITONYMS) के संदर्भ में है। इन शब्दों की मात्रा बहुत ही कम पाई जाती है। जिसका मूल कारण इन शब्दों का गुण है। लिपि परिवर्तक अनेकार्थक शब्द रोमन भाषाओं या उन भाषाओं में पाये जाते हैं; जिन भाषाओं की लिपि में वाक्य, शब्द या नाम में किसी भी शब्द का प्रथम अक्षर ह्रस्व हो जाता है। ऐसी भाषाओं में ह्रस्व, दीर्घ के परिवर्तन के कारण अर्थ में भिन्नता उत्पन्न हो जाती है। देवनागरी लिपि में इस प्रकार का कोई नियम व्याकरणिक और भाषावैज्ञानिक रुप से नहीं है। जिसके कारण हिंदी में ऐसे शब्द नहीं पाये जाते। इन सभी विभाजित प्रकारों में अर्थ की भिन्नता सबसे प्रमुख है। कुछ स्थानों पर एकाधिक अर्थों वाले शब्द (HETERONYMS) के अपवादों को छोड़ दिया जाए तो उच्चारण में भी समानता होना अनिवार्य हो जाता है जिसमें लिपि का कोई बंधन नहीं है। संरचनात्मक अस्पष्टता एवं संदर्भगत अस्पष्टता में शब्दों और पदों के संबंध का अधिक महत्व होता है। शाब्दिक अस्पष्टता में व्याकरण और व्याकरणिक कोटियों का अधिक महत्व है। "शाब्दिक अस्पष्टता के निर्माण में और अनेकार्थक शब्दों की उत्पत्ति में व्याकरणिक विकास की महत्वपूर्ण भूमिका होती है। जैसे:- 'जीता' शब्द 'जीना' शब्द का

भूतकालिक रुप है, जिसका अर्थ 'जीवित' होता है। 'जीतना' 'जीता' का भूतकालिक रुप 'जीता' होता है। 'दिया' शब्द के भी दो अर्थ होते हैं। एक 'दिया' का अर्थ 'दीपक' दूसरा अर्थ 'देना' का भूतकालिक रुप 'दिया'।" Grammatical developments also sometimes lead to the formation of homonyms. Jita as the past tense of jina, meaning alive, is also the past tense of jitna, meaning conquered; and diya is both a lamp and the past tense of dena or give.[16] इस प्रकार देखा जाए तो कालों के परिवर्तन से भी इन शब्दों का निर्माण होता है। जो पूर्णत: व्याकरणिक है। अनेकार्थक शब्दों के निर्माण में व्याकरण के नियमों की महत्वपूर्ण भूमिका दिखाई देती है। व्याकरण केवल अनेकार्थक शब्दों के निर्माण में सहायता ही नहीं करता बल्कि इन शब्दों के निराकरण में भी महत्वपूर्ण भूमिका निभाता है। जैसे:- 'दिया' का उपयोग नाम के रुप में किया गया हो तो सामान्य रुप से 'दिया' का अर्थ 'दीपक' होगा। जिस वाक्य में 'दिया' शब्द का उपयोग क्रिया के रुप में उपयोग किया गया हो उस जगह 'दिया' शब्द का अर्थ 'दिया हुआ' या 'देना' के रुप में होगा। इस प्रकार व्याकरण का उपयोग दोनों रुपों में किया जाता है।

इन शब्दों को भाषावैज्ञानिक रुप से देखा जाए तो भाषा में तीन प्रकार की अस्पष्टता पायी जाती है। जिसे निम्न रुप से देखा जा सकता है।

1. संरचनात्मक अस्पष्टता (Structural Ambiguity) - वाक्य का निर्माण एक विशिष्ट संरचना से होता है, जिसके अनुसार वाक्य का निर्माण होता है। संरचना में होने वाले बदलाव के कारण वाक्य में अस्पष्टता उत्पन्न हो जाती है जिसे संरचनात्मक अस्पष्टता कहा जाता है।

2. शाब्दिक अस्पष्टता (Lexical ambiguity):- पाठ में किसी विशिष्ट शब्द के कारण निर्माण होने वाली अस्पष्टता को शाब्दिक अस्पष्टता कहा जाता है। इस अस्पष्टता का निराकरण इसलिए भी महत्वपूर्ण है कि कंप्यूटर वाक्य की संरचना को समझ सकता है लेकिन शब्द के अर्थ को समझना उसके लिए अधिक कठिन कार्य है। जैसे हम अंग्रेजी का उदाहरण देख सकते है।

1. He chairs the session.

---

[16] http://www.tribuneindia.com/2000/20000819/windows/roots.htm

2. The chairs in this room are comfortable.

पहले वाक्य में chairs शब्द का उपयोग क्रिया के रुप में किया गया है और दूसरे वाक्य में chairs का उपयोग नाम के लिए किया गया है। इस अवस्था में chairs का अर्थ प्रत्येक वाक्य के संदर्भ के अनुसार लिया जाना चाहिए। इस कार्य का सामान्य निराकरण POS tags के माध्यम से किया जा सकता है।

3. **व्याकरणिक अस्पष्टता(Grammatical ambiguity)** :- प्रत्येक भाषा का एक व्याकरण होता है। जिसके अनुसार भाषा का निर्माण होता है। जब भाषा व्याकरणिक नियमों के अनुसार न हो तो वाक्य में अस्पष्टता उत्पन्न हो जाती है। जिसे व्याकरणिक अस्पष्टता कहा जाता है।

भाषा में निर्माण होने वाली इन तीन अस्पष्टताओं में **अनेकार्थक शब्दों के कारण उत्पन्न होने वाली अस्पष्टता शाब्दिक अस्पष्टता कहलाती है।** शाब्दिक अस्पष्टता संचनात्मक अस्पष्टता एवं व्याकरणिक अस्पष्टता से अधिक प्रभावी दिखाई देती है। जिसके कारण प्रत्येक भाषा में शाब्दिक अस्पष्टता आसानी से दिखाई पड़ती है। भाषा में शाब्दिक अस्पष्टता भी कई प्रकार की होती है।

समान उच्चारण वाले अस्पष्ट शब्द यह एक शाब्दिक अस्पष्टता है।(The Homonymy is a lexical ambiguity.) [17] इस अस्पष्टता के निराकरण के मशीनी अनुवाद में Word Sense Disambiguation (WSD) के माध्यम से दूर करने का काम किया जाता है, जिसके लिए प्रत्येक मशीनी अनुवाद यंत्र में स्रोत भाषा एवं लक्ष्य भाषा के व्याकरण एवं भाषिक नियमों के आधार पर इस भाषिक संसाधन का निर्माण किया जाता है। यह कार्य भाषिक संसाधन निर्माण की प्रक्रिया में सबसे जटिल कार्य माना जाता है।

इस प्रकार देखा जाए तो भाषा में अस्पष्टता को दूर करने का सबसे महत्वपूर्ण योगदान व्याकरण का ही है। भृर्तृहरि ने भी अनेकार्थक शब्दों की अस्पष्टता को दूर करने के लिए व्याकरण के ही आधार पर छह नियमों का निर्माण किया था।

4. मानव द्वारा होने वाले अनुवाद एवं मशीनी अनुवाद में समान उच्चारण वाले शब्दों की समस्याएँ

---

[17] . Kalra.Sachin. June 2004. *Punjabi Text Generation using Interlingua approach in Machine Translation* (A thesis Submitted fulfilment of Master of Engineering) page No..- 39-40

मानव द्वारा होने वाले अनुवाद एवं मशीनी अनुवाद में अनेकार्थक शब्दों की समस्या कई जगह समान रुप से दिखाई देती है। इस समस्या से मानव-मशीन, छोटे-बड़े, शिक्षित-अशिक्षित, सरकारी कर्मचारी से व्यवहारिक-दुकानदारों और भाषाविदों से लेकर सामान्य व्यक्तियों को भी अनेकार्थक शब्दों के कारण कई समस्याओं का सामना करना पड़ता है। सामान्य रुप से भाषा का उपयोग संप्रेषण के लिए किया जाता है लेकिन भाषा भिन्नता के कारण एक भाषा से दूसरी भाषा में संप्रेषण स्थापित करने में कठिनाई होती है, जिसके लिए अनुवाद का उपयोग किया जाता है। अनुवाद के प्रमुख प्रकारों पर विचार किया जाए तो भाषांतरण(Interpretation) एवं मानव द्वारा लिखित रुप में किए जाने वाले अनुवाद पर अधिक जोर दिया जाता है। जबकी भाषांतरण अधिकतर व्यवहारिक रुप में उपयोग में लाया जाता है। लिखित अनुवाद सामान्य से सामान्य जगहों पर भी उपयोग में लाया जाता है। 21वीं शताब्दी में कंप्यूटर का उपयोग बढ़ने से एवं इंटरनेट की सुविधा के कारण मशीनी अनुवाद का प्रचलन बढ़ने लगा है लेकिन मशीनी अनुवाद में उत्पन्न होने वाली समस्याओं एवं कंप्यूटर की अज्ञानता के कारण मशीनी अनुवाद एक सीमा तक ही प्रभावी है।

आज कई संस्थानों, अकादमियों, प्रकाशकों द्वारा एक भाषा से दूसरी भाषाओं में अनुवाद का कार्य हो रहा है। यह अनुवाद अधिकतर मानव द्वारा होने वाले अनुवाद होते हैं। भारत जैसे बहुभाषिक देश में मानव द्वारा किया जाने वाला अनुवाद केवल संप्रेषण का ही साधन नहीं वरन संस्कृति, समाज, और संपूर्ण भारतवर्ष को एक संघ के रुप में बनाये रखने का भी साधन है। इसीलिए भारत के दिल्ली स्थित "साहित्य अकादमी संस्थान" विश्व की सबसे ज्यादा भाषाओं में एक साथ अनुवाद कराने वाली सबसे बड़ी संस्था के नाम से जानी जाती है। इस कार्य की महत्ता को समझने एवं उसकी अनुसंधानात्मता को बनाये रखने की जरुरत है। इसलिए मानव द्वारा होने वाले अनुवाद में अनेकार्थक शब्दों के कारण अनुवाद के निर्माण में होने वाली समस्याओं को दूर करना एवं अनुवादकों को इन गलतियों से परिचित कराना भी आवश्यक है।

अनेकार्थक शब्दों के कारण अनुवादक अनुवाद में एक प्रकार की गंभीर गलती कर सकता है। जिसके कारण अनुवाद का अर्थ भिन्न हो सकता है। मानव अनुवाद करते समय

प्रथमत: एक लय में मन-ही-मन में शब्दों को अलिखित रुप में अनूदित करता है। इसके उपरांत मन में किए गए अनूदित शब्दों को विचारपूर्वक लिख देता है। यह प्रक्रिया निरंतर रुप से चलती रहती है। अनुवाद के समय कुछ अन्य तत्व भी मानव मस्तिष्क में कार्य करते रहते हैं। जो शब्द निरसण की प्रक्रिया में बाधा उत्पन्न करते हैं। प्रमुख रुप से तब जब द्विअर्थक शब्द, अनेकार्थक शब्द, अस्पष्ट शब्द पाठ में आते हों। ऐसे समय अनुवादक या तो इन शब्दों को अनदेखा कर अनुवाद करना छोड़ देता है, या गलत अनुवाद कर देता है। यह क्रिया अनुवाद के अनदेखेपन का भी नतीजा हो सकती है, जान बूझकर किया गया काम या फिर शब्द की परख न होने की अज्ञानता का नतीजा भी हो सकती है। मशीन किसी भी शब्द को अनदेखा नहीं कर सकती। उसे प्रत्येक शब्द का अनुवाद करना ही पड़ता है।

4.1 मानव द्वारा होने वाले अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं के उदाहरण

मानव द्वारा अनुवाद की प्रक्रिया में अनेकार्थक शब्दों का किस प्रकार निरसण किया जाता है। यह स्पष्ट करने के लिए हिंदी से मराठी में अनूदित उपन्यास और कहानियों को लिया गया है। जिसमें हिंदी के अग्रणी साहित्यकार 'प्रेमचंद' की कहानियाँ ली गई हैं। 'कमलेश्वर' द्वारा लिखित "कितने पाकिस्तान" उपन्यास को चुना गया है। प्रेमचंद की कहानियों का कई भाषाओं में अनुवाद हो चुका है। यह अनुवाद काफी अच्छे हैं। 'कितने पाकिस्तान' उपन्यास हिंदी साहित्य में बीसवीं सदी की उपलब्धि मानी जाता है। जिसका कई भारतीय भाषाओं में अनुवाद हो चुका है। स्रोत पाठ में आये अनेकार्थक शब्दों को अनूदित रचना में बड़े ही उत्कृष्ट ढंग से अनुवादित किया गया है। जिसका उपयोग अस्पष्टता के निराकरण के लिए होगा।

मूल हिंदी उपन्यास का नाम - कितने पाकिस्तान

लेखक - कमलेश्वर

अनूदित मरठी उपन्यास का नाम - किती पाकिस्तानं

अनुवादक का नाम - पद्माकर जोशी

प्रेमचंद की चुनिंदा कहानियाँ

मराठी कहानी संग्रह का नाम - निवडक प्रेमचंद

अनुवादक - दिनकर साविरकर

हिंदी 1.वह न्याय का दरबार था, किंतु पक्षपात का नशा छाया हुआ था।[18]

मराठी 2.तो न्यायाचा दरबार होता परंतु सर्व कर्मचारांवर पक्षपाताची नशा चढली होती [19]

हिंदी 1. वकीलों ने यह फैसला सुना और उछल पड़े।[20]

मराठी 2. हा निकाल ऐकताच वकील मंडळी तर आनंदाने नाचू लागली.[21]

'पर' 'जाती' शब्द का अनुवाद कई जगह भिन्न भिन्न रुप में किया गया है। कई जगह 'पर' और 'जाती' शब्द के मुख्य अर्थ से हटकर भी अनुवाद किया गया है। लेकिन शब्द के अर्थ को वाक्य में सही प्रमाणित करने में अनुवादक ने यश प्राप्त किया है।

जैसे :-

हिंदी 1.इसी समय उनके द्वार पर सजा हुआ रथ आकर रूका।

मराठी 2. त्याच वेळी एक रथ त्यांच्या दरवाजा समोर येऊन उभा राहिला.

हिंदी 1. अफसर लोग उन पर बहुत विश्वास करने लगे।

मराठी 2. अधिका-यांचा त्यांच्यावर फार विश्वास बसला होता।

हिंदी 1. चाहे कोई मर ही जाए, पर उठने का नाम नहीं लेते।[22]

मराठी 2. कुणी जीवानिशी मेलं तरी जागचे उठनार नाहीत।[23]

हिंदी 1. इसलिए वे इलाहाबाद स्टेशन पर मिल ही जाते थे।

---

[18] नमक का दरोगा - प्रेमचंद की चुनिंदा कहानियाँ

[19] मिठाचा फौजदार- निवडक प्रेमचंद - अनुवादक - दिनकर साविरकर

[20] नमक का दरोगा - प्रेमचंद की चुनिंदा कहानियाँ

[21] मिठाचा फौजदार- निवडक प्रेमचंद - अनुवादक - दिनकर साविरकर

[22] शतरंज के खिलाड़ी-प्रेमचंद की चुनिंदा कहानियाँ

[23] बुद्धिबळाचे खेळाडू-निवडक प्रेमचंद - अनुवादक - दिनकर साविरकर

मराठी 2. त्या मुळे त्यांच्या भेटी अलाहाबाद स्टेशनवर होत असत.

हिंदी 1. अनबूझी इच्छाएँ आती और चली जाती हैं।

मराठी 2. अज्ञात इच्छा येतात आणि निघून जातात

हिंदि 1. विया 'अच्छा' कहकर पुल पर चढ़कर अपनी गाड़ी वाले प्लेट फॉर्म पर चली जाती थी

मराठी 2. विया 'अच्छा' मणून पुलावर जाऊन आपल्या गाडीचा प्लॅटफॉर्मवर जात असे.

अनेकार्थक शब्दों में समान उच्चारण वाले शब्दों का अधिकतर अनुवादक सही अनुवाद करते है। यह इसलिए भी आसान हो जाता है कि अनुवादक सभी शब्दों से परिचित होता है और वह लिपि के आधार पर शब्दों के अर्थ समझता है। यहीं नियम संगणक पर भी होगा। मशीनी अनुवाद में मशीन समान उच्चारण वाले शब्दों का सही अनुवाद करती है। मानव द्वारा किए कुछ अनुवाद के सही उदाहरण हम निम्न रुप से देख सकते हैं।

हिंदी 1. हल्कू ने आकर स्त्री से कहा।[24]

मराठी 2. हल्कूने येऊन बायकोला सांगितले।[25]

हिंदी 1. यह प्रकाश एक नौका के समान हिलता मचलता हुआ, जान पड़ता था।[26]

मराठी 2. हा प्रकाश एकाद्या नौके सारखा हलत डोलत तरलत आहे असे वाटत होते।[27]

हिंदी 1. रात की ठंड में यहाँ सोना तो न पडेगा?

मराठी 2. निदान रात्री थंडीत इथं झोपावं तर लागणार नाही.

हिंदी 1. वह हार का बदला चुकाने के लिए अधीर हो रहे थे।

मराठी 2. पराभवाचा बदला घेण्यासाठी ते अधीर झाले होते.

हिदी 1. मुसाफिरों में चलती आपसी बातचीत के दौरान पता चला था कि वे किसान खानदान पहले अलीगढ़ जा रहे थे।

---

[24] पूस की रात- प्रेमचंद की चुनिंदा कहानियाँ

[25] पौषातील रात्र - निवडक प्रेमचंद - अनुवादक - दिनकर साविरकर

[26] पूस की रात- प्रेमचंद की चुनिंदा कहानियाँ

[27] पौषातील रात्र - निवडक प्रेमचंद - अनुवादक - दिनकर साविरकर

मराठी 2. प्रवाशांच्या आपाअपसांतल्या बोलण्यावरून समजलं की ती शेतकरी कुटुंब प्रथम अलीगडला जाणार होती।

हिंदी-मराठी की लिपि समान होने के कारण हिंदी-मराठी भाषा में संस्कृत और अन्य भाषाओं का समान रुप से प्रयोग होता है। जिसके कारण अधिकतर जगह यह देखा गया है कि जहाँ अनुवादक को अनुवाद करने में कठिनाई हुई हो, समानार्थी शब्द न मिल पाया हो ऐसी जगह पर अनुवादक ने हिंदी शब्द ही लिया है। जैसे:-

हिंदी - 1. जब इस बार भी कोई उत्तर न मिला।

मराठी - 2. या वेळी सुद्धा उत्तर मिळाले नाही.

हिंदी - 1.इधर छोटे से बड़े कौन ऐसे थे जो उनके ऋणी न हों।

मराठी - 2. त्यांचा ऋणी नसलेला माणूस पंच क्रोशीत विराळा होता.

हिंदी 1. वंशीधर ने कठोर स्वर में कहा।

मराठी 1. वंशीधर कठोर स्वरात म्हणाला.

इस प्रकार के अनुवाद का एक महत्वपूर्ण उदाहरण 'कितने पाकिस्तान' उपन्यास के भाग दो में 'सर' शब्द का दो पन्नों के भीतर(पृष्ठ संख्या 14-15) में 10 बार उपयोग किया गया है। जिसका अर्थ गुरुजी, मास्टरजी, अधिकारीजी, साहब है। जिसका मराठी अनुवाद गुरूजी, मास्तरजी, अधिकारीजी, साहेब होता है। सर शब्द का दूसरा अर्थ सर - तालाब, ताल, सरोवर, पुष्कर, तोयाधार भी होता है। इन शब्दों के लिए मराठी शब्द तलाव, तळ, पोखर, सरोवर, पुष्कर, तोयाघर आदि है। सर - शब्द का तीसरा अर्थ - सिर, शीश, मुंड, मुंडक, मूँडी के रुप में भी उपयोग किया जाता है। जिनके मराठी में डोकं, शीश, मुंड, मुंडके, मुंडी शब्द हैं। लेकिन अनुवादक ने किसी भी जगह तालाब, डोकं या गुरुजी नहीं किया है। जो अनुवाद की दृष्टि से अत्यंत सही अनुवाद है। लेकिन कई जगह अनुवादक 'साहब' शब्द का भी प्रयोग कर सकता था। जिससे अनुवाद पर किसी भी प्रकार का विपरीत परिणाम नहीं होता। यह प्रकिया मशीनी अनुवाद में उपयुक्त हो सकती है। उदाहरण -

1) हिंदी 1. हुआ यह नहीं था... सर ! पहले यह सुनिए कि हुआ क्या है...[28]

---

[28] कितने पाकिस्तान - कमलेश्वर, पृष्ठ संख्या - 12

मरठी 2. झालं होतं असं नाही...सर ! अगोदर काय झालंय ते ऐका...[29]

2) हिंदी 1. सर ! इस वक्त तो आपके फ्रंट पेज ऐडीटोरियल की जरूरत है...[30]

मराठी 2. सर ! आता तर आपल्या फ्रंट पेज एडिटोरियलची अतिशय आवश्यकता आहे...[31]

3) हिंदी 1. अभी आप डिक्टेट कर दें सर तो पहले एडीशन में चला जाएगा।

मराठी 2. आताच आपण डिक्टेट कराल सर, तर पहिल्या एडिशन मध्ये जाईल.

4) हिंदी 1. इतना ही नहीं सर ! समाचार संपादक ने कहा ।

मराठी 2.एवढंच नाही सर ! वृत्तसंपादक म्हणाला.

5) हिंदी 1. लेकिन... सर नजम सेठी इसमें क्या करेगा ?

मराठी 2. सर नजम सेठी यात काय करणार ?

इसी के साथ हिंदी के मूल उपन्यास में 'सर' शब्द का उपयोग अन्य वाक्यों भी किया गया है।

6) सर ! पाकिस्तानी प्रधानमंत्री और उनके विदेश मंत्री ने कहा है।

7) सर ! यह हामला युद्ध में बदल गया तब तो बड़ा नुकसान होगा।

8) सर ! यह है हमारे अब तक के शहीद सिपाहियों और वायु सैनिकों की लिस्ट।

9) यही तो सर !

10) यस सर ! महमूद ने हाजरी दी।

इस प्रकार हिंदी उपन्यास में लिए गए इन शब्दों को अनूदित मराठी उपन्यास में सभी जगह अनुवाद के रुप में 'सर' का ही उपयोग किया गया है और अनुवाद में भी कोई गडबड़ी नहीं है। इससे स्पष्ट है कि मानव अनुवाद में इस सूत्र से अनुवाद गलती की गुंजाइश कम दिखाई देती है। इसी प्रकार अनूदित उपन्यास में और कई उदाहरण मिलते हैं।

शब्द चयन की समस्या।

हिंदी – 1. हार की चोट बुरी होती है

मराठी – 2. हार खाण्याची जखम असह्य असते

---

[29] मराठी अनुवाद- किती पाकिस्तानं - अनुवादक - पद्माकर जोशी, राजपाल एंड सन्ज प्रकाशन दिल्ली, पृष्ठ संख्या - 16

[30] कितने पाकिस्तान - कमलेश्वर, पृष्ठ संख्या - 12

[31] मराठी अनुवाद- किती पाकिस्तानं - अनुवादक - पद्माकर जोशी, राजपाल एंड सन्ज प्रकाशन दिल्ली, पृष्ठ संख्या - 16

हिंदी – 1. वह हार का बदला चुकाने के लिए अधीर हो रहें थे।

मराठी – 2. पराभवाचा बदला घेण्यासाठी ते अधीर झाले होते.

अनुवाद में शब्द चयन एक प्रमुख कार्य होता है। केवल किसी शब्द के लिए दूसरी भाषा का शब्द रखना अनुवाद नहीं होता। यही नियम अनेकार्थक शब्दों के संदर्भ में भी दिखाई पड़ता है। अनुवादक ने प्रथम वाक्य में हिंदी के 'हार' शब्द के लिए मराठी अनुवाद में भी 'हार' शब्द का प्रयोग किया है। हिंदी के 'हार' शब्द के लिए मराठी में भी 'हार' अर्थ होता है लेकिन इस वाक्य में इस शब्द का दूसरा समानार्थी शब्द होना आवश्यक है जिससे अनुवाद सुंदर बन सके। जैसे अनुवादक ने दूसरे वाक्य में इस गलती को न दोहराते हुए, हिंदी के 'हार' शब्द के लिए ही 'पराभव' शब्द दिया है। जो वाक्य में अर्थ की दृष्टि से स्वीकृत है।

मानव अनुवाद में अनुवाद को वैश्विक ज्ञान की अधिक सहायता मिलती है। हम देख सकते है कि अनुवादक अपनी बुद्धि की सहायता से किस प्रकार अनुवाद के लिए शब्दों को तोड़-मरोड कर शब्दों का अनुवाद करता है। जैसे –

हिंदी 1. नाश्ता किया और हाथ धो के बाहर चल दिया।

मराठी 2. नाश्ता केला आणि हाथ धुऊन बाहेर पडला।

हिंदी 1. यानी उसे यकीन हो गया कि लगी हुई गोली खाकर वह बचेगा नहीं
तो मौत से डर कर वह रोया या अस्पताल ले जाने के लिए
गिड़गिड़ाया नहीं।[32]

मराठी 2. गोळी खाल्यावर तो जगणार नाही अशी त्याचि खात्री झाल्यावर
मरणाला घाबरून तो रडला नाही की हॉस्पिटल मध्ये घेउन
जाण्यासाठी गयावया केला नाही.[33]

पहले वाक्य में 'चल' और 'दिया' दो भिन्न शब्द भी हो सकते हैं। यह शब्द अनेकार्थक शब्दों की कोटि में भी आते हैं लेकिन अनुवादक ने इन दोनों शब्दों को शब्दानुवाद की दृष्टि से

---

[32] कितने पाकिस्तान - कमलेश्वर, पृष्ठ संख्या - 59

[33] मराठी अनुवाद- किती पाकिस्तानं - अनुवादक - पद्माकर जोशी, राजपाल एंड सन्ज प्रकाशन दिल्ली, पृष्ठ संख्या - 69

न देख कर भावानुवाद के रुप में अनूदित किया है। दूसरे वाक्य में 'गोली' एवं 'जाने' शब्द का प्रयोग किया गया है। जिसका अनुवाद अनुवादक ने शब्दानुवाद के रुप में ही किया है लेकिन वाक्य अधिक बड़ा होने के कारण अनुवादक ने अनुवाद को केवल शब्दानुवाद तक ही सीमित नहीं रखा। जिसके कारण अनेकार्थक शब्दों के अनुवाद पर भी प्रभाव दिखाई देता है। अनेकार्थक शब्दों के अनुवाद में एक प्रमुख भाषा वैज्ञानिक समस्या दिखाई देती है। जैसे - हिंदी 1. हकीम साहब ने बिगड़ कर कहा ! ई का कर रहा है बे हरामज़ादे?

मराठी 2. हकीम साहेब संतापून म्हणाले - "हरामजादे, हे तू काय करतोस?

इस वाक्य में 'कर' शब्द का उपयोग दो बार किया गया है। जिसका अर्थ कर - करना, कर - हाथ, कर - टैक्स आदि होता है। मराठी अर्थ कर-करने, कर - हात, कर-महसूल, टैक्स है। लेकिन अनुवादक ने भाषा की उपयोगिता की दृष्टि से अनुवाद कर अनुवाद में प्रथम कर का 'संतापून' तथा दूसरे 'कर' का 'कर' शब्द के साथ 'तोस' शब्द को जोड़कर 'करतोस' शब्द बनाया है। यह क्रिया मशीन के लिए अत्यंत कठिन है। क्योंकि अनेकार्थक शब्दों में मानव अनुवाद करते समय शब्दों के रुपों को जिस प्रकार बदलकर प्रस्तुत करता है जैसे- कर के- करुन, दे कर-देऊन, जाते-जात, जाती-जाणारी, बिगड़ कर-संतापून, कर रहे-करणारे आदि। इस प्रकार के शब्दों का निर्माण मानव अनुवाद करते समय करता है। इस कार्य के लिए मशीनी अनुवाद यंत्र में 'शब्द निर्माण यंत्र' का निर्माण करना आवश्यक है जिससे अनेकार्थक शब्दों के निराकरण के साथ ही अनुवाद के अन्य शब्दों के अनुवाद की समस्या का भी समाधान हो सके।

भारत में अधिकतर मशीनी अनुवाद अंग्रेजी भाषा से हिंदी में अनुवाद के लिए कार्य कर रहे हैं। जिसमें से कुछ अनुवाद यंत्र उपयोग के लिए इंटरनेट पर भी उपलब्ध हैं। इन अनुवाद यंत्रों को अनुवाद के लिए कुछ अनेकार्थक शब्द युक्त वाक्य दिए गए। अनुवाद के उपरांत यह देखा गया कि कितने अनुवाद यंत्रों ने सही अनुवाद किया है। यह अनुवाद यंत्र मानव अनुवाद के समान ही गलतियाँ करते हैं? या इसमें भिन्नता पाई जाती है? क्योंकि मशीनी अनुवाद एक जटिल यंत्र है जिसमें अलग-अलग स्तरों पर पाठ की प्रक्रिया होती है। इसलिए इसमें कई गलतियों की संभावना है। जैसे - बी.कुमार संमुगम (B Kumara Shanmugam) के अनुसार "मशीनी अनुवाद एक जटिल प्रक्रिया है जिसमें भिन्न-भिन्न

स्तरों पर पाठ की प्रक्रिया होती है। अन्य भाषाओं की तरह ही तमिल भाषा में भी समान उच्चारण वाले शब्द पाए जाते हैं। अवश्यकता है कि इन शब्दों का कोश बना कर इन शब्दों को स्पष्ट किया जाए।" "Machine Translation is a complex system that does different kinds of processing over the text at various levels. As in other languages Tamil also contains polysemous words that require a collocation dictionary to disambiguate them."[34] ऐसी कई समस्याएँ मशीनी अनुवाद में पायी जाती है। जिसके निराकरण का कार्य अलग-अलग दृष्टि से करना आवश्यक है।

4.2 मशीन द्वारा होने वाले अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं के उदाहरण

मशीनी अनुवाद के निर्माण में होने वाली अनेकार्थक शब्दों की समस्या को देखने के लिए मशीनी अनुवाद यंत्र को अनेकार्थक शब्द के उदाहरण देकर मशीनी अनुवाद के द्वारा मिलने वाले लक्ष्य पाठ को देख कर ही दूर किया जा सकता है। इस कार्य के लिए कुछ उदाहरण एकत्रित किये गये हैं जो कि मशीनी अनुवाद के द्वारा अनूदित किए गए हैं। जिसका प्रमुख उद्देश्य मशीनी अनुवाद के निर्माण में होने वाले अनेकार्थक शब्दों की समस्या के संबंध में अनुसंधान करना एवं समस्याओं पर विचार करना है।

गूगल अंग्रेजी-हिंदी मशीनी अनुवाद यंत्र को दिए गए 'समान उच्चारण वाले भिन्नार्थक' लेकिन 'समान लिपि वाले शब्द' युक्त वाक्य निम्न रुप से हैं।

1.इन उदाहणों में गूगल अनुवाद यंत्र के साथ भारत में निर्माण किए गए मशीनी अनुवाद यंत्र भी हैं। यह अनुवाद मार्च2009 में मशीन से करवाया गया है।

1.The night looked dashing in his armour

रात अपने कवच में तेज नजर

The knight looked dashing in his armour

इस शूरवीर अपने कवच में तेज नजर

---

[34] B Kumara Shanmugam. 2002. *Machine Translation as related to Tamil*. AU-KBC Research Centre. M.I.T., Anna University, Chennai, India.

2. I have blond hair and blew eyes.

मैं ब्लॉन्ड बाल हैं और आँखें फूंका.

I have blond hair and blue eyes.

मैं ब्लॉन्ड बालों और नीली आँखों है.

3. Bambi's mother was a dear

बांबी की माँ एक बेटा था

Bambi's mother was a deer

बांबी की माँ एक हिरण था

4. He said he knew where the place was.

उसने कहा कि वह जहां जगह पता था कि कहा.

He said he new where the place was.

उसने कहा कि वह जहां जगह थी नया है.[35]

2.अनुसारका अंग्रेजी-हिंदी मशीनी अनुवाद यंत्र को दिए गए (अनेकार्थक शब्द) लेकिन 'समान उच्चारण वाले शब्द' युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च2009 को लिया गया है। इस अनुवाद यंत्र का फॉन्ट ठीक तरह से न आने के कारण इस अनुसंधान में अनुसारका मशीनी अनुवाद यंत्र के उदाहरण नहीं दिए जा रहे हैं। जैसे:-

The night looked dashing in his armour

1.1 à¤†à¤ªà¤•à¤¾ à¤µà¤¾à¤•à¥ à¤¯

2.1 à¤°à¤¾à¤¤ à¤‰à¤¸à¤•à¥‡ à¤•à¤µà¤š à¤®à¥‡à¤‚ à¤¦à¤¿à¤–à¥€ à¤¦à¥‡
à¤®à¤¾à¤°à¤¤à¥‡ à¤¹à¥ à¤

[35] http://translate.google.com/translate_t#en|hi|He

3.शक्ति अंग्रेजी-हिंदी मशीनी अनुवाद यंत्र को दिए गए अनेकार्थक लेकिन समान उच्चारण(Homophon)वाले शब्द युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च2009 को लिया गया है।

1.The night looked dashing in his armour
! r^aata ne usake kavacha men' de maaranaa dekhaa

The knight looked dashing in his armour
! yoddhe ne usake kavacha men' de maaranaa dekhaa

2.I have blond hair and blew eyes.
! mein' sunaharaa kesha huuZ oora aazkhen' ud'Zaa

I have blond hair and blue eyes.
! mein' sunaharaa kesha oora niilaa ran'ga aazkhen' huuZ

3. Bambi's mother was a dear
! bambi kii maataa priya thaa

Bambi's mother was a deer
! bambi kii maataa hirand-a thaa[36]

4.मंत्रा मशीनी अनुवाद यंत्र को दिए गए अनेकार्थक शब्द लेकिन 'समान उच्चारण वाले शब्द' युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च2009 में लिया गया है।

1.The night looked dashing in his armour

उनके कवच में तेज़ी से दौड़ना देखागया रात

The knight looked dashing in his armour

उनके कवच में तेज़ी से दौड़ना देखागया बहादुर

2.I have blond hair and blew eyes.

मैं सुनहला बाल और बलेव आँखें हैं।

[36] http://shakti.iiit.net/~shakti/cgi-bin/mt/translate.cgi

I have blond hair and blue eyes.

मैं सुनहला बाल और नील आँखें हैं|

3.Bambi's mother was a dear

बमबि के माँ प्रिय थे

Bambi's mother was a deer

बमबि के माँ हिरण थे[37]

5.मंत्र मशीनी अनुवाद यंत्र को दिए गए अनेकार्थक शब्द "समान उच्चारण"(homophon) वाले शब्द युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च2009 में लिया गया है।

1.The night looked dashing in his armour

रात उसका कवच में पटकती है

The knight looked dashing in his armour.

योद्धा उसका अर्मूरर्मर में पटकता है

2.I have blond hair and blew eyes.

मेरा सुनहरा भूरा केश हूँ और आँख ब्लू किये

I have blond hair and blue eyes.

मैं हैं सुनहरा भूरा केश और नीला आँखें

3.Bambi's mother was a dear

बैम्बी का माता एक प्रिय था

Bambi's mother was a deer

बैम्बी का माता एक हरिण था

4. He said he knew where the place was.

वह कहे वह क्ने किये वेर स्थान थे

[37] http://mantra-rajbhasha.cdac.in/mantrarajbhasha/index.html

He said he new where the place was.

उन्हो ने वह नया वेर स्थान था कहा[38]

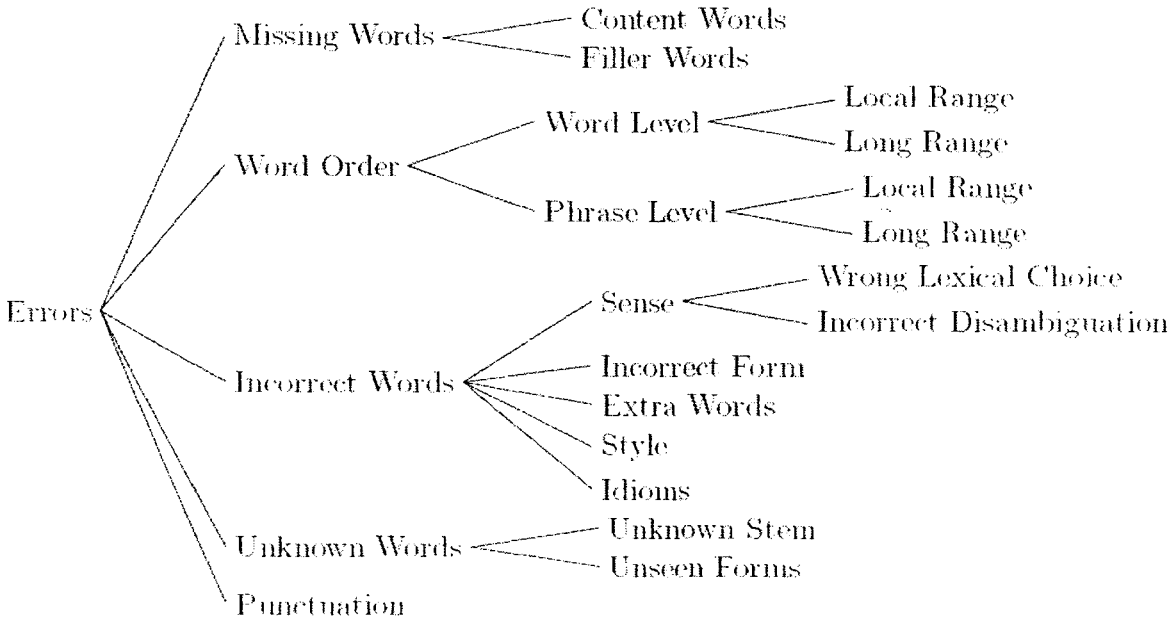
मशीनी अनुवाद में समान ध्वनि वाले शब्दों का सही अनुवाद किया जाता है, क्योंकि मशीन अधिकतर जगह आकृतियों को पहचानने में गलतियाँ नहीं करती। समान ध्वनि वाले शब्दों की अधिकतर आकृतियाँ यानी लिपि भिन्न-भिन्न होती है जिसके आधार पर मशीन शब्द की लिपि के अनुसार शब्द का अर्थ देकर अनुवाद कर सकती है। लेकिन समस्या उस समय उत्पन्न होती है जब समान ध्वनि वाले शब्दों को वाक्य में आये शब्दों के अनुसार अर्थों को स्पष्ट करना होता है। अंततः व्याकरणिक और अर्थ की दृष्टि से देखा जाए तो मशीनी अनुवाद के लक्ष्य पाठ में कई समस्याएँ दिखाई देती है। इन समस्याओं में से कुछ प्रमुख समस्याएँ निम्न रुप से देख सकते हैं। 1.शब्द चयन की गलतियाँ (Lexical inaccuracies) 2. समान अर्थ वाले शब्दों के अनुवाद में गलतियाँ 3. अनेकार्थक शब्दों की रचनात्मकता का उल्लंघन(Violated compositionality of multiword units) 4.अनेकार्थक शब्दों की इकाईयों का अशुद्ध अनुवाद(Mistranslated multiword units) 5. अज्ञात समकक्ष अनेकार्थक शब्दों के गलत अनुवाद(Wrong equivalents due to unidentified homonymy) 6. बगैर अनुवाद किए गए शब्दों(Untranslated words) का लक्ष्य भाषा में मिलना 7.व्याकरणिक गलतियाँ(Grammatical inaccuracies) 8. शब्दों या पदों की गलत सूचना (Wrong formation of words or phrases) 9. शब्दों की गलत संरचना (Wrong word form agreement) 10. पूर्वसर्गों का शून्य या गलत अनुवाद(Wrong or missing prepositions) 11. शब्दों का गलत क्रम (Inaccurate word order) 12. (Inadequate clause connection) 13. लक्ष्य भाषा के अतिरिक्त व्याकरणिक कार्य (Unusual grammatical form in target language) 14.निर्माण कार्य में गलतियाँ(Mistakes in government) सामान्य रुप से यह समस्याऐं दिखाई देती हैं।

इन समस्याओं पर विचार किया जाए तो मशीन गलतियाँ अधिक करती है। जैसे :- शब्द चयन में गलती करना। समस्याओं में यह नहीं कहा गया है कि मशीन पूर्णतः गलत अनुवाद कर रही है। मात्र मशीन के चयन में गलती है ऐसे समय निर्माण कर्ताओं का

[38] http://www.cdacmumbai.in/matra/index.jsp

काम यह होगा की चयन प्रक्रिया में सुधार लाया जाए जिससे शब्द का चयन अधिक प्रभावी रुप से हो सके। दूसरी समस्या पर ध्यान दिया जाए तो अनूदित शब्द का गलत क्रम में आना और वाक्य निर्माण में गलतियाँ, देखा जाए तो यह कार्य पूर्ण रुप से व्याकरणिक एवं भाषा वैज्ञानिक है। जिसका निराकण संपूर्ण मशीनी अनुवाद के व्याकरण निर्माण के कारण होता है। इस लिए यह कार्य तब तक पूर्ण होना असंभव हो जाता है जब तक संबंधित मशीन का व्याकरणिक स्वरुप लक्ष्य पाठ के अनुरुप न बनाया जाए।

शाब्दिक समस्याओं को अधिक सटीक रुप से स्पष्ट कर निराकरण पर विचार किया जाए तो वेलर एट एल (Vilar et al) एवं रेमकुट(Rimkut`e) / कोवलेव्सकाटे (Kovalevskat`e) ने निम्न वृक्ष संपादन के द्वारा मशीन में उत्पन्न होने वाली शब्दिक समस्याओं को स्पष्ट किया है। जिसका उपयोग अनेकार्थक शब्दों के अनुवाद में उत्पन्न होने वाली समस्याओं के निराकरण में सहायता करेगा।



चित्र संख्या - १[39]

[39] Vilar et al.& Rimkut`e and Kovalevskat`e, 2007,*Machine translation evaluation. Error analysis. Frameworks – example 1. (English to Lithuanian)*

मशीनी अनुवाद में अनेकार्थक शब्दों के अनुवाद में इन शब्दों की तुलना में समान लिपि वाले भिन्नार्थक शब्दों का अनुवाद करना सबसे अधिक चुनौती भरा कार्य है। जिसके कुछ उदाहरण निम्न रुप से है।

2. समान लिपि वाले भिन्नार्थक शब्दों के उदाहरण (Homographs example) -

1.गुगल हिंदी-अंग्रेजी मशीनी अनुवाद यंत्र को दिए गए अनेकार्थक लेकिन समान लिपि वाले शब्द युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च 2009 में लिया गया है।

1. Water travelled through ancient Rome through lead pipes.
   जल प्राचीन रोम के माध्यम से नेतृत्व पाइप के माध्यम से यात्रा की.
   The mother duck can lead her ducklings around.
   माँ बतख ducklings उसके आसपास का नेतृत्व कर सकते हैं.
2. The baby sat in awe at the bright colors on the mobile.
   शिशु रोब में मोबाइल पर चमकदार रंगों में शनिवार

Although most animals are mobile, the sponge is sessile.
यद्यपि अधिकांश पशु मोबाइल हैं, स्पंज बिना डंठल है.

3. All need to be present for a unanimous vote.
   सभी की जरूरत है एक सर्वसम्मत वोट के लिए उपस्थित होने के लिए.

I need to buy my sister a present for her birthday.

मैं अपनी बहन को उसके जन्मदिन के लिए एक उपहार खरीद करने की आवश्यकता है.

He who neglects the present moment throws away all he has

वह जो वर्तमान क्षण neglects दूर सब वह फेंकता है

He will present his ideas to the Board of Directors tomorrow.

उन्होंने कहा कि निदेशक मंडल की कल अपने विचार प्रस्तुत करेंगे.

4.The sow suckled her newborn piglets.

यह उसके नवजात piglets suckled बोना.

The farmer will sow oats in the back forty.

किसान बोना होगा पीछे चालीस में जई.

5.speaker please wind up your speech
वक्ता अपने भाषण को हवा कृपया
The wind blew from the northeast.
हवा के उत्तर पूर्व से उड़ा दिया.[40]

2.शक्ति मशीनी अनुवाद यंत्र को दिए गए अनेकार्थक लेकिन समान लिपि वाले शब्द युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च2009 में लिया गया है।

1.Water travelled through ancient Rome through lead pipes.
! paanii ne netRtva paaipon' men' se praachiina rome men' se yaatraa kiyaa

The mother duck can lead her ducklings around.

! maataa batakha ve batakha kaa bachce maar^a dikhaa sakataa hei idhara udhara

2.The baby sat in awe at the bright colors on the mobile.
! shishu mobile para chamakiile ran'gon' para vismayapuurnd-a aadara men' beit'haa

Although most animals are mobile, the sponge is sessile.
! haalaan'ki sabse adhika pashu phira bhii chalataa phirate phira bhii hein' phira bhii span'ja sessile Hei

3.All need to be present for a unanimous vote.
! sabhii zaruurata honaa ekamata mata kaa honaa

I need to buy my sister a present for her birthday.
! mein' zaruurata honaa merii bahana vaha janmadina kaa khZariidanaa

4.The sow suckled her newborn piglets.
! sov ne ve navajaata ghen't'e stana se duudha pilaaye

[40] http://translate.google.com/translate_t#en|hi|speaker

The farmer will sow oats in the back forty.
! kisaana oat boegaa mein' piiche kaa forty

5. speaker please wind up your speech.
! vaktaa aapakii bolii para havaa khusha karataa hei

The wind blew from the northeast.
! havaa ud'Zii se uttara puurva[41]

3.मंत्रा मशीनी अनुवाद यंत्र को दिए गए अनेकार्थक शब्द लेकिन समान लिपि वाले शब्द युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च2009 में लिया गया है।

1 Water travelled through ancient Rome through lead pipes.

प्राचीन रोम के माध्यम से यात्रा किया गए जल के माध्यम से पाइपों को नेतृत्व करते हैं|

The mother duck can lead her ducklings around.

माँ बतख करीब उनके बत्तखों नेतृत्व कर सकते हैं|

2.The baby sat in awe at the bright colors on the mobile.

मोबाइल पर चमकीला रंग में बच्चा शनिवार में रोब|

Although most animals are mobile, the sponge is sessile.

सबसे जानवरों यद्यपि मोबाइल, स्पंज हैं अवृन्त है|

3.All need to be present for a unanimous vote.

सब एकमत मत के लिए वर्तमान आवश्यकता होगी|

I need to buy my sister a present for her birthday.

अपनी बहन खरीदाने के लिए मैं जरूरत उनके जन्मदिन के लिए को वर्तमान|

4.The sow suckled her newborn piglets.

उनके नवजात शिशु सूअर का बच्चे को स्तन पिलाने वाली बोना|

[41] http://shakti.iiit.net/~shakti/cgi-bin/mt/translate.cgi

The farmer will sow oats in the back forty.

कृषक पीठ फोर्ती में जइयाँ बोताेंगे|

5.speaker please wind up your speech

अध्यक्ष हवा कृपया आपके भाषण बढ़ाकर

The wind blew from the northeast.

उत्तरपूर्वी से बंद बलेव|[42]

४.मंत्र मशीनी अनुवाद यंत्र को दिए गए अनेकार्थक लेकिन समान लिपि वाले शब्द युक्त वाक्य और अनूदित वाक्य निम्न रुप से हैं। यह अनुवाद मार्च 2009 में लिया गया है।

1. Water travelled through ancient Rome through lead pipes.

वैटर ने प्राचीन रोम से नेतृत्व नलिके से यात्रा किया

The mother duck can lead her ducklings around.

माता बतख उसका बतख का बच्चों का लगभग का नेतृत्व कर सकते हैं

2. The baby sat in awe at the bright colors on the mobile.

शिशु ने श्रध्दायुक्त भय में चमकीला कलौर्ज़ पर चलता फिरता पर सैट किया

Although most animals are mobile, the sponge is sessile.

अल्दो मोस्ट पशुएं चलता फिरता स्पंजें हैं सेसाइल कर रहे हैं

3.All need to be present for a unanimous vote.

सब आवश्यकता एक एकमत चुनाव के लिये उपस्थित है

I need to buy my sister a present for her birthday.

मैं मेरा बहन उसका जन्मदिन के लिये एक वर्तमानकाल खरीदती हूँ

He who neglects the present moment throws away all he has

वह कौन उपेक्षा करते हैं उपस्थित क्षण फेंकते हैं दूर सब वह हैं

[42] http://mantra-rajbhasha.cdac.in/mantrarajbhasha/index.html

He will present his ideas to the Board of Directors tomorrow.

वह बोर्ड को निर्देशक का उसका विचार कल पेश करेंगे

4. The sow suckled her newborn piglets.

बो सक्ल्ड उसका नवजात घेंटे बो सक्ल्ड उसका नवजात घेंटे

The farmer will sow oats in the back forty.

किसान पीछे का चालीस में ओट्ज़ बोएगा

5.speaker please wind up your speech

स्पीकर प्लीस करते हैं करते हैं तुम्हारा वाणी

The wind blew from the northeast.

हवा ने उत्तर पूर्व से ब्लू किया [43]

उपर्युक्त वाक्यों में प्रमुखत: पाँच शब्दों को लिया गया है। जिनके एकाधिक अर्थ है। इन वाक्यों को पाँच भिन्न मशीनी अनुवाद यंत्रों को दिया गया। इन मशीनी अनुवादों को पाठक की दृष्टि से देखा जाए तो किसी भी मशीनी अनुवाद का लक्ष्य पाठ अधिक प्रभावी नहीं है। सभी अनुवाद यंत्रों की वाक्य संरचना, शब्द चयन और अर्थ में भिन्नता पायी जाती है। इसके बावजूद कोई भी मशीनी अनुवाद यंत्र लक्ष्य पाठ का मूल अर्थ देने में सफल नहीं हो सका। इस समस्या के कारण यह भी स्पष्ट नहीं हो पाता कि किस मशीनी अनुवाद यंत्र ने सबसे अच्छा अनुवाद किया है। लेकिन अनुसंधान ने इस कार्य में अनुवाद की योग्यता पर अधिक विचार न कर अधिकतर विचार अनेकार्थक शब्दों के अर्थ पर केंद्रित किया है। अनुवाद में प्रथम बात यह स्पष्ट है कि मशीन को दिए गए दो वाक्यों में प्रथम का अर्थ सही है। लेकिन मशीन जब दूसरे अनेकार्थक शब्दों के दूसरे वाक्य का अनुवाद करती है तो वाक्य के साथ अनुवाद किए गए शब्द का अर्थ नहीं बदलता। जिसके कारण अनुवाद गलत होता है।

जैसे :- lead – शब्द का अनुवाद अधिकर म.अनु.यंत्रों ने 'नेतृत्व' ही किया है जबकी प्रथम वाक्य में नेतृत्व यह अनुवाद सही है। लेकिन दूसरे वाक्य में 'मार्ग दिखाना' 'या रास्ता

[43] http://www.cdacmumbai.in/matra/

दिखाना' होना चाहिए। दूसरा उदाहरण 'wind' शब्द का अनुवाद 'हवा' है लेकिन 'wind' के बाद जब 'up' आता है तो शब्द का अर्थ बदल कर 'समेटना' हो जाता है। जो मशीन नहीं कर पाई है। अर्थ को स्पष्ट करने से पहले यह देखना जरुरी हो जाता है कि दिए गए पाठ में कितने अनेकार्थक शब्द आये हैं और किन शब्दों के साथ आये हैं। दूसरी समस्या पर विचार किया जाए तो 'present' शब्द का अनुवाद 'प्रस्तुत', "वर्तमान", 'उपहार' है। इस शब्द के तीन वाक्यों में 'मंत्र' मशीनी अनुवाद ने एक अनुवाद सही किया है। दूसरे और तीसरे वाक्यों के अनुवाद में मशीन ने तीसरे शब्द का अनुवाद दूसरे वाक्य के लिए किया है। और दूसरे शब्द का अनुवाद गलत किया है। इससे स्पष्ट है कि मशीन तीसरा वाक्य भी अनूदित करने में सफल हो जाती अगर वाक्य चयन में गलती न होती। लेकिन गूगल मशीनी अनुवाद में चार वाक्य दिए गए जिसमें से पहला, तीसरा और चौथा सही अनूदित है लेकिन दूसरा वाक्य गलत अनूदित किया गया है। शब्दकोश में गलत शब्दों को दिए जाने की वजह से भी अनुवाद गलत हुआ है। जैसे – मोबाईल शब्द के लिए 'चलत फिरता' का उपयोग किया गया। तो कुछ यंत्रों में शब्द संचयन ठीक तरह नहीं हो पाया है। जैसे:- 'sow' शब्द के लिए 'बोताेंगे' इस प्रकार की गलती भी हो गई है। इन तकनीकी समस्याओं से भी म.अनु.यंत्र को उबारना जरुरी है। लगभग सभी मशीनी अनुवाद यंत्रों में यह तकनीक अपनाई जाती है कि जो शब्द मशीनी अनुवाद यंत्र के शब्द संचयन में उपलब्ध नहीं होते तब उस शब्द का लिप्यंतरण किया जाता है। लेकिन अनेकार्थक शब्दों के संबंध में यह तकनीक उपयुक्त नहीं है। अनेकार्थक शब्द अन्य शब्दों की तुलना में अर्थ की दृष्टि से अधिक संवेदनशील एवं महत्वपूर्ण होते हैं। जिससे लिप्यंतरण की तकनीक को इन शब्दों से अलग रखना ही अधिक उपयुक्त होगा। इन समस्याओं के साथ ही 'समान उच्चारण में आने वाली' अधिकतर समस्याऐं 'समान लिपि वाले भिन्नार्थक शब्दों के संदर्भ में भी आते हैं।

मानव अनुवाद एवं मशीनी अनुवाद के द्वारा होने वाले अनेकार्थक शब्दों के अनुवाद में स्पष्ट रुप से भिन्नता दिखाई देती है। जहाँ मानव विचार पूर्वक इन शब्दों का आकलन कर अनुवाद करता है। वहा मशीन को मशीन में दिए गए डेटाबेस, कार्पस और प्रोग्रामिंग पर ही निर्भर(depend) रहना पड़ता है। मानव अनुवाद के दौरान शब्दों का निर्माण कर

सकता है लेकिन यह कार्य मशीन के लिए अब तक सम्भव नहीं हो पाया है। जिससे अनुवाद में गंभीर रुप की समस्याएँ दिखाई देती हैं। अधिकतर जगह मानव अनुवाद में समान उच्चारण वाले शब्द आसानी से अनूदित किए जाते हैं, वही कार्य करने के लिए मशीनी अनुवाद में इस कार्य में गलतियाँ दिखाई देती हैं। मानव अनुवाद में किसी भी जगह अनेकार्थक शब्दों का अनुवाद उसी भाषा में यानी हिंदी भाषा का अनुवाद हिंदी भाषा में ही नहीं मिलता। लेकिन मशीनी अनुवाद में यह समस्या स्पष्ट रुप से दिखाई देती है। मशीन को जो शब्द नहीं मिल पाता उस शब्द को मशीन उसी भाषा की लिपि में प्रस्तुत करती है। यह मशीन की अकुशलता के सबसे बड़े उदाहरण के रुप में सामने आता है। मशीनी अनुवाद की इन समस्याओं के अनुसंधान के पश्चात यह बात स्पष्ट रुप से सामने आती है कि जिस प्रकार एक मानव अनेकार्थक शब्दों का अनुवाद करता है। उसी प्रकार मशीन भी अनुवाद कर सकती है। मात्र निर्देश एवं शब्द स्पष्टता एवं मशीन में व्याकरणिक सिद्धांतों का योगदान उचित होना चाहिए।

तीसरा अध्याय

हिंदी-मराठी मशीनी अनुवाद में समान उच्चारण वाले शब्दों की समस्याओं का भाषिक सूत्र

तीसरे अध्याय का सार

1.हिंदी में समान उच्चारण वाले शब्द का प्रचलन और समान उच्चारण वाले शब्द का विभाजन

2.समान लिपि, समान उच्चारण, भिन्नार्थक शब्द

3. भिन्न लिपि, समान उच्चारण, भिन्नार्थक शब्द

4.अंग्रेजी से हिंदी में आये समान उच्चारण वाले शब्द

तीसरे अध्याय का सार

अध्याय के प्रथम चरण में यह स्पष्ट किया गया है कि हिंदी में अनेकार्थक शब्दों का होना नवीन नहीं है। यह परंपरा कबीर, तुलसी के समय से ही यानी हिंदी के शैशव अवस्था में ही सामने आ गयी थी। जिसका नवीनतम रुप हिंदी में अन्य भाषाओं से आये ऐसे शब्द हैं जो हिंदी में अनेकार्थकता का निर्माण करते है। जिसके कई उदाहरण स्पष्ट रुप से यहां दिए गए हैं। यह बात भी स्पष्ट है कि इन शब्दों का उपयोग मात्र भाषा की व्यंजना एवं भाषा में रुची निर्माण करने के उद्देश्य से ही किया जाता था। जिस प्रकार अन्य भाषाओं में अनेकार्थक शब्दों के निराकरण के लिए कोई विशिष्ट सूत्र का निर्माण नहीं हो पाया उसी प्रकार हिंदी में भी अनेकार्थक शब्दों के निराकरण के लिए पुराने ग्रंथों(व्याकरणिक) में कोई सूत्र नहीं मिलता है। इस अध्याय में समान उच्चारण वाले भिन्नार्थक शब्दों का विभाजन चार भागों में किया गया है। 1.हिंदी में समान उच्चारण वाले शब्द का प्रचलन और समान उच्चारण वाले शब्द का विभाजन 2.समान लिपि, समान उच्चारण, भिन्नार्थक शब्द और 3. भिन्न लिपि, समान उच्चारण, भिन्नार्थक शब्द। 4. अंग्रेजी से हिंदी में आये समान उच्चारण वाले शब्द। इस विभाजन के आधार पर हिंदी में उत्पन्न होने वाले समान उच्चारण वाले शब्दों का हिंदी-से-मराठी में अनुवाद किया गया है। जिसका प्रमुख उद्देश्य हिंदी-मराठी मशीनी अनुवाद यंत्र में अनेकार्थक शब्दों का अनुवाद करने की संकल्पना है। इन शब्दों के संकलन पर विस्तृत चर्चा भी की गई है। इसी के साथ हिंदी से मराठी में अनुवाद कर समान अर्थ वाले शब्दों को दे कर अर्थों को स्पष्ट किया गया है। समान उच्चारण वाले शब्दों को व्याकरणिक कोटियों की सहायता से स्पष्ट किया गया है, जिससे शब्द का व्याकरणिक अर्थ भी स्पष्ट हो गया है। मशीनी अनुवाद में POS टैगर की आवश्यकता को भी स्पष्ट किया गया है। यह पाठ हिंदी-मराठी मशीनी अनुवाद में प्राकृतिक भाषा संसाधन के भाषिक संसाधनों में अनेकार्थक शब्दों की अस्पष्टता को दूर करने के लिए भाषा साधन(Language Tool) के निर्माण में लघु शब्दकोश के रुप में महत्वपूर्ण योगदान देगा। इस प्रकार देखा जाए तो यह पाठ इस अनुसंधान का महत्वपूर्ण अंग है।

1.हिंदी में समान उच्चारण वाले शब्द का प्रचलन और समान उच्चारण वाले शब्द का विभाजन

हिंदी से मराठी अनुवाद में किस प्रकार अनेकार्थक शब्द अस्पष्टता उत्पन्न करते हैं और भाषा में आसानी के साथ घुल-मिल जाते है। इसका उदारण हिंदी और मराठी में मिलने वाले समान लिपि, समान उच्चारण, भिन्न एवं समान अर्थ वाले शब्दों के रुप में देख सकते है। इसके तीन प्रमुख कारण दिखाई देते हैं। 1.हिंदी एवं मराठी का एक ही भाषा परिवार से उत्पन्न होना 2. दोनों भाषाओं की समान लिपि एवं समान ध्वनि उच्चारण व्यवस्था 3. संस्कृति की अधिकतर समानता। यह कार्य भाषा -विज्ञान के व्यतिरेकी भाषा-विज्ञान के अंर्तगत आता है। लेकिन इस भाषिक विश्लेषण की समानता अनेकार्थक शब्दों की समस्याओं से अधिक भिन्न नहीं दिखाई देती। इस भाषिक विश्लेषण में केवल अर्थ में समानता और भिन्नता के कारण दोनों कार्यों की अनुसंधान प्रक्रिया भिन्न हो जाती है। यह स्पष्ट करना इसलिए भी महत्व पूर्ण है कि कई बार व्यक्ति अनेकार्थक शब्दों पर हो रहे अनुसंधान को व्यतिरेकी भाषा विज्ञान की कोटि में प्रविष्ट कर देखता है लेकिन यह कार्य व्यतिरेकी भाषाविज्ञान के कार्य से भिन्न है। अनेकार्थक शब्द केवल द्विअर्थकता ही निर्माण नहीं करते बल्कि कई विद्वानों का यह भी मानना हैं कि इन शब्दों के कारण भाषा में एक लचीले पन निर्माण होता है।[1] इन शब्दों के कारण भाषा में एक प्रकार की भाषिक शैली का निर्माण होता है। जैसे इन शब्दों के कारण कई व्यंग्यकार अपनी बात व्यंग्यात्मक रुप से कहने में सफल हो जाते है। जिससे अनेकार्थक शब्दों से व्यंग्यात्मक शैली को अधिक से अधिक फायदा होता है। भारत में मशीनी अनुवाद पर कार्य कर रहीं अक्षर भारती टीम ने मशीनी अनुवाद में निर्माण होने वाले अनेकार्थक शब्दों को दो भागों में विभाजित किया है। 1.पूर्णत: अनेकार्थक शब्द (complet) 2. आंशिक(partial) अनेकार्थक शब्द।[2] समाज में सामान्यत: अधिकतर जगह द्वंदात्मकता के समय ऐसे शब्दों का उपयोग देखा जाता है। जिसमें यह भी कहा जाता है कि मेरे कहने का अर्थ यह नहीं था आप दूसरा अर्थ समझ रहे हो। जैसे:- हाल ही के दिनों में क्रिकेट जगत में हुआ भारतीय क्रिकेटर हरभजन सिंह और(आस्ट्रेलियन क्रिकेटर) एंड्रयू सायमंड के बीच हुआ झगड़ा। जिसमें हरभजन सिंह ने एंड्रयू सायमंड को मंकी कहा। लेकिन

---

[1] सिन्हा, राजेंद्र प्रसाद, *शुध्द हिंदी कैसे सीखें,* भराती भवन प्रकाशन, पटना, पृष्ठ सं.- 431-432

[2] Akshr Bharati - Dipti Misra-sharma, Amba Kulkarani, Vinit Chaitanya, Anusaraka:an approach to machine Translation

बाद में यह कहा गया की हरभजन सिंह 'मंकी' नहीं 'माँ की' कह रहे थे।[3] पंजाबी भाषा के उच्चारण के कारण यह 'मंकी' सुनाई दिया। जो एक प्रकार का होमोफोन है। कई नेता गण अपने भाषण के समय भी इन शब्दों का जान बूझकर उपयोग करते है। लेकिन जब बात नहीं बन पाती तो स्पष्ट रुप से कह देते है कि मेरे कहने का अर्थ यह नहीं था। आप दूसरा अर्थ समझ रहे हैं। इस प्रकार देखा जाए तो अनेकार्थक शब्दों का उपयोग चाहे अन चाहे रुप से समाज के कई लोग करते हैं। शास्त्रीय पद्धति से इन शब्दों का निरपेक्ष रुप से संशोधन करने के लिए मशीन से अधिक निरपेक्ष वस्तु कोई नहीं हो सकती। इस कार्य के लिए द्वि अर्थक शब्द स्पष्टता के भाषिक संसाधन का निर्माण करना आवश्यक है। यह स्थिति मशीनी अनुवाद से भिन्न नहीं है। इस कार्य का प्रयोग मशीनी अनुवाद के निर्माण के साथ करना इस लिए उपयोगी होगा कि अनुवाद का प्रथम नियम यही है कि किसी भी पाठ का अनुवाद निरपेक्ष रुप से होना चाहिए।[4] अनुवादक को किसी भी प्रकार का पक्षपाती काम नहीं करना चाहिए। मानव एक संवेदन शील प्राणी है, धर्म, रिश्ते-नाते, समाजिक दबाव के कारण अनुवादक पक्षपाती काम करने पर बाध्य हो जाता है। लेकिन मशीन एक असंवेदन शील वस्तु है अतः महत्वपूर्ण एवं संवेदन शील कार्यों में निष्पक्ष रुप से किसी पाठ का विश्लेषण एवं अनुवाद करने के लिए मशीन ही सर्वोत्तम साधन है।

हिंदी में अनेकार्थक शब्दों का मिलना साधारण बात है। हिंदी के कई पुराने ग्रंथों एवं कविओं की कविताओं में अनेकार्थक शब्द मिलते हैं। इसी आधार पर हिंदी अलंकारों में अनेकार्थक शब्दों के संदर्भ में चर्चा करने वाला एक अलंकार मिलता है। इस अलंकार का उपयोग केवल भाषा के सौंदर्य को बढ़ाना है। इस अलंकार में अनेकार्थक शब्दों के निराकरण की कोई चर्चा नहीं की जाती। यमक अलंकार "हिंदी में समान ध्वनि और समान लिपि लेकिन भिन्न अर्थ को लेकर एक विशिष्ट अलंकार का निर्माण करता है। " 'यमक' अलंकार की परिभाषा निम्न रुप से है। "यमक अलंकार में एक ही आकार वाला शब्द(स्वर-व्यंजन-समुदाय) कम से कम दो बार आता है, पर भिन्न-भिन्न अर्थो में इन शब्दों का अलंकार में उपयोग होता है।[5] पाठक को चकित करना और उसका ध्यान शब्द के अर्थ की

---

[3] http//www.srajaram.com/2008/01/did-harabhajan-say-maaki-or-monkey-to.html

[4] अनुवाद विज्ञान सिद्धांत और अनप्रयोग - डॉ.नगेंद्र

[5] सिन्हा, राजेंद्र प्रसाद, *शुध्द हिंदी कैसे सीखें,* भराती भवन प्रकाशन, पटना, पृष्ठ सं - 431

भिन्नता या सुन्दरता की और आकृष्ट करना इन शब्दों का उद्देश्य होता है।" इस अलंकार के उदाहरण हम निम्न रुप से देख सकते है।

"कबिरा सोई पीर है, जे जाने पर पीर।

जे पर परी न जानई, सो काफिर बेपीर॥"[6]

इस साखी में 'पीर' शब्द का अर्थ प्रत्येक समय भिन्न-भिन्न अर्थो के साथ प्रकट होता है। पहले 'पीर' का अर्थ 'पैगम्बर' के अर्थ में आता है। दूसरे 'पीर' का अर्थ 'दुख:' के संदर्भ में आता है और तीसरे 'पीर' का अर्थ 'निर्दय' के अर्थ में आता है। इस प्रकार एक ही 'पीर' शब्द का तीन भिन्न अर्थों के साथ उपयोग ही इस अलंकार का निर्माण करता है।

"कनक कनक ते सौ गुनी मादकता अधिकाय।

उहि खाये बौराय नर, इहि पाये बौराय ॥"[7]

इस उदाहरण में 'कनक' शब्द का उपयोग दो बार भिन्न-भिन्न अर्थों में किया गया है। पहले 'कनक' शब्द का उपयोग 'धतूरा' धतूरे का फूल इस अर्थ को प्रकट करता है। दूसरे 'कनक' शब्द का अर्थ 'सोना' सोना धातु के संदर्भ में किया गया है। इस प्रकार देखा जाए तो हिंदी में समान उच्चारण वाले शब्दों पर अलंकार के माध्यम से पहले ही इन शब्दों के संबंध में सोचा गया लेकिन केवल इसका उपयोग मात्र भाषा के सौंदर्य के रुप में ही किया गया। आज भी कविताओं में 'यमक' अलंकार का उपयोग किया जाता है। इन शब्दों के निराकरण के लिए कोई विशेष नियम प्रस्थापित नहीं किया गया है। भाषा सौंदर्य शास्त्र में केवल इन शब्द अलंकारों के उपयोग में आये शब्दों पर टिप्पणी देकर शब्दों के दोनों अर्थों को स्पष्ट किया जाता है। यह कार्य मात्र स्पष्टीकरण के रुप में किया जाता है। हिंदी में निर्माण होने वाली यह समस्या केवल हिंदी तक ही सीमित नहीं रह जाती। जब किसी भाषा का विकास होता है तो एक भाषा में लिखा हुआ साहित्य दूसरी भाषाओं में अनुवाद के माध्यम से आता है। अनुवाद की प्रक्रिया के दौरान ऐसे शब्दों के कारण विशेष समस्या होती है। इसलिए साहित्यिक दृष्टि से भी अनेकार्थकता की समस्या का समाधान होना आवश्यक है। इस अनुसंधान में अनुवाद के लक्ष्य भाषा के रुप में मराठी को रखा गया है इसलिए इस प्रक्रिया

---

[6] वहीं पृष्ठ संख्या-431

[7] वहीं पृष्ठ संख्या-432

में हिंदी से मराठी मशीनी अनुवाद पर हो रहे अनुसंधान में यह जानना महत्वपूर्ण होगा कि मराठी भाषा की पृष्टभूमि क्या रही है।

देश के ही नहीं बल्कि विश्व मंच पर मराठी भाषा का 15वां स्थान है।[8] मराठी महाराष्ट्र राज्य की राजभाषा है। आज महाराष्ट्र नाम से जिस प्रदेश को हम जानते हैं उसकी व्याप्ति प्राचीन समय से बदलती रही है। दंडकारण्य, महाकान्तार, महाकवि उसके प्राचीन नाम है।[9] सन 300 के पूर्व इस महान अरण्यमय प्रदेश में लोग बसने लगे। सन 505 में लिखे गये वराहमिहिर के ग्रंथ में तथा सन 611 में सत्याश्रम पुलकेशी द्वारा स्थापित बद्दमीके शिलालेख में महाराष्ट्र का प्रथम उल्लेख मिलता है।[10] जिस प्रदेश को आज महाराष्ट्र कहा जाता है, उसे दक्षिण पथ भी कहा गया है। मातृभाषा की लोक जनगणना के अनुसार मराठी विश्व की 15वीं व भारत की 4थे नंबर की सबसे अधिक बोली जाने वाली भाषा है।[11] मराठी भाषा के उत्पत्ति के संबंध में कई मत प्रचलित हैं। प्राकृत भाषाओं का निर्माण काल दूसरी या तीसरी शताब्दी है। इन्हीं प्रकृतों, महाराष्ट्री, अर्धमागधी, शौरसेनी, पैशाची में से मराठी की उत्पत्ति हुई, ऐसी कुछ लोगों की धारणा है।[12] सन 1188 में मराठी के अदि ग्रंथकार मुकुंन्दराज का विवेकसिंधु मिलता है। उसके बाद सन 1290 में ज्ञानेश्वरजी ने 'ज्ञानेश्वरी' लिखी। अतएव भाषा तत्व की दृष्टि से हम यह अवश्य कह सकते हैं कि ग्यारहीं शताब्दी के पूर्व ही मराठी भाषा लोक व्यवहार की भाषा बन गई थी। शिलालेखों तथा ताम्र पठों में खुदे हुए वाक्यों की भाषा से और शब्दों से यह स्पष्ट होता है। यह ताम्रपट शक - 602, 675, 660, 735, 757, 810, 832 और इ.स.900 तक के मिले है। जिसमें 'पन्नास, प्रिथिवी, देऊळवाडे, बारिखेड, खळखळ, येज्ञ, स्थिति, वदुरें, और साम्राज्य' जैसे मराठी शब्द मिलते है।[13] शक 905 (सन 983) में मैसूर रियासत के 'श्रवणबेळगोळ' नामक स्थान में जो शिलालेख मिला है, उनमें मराठी का रुप इस प्रकार है।

---

[8] www.wikipedeia/marathi-lang/

[9] जोगलेकर, न.चि., प्रथम संस्करण-1951,*मराठी का वर्णनात्मक व्याकरण,* प्रकाशक-राष्ट्रभाषा प्रचार समिति, वर्धा,पृष्ठ सं.-11

[10] वहीं पृष्ठ संख्या-11

[11] जोगलेकर, न.चि., प्रथम संस्करण-1951,*मराठी का वर्णनात्मक व्याकरण,* प्रकाशक-राष्ट्रभाषा प्रचार समिति, वर्धा,पृष्ठ सं.-11

[12] वहीं पृष्ठ संख्या - 11

[13] वहीं पृष्ठ संख्या - 12

श्री चामुंडराये करवियले ।

श्री गंगाराये सुत्ताले करवियले ।।

'ज्ञानेश्वरी' के एक वर्ष पूर्व सन 1289 में खुदा हुआ एक शिलालेख निजाम रियासत के उनकेश्वर नामक स्थान में मिलता है। इसके लेख की भाषा संस्कृत मिश्रित है। जैसा की हमें इस संकल्पना से ज्ञात है कि भारत के सभी क्षेत्रों में बोलियों का विशेष महत्व होता है, मराठी में भी अनेक प्रकार की बोलियाँ मिलती है। जैसे एक कहावत है "चार कोस पर पानी बदले आठ कोस पर वाणी" इसी के अनुरुप महाराष्ट्र में कई बोलियाँ मिलती है। जिसे हम इस प्रकार देख सकते हैं। 1.पुणे-सतारा के आसपास की मराठी जो मध्य मराठी कहलाती है। 2. विदर्भ-नागपुर के उत्तरी प्रदेश में बोली जानेवाली 'वऱ्हाडी' 3.दक्षिण में विशेषता: कोंकण में बोली जाने वाली कोंकणी भाषा है। इसी के साथ खानदेशी, नागपुरी, चित्पाननी, कोल्हापुरी, कुडाळी, कुणबाऊ, परभी, ठाकरी, धनगरी और मावळी आदि और भी बोलियाँ हैं। इस प्रकार हमने देखा कि अपनी-अपनी बोलियों से विकसित मराठी महाराष्ट्र की राजभाषा भाषा है। जहाँ तक मराठी के साहित्यिक भाषा एवं प्रशासनिक भाषा का सवाल है पुणे के आसपास बोली जाने वाली मराठी को ही साहित्यिक एवं प्रशासनिक भाषा के रुप में उपयोग किया जाता है। मराठी भाषा का इतिहास और साहित्य समृद्ध और परिपूर्ण है। इक्कीसवीं शताब्दी के दौर में मराठी भाषा का तकनीकी विकास तेजी से हो रहा है। भारतीय भाषाओं में मराठी के अन्यन्य साधारण महत्व को देखते हुए हिंदी-मराठी में तकनीकी विकास की आवश्यकता पर भारत सरकार भी कार्य कर रही है।

भाषा हमेशा अपने युग के अनुसार समाज के वैचारिक, सामाजिक तथा राजनीतिक उत्थान और पतन से प्रभवित होती रहती है। इसकी रुपरेखा हमें जनसमाज की भाषाओं, विचारों तथा भाषा की शैली के दर्पण से मिलती है। यह दर्पण रुपी प्रतिबिंब ही हमें भाषा के व्याकरण और भाषा विज्ञान की भाषिक प्रक्रिया के संबंध में अनुसंधान का विषय बना देता है। इसलिए प्रत्येक कालखंड और उसकी प्रमुख विशेषताओं से परिचित होना आवश्यक हो जाता है। जिससे हम उस भाषा की बानगी देखने में समर्थ हो सकेंगे। इस कार्य पर यह सवाल उठ सकता है कि हिंदी-से-मराठी में मशीनी अनुवाद क्यों? भारत में कई भाषायें एक साथ बोली और पढ़ाई जाती हैं। जिनमें कुछ ऐसी भी भाषाऐं हैं जो एक ही भाषा परिवार से आती हैं। जिसमें कुछ आर्य भाषा परिवार की और कुछ दक्षिण भारत

की भाषायें आती हैं। मराठी एवं हिंदी ऐसी भाषायें हैं जिनकी उत्पत्ति एक ही भाषा परिवार से हुई है। केवल यह एक ही कारण नहीं जो इस अनुसंधान को आगे बढ़ाने में मदद करेगा बल्कि दोनों भाषाओं की लिपि भी समान है। भाषा के अक्षरों के उच्चारण में अधिकतर समानता मिलती है। दोनों भाषाओं का व्याकरणिक ढाँचा भी अधिकतर समान है। संस्कृति में मिलने वाली अधिकतर समानता भी एक महत्वपूर्ण कारण है। इसी के साथ भारत एक संघ राज्य है जिससे संविधान में निर्माण होने वाले कानून के पालन की समान धाराओं को भी सहायता के रुप में देख सकते हैं। जिसका दोनों भाषाओं पर समान रुप से प्रभाव दिखाई देता है। आज जिस प्रकार हिंदी ने कई अंग्रेजी शब्दों को अपनी भाषा में जगह दी है और अन्य भाषाओं के शब्दों को अपनाया है, उसी प्रकार मराठी में भी कई अंग्रेजी एवं अन्य भाषाओं के शब्द पाये जाते हैं। इन समानताओं के कारण ही आई.आई.टी.मुंबई की संस्था को हिंदी एवं मराठी के तकनीकी पक्ष को समान रुप से देखने के लिए एक ही जगह काम करने की अनुमती दी गई है।[14] जिससे इस कार्य में सहायता हो सके। इसकी कई सफलताओं को आज हम इंटरनेट पर भी देख सकते हैं। इन कुछ महत्व पूर्ण कारणों से हिंदी से मराठी में हो रहे कार्य को आगे बढ़ाने की दृष्टि से हिंदी-मराठी मशीनी अनुवाद की अनेकार्थक शब्दों की समस्याओं के निराकरण का कार्य महत्व पूर्ण हो जाता है। इस कार्य के लिए यह जानना भी महत्व पूर्ण हो जाता है कि मराठी भाषा का तकनीकी दृष्टि से कितना विकास हो चुका है।

अन्य भाषाओं की तुलना में मराठी के भाषा संसाधनों का अधिक विकास हुआ है। भारत सरकार की कई योजनाओं के कारण प्रौद्योगिकी के कई नामी संस्थाओं में यह कार्य नियमित रुप से आकार ले रहा है। (TDIL) भारतीय भाषाओं के लिए प्रौद्योगिकी विकास की विश्वभारत पत्रिका के अनुसार एवं आई.आई.टी.मुंबई की वेब साईट पर मिलने वाली जानकारी के अनुसार मराठी भाषा पर काम करने के लिए आई.आई.टी.मुंबई(2000), सी-डॅक,मुंबई, सी-डॅक,पुणे को काम दिया गया IIT Bombay, CDAC Mumbai and CDAC Pune from Maharastra are playing leading roles in these endeavours.[15] जिसमें प्रमुख रुप से (i) शब्दकोश निर्माण जैसे शब्दजाल (lexical resources like *wordnets), ऑन्टोलॉजी एवं अर्थ अधारित मशीन वाचक शब्दकोश (ontologies* and

---

[14] http://www.cfilt.iitb.ac.in

[15] Dr. Bhattacharyya, Pushpak, *White Paper on Cross Lingual Search and Machine Translation* Department of Computer Science and Engineering, IIT Bombay

semantics-rich machine readable dictionaries (MRD)), (ii) यंत्रभाषिक मशीनी अनुवाद का निर्माण हिंदी एवं मराठी के लिए interlingua based machine translation softwares for *Hindi* and *Marathi* and (iii) मराठी शोध यंत्र एवं वर्तनी जांचकर्ता (Marathi search engine and spell checker)[16] आदि का निर्माण करने का काम सौपा गया है। इसी के साथ सी-डॅक,पुणे और आई.आई.टी हैदराबाद ने 'शक्ति' नामक अंग्रेजी से मराठी अनुवाद यंत्र निर्माण करने का काम किया है। मुख्य रुप से यह देखना भी आवश्यक है कि मशीनी अनुवाद के साथ ही किन भाषिक संसाधनों का विकास मराठी में हो चुका है एवं कौन से संसाधनों पर काम किया जा रहा है। 1.शब्दकोश(**Lexicon**):- मराठी से मराठी, मराठी-हिंदी, अंग्रेजी-मराठी के साथ अन्य भाषाओं में भी मराठी को ध्यान में रखते हुए शब्दकोशों का निर्माण किया गया है। महाराष्ट्र सरकार ने भी एक परियोजना के तहत मराठी के संपूर्ण शब्दों का संगणकीकृत शब्द कोश तैयार किया है।[17]

2. शब्दजाल : - मराठी भाषा का स्वतंत्र शब्द-जाल(शब्दबंध) विकसित किया गया है। जिसका उपयोग इंटरनेट के माध्यम से किया जा रहा है, यह कार्य हिंदी शब्द-जाल के अधार पर ही किया गया है। 3.मराठी विश्लेषक(Marathi Analysis) 4. शब्द Word Sense Disambiguation 5. मराठी भाषा के लिए वाक विश्लेषण प्रणाली(Speech Synthesis for Marathi Language) 6. प्रोजेक्ट तुकाराम(Project Tukaram): - तुकाराम यह नाम महाराष्ट्र में सर्वसाधारण है। प्रोजेक्ट तुकाराम का प्रमुख उद्देश्य तुकाराम द्वारा रचित अभंगों का कंप्यूटरी कृत संकलन तैयार कर शोधयंत्र के साथ जोड़ना है।[18] 7. देवनागरी फॉन्ट का निर्माण करना(Designing Devanagari Fonts) 8. आई.टी का स्थानीयकरण(IT Localization) आदि भाषिक संसाधनों पर काम हुआ है तथा कुछ और भाषिक संसाधनों पर भी काम हो रहा है। यह मराठी भाषा की एक उपलब्धि है कि मराठी भाषा के लिए देश की नामी संस्थाओं मे काम हो रहा है। इन सभी जानकारियों के बाद यह समझना आसान हो जाता है

---

[16] paradigm Shift Of Language Technology initiatives Under Programme (त्रैमासिक पत्रिका) विश्वभारत@TDIL सूचना प्रौद्योगिकी मंत्रालय, भारत सरकार - page 5

[17] वहीं पृष्ठ संख्या - 6

[18] paradigm Shift Of Language Technology initiatives Under Programme (त्रैमासिक पत्रिका) विश्वभारत@TDIL सूचना प्रौद्योगिकी मंत्रालय, भारत सरकार - page 5

कि मराठी भाषा में हो रहा यह कार्य केवल मराठी तक ही सीमित न रहकर हिंदी एवं अन्य भाषा के अनुसंधानों में भी अपनी भूमिका को स्पष्ट करता है। हिंदी में निर्माण होने वाले अनेकार्थक शब्दों का उचित अर्थ के साथ अनुवाद करना भी इस कार्य का ही एक भाग है।

भाषा की लघुत्तम इकाई शब्द है जैसे-दूध, गोरा, बचपन, बाल आदि। किसी विकसशील या विकसित भाषा के शब्द-भंडार के सभी शब्दों की पूरी गणना करना संभव नहीं तो सहज भी नहीं है। जिस प्रकार बचपन से लेकर मृत्यु तक किसी व्यक्ति के शब्द-भंडार में परिवर्तन होता रहता है उसी प्रकार जीवित भाषाओं के शब्दों के भंडार में भी यह परिवर्तन दिखाई देता है इस प्रक्रिया में शब्द परिवर्तन का केंद्र व्याकरणिक कोटियाँ ही होती है। इसी आधार पर शब्द-भेद किया जाता है और शब्दों को भाषा में स्थापित किया जाता है। इसी अधार पर अनेकार्थक शब्दों की शब्द-भेद के इन व्याकरणिक कोटियों में विभाजित किया गया है।[19]

क्र.	व्याकरणिक कोटि	संक्षिप्ताक्षर	उप-व्याकरणिक कोटि
1.	संज्ञा	N	1. व्यक्तिवाचक संज्ञा
			2. जातिवाचक संज्ञा
			3. समूहवाचक संज्ञा
			4. द्रव्यवाचक संज्ञा
			5. भाववाचक संज्ञा
	संज्ञा	N	1. संज्ञा-लिंग व्यवस्था
			2. संज्ञा-वचन व्यवस्था
			3. संज्ञा-कारक व्यवस्था
2.	सर्वनाम	PRN	1. पुरुषवाचक
			2. निश्चयवाचक
			3. अनिश्चयवाचक
			4. प्रश्नवाचक
			5. संबंधवाचक
			6. निजवाचक

[19] डॉ. राणा, महेंद्र सिंह, *हिंदी व्याकरण-प्रकाश-लेखन,* हर्षा प्रकाशन, आगरा, पृष्ठ सं.-118-216

	सर्वनाम	PRN	1. सर्वनाम-पुरुष व्यवस्था
			2. सर्वनाम-लिंग व्यवस्था
			3. 3सर्वनाम-लिंग व्यवस्था
			4. सर्वनाम-वचन व्यवस्था
			5. सर्वनाम-कारक व्यवस्था
3.	विशेषण	ADJ	1. गुण बोधक
			2. संख्या बोधक
			3. परिणाम बोधक
			4. सर्वनामिक

क्र.	व्याकरणिक कोटि	संक्षिप्ताक्षर	उप-व्याकरणिक कोटि
4.	क्रिया	V	1. प्रेरणार्थक क्रिया
			2. संयुक्त क्रिया
			2.1. सहायक क्रिया
			2.2. अनुकरणात्मक क्रिया
			3. अनुकरणात्मक
			4. नाम धातु
			5. अकर्मक क्रिया
			6. सकर्मक क्रिया
	क्रिया	V	1. क्रिया- लिंग
			2. क्रिया- वचन
			3. क्रिया- वाच्य
			4. क्रिया- वचन
			5. क्रिया- वृत्ती
			6. क्रिया- पक्ष
			7. क्रिया- पुरुष
			8. क्रिया- काल
			8.1. क्रिया- वर्तमानकाल
			8.2. क्रिया- भविष्यकाल

			8.3 क्रिया- भूतकाल

अविकारी शब्दों के शब्द-भेद की कोटियों को निम्न रुप से देख सकते है।[20]

क्र.	व्याकरणिक कोटि	संक्षिप्ताक्षर	उप-व्याकरणिक कोटि
1.	कालवाचक		1. समय सूचक
			2. अवधि
			3. बारंबारता
2.	स्थानवाचक		1. स्थिति बोधक
			2. दिशा बोधक
			3. प्रश्नवाचक
3.	क्रिया विशेषण		1. परिमाणवाचक
			2. रीतिवाचक
4.	प्रश्नवाचक		--------
5.	संबंध सूचक		1. कालवाचक
			2. स्थानवाचक
			3. दिशावाचक
			4. साधनवाचक
			5. उद्देश्यवाचक
			6. पार्थक्यवाचक
			7. व्यतिरेकवाचक
			8. विषयवाचक
6.	समुच्चय बोधक	EXM	1. समानाधिकारण
			2. संयोजक
			3. विभाजक
			4. विरोधदर्शक
			5. परिणामदर्शक
			6. व्याधिकरण

[20] डॉ. राणा, महेंद्र सिंह, *हिंदी व्याकरण-प्रकाश-लेखन,* हर्षा प्रकाशन, आगरा, पृष्ठ सं.-118-216

			7. व्याख्या सूचक
			8. कारणसूचक
			9. उद्देश्यसूचक
			10. प्रतिबंधसूचक
7.	मनोभावसूचक	PSY	1. विस्मय सूचक
			2. हर्ष सूचक
			3. शोक सूचक
			4. संबोधन सूचक
			5. तिरस्कार सूचक
			6. आर्शीवाद सूचक
			7. स्वीकार सूचक
			8. अनुमोदन सूचक
8.	निपात	NP	---- ---

मशीनी अनुवाद में कार्य कर रही संस्थाओं के पास शब्द भेद की अति सूक्ष्म कोटियाँ तैयार की हैं। जिसके आधार पर डेटा बेस में दिए जाने वाले शब्दों को शब्द-भेद के अधार पर अधिक तर्क संगत रुप से विभाजित किया जा सकता है।[21]

मायक्रोसॉफ्ट इंडीक टैग सेट-भारतीय एवं पाश्चात्य भाषाविदों द्वारा निर्माण किया गया मायक्रोसॉफ्ट इंडीक टैग सेट का उपयोग आठ भारतीय भाषाओं में दिए गए पाठ के व्याकरणिक कोटियों को विश्लेषित करने के लिए उपयोग में लाया जाता है। इस अनुसंधान में इस टैग सेट का उपयोग शब्दों को विश्लेषित करने के लिए किया गया है। मायक्रोसॉफ्ट इंडीक टैग सेट में भारतीय भाषाओं के विश्लेषण के लिए कुछ प्रमुख व्यकरणिक कोटियों को भी चुना है, इसी के साथ वचन, लिंग, क्रिया की। सकर्मक, अकर्मक क्रियाओं पर भी ध्यान दिया गया है। भाषा विश्लेषण का काम तीन स्तरों में होता है जिसे NLP एवं संगणक वैज्ञानिकोंकी सहायता से बनाया गया है। इस संसाधन में विविध श्रेणियों, रुपरेखाओं और अस्पष्ट शब्दों पर विचार कर संसाधन का निर्माण किया गया है। जिसका उपयोग इस अनुसंधान को हुआ है। इस टैग सेट का उपयोग कई जगह किया जा रहा है।

---

[21] paradigm Shift Of Language Technology initiatives Under Programme (त्रैमासिक पत्रिका) विश्वभारत@TDIL सूचना प्रौद्योगिकी मंत्रालय, भारत सरकार - page 11

2.समान लिपि, समान उच्चारण , भिन्नार्थक शब्द

इन शब्दों को अंग्रेजी में (Homograph) कहा गया है। अनेकार्थक शब्दों की परिभाषाओं को देखा जाए तो परिभाषाओं के अनुसार इन शब्दों में सभी शब्द समान वर्तनी के होना आवश्यक हैं। अपवाद के रुप में मिलने वाले कुछ शब्दों में लिपि समान होते हुए भी उन शब्दों का उच्चारण भिन्न हो सकता है। लेकिन अधिकतर जगह उच्चारण समान ही होता है। ऐसे समय यह शब्द समान उच्चारण वाले शब्द भी हो सकते हैं। मुख्यत: मशीनी अनुवाद की दृष्टि से देखा जाए तो उच्चारण की समस्या सभी मशीनी अनुवाद के लिए महत्व पूर्ण नहीं हैं क्योंकि भारतीय मशीनी अनुवाद कार्यक्रम में स्पीच टू स्पीच(STS) या स्पीच टू टेक्स्ट(STT) मशीनी अनुवाद का कोई भी कार्यक्रम शुरु नहीं हो पाया है। जब भी यह कार्य शुरु होगा अनेकार्थक शब्दों में निर्माण होने वाली यह समस्याऐं स्पीच टू स्पीच या स्पीच टू टेक्ट मशीनी अनुवाद मशीन के लिए सबसे महत्वपूर्ण होंगी।

संपूर्ण अनुसंधान में सर्वाधिक महत्वपूर्ण भाग समान लिपि, समान उच्चारण, भिन्न अर्थ वाले शब्द हैं। एवं इन्हीं शब्दों के निराकरण में सर्वाधिक कठिनाई भी उत्पन्न होती है। इसी के साथ काल एवं संदर्भ के अनुसार भी महत्व पूर्ण कार्य हो सकता है। यह समस्या तब निर्माण हो सकती है जब शब्दों की व्याकरणिक कोटि समान हो एवं शब्द का अर्थ काल के अनुसार ही बदलता हो। ऐसे शब्दों के समान अर्थ वाले शब्द अधिक न मिलने के कारण यह कार्य और भी कठिन हो जाता है। कुछ शब्द व्याकरण की क्रिया कोटि में आते है जो काल एवं संदर्भ के अनुसार ही बदलते हैं। इसलिए इन शब्दों पर भिन्न रुप से विचार होना आवश्यक है। अनेकार्थक शब्दों में विशेषण के रुप में आने वाले शब्द अनुसंधान की दूसरी एवं सबसे महत्वपूर्ण समस्या है। यह समस्या अधिकतर संदर्भ के साथ होती है। इस समस्या के समाधान के लिए कार्पस की आवश्यकता होगी। न केवल कार्पस बल्कि अनुवाद के उपरान्त मिलने वाले लक्ष्य पाठ पर भी विशेष ध्यान देना आवश्यक है। अनेकार्थक शब्दों में संज्ञा/नामों की सर्वाधिक संख्या पायी जाती है। इसका प्रमुख कारण नामों का अर्वाधिक उपयोग है। यह कार्य न केवल साहित्यिक पाठ में बल्कि प्रशासनिक कार्यों के पाठों में भी उत्पन्न हो सकता है। साहित्यिक भाषा की तुलना में प्रशासनिक कार्यों में नाम/संज्ञा कम मिलते हैं। इसका प्रमुख कारण प्रशासनिक कार्यों में नामों के उपयोग को अधिक महत्व नहीं दिया जाता है। प्रशासनिक कार्यों के

उपयोग में लाए जाने वाले शब्द तकनीकी होते हैं। हिंदी में अनेकार्थकता उत्पन्न करने वाले शब्दों की सूची एवं उसका मराठी अनुवाद निम्न रुप से दिया जा रहा है।

1. हिंदी शब्द - द्विज (नाम) जिसकी उत्पत्ति अंडे से हुई हो।

उदाहरण - साँप एक द्विज प्राणी है।

मराठी शब्द - अंडज (विशेषण) (समान अर्थ वाले शब्द - अंडज)

मराठी अनुवाद - साप एक अंडज प्राणी आहे.

1. हिंदी शब्द - द्विज (नाम) हिन्दुओं के चार वर्णों में पहला और सबसे श्रेष्ठ वर्ण।

उदाहरण - द्विज को अपने कर्मों का पालन करना चाहिए।

मराठी शब्द - द्विज (नाम) (समान अर्थ वाले शब्द - ब्राह्मण, विप्र,द्विजन्मा)

मराठी अनुवाद - द्विजाने आपली कर्तव्ये पाळायला हवीत.

2.हिंदी शब्द - द्विज (नाम) दाँत (नाम)

उदाहरण - द्विज धारी नाग विषैले होते हैं।

मराठी शब्द - दात (नाम) (समान अर्थ वाले शब्द - दंत, दंतावली)

मराठी अनुवाद - दंतधारी नाग विषारी असतात.

इन शब्दों में प्रथम शब्द नाम समूह वाचक है और दूसरा समूह वाचक विशेषण है। तीसरा शब्द वस्तु वाचक नाम है। अत: इन शब्दों का निराकरण व्याकरणिक कोटियों से स्पष्ट किया जा सकता है।

1. हिंदी शब्द - अंकुश (नाम) गजांकुश (नाम)

उदाहरण - सारथी ने अंकुश से हाथी को आगे बढ़ने से रोका।

मराठी शब्द - गजांकुश(नाम) (समान अर्थ वाले शब्द - टोचणी)

मराठी अनुवाद - सारथी ने गजांकुशाने हत्तीला पुढे जाण्या पासून थांबवले.

2.हिंदी शब्द - अंकुश (नाम) रोक (नाम)

उदाहरण - पानी की तरह बह रहे पैसे पर अंकुश लगाना जरुरी है।

मराठी शब्द - अंकुश (नाम)

मराठी अनुवाद - पाण्या प्रमाणे वाहणा-या पैश्यावरती अंकुश लावणे आवश्यक आहे.

प्रथम शब्द वस्तु वाचक नाम है द्वितीय शब्द नाम विशेष है। ऐसे शब्द अनुवाद में सबसे अधिक समस्याऐं निर्माण करते हैं। प्रथम शब्द अंकुश का अर्थ 'गजांकुश 'है। मराठी में इस शब्द का अर्थ गजांकुश है। दूसरे शब्द 'अंकुश' अपने आप में स्पष्ट नहीं है कि किस प्रकार के 'अंकुश' अर्थ दे रहा है। इसलिए मराठी में अनुवाद करना कठिन हो जाता है। अंकुश शब्द का कोई अन्य समानार्थक शब्द भी नहीं मिल पाता। इसलिए सही रुप में अनुवाद करना अधिक कठिन हो जाता है। संदर्भ के अनुसार देखा जाए तो हिंदी में इस वाक्य का अर्थ पैसों पर रोक लगाने के संदर्भ में है। जिसके आधार पर मराठी अनुवाद हो सकता है।

1.हिंदी शब्द - अकड़ना (नाम) घमंड (नाम)

उदाहरण - राजेंद्र का इतना अकड़ना सबको अच्छा नहीं लगता।

मराठी शब्द - घमेंड (नाम)(ऐट,अभिमान)

मराठी अनुवाद - राजेंद्रचे इतके अकड़ने सर्वांना रुचत नाही.

2.हिंदी शब्द - अकड़ना(क्रिया) ठिठुरना(क्रिया)

उदाहरण - ठंड में हाथ, पैर अकड़ना आम बात है।

मराठी शब्द - आखडणे (क्रिया) (आकुंचन)

मराठी अनुवाद - थंडी मध्ये हात, पाय अकडणे साधारण गोष्ट आहे.

जिन अनेकार्थक शब्दों की व्याकरणिक कोटि पूर्णत: भिन्न होती हैं, ऐसे शब्दों के अनुवाद में अधिक समस्या नहीं होती है। जिन शब्दों की व्याकरणिक कोटि भिन्न होगी उन शब्दों की टैगर के द्वारा व्याकरणिक कोटि को स्पष्ट कर व्याकरणिक कोटियों के आधार पर अनुवाद करना संभव है। जैसे :- जिस जगह 'अकड़ना' नाम के रुप में आयेगा। उस जगह 'घमंड' का अर्थ आयेगा जिसका मराठी अनुवाद 'घमंड' के रुप में होगा। और जहाँ 'अकड़ना' क्रिया के रुप में आता है वहाँ उसका मराठी अनुवाद अपने आप 'अकडणे' आयेगा। यह प्रक्रिया उन सभी अनेकार्थक शब्दों के साथ होगी जिनकी

व्याकरणिक कोटि भिन्न है। कुछ जगहों पर क्रिया के अनुसार शब्दों की संरचना बदल जाती है। ऐसी जगह कुछ शब्दगत बदलाव की भी आवश्यकता होती है। यह उदाहरण बगैर किसी बदलाव के केवल व्याकरणिक कोटियों के माध्यम से हो सकता है।

1.हिंदी शब्द - सजना (नाम) पति(नाम)

उदाहरण - मेरे सजना गाँव गये हैं।

मराठी शब्द - पती(नाम) (समान अर्थ वाले शब्द- नवरा,स्वामी, धनी, मालक, भ्रतार, यजमान, कारभारी)

मराठी अनुवाद - माझे पती गावी गेले आहेत.

2.हिंदी शब्द - सजना(क्रिया) श्रगांर करना (क्रिया)

उदाहरण - दुल्हन विवाह मंडप में जाने से पूर्व सजना चाहती है।

मराठी शब्द - सजणे(क्रिया) (नटने,उद्योक्त)

मराठी अनुवाद - वधू विवाह मंडपात जाण्या अगोदर शृगार करु इच्छिते.

1.हिंदी शब्द - तना(नाम) डाली (नाम)

उदाहरण - इस वृक्ष का तना बहुत छोटा है।

मराठी शब्द - फांदी(नाम) (समान अर्थ वाले शब्द-शाखा,फाटी,फाटा)

मराठी अनुवाद - या वृक्षाची फांदी फार लहान आहे.

2.हिंदी शब्द - तना(क्रिया) तना हुआ(क्रिया)

उदाहरण - रस्सी को इतना मत, तनाओ टूट जाएगी।

मराठी शब्द - तानलेला(क्रिया) (खेचलेला, खेचलेली,खेचू)

मराठी अनुवाद - दोरीला इतके तानू नका तूटेल.

जैसे कि हमने देखा व्याकरणिक कोटि भिन्न होने के बावजूद कुछ संरचनागत बदलाव की भी आवश्यकता होती है। जिसे हम इस उदाहरण के माध्यम से देख सकते हैं। 'तना' यह शब्द नाम के रुप

में आ जाए तो स्पष्ट रुप से इसका मराठी अनुवाद 'फांदी' के रुप में हो जाएगा। लेकिन जब तना क्रिया के रुप में हिंदी में भिन्न-भिन्न रुप में आ जाए तो 'तना' इस क्रिया की संरचना भी वाक्य के बदलाव के साथ अनुदित वाक्य में भी बदलाव की गुंजाइश रखेगी।

जैसे :- 1. हिंदी - रस्सी को इतना मत, तनाओ टूट जाएगी।

मराठी - या दोरीला इतके तानू नका तूटेल.

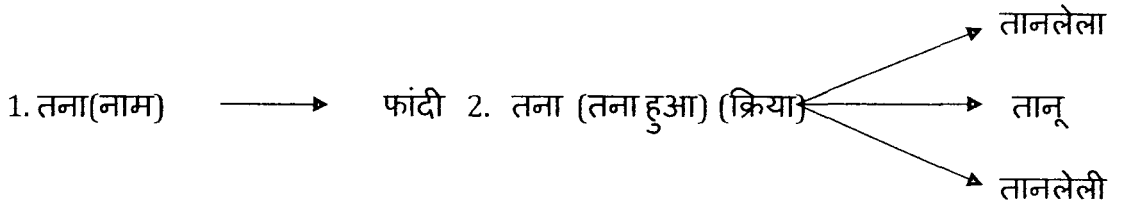
2. हिंदी – तनी हुई रस्सी।

मराठी - तानलेली दोरी.

3. हिंदी - तना हुआ लड़का।

मराठी - तानलेला मुलगा.

इन छोटे वाक्यों में लिंग के अनुसार अनेकार्थक शब्दों के अनुवाद में "तानू, तानलेली, तानलेला यह बदलाव अनुवाद में दिखाई दे रहा है। यह भेद वाक्य की संरचना के कारण यह दिखाई दे रहा है। ऐसी परस्थिति में यह महत्वपूर्ण हो जाता है कि अनुवाद में किस प्रकार शब्द में बदलाव किया जाए । इस कार्य के लिए अलग से अनुसंधान की आवश्यकता है।



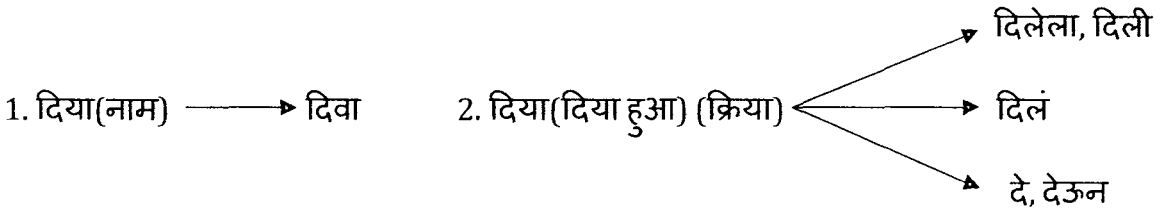
इस प्रक्रिया में क्रिया के अर्थ की संरचना में कर्म के लिंग के अनुसार संरचना बदलती है। अत: ध्यान देने वाली बात यह है की लिंग के साथ क्रिया की संरचना कैसे बदली जाए। जिसे हम निम्न शब्द को देखकर स्पष्ट कर सकते है।

1.हिंदी शब्द - दिया(क्रिया) देना (क्रिया)

उदाहरण - राम ने श्याम को फल दिया।

मराठी शब्द - दिलं (क्रिया) (समान अर्थ वाले शब्द– देणे, प्रदान करणे)

मराठी अनुवाद - राम ने श्याम ला फळ दिले.



(यह रुप व्याकरणिक होते है।)

2.हिंदी शब्द - दिया (नाम) दीपक(नाम)

उदाहरण - शाम होते ही घर में दिया जलता है।

मराठी शब्द - दिवा(नाम) (समान अर्थ वाले शब्द- दीप, दिपक)

मराठी अनुवाद - रात्र होताच घरात दिवा जळतो.

'दिया' का एक अर्थ 'देना' है जो एक क्रिया है। 'दिया' का दूसरा अर्थ दीपक भी है जो एक नाम के रुप में उपयोग में आता है। दूसरे शब्द की अर्थ की संरचना सभी जगह एक ही होगी लेकिन पहले शब्द 'दिया' क्रिया की शब्द संरचना में बदलाव दिखाई देगा। इस बदलाव के कारण यह शब्द अनेकार्थक शब्द की कोटि से भी बाहर जाने की संभावना पैदा कर देता है।

जैसे : - हिंदी शब्द - राम ने श्याम को फल दिया।

मराठी अनुवाद - राम ने श्याम ला फळ दिले.

हिंदी - राम ने श्याम को किताब दी।

मराठी - राम ने श्याम ला पुस्तक दिले.

इन दो वाक्यों में हम स्पष्ट रुप से देख सकते हैं कि वाक्य के कर्म में हुए बदलाव के कारण अनेकार्थक शब्द पर प्रभाव पड़ा है। प्रथम वाक्य में कर्म 'फल' के बाद क्रिया 'दिया' का उपयोग होता है लेकिन जब 'फल ' (पुल्लिंग) कर्म की जगह 'किताब'(स्त्रीलिंग) के आने से क्रिया 'दिया' की जगह 'दी' क्रिया का उपयोग किया गया है। जो एक प्रकार से अनेकार्थक शब्द को लुप्त करता है। इस प्रकार के शब्दों का कोई समान अर्थ वाला शब्द भी नहीं होता। इस स्थिति में यह भी विचार किया जाना चाहिए की क्रिया के रुप में आने वाले शब्दों में बदलाव की स्थिति आती हो तो उस शब्द का अधिकाधिक निकटतम अर्थ देनेवाला शब्द ही डेटा बेस में दिया जाए जिससे अनुवाद में अधिक समस्या उत्पन्न न हो सके। एक समाधान यह भी हो सकता है कि शब्द के सभी

क्रिया रुप एक साथ दिए जाए एवं (user) उपयोग कर्ता को शब्द चयन की अनुमती दी जाए। जैसे = दिला, दिली, दिले.

1.हिंदी शब्द - भाग (नाम) भाग(नाम)

उदाहरण - एक कार्यालय के कई भाग होते हैं।

मराठी शब्द - विभाग (नाम) (समान अर्थ वाले शब्द– खाते,भाग)

मराठी अनुवाद - एका कार्यालयाचे कित्तेक विभाग असतात

2.हिंदी शब्द - भाग (नाम) भाग्य (नाम)

उदाहरण - उसके भाग फूटे हैं जो यहां आयेगा।

मराठी शब्द - नशीब (समान अर्थ वाले शब्द -भाग्य, नियती, दैव, अदृष्ट, प्राक्तन, विधिलिखित)

मराठी अनुवाद - त्याच नशीब फुटलं आहे जो इथे येईल.

मशीनी अनुवाद में प्रमुख रुप से अनेकार्थक शब्दों के निराकरण के लिए एक प्रभावी कोश के रुप में Ontological(ऑंटोलॉजीकल) शब्दकोश को जाना जाता है।[22] इस कोश में व्याकरणिक कोटियों को कई भागों में विभाजित किया गया है। जिसका सामान्य रुप इस प्रकार देखा जा सकता है। Ontological Categories (अंटोलॉजीकल श्रेणियाँ) :- यह एक पदानुक्रमित संगठन है। यह क्षेत्र विशिष्ट के लिए महत्वपूर्ण है, लेकिन यह कार्य एन.एल.पी यंत्र के लिए कठिन है। इसके लिए नामों को चेतन एवं अचेतन श्रेणियों में बांटा गया है। चेतन वर्ग को पुन: पशुओं एवं जानवरों की श्रेणियों में विभाजित किया गया है। अचेतन नामों को पुन: कई भागों में विभाजित किया गया है। जिसमें से प्रमुख है वस्तु, स्थान, घटना के रूप, सार इकाई आदि अवधारणा, समय संकेत, भावना, राज्य, आदि को शामिल किया गया है। इसी प्रकार क्रिया कोटि को कई भागों में विभाजित किया गया है जिसमें प्रमुख है इच्छार्थक, करना, स्थान, स्थायी एवं अस्थायी इसी के साथ अन्य भी है। विशेषण कोटि को वर्णनात्मक रूप से वर्गीकृत किया गया है। वजन, रंग, गुणवत्ता, मात्रा, आदि इसके प्रमुख रूप हैं। क्रिया-विशेषण रुपों को निम्न रुपों में वर्गीकृत किया गया है समय, जगह, ढंग, मात्रा, एवं कारण। इस प्रकार देखा जाए तो केवल व्याकरणिक कोटियों के सूक्ष्म विभाजन से भी अनेकार्थकता को दूर किया जा सकता है। जिसका आधार अंटोलॉजीकल कोश (Ontological Dicationary) के रुप में

---

[22] paradigm Shift Of Language Technology initiatives Under Programme (त्रैमासिक पत्रिका) विश्वभारत@TDIL सूचना प्रौद्योगिकी मंत्रालय, भारत सरकार - page 11

किया जा रहा है। जिसे हम नामों से होने वाले अनेकार्थक शब्दों के निराकरण के लिए उपयोग में ला सकते हैं।

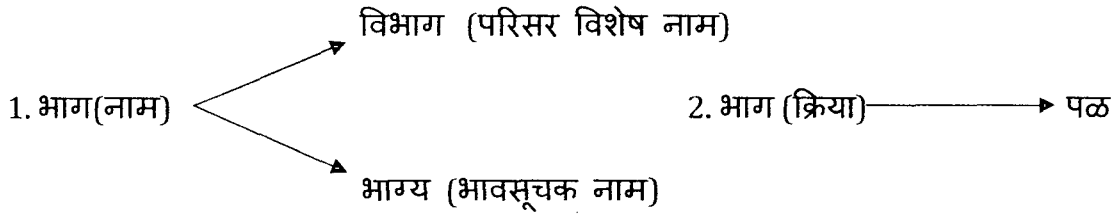
2.हिंदी शब्द - भाग (नाम) भगाना(क्रिया)

उदाहरण - भाग यहाँ से, नहीं तो खूब मार पड़ेगी।

मराठी शब्द - पळ(क्रिया) (दौड,पसार होणे,पळ काढणे)

मराठी अनुवाद - पळ इथून नाही तर खूप मार खाशील.

इस प्रकार के शब्दों के लिए अन्टोलॉजिकल शब्दकोश के माध्यम से समाधान खोजा जा सकता है। आई.आई.टी.मुंबई के अन्टोलॉजिकल शब्दकोश में नामों को 57 भागों में विभाजित किया गया है।[23] जिसके आधार पर यह कार्य आसानी के साथ किया जा सकता है। जैसे:- भाग(नाम) एक घटित क्रिया का काम और भाग(नाम) एक परिसर विशेष का नाम है। जिसका निर्धारण ऑन्टोलॉजिकल शब्दकोश करता है। भाग(क्रिया) जिसका अनुवाद मराठी अनुवाद 'पळ' होता है। जिसका निर्धारण टैगर आसानी से कर सकेगा। जैसे।



व्याकरण में काल का अन्यन्य साधारण महत्व है जिसके आधार पर भाषा के भविष्य, वर्तमान एवं भूत की घटनाओं का सरल आकलन होता है। मानव जिस सहजता से कालों का उपयोग करता है उससे वह भाषा में आये अनेकार्थक शब्दों को भी आसानी के साथ समझ सकता है। जैसे 'कल' शब्द के दोनों अर्थों को मनुष्य आसानी के साथ समझता है लेकिन मशीन के लिए यह एक महत्वपूर्ण और जटिल समस्या है। जिसके निराकरण के लिए मशीन को काल के आधार पर निर्देश देने वाले प्रोग्राम की आवश्यकता है। जैसे – कल शब्द का उपयोग भविष्य काल एवं वर्तमान काल के संदर्भ के रुप में एवं कल शब्द का उपयोग भूतकाल के वाक्यों में विशेष रुप से दिखाई देता है। वाक्य के अंत में 'कल' का उपयोग जब गा,गी,गे के साथ आता है तो मराठी अनुवाद 'उद्या' यह होता है। जब वाक्य भविष्य काल में हो तो कल शब्द का अनुवाद 'उद्या' होना चाहिए एवं वाक्य के अंत में था,थी,थे यानी वाक्य भूतकाल में हो तो अनुवाद 'काल' में होना चाहिए। जिसका निर्धारण पूर्णत: वाक्य के काल पर ही निर्भर करता है। इस प्रकार देखा जाए तो अनेकार्थक शब्दों में काल पर आधारित शब्दों के अर्थ

---

[23] वही पृष्ठ संख्या - 4

परिवर्तन से भी मशीनी अनुवाद में समस्या उत्पन्न हो सकती है। जिसका निराकरण भी उसी के आधार पर होना चाहिए। इसी प्रकार काल संबधित अन्य शब्दों पर कार्य किया जा सकता है। जिसका विवरण निम्न रुप से देख सकते है।

1. कल (adverb) (अगला दिन) (क्रियाविशेषण)

उदाहरण - तुम मुझसे कल मिलना आज समय नहीं है।

मराठी शब्द - उद्या (adverb)

मराठी अनुवाद - तू मला उद्या भेट आज माझ्या कडे वेळ नाही.

2. कल (adverb) (पिछला दिन)

उदाहरण - कल का दिन तो मेरे लिए खुशी का दिन था।

मराठी शब्द – काल (adverb)

मराठी अनुवाद - कालचा एकच दिवस तर माझ्या साठी आनंदाचा दिवस होता।

एक संभावना यह भी बनती है कि वर्तमान काल में वाक्य के अंत में गा,गी,गे की जगह 'है' का उपयोग कर 'कल' शब्द के उपयोग में आने वाले दिन के संदर्भ में किया जाता है ऐसे समय में हम मशीन को यह निर्देश दे सकते हैं कि पहले यह देखे की वाक्य में था,थी,थे का उपयोग किया गया है या नहीं अगर था,थी,थे का उपयोग नहीं किया गया हो तो 'कल' अनुवाद का उपयोग 'उद्या' के रुप में ही (आने वाल दिन) करे नाकी 'काल' (बीता हुए दिन) के रुप में। यह काल संबंधी सामान्य सूत्र है। इन सूत्रों के साथ आगे हिंदी में आने वाले अन्य शब्दों को मराठी अनुवाद के साथ स्पष्ट किया जा रहा है। प्रत्येक शब्द को स्पष्ट कर उनका वाक्य में उपयोग भी किया गया है। साथ ही मराठी में अनुवाद कर मराठी अनूदित शब्द के अन्य समान अर्थ वाले शब्द भी दिए गए है।

1.हिंदी शब्द - अकेला(विशेषण) एकमात्र(विशेषण)

उदाहरण - वह घर पर अकेला है।

मराठी शब्द - एकटा (विशेषण) (समान अर्थ वाले शब्द - एकाकी ,एकलकोंडा)

मराठी अनुवाद - तो घरी एकटा आहे.

2.हिंदी शब्द - अकेला (विशेषण) अविवाहित(विशेषण)

उदाहरण - राजेश की उम्र अट्ठाइस है फिर भी वह अकेला है।

मराठी शब्द - अविवाहित (समान अर्थ वाले शब्द - सडा ,अपरिणित,कुमार,ब्रम्हचारी,कौमार्य)

मराठी अनुवाद - राजेशचे वय आठ्ठावीस आहे तरी ही तो अविवाहित आहे.

1.हिंदी शब्द - अचल (नाम) - पर्वत(नाम)

उदाहरण - अचल गिरि के शिखरों पर।

मराठी शब्द - पर्वत (पु) (नाम) (डोंगर,पहाड़,शैल्य,गिरि,डोंगरमाथा,घाटमाथा,पठार,गिरिवर)

मराठी अनुवाद - उंच पर्वत शिखरांनवरती.

2.हिंदी शब्द - अचल(विशेषण) स्थिर (विशेषण)

उदाहरण - बच्चे एक जगह स्थिर नहीं रह सकते।

मराठी शब्द - स्थिर (विशेषण) (समान अर्थ वाले शब्द - अप्रवाही,गतिहीन,स्थावर, अचल, गंभीर, शांत,अढळ, दृढ, अविचल,स्थिरचित्त, शांत,धीराचा)

मराठी अनुवाद - मुलं एका जागी स्थिर राहू शकत नाहीत.

1. हिंदी शब्द - अदा(विशेषण) चुकता(विशेषण)

उदाहरण - शंकर ने आपका पैसा अदा किया।

मराठी शब्द - फेडणे (क्रिया) (समान अर्थ वाले शब्द – चुकवणे, चुकता करणे, परिहारणे, सोडवणे, परत करणे)

मराठी अनुवाद - शंकरने तुमचे पैसे फेडले

2.हिंदी शब्द - अदा(नाम) हावभाव(नाम)

उदाहरण - इसकी अदा तो देखो कैसी है।

मराठी शब्द - नखरा(नाम)(पु) (समान अर्थ वाले शब्द -नट्टापटा, टाकमटिका, डामडौल, थाटमाट, नटवेपणा, नखरेलपणा, टिवल्या बावल्या, ऐट, तोरा, इमाज, ठमकारा, विलास, भमका)

मराठी अनुवाद - तिचा नखरा तर पहा कसा आहे.

1.हिंदी शब्द - अलंकार(नाम) काव्यालंकार (नाम)

उदाहरण - बच्चो यह काव्य अलंकार परीक्षा में पूछा जायेगा।

मराठी शब्द - अलंकार (नाम) अलंकार (नाम)

मराठी अनुवाद - मुलांनो हा काव्य अलंकार परीक्षेत विचारला जाईल.

2.हिंदी शब्द - अलंकार (नाम) आभूषण(नाम)

उदाहरण - वह स्त्री काफी अलंकार पहनी हुई है।

मराठी शब्द - आभूषण (नाम) (समान अर्थ वाले शब्द – डाग, डागिना, लेणे, भूषण, अलंकार, विभूषण, शृंगार, रत्न, जडजवाहीर, दागदागिणे)

मराठी अनुवाद - त्या स्त्री ने खूप अभूषण परिधान केले आहेत.

1.हिंदी शब्द - आना(नाम) पैसे(नाम)

उदाहरण -भिखारी के कटोरे में आठ आना है।

मराठी शब्द - आना(नाम) (समान अर्थ वाले शब्द–आने)

मराठी अनुवाद - भिका-याच्या वाटी मध्ये आठ आने आहेत.

1.हिंदी शब्द - भवज(नाम) भाभी (नाम)

उदाहरण - मेरी भाभी अध्यापिका है।

मराठी शब्द - वहिनी (नाम) (समान अर्थ वाले शब्द– भावजय)

मराठी अनुवाद - माझी वहिनी अध्यापिका आहे.

2.हिंदी शब्द - भवज(नाम) तल(नाम)

उदाहरण - किले के भवज में एक सुरंग अवश्य होती है।

मराठी शब्द - तळ (नाम)(पु) (समान अर्थ वाले शब्द– पाया, तच, अड्डा, आधिवास)

मराठी अनुवाद - किल्याच्या तळात एक सूरंग अवश्य असते.

1. हिंदी शब्द - गम(नाम) दुख(नाम)

उदाहरण - बच्चे के बीमार होने पर सभी लोग गम में डूब गए।

मराठी शब्द - दुख (नाम) (समान अर्थ वाले शब्द-दिलगिरी, हळहळ, असंतोष, नाखुशी, यातना, ताव, पीडा, वेदना, कष्ट, व्याथा, खेद, शोक, विलाप, रज)

मराठी अनुवाद - बाळ अजारी पडल्याने सर्व लोक दुखी झाले.

2.हिंदी शब्द - गम (नाम) गोंद(नाम)

उदाहरण - कागज चिपकाने के लिए गम ले आओ।

मराठी शब्द - डिंक (नाम) (पु) (समान अर्थ वाले शब्द- गोंद, सरस, खळ, गंजी, चीक, पेज, पाजणी,राळ, लांबी, लुकण)

मराठी अनुवाद - कागद चिटकावण्यासाठी डिंक घेऊन या.

1.हिंदी शब्द - पीर (नाम) फकीर(नाम)

उदाहरण - मुस्लिम समाज में पीर भगवान के दूत होते हैं।

मराठी शब्द - पीर(नाम) (पु) फकीर(नाम)(साधू, अंतेष्ठे)

मराठी अनुवाद - मुस्लिम समाजात पीर देवा प्रमाणे असतात.

2.हिंदी शब्द - पीर(नाम) दुःख (नाम)

उदाहरण - वैष्णव जन तो तेणे कहिए, पीर पराई जाने रे।

मराठी शब्द - दुख: (नाम) (समान अर्थ वाले शब्द-दिलगिरी, हळहळ, असंतोष, नाखुशी, यातना, ताव,पीडा, वेदना, कष्ट, व्याथा, खेद, शोक, विलाप, रज)

मराठी अनुवाद - वैष्णव जन त्यांना म्हणावं, जो दुख: दुस-यांचे जाने रे।

1.हिंदी शब्द - माल(नाम) माला(नाम)

उदाहरण - नेता ने कहा दूसरे की माल मेरे गले में डाल रहे हो।

मराठी शब्द - माळ(नाम) (समान अर्थ वाले शब्द – हार, माला, मालिका, पुष्पहार, सरणी, फुलमाळ, गाजरा, वेणी)

मराठी अनुवाद – नेता म्हणाला दूस-याची माळ माझ्या गळ्यात घालत आहात.

2.हिंदी शब्द - माल(नाम) सामान(नाम)

उदाहरण - सारा माल एक ही दिन में बिक गया।

मराठी शब्द - माल(नाम) (समान अर्थ वाले शब्द - संपत्ती, वस्तू, जिन्नस, सामान)

मराठी अनुवाद - संपूर्ण माल एकाच दिवसात विकला.

1.हिंदी शब्द - राय(नाम) विचार(नाम)

उदाहरण - मेरी राय कोई माने तब तो मैं बोलूँ।

मराठी शब्द - म्हंणण(नाम) (समान अर्थ वाले शब्द - बोलणं, विचार)

मराठी अनुवाद - माझं म्हंणण कोणी मानेल तर मी काही बोलू.

2.हिंदी शब्द - राय(नाम) राजा(नाम)

उदाहरण - राय साहब अंग्रेजों का जमाना मुझे अब भी याद है।

मराठी शब्द - राजा(नाम) (समान अर्थ वाले शब्द - नृप, नृपती, नृपाल, नरेश, नरेंद्र, नरपती, नराधिप, भूप, भूपती, भूपाल, महीपाल,राज्यकर्ता, शासक)

मराठी अनुवाद - राजा साहेब ब्रिटिशांचा काळ मला आज ही आठवतो.

1.हिंदी शब्द - शुक्र(नाम) शुक(नाम)

उदाहरण - शुक्र मानो बारिश नहीं हुई।

मराठी शब्द - आभार(नाम) (समान अर्थ वाले शब्द - धन्यवाद)

मराठी अनुवाद - आभार मान पाऊस नाही आला.

2.हिंदी शब्द - शुक्र(नाम) शुक्र तारा(नाम)

उदाहरण - शुक्र तारा वैज्ञानिकों की पुरानी खोज है।

मराठी शब्द - शुक्रतारा (नाम) (समान अर्थ वाले शब्द - शुक्र)

मराठी अनुवाद - शुक्रतारा वैज्ञानिकांची जुना शोध आहे.

3.हिंदी शब्द - शुक्र (नाम) शुक्रवार(नाम)

उदाहरण - सर ने कहा था, कि शुक्र को क्लास नहीं है।

मराठी शब्द - शुक्रवार (नाम) (समान अर्थ वाले शब्द - शुक्रवार)

मराठी अनुवाद - सरांनी म्हंटलं होत शुक्रवारी क्लास नाही.

4.हिंदी शब्द - शुक्र(नाम) वीर्य(नाम)

उदाहरण - इनके शुक्राणु में समस्या होने के कारण इन्हे संतान नहीं हो सकती।

मराठी शब्द - शुक्राणू (नाम) (समान अर्थ वाले शब्द - रेत, धातू)

मराठी अनुवाद – यांच्या शुक्राणू मध्ये समस्या असल्या मुळे यांना मूल होऊ शकत नाही.

1.हिंदी शब्द - सर(नाम) सिर(नाम)

उदाहरण - इस बच्चे का सर कैसे फूटा।

मराठी शब्द - डोके(नाम) (समान अर्थ वाले शब्द- शिर, शिरकमल, शिरोभाग, माथा, माथे, शिर्ष, मुंड, डोई, डुई, डोकसे, डोसके, टकुरे, टकोरे, टाळके, टकले, कवटी, मानशील)

मराठी अनुवाद - या मूलाचे डोके कसे फूटले.

2.हिंदी शब्द - सर(नाम) तालाब(नाम)

उदाहरण - बरस धरा में, बरस सरत गिरि सर सागर में।

मराठी शब्द - तळे (नाम) (समान अर्थ वाले शब्द- पुष्कर, सरोवर, पाणवठा, तलाव, जलाशय, जलकुंड, डोह, पाझर तलाव)

मराठी अनुवाद - कोसळणा-या धारे मध्ये, कोसळत आहे आभाळ तलावी सागरा मध्ये.

1.हिंदी शब्द - सर(नाम) मास्टर(नाम)

उदाहरण - सर ने इतिहास काफी अच्छी तरह से पढ़ाया है।

मराठी शब्द - गुरुजी (नाम) (समान अर्थ वाले शब्द - मास्तर, अध्यापक, शिक्षक, गुरुवर्य, शिक्षक, शाळामास्तर, उस्ताद, वस्ताद)

मराठी अनुवाद - गुरुजींनी इतिहास खूप चांगल्या पद्धतीने शिकविला आहे.

2.हिंदी शब्द - अमल (नाम) आचरण(नाम)

उदाहरण - गुरुजी के उपदेशों पर अमल भी किया करो।

मराठी शब्द - पालन (नाम) (समान अर्थ वाले शब्द - सांभाळ, संगोपन, पोषन, संवर्धन, परिपालन, पालनपोषन, जोपसना, जतन, जतणूक, प्रतिपालन, प्रतिपाळ)

मराठी अनुवाद - गुरुजींच्या उपदेशांचे पालन ही करत जा.

1.हिंदी शब्द - अमल (विशेषण) निर्मल (विशेषण)

उदाहरण - अमल गिरि के शिखरों पर बादल को घिरते देखा है।

मराठी शब्द - निर्मळ(विशेषण) (समान अर्थ वाले शब्द - चखोट, साफ, स्वच्छ, पवित्र, प्रांजळ, निर्लेप, पक्षाळित, निरभ्र)

मराठी अनुवाद - निर्मळ पर्वतांच्या शिखरांनवर आकांशांना पडताना पाहिलं आहे.

2.हिंदी शब्द - अजान(नाम) अज़ान(नाम)

उदाहरण - अजान की आवाज सुनते ही नींद टूट जाती है।

मराठी शब्द - अजान(नाम) (समान अर्थ वाले शब्द - बांग)

मराठी अनुवाद -आजानचा आवाज एकताच झोप तूटते.

1.हिंदी शब्द - अजान(विशेषण) अनजान (विशेषण)

उदाहरण - एक अजान सौ सुजान।

मराठी शब्द - भोळा(विशेषण)(समान अर्थ वाले शब्द सरळ,साधाभोळा,निष्कपट,निष्कपटी,साधा)

मराठी अनुवाद - एक भोळा तर शेकडो जानकार.

2.हिंदी शब्द - कालीन (नाम) गालीचा(नाम)

उदाहरण - घर में कालीन बिछा है।

मराठी शब्द - कालीन(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - घरामध्ये कालीन पसरुण आहे.

1.हिंदी शब्द -कालीन (नाम) काल संबंधी

उदाहरण - मौर्यकालीन शासन की कई बातें आज भी उल्लेखनीय हैं।

मराठी शब्द - काळातील (समान अर्थ वाले शब्द - काळ, कालखंड, युग, समकाल, काळवेळ, पूर्वकाळ, हंगाम, मोसम, इसवी, जमाना, शके)

मराठी अनुवाद - मौर्य काळातील शासनाच्या कितेक गोष्टि आज ही उल्लेखनिय आहेत

2.हिंदी शब्द - गौर (नाम) विचार(नाम)

उदाहरण - बड़ों की बातों पर गौर किया करो।

मराठी शब्द - विचार(नाम) (समान अर्थ वाले शब्द -चिंतन)

मराठी अनुवाद - मोठ्यांच्या गोष्टिंनवर विचार करत जा.

1.हिंदी शब्द - जाली (विशेषण) नकली (विशेषण)

उदाहरण - यह प्रमाण-पत्र जाली है।

मराठी शब्द - खोटा (विशेषण) (समान अर्थ वाले शब्द - खोटारडा, असत्य,नकली, कुटिल, कपटी, ढोंगी, लुच्चा, लटका)

मराठी अनुवाद - हे प्रमाण-पत्र खोटं आहे.

2.हिंदी शब्द - जाली (नाम) जाली (नाम)

उदाहरण - यह जाली खिड़की में लगवा दो।

मराठी शब्द - जाळी(नाम) (समान अर्थ वाले शब्द -बनावट, विखाऊ, नाटकी, पोकळा, बंडल, लबाड, फसवा, अप्रामाणिक)

मराठी अनुवाद - ही जाळी खिडकी मध्ये लावा.

1.हिंदी शब्द - पाक (नाम) भोजन बनाने की क्रिया

उदाहरण - हमारी बेटी पाक कला में निपुण है।

मराठी शब्द - पाककला(नाम) (समान अर्थ वाले शब्द - स्वयंपाक)

मराठी अनुवाद – आमची मुलगी पाक कलेमध्ये निपुण आहे.

2.हिंदी शब्द - पाक (नाम) पाकिस्तान का संक्षिप्त रुप पाक (नाम)

उदाहरण - पाक समाचार पत्र भारत के विरोध में लिखते हैं।

मराठी शब्द - पाकिस्तान(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - पाकिस्तानातील समाचार पत्र भारताच्या विरोधात लिहितात.

3.हिंदी शब्द - पाक (विशेषण) पवित्र (विशेषण)

उदाहरण - मंदिरों में पंडित भी पाक होते हैं।

मराठी शब्द - स्वच्छ(विशेषण) (समान अर्थ वाले शब्द - साफ,पवित्र, निर्मळ, निर्मल, चखोट)

मराठी अनुवाद - मंदिरांन मध्ये पंडित ही पवित्र असतात.

4.हिंदी शब्द - पाक (नाम) चीनी का पाक(नाम)

उदाहरण - हलवाई जलेबियों को छानकर पाक में डालता जा रहा था।

मराठी शब्द - पाक(नाम) (समान अर्थ वाले शब्द -चाचणी, चासणी)

मराठी अनुवाद - हलवाई जलेबी शोधून पाकामध्ये घालत होत.

हिंदी में इस प्रकार के शब्द बहुत ही कम दिखाई पड़ते हैं। जिन के चार भिन्न अर्थ मिलते हों। इस प्रकार के शब्द अर्थ की दृष्टि से अधिक संख्या में होने के बावजूद संदर्भ में अधिक संवेदन शील होते हैं। प्रत्येक शब्द अपना भिन्न अर्थ रखने में कामयाब रहे हैं। हिंदी में मिलने वाले अधिकतर समान लिपि, समान उच्चारण, भिन्न अर्थ वाले शब्दों की संपूर्ण सूची देने का प्रयास किया गया है। जिनका निराकरण डेटा बेस के माध्यम से किया जा रहा है।

3. भिन्न लिपि, समान उच्चारण, भिन्नार्थक शब्द

इन शब्दों को अंग्रेजी में होमोफोन (Homophon) कहा गया है। भिन्न लिपि, समान उच्चारण, भिन्नार्थक शब्दों की अधिक समस्या मशीनी अनुवाद में 'पाठ से पाठ मशीनी अनुवाद'(Text to Text Machine Translation(TTMT)) के दौरान नहीं आनी चाहिए। मशीन प्रत्येक शब्द को शब्द संचयन के दौरान वर्णमाला के अनुसर ही संग्रहीत करती है। इस लिए जिन शब्दों की संरचना भिन्न हो एवं जिन शब्दों के अर्थ भिन्न हो ऐसे शब्द संगणक के लिए पूर्णत: भिन्न होते हैं। अनेकार्थक शब्दों में मिलने वाले भिन्न लिपि, समान उच्चारण, भिन्न अर्थ वाले शब्दों में इसी प्रकार के शब्द मिलते हैं । अत: यह शब्द अनेकार्थक शब्दों के अंतर्गत आने के बावजूद अधिक समस्या उत्पन्न नहीं करते हैं। यह शब्द अनेकार्थक शब्दों की कोटियों में आने से इन शब्दों पर विचार होना आवश्यक है। जिसका विवरण प्रत्येक शब्द के साथ निम्न रुप से दिया जा रहा है।

जैसे : - 1.'जरिया' 2.'ज़रिया'

1.हिंदी शब्द - जरिया(नाम) (नाम) जरीचा(नाम)

उदाहरण - वाहन यात्रा का साधन है।

मराठी शब्द - साधन(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - वाहन यात्रेचे साधन आहे.

2.हिंदी शब्द - ज़रिया(नाम) माध्यम(नाम)

उदाहरण - कोई ऐसा उपाय बताइए जिससे यह काम आसानी से हो जाए।

मराठी शब्द - माध्यम(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - कोणताही असा उपाय सांगा ज्याने हे काम सोप्या पद्धतीने होऊन शकेल.

प्रथम शब्द और द्वितीय शब्द इन दोनों शब्दों में उच्चारण अधिकतर समान ही है लेकिन 'ज' एवं' 'ज़' इन दो अक्षरों में स्पष्ट रुप से आकृतीगत भिन्नता दिखाई देती है। प्रथम 'ज' के नीचे नुक्ता नहीं लगाया गया है। वही दूसरे 'ज़' अक्षर के नीचे नुक्ता लगाया गया है। इस भिन्नता को मशीन आसानी के साथ समझ सकती है। जिसके आधार पर शब्द संचयन में दिए गए उचित अर्थ को दे सकती है। यह प्रक्रिया मशीन निर्माण करने वाले अन्य नियमों से अधिक सरल है। लेकिन समस्या तब उत्पन्न हो सकती है जब अनुवाद की प्रक्रिया में उपयोक्ता (user) पाठ देने में गलती करे जैसे 'जरिया' शब्द का अर्थ वह 'साधन' चाहता है लेकिन मशीन को पाठ देते समय गलती से' 'ज़रिया' शब्द दिया हो ऐसे

समय मशीन शब्द का अर्थ 'साधन' न देकर 'माध्यम' ही देगी यह गलती मशीन की न होकर पूर्णत: मानव की ही होगी। यह समझना इसलिए भी महत्वपूर्ण हो जाता है कि मशीन समान भाषा का उपयोग मात्र एक भाषिक साधन के रुप में ही करती है। इसलिए भाषा की छोटी-छोटी गलतियों के उपर अधिक ध्यान न देते हुए अनेकार्थक शब्दों का उपयोग संदर्भ के अनुसार करने से चूक जाते हैं। यह समस्या सामान्य व्यक्ति की अपेक्षा एक भाषा वैज्ञानिक को अधिक काबिलियत से समझना आवश्यक हो जाता है। इस प्रकार के शब्दों के लुप्त हो जाने की संभावना भी उत्पन्न हो सकती है। भले ही यह शब्द भाषा में समस्या उत्पन्न करते हों लेकिन इन शब्दों का भाषा में रहना भी आवश्यक है।[24]

1.हिंदी शब्द - जाति(नाम) जात(नाम)

उदाहरण - हिंदुओं में अपनी ही जाति में शादी करने का प्रचलन है।

मराठी शब्द - जात (नाम) (समान अर्थ वाले शब्द -प्रकृती, ज्ञाती)

मराठी अनुवाद - हिंदूनी आपल्याच जाति मध्ये लग्न करण्याचे प्रचलन आहे.

2.हिंदी शब्द - जाती(क्रिया) जानेवाली(क्रिया)

उदाहरण - वह लड़की रोज घर जाती है।

मराठी शब्द - जाते(क्रिया) (समान अर्थ वाले शब्द - जाणारी)

मराठी अनुवाद - ती मुलगी रोज घरी जाते.

1. हिंदी शब्द - जितना (विशेषण) जिस परिणाम में/मात्रा में(विशेषण)

उदाहरण - मैं आपसे जितना पैसा लेता हूँ, उसका हिसाब रखिये।

मराठी शब्द - जेवढे(विशेषण) (समान अर्थ वाले शब्द -किती, केवढा)

मराठी अनुवाद - मी तुमच्याकडून जेवढे पैसे घेतो, त्याचा हिशोब ठेवा.

2.हिंदी शब्द - जीतना(क्रिया) विजयी होना(क्रिया)

उदाहरण - आज भारतीय संघ को जीतना आवश्यक है।

मराठी शब्द - जिंकणे (समान अर्थ वाले शब्द -)

---

[24] सिन्हा, राजेंद्र प्रसाद, *शुध्द हिंदी कैसे सीखें,* भराती भवन प्रकाशन, पटना, पृष्ठ सं.--431

मराठी अनुवाद - आज भारतीय संघला जिंकावेच लागेल.

1.हिंदी शब्द - ज़ीन(नाम) जीन(नाम)

उदाहरण - उसने घोड़े का ज़ीन उतार कर नीचे रख दिया।

मराठी शब्द - जीन (नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - त्याने घोड्याचे ज़ीन काढून खाली ठेवले.

2.हिंदी शब्द - जीन(नाम) जीन्स (नाम)

उदाहरण - श्याम जीन्स और कुर्ता पहना है।

मराठी शब्द - जिन्स (नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - श्याम जिन्स आणि कुर्ता घातला आहे.

3.हिंदी शब्द - जिन(नाम) पिशाच(नाम)

उदाहरण - आधुनिक युग में भूतों का कोई अस्तित्व नहीं है।

मराठी शब्द - भूत (नाम) (समान अर्थ वाले शब्द - पिशाच)

मराठी अनुवाद – आधुनिक युगात भूताचे काही अस्तित्व नाही.

1. हिंदी शब्द - कृति(नाम) रचना(नाम)

उदाहरण - गीतांजली विश्व की सर्वोत्तम कृतियों में से एक है।

मराठी शब्द - कृति(नाम) (समान अर्थ वाले शब्द - पुस्तक, ग्रंथ,रचना)

मराठी अनुवाद - गितांजली विश्वातल्या सर्वोत्तम कृतिनमधून एक आहे.

2.हिंदी शब्द - कृति(नाम) कार्य(नाम)

उदाहरण - ताजमहल विश्व की सर्वोत्तम कृतियों में से एक है।

मराठी शब्द - कार्य (नाम) (समान अर्थ वाले शब्द -काम, कर्म, कृत्य, कृती, क्रिया काम, कार्य, कामकाज, कामाठी)

मराठी अनुवाद - ताजमहल विश्वातल्या सर्वोत्तम कार्यां मधून एक आहे.

1.हिंदी शब्द - तरनी(नाम) सूर्य(नाम)

उदाहरण - सूर्य पश्चिम में अस्त होता है।

मराठी शब्द - सूर्य(नाम) (समान अर्थ वाले शब्द -सूर्य, भानु, दिवाकर, भास्कर, प्रभाकर, दिनकर, मित्र, मिहिर, रवि, दिनेश, अर्क, सविता, गभस्ति, चंडांशु, दिनमणी, सहस्ररश्मी, रश्मीकर, सूर्यनारायण, आदित्य, अंशुमान)

मराठी अनुवाद - सूर्याचा अस्त पश्चिमेला होतो.

2.हिंदी शब्द - तरणि(नाम) नाव(नाम)

उदाहरण - प्राचीन काल में नौका यातायात का प्रमुख साधन थी।

मराठी शब्द - नाव(नाम) (समान अर्थ वाले शब्द -नौका, होडी, तरांडे, तारू, मचवा, पडाव, शिबाड, होडगे, तर, डोंगी, डोणी, डोंगा, तराफा)

मराठी अनुवाद - प्राचीनकाळी नाव प्रवासाचे प्रमुख साधन होती.

1.हिंदी शब्द - दिन(नाम) दिवस(नाम)

उदाहरण - आज का दिन मेरे लिए बहुत ही अच्छा रहा।

मराठी शब्द - दिवस (नाम) (समान अर्थ वाले शब्द -दिनमान, वार)

मराठी अनुवाद - आजचा दिवस माझ्यासाठी खूप छान राहिला.

2.हिंदी शब्द - दीन(नाम) /(विशेषण) गरीब(नाम)

उदाहरण - निर्धन व्यक्ति कड़ी मेहनत करके धनी हो सकता है।

मराठी शब्द - दिन (समान अर्थ वाले शब्द -गरीब, निर्धन, अकिंचन, द्रव्यहीन, दरिद्री, रंक, खंक)

मराठी अनुवाद - दिन व्यक्ती मेहनत करुन श्रीमंत होऊ शकतो.

1.हिंदी शब्द - पुरा(नाम) बस्ती(नाम)

उदाहरण - खड़गपुरा में चोरों की कमी नहीं है।

मराठी शब्द - पूर (नाम) (समान अर्थ वाले शब्द -ग्राम,गाव,वस्ती)

मराठी अनुवाद - खड़गपूर मध्ये चोरांची कमी नाही.

2.हिंदी शब्द - पूरा(विशेषण) पूर्ण(विशेषण)

उदाहरण - महेश का इस कार्यालय पर पूरा नियंत्रण है।

मराठी शब्द - संपूर्ण(विशेषण) (समान अर्थ वाले शब्द - सगळा, अखिल, सर्व, सबंध, अख्खा, आख्खा, उभा, समग्र, समस्त, सारा, अवघा, तमाम पूर्ण, पुरेपूर)

मराठी अनुवाद - महेशचे या कार्यालयावर संपूर्ण नियंत्रण आहे.

1.हिंदी शब्द - फ़क(विशेषण) फीका(विशेषण)

उदाहरण - यह सब्जी तो फ़क है।

मराठी शब्द - आळणी (विशेषण) (समान अर्थ वाले शब्द -सपक, फिकी)

मराठी अनुवाद - ही भाजी तर आळणी आहे.

2.हिंदी शब्द - फ़क्क(विशेषण) फक रंग(विशेषण)

उदाहरण - राजीव फ़क्क रंग के कपड़ों में ही अच्छा दिखाई देता है।

मराठी शब्द - पांढरा (विशेषण) (समान अर्थ वाले शब्द -श्वेत, धवल, शुभ्र, सफेत, सफेद)

मराठी अनुवाद - राजीव पांढ-या रंगाच्या कपड्यांनमध्येच चांगला दिसतो.

1.हिंदी शब्द - फन(नाम) सर्पफण(नाम)

उदाहरण - नाग बीन की आवाज सुनकर अपना फन इधर-उधर घुमाने लगा।

मराठी शब्द - फना (नाम) (समान अर्थ वाले शब्द -फणी, फडा)

मराठी अनुवाद - नाग पुंगीचा आवाज ऐकुण आपला फना इकडे-तिकडे फिरऊ लागला.

2.हिंदी शब्द - फ़न(विशेषण) पारंगत(विशेषण)

उदाहरण - वह गाने में फ़नकार है।

मराठी शब्द - तरबेज (विशेषण) (समान अर्थ वाले शब्द -प्रवीण, निपुण, हुशार, कुशल, जाणता, पारंगत, जाणकार, निष्णात, तज्ज्ञ, तरबेज, वाकबगार, हातखंडा, पटाईत, तज्ञ, फरडा, विशारद, कसबी)

मराठी अनुवाद - तो गाण्या मध्ये तरबेज आहे.

1. हिंदी शब्द - बलि(नाम) निछावर(नाम)

उदाहरण - मैं बलि जाऊँ पिया पर।

मराठी शब्द - अनुरक्त(नाम) (समान अर्थ वाले शब्द -मोहित, लुब्ध, आकृष्ट, लट्टू, लोलुप, फिदा, आसक्त,)

मराठी अनुवाद - मी अनुरूक्त होऊन जाइन प्रियकरावर.

2.हिंदी शब्द - बली(विशेषण) वीर(विशेषण)

उदाहरण - हनुमान बजरंग बली थे।

मराठी शब्द - बलवान(विशेषण) (समान अर्थ वाले शब्द -शक्तिमान, शक्तिवान, शक्तिवंत, बलशाली, जबर, बलाढ्य,शक्तिशाली)

मराठी अनुवाद - हनुमान बजरं बली होते.

1.हिंदी शब्द - बहु(विशेषण) अधिक(विशेषण)

उदाहरण - बहुत लोगों का जमावडा लगा था।

मराठी शब्द - कित्येक (विशेषण) (समान अर्थ वाले शब्द - अनेक, पुष्कळ, नाना, कैक,)

मराठी अनुवाद - कित्येक लोकांचा जमावडा लागला होता.

2.हिंदी शब्द - बहू(नाम) पुत्रवधू(नाम)

उदाहरण - मेरी बहू बहुत सुशील है।

मराठी शब्द - सून(नाम) (समान अर्थ वाले शब्द - सूनबाई, स्नुषा)

मराठी अनुवाद - माझी सून खूप सुशील आहे.

1.हिंदी शब्द - बाजि(नाम) घोड़ा(नाम)

उदाहरण - इस बाजि की रेस पर कभी विश्वास नहीं करना चाहिए।

मराठी शब्द - घोडा (नाम) (समान अर्थ वाले शब्द - अश्व, तुरग, तुरंग, तुरंगम, हय, वारू)

मराठी अनुवाद - या घोड्यांच्या रेस वरती कधी विश्वास नाही ठेवाला पाहिजे।

2.हिंदी शब्द - बाज़ी(नाम) दाँव(नाम)

उदाहरण - बाजी जीतनेवाला जुआरी बहुत प्रसन्न था।

मराठी शब्द - बाजी(नाम) (समान अर्थ वाले शब्द -पाळी, खेप, बारी, डाव)

मराठी अनुवाद - बाजी जिंकणारा जूगारी खूप प्रसंन्न होता.

1.हिंदी शब्द - साधु(नाम) सन्त(नाम)

उदाहरण - साधु का जीवन परोपकार में ही व्यतीत होता है।

मराठी शब्द - संत(नाम) (समान अर्थ वाले शब्द -ऋषी, तपस्वी, साधू)

मराठी अनुवाद - संतांचे जीवन परोपकारा मध्येच व्यतीत होते.

2.हिंदी शब्द - साधू(नाम) महात्मा(नाम)

उदाहरण - साधु का जीवन परोपकार में ही व्यतीत होता है।

मराठी शब्द - साधु (नाम) (समान अर्थ वाले शब्द -ऋषी, तपस्वी,संत)

मराठी अनुवाद - साधूचे जीवन परोपकारा मध्येच व्यतीत होते.

1.हिंदी शब्द - सूर(विशेषण) अंध(विशेषण)

उदाहरण - सूर बनके काम करना सबसे कठिन है।

मराठी शब्द - अंध(विशेषण) (समान अर्थ वाले शब्द -आंधळा, नेत्रहीन, दृष्टिहीन)

मराठी अनुवाद - अंध होऊन काम करने सर्वात कठिन आहे.

2.हिंदी शब्द - सुर(नाम) स्वर(नाम)

उदाहरण - षडज,ऋषभ,गांधार,मध्यम,पंकज,धैवत और निषाद ये सात संगीत स्वर हैं।

मराठी शब्द - सूर (समान अर्थ वाले शब्द -सूरावट,स्वर)

मराठी अनुवाद - षडज,ऋषभ,गांधार,मध्यम,पंकज,धैवत आणि निषाद हे सात संगीत स्वर आहेत.

1.हिंदी शब्द - पूछ(क्रिया) पूछताछ(क्रिया)

उदाहरण - इतनी पूछताछ का भी कोई फायदा नहीं हुआ।

मराठी शब्द - चौकशी(क्रिया) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - इतक्या चौकशीचा ही काही उपयोग झाला नाही.

2.हिंदी शब्द - पूँछ(नाम) दूम(नाम)

उदाहरण - गाय,भैंस आदि की पूँछ होती है।

मराठी शब्द - शेपूट(नाम) (समान अर्थ वाले शब्द -शेपटी, शेप)

मराठी अनुवाद - गाय, म्हैस आदिंना ही शेपूट असते.

1.हिंदी शब्द - है(क्रिया) है(क्रिया)

उदाहरण - रमा उस कमरे में है।

मराठी शब्द - आहे (क्रिया) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - रमा त्या खोलीत आहे.

2.हिंदी शब्द - हैं (क्रिया) हैं (क्रिया)

उदाहरण - हम चोर नहीं हैं।

मराठी शब्द - आहेत (क्रिया) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - आम्ही चोर नाही.

1.हिंदी शब्द - बिन(सर्वनाम) बगैर(सर्वनाम)

उदाहरण - उसके बिन तुम्हारा काम नहीं हो सकता।

मराठी शब्द - विना(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद – त्याच्या विना तुझं काम होऊ शकणार नाही.

2.हिंदी शब्द - बीन(नाम) बीन(नाम)

उदाहरण - बीन बजाते ही साँप अपना सिर हिलाने लगा।

मराठी शब्द - पुंगी (नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - पुंगी वाजवताच साप आपले डोके हलवू लागला.

1.हिंदी शब्द - चुना(विशेषण) चयनित(विशेषण)

उदाहरण - इस पुरस्कार के लिए राम को चुना गया है।

मराठी शब्द - चयनित (विशेषण) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - या पुरस्कारासाठी रामाचे चयन करण्यात आले आहे.

2.हिंदी शब्द - चूना(नाम) चूना(नाम)

उदाहरण - चूने का अधिकतर प्रयोग दीवारों की पुताई करने में किया जाता हैं।

मराठी शब्द - चूना(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - चून्याचा अधिकतर उपयोग भिंती रंगवण्यासाठी केला जातो.

1.हिंदी शब्द - कही(नाम) कहा(नाम)

उदाहरण - हमारी कही मानो।

मराठी शब्द - म्हणणे (समान अर्थ वाले शब्द -)

मराठी अनुवाद - आमचे म्हणंणे ऐक.

2.हिंदी शब्द - कहीं(क्रिया-विशेषण) कहूँ(क्रिया-विशेषण)

उदाहरण - राम यहीं कहीं होगा।

मराठी शब्द - कुठे तरी(क्रियाविशेषण) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - राम इथेच कुठे तरी असेल.

1.हिंदी शब्द - सूना(नाम) निर्जन(नाम)

उदाहरण - कुछ लोग निर्जन स्थान में निवास करना पसंद करते हैं।

मराठी शब्द - ओसाड (नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - काही लोक ओसाड जागी राहणे पसद करतात.

2.हिंदी शब्द - सुना(विशेषण) सुना हुवा(विशेषण)

उदाहरण - मेरी अनपढ़ दादी हमें श्रुत कथाएँ सुनाती थी।

मराठी शब्द - ऐकीव (विशेषण) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - माझी अनपड आजी आम्हाला ऐकिव गोष्टि सागायची.

1.हिंदी शब्द - काटा(क्रिया) काटा(क्रिया)

उदाहरण - कुत्ते ने श्याम को काटा।

मराठी शब्द - चावला (क्रिया) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - कुत्र्याने श्यामला चावले .

2.हिंदी शब्द - काँटा(नाम) कंटक(नाम)

उदाहरण - जंगल से गुज़रते समय उसके पैर में काँटा चुब गया।

मराठी शब्द - काटा(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - जंगलातून जाताना त्याच्या पायामध्ये काटा टोचला.

3.हिंदी शब्द - काँटा(नाम) तराजू (नाम)

उदाहरण - किसान अनाज़ आदि तौलने के लिए काँटा रखते हैं।

मराठी शब्द - तराजू (नाम) (समान अर्थ वाले शब्द - काटा)

मराठी अनुवाद - शेतकरी धान्य आदि तोलण्यासाठी तराजू ठेवतात.

1.हिंदी शब्द - चाँटा (नाम) थप्पड़ (नाम)

उदाहरण - उसने मुझे एक चाँटा मारा।

मराठी शब्द - चापट (नाम) (समान अर्थ वाले शब्द -थप्पड, चपराक)

मराठी अनुवाद - त्याने मला एक चापट मारली.

2.हिंदी शब्द - चाटा(विशेषण) चाटा हुआ(विशेषण)

उदाहरण - वह बिल्ली का चाटा हुआ ग्लास है।

मराठी शब्द - चाटलेला(विशेषण) (समान अर्थ वाले शब्द –चाटलेले, चाटलेली)

मराठी अनुवाद - ते मांजराने चाटलेला ग्लास आहे.

1.हिंदी शब्द - जाए(क्रिया) जाइए(क्रिया)

उदाहरण - आप घर जाए हम युद्धक्षेत्र विजयी कर लौटते हैं।

मराठी शब्द - जावं (क्रिया) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - आपण घरी जावं आम्ही युद्धक्षेत्र जिंकून येत आहोत.

2.हिंदी शब्द - जाँए (क्रिया) गमन, जाना(क्रिया)

उदाहरण - हम घर जाँए।

मराठी शब्द - जाऊ (क्रिया) (समान अर्थ वाले शब्द -जा)

मराठी अनुवाद - आम्ही घरी जाऊ.

1.हिंदी शब्द - सास(नाम) सासु(नाम)

उदाहरण - कौशल्या सीता की सास थी।

मराठी शब्द - सासू(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - कौशल्या सीताची सासू होती.

2.हिंदी शब्द - साँस(नाम) सांस(नाम)

उदाहरण - साँस लेना सजीव प्राणियों का लक्षण है।

मराठी शब्द - श्वास(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - श्वास घेणे सजीव प्राण्यांचे लक्षण आहे.

1.हिंदी शब्द - खास(विशेषण) असाधारण(विशेषण)

उदाहरण - मैं यहाँ एक खास काम से आया हूँ।

मराठी शब्द - खास (विशेषण) (समान अर्थ वाले शब्द -असामान्य, विशेष, असाधारण)

मराठी अनुवाद - मी इथे एका खास कामासाठी आलो आहे.

2.हिंदी शब्द - खाँस(क्रिया) खांसना(क्रिया)

उदाहरण - वह रोज घर के सामने से खाँस कर गुजरता है।

मराठी शब्द - ठसका(क्रिया) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - तो ठसका घेत रोज घरा समोरुन निघिन जातो.

4. अंग्रेजी भाषा से आये अनेकार्थक शब्द

भाषिक दृष्टि से देखा जाए तो हिंदी में उत्पन्न हुई यह एक पूर्णत: नवीन समस्या है। जिसपर हिंदी भाषा शास्त्रियों को ध्यान देने की आवश्यकता है। जैसे-जैसे अंग्रेजी भाषा का हिंदी पर प्रभाव बढ़ता जा रहा है और अंग्रेजी के शब्दों को हिंदी में स्थान मिलता जा रहा है। हिंदी में यह समस्या बढ़ती जा रही हैं। भारत में अंग्रेजी राजभाषा के रुप में उपयोग में लाई जाती है। इसके कारण अनुवाद या उपयोग कर्ता अंग्रेजी के ऐसे शब्दों को जिनका अनुवाद नहीं हो सकता। अनुवाद करने से अधिक अंग्रेजी शब्द को ही रखना उपयोगी समझता है। ऐसे समय अंग्रेजी के शब्दों को लिप्यंतरित कर के रखा जाता है, जो कुछ समय बाद अनेकार्थकता उत्पन्न करते हैं। जिसका निराकरण करना भी मशीनी अनुवाद की एक महत्वपूर्ण समस्या बन जाती है। खास कर उन मशीनों के लिए जो प्रशासनिक भाषा के अनुवाद पर काम कर रही हैं। दूसरी भाषाओं से शब्दों को आयातित करना भाषाविज्ञान की दृष्टि से गलत काम नहीं है। इससे भाषा समृद्ध होती तो है, परंतु इन शब्दों के कारण अन्य समस्याऐं उत्पन्न न हो पाये इसपर भी ध्यान देना आवश्यक है। कई बार अन्य भाषाओं से शब्द आयातित करने का यह नुकसान उठाना पड़ता है कि अन्य भाषाओं से लिया गया शब्द इतना लोकप्रिय हो जाता है। और अपनी भाषा में होने वाला उस शब्द का समानार्थी शब्द लोग भूल जाते हैं। इस प्रकार के कई उदाहरण मिलते हैं। उदाहरण के रुप में हम निम्न शब्दों को देख सकते हैं

। जिनके निराकरण की चर्चा भी की गई है। अंग्रेजी से हिंदी में आए शब्द लेकिन हिंदी में अनेकार्थकता निर्माण करने वाले शब्दों की सूची देकर इन शब्दों का मराठी में अनुवाद किया जा रहा है। इस संकलन में अधिक से अधिक शब्दों को देने का प्रयास किया गया है।

1.हिंदी शब्द - टेस्ट(नाम) स्वाद(नाम)

उदाहरण - वाह क्या टेस्ट है।

मराठी शब्द - स्वाद (नाम) (समान अर्थ वाले शब्द -अभिरुची,चव, लज्जत, स्वाद, गोडी,रुची,खुमारी)

मराठी अनुवाद - वाह काय स्वाद आहे.

2.हिंदी शब्द - टेस्ट(नाम) इम्तहान(नाम)

उदाहरण - आजका टेस्ट काफी खराब रहा।

मराठी शब्द - चाचणी (नाम) (समान अर्थ वाले शब्द - कसोटी,परीक्षा, पारख,तपासणी, परीक्षण,सत्यपरीक्षा,निर्णय, निवडचाचणी,चाचणीपरीक्षा)

मराठी अनुवाद - आजची चाचणी फार खराब राहिली.

1.हिंदी शब्द - इंटरेस्ट(नाम) दिलचस्पी(नाम)

उदाहरण - इस विषय में तुम्हारा भी इंटरेस्ट दिखाई देता है।

मराठी शब्द -आवड (नाम) (समान अर्थ वाले शब्द -पसंत,हौस, शौक,पसंती,आंबटशोक,छंद,ओत्सुक्य)

मराठी अनुवाद - या विषयात तुझी ही आवड दिसते.

2. हिंदी शब्द -इंटरेस्ट(नाम) ब्याज(नाम)

उदाहरण - एक लाख का कितना ब्याज होगा।

मराठी शब्द - व्याज (नाम) (समान अर्थ वाले शब्द -ऋण,कर्ज, उधारी, उधारी, बोजा,उसनवट, उचल, उधारी, थकबाकी,येणे,तगाई,सावकारी,व्याजबट्टा)

मराठी अनुवाद - एक लाखाचे किती व्याज होइलं.

1.हिंदी शब्द - कंडीशन(नाम) हालत(नाम)

उदाहरण - मुझे इस मरीज की कंडीशन ठीक नहीं लगती.

मराठी शब्द - हाल (नाम) (समान अर्थ वाले शब्द -दुर्दशा, दुर्गती, दुःस्थिती,हालअपेष्टा, दैना, धूळधाण, वाताहत,गैरसोय, अडचण, कुचंबणा,त्रास, हालत, हैराणी,कष्ट,धकाधकी)

मराठी अनुवाद - मला या रोग्याचे हाल ठिक वाटत नाहीत.

2.हिंदी शब्द - कंडीशन(नाम) शर्त(नाम)

उदाहरण - यह कंडीशन मुझे मंजूर नहीं है।

मराठी शब्द - प्रतिबंध (नाम) (समान अर्थ वाले शब्द -मज्जाव, प्रतिषेध, अटकाव, प्रतिषेध , स्तंभण)

मराठी अनुवाद - मला हा प्रतिबंध मंजूर नाही.

1.हिंदी शब्द - कमीशन (नाम) आयोग(नाम)

उदाहरण - यू.जी.सी. के कमीशन से इस विषय को मंजूरी मिली है।

मराठी शब्द - आयोग (नाम) (समान अर्थ वाले शब्द –कमिशन, आधिकारीवर्ग, शासक,प्रशासक)

मराठी अनुवाद - यू.जी.सी. च्या आयोगा द्वारे या विषयाला मंजूरी मिळाली आहे

2.हिंदी शब्द - कमीशन(नाम) दलाली(नाम)

उदाहरण - तुम्हे कितना कमीशन मिल रहा है।

मराठी शब्द - दलाली (नाम) (समान अर्थ वाले शब्द -अडत, कमिशन, मध्यस्थी, व्यापारउद्यमी, गुमास्तेगिरी)

मराठी अनुवाद - तुला किती दलाली मिळतं आहे.

1.हिंदी शब्द - टाइप(क्रिया) टंकन(क्रिया)

उदाहरण - हिन्दी में टाइप करना आज आम बात है।

मराठी शब्द - टंकन (क्रिया) (समान अर्थ वाले शब्द -टाइप)

मराठी अनुवाद - हिंदी मध्ये टंकन करने आज साधारण गोष्ट आहे.

2.हिंदी शब्द - टाइप(नाम) प्रकार(नाम)

उदाहरण - यह किस टाइप का लड़का है।

मराठी शब्द - प्रकार (नाम) (समान अर्थ वाले शब्द -पद्धत, पद्धती, तंत्र, रीत, तऱ्हा, धाटणी, ढंग, ढब, शैली)

मराठी अनुवाद - हा कोणत्या प्रकारचा मुलगा आहे .

1.हिंदी शब्द - टेबल(नाम) तालिका(नाम)

उदाहरण - यह बच्चों के नामों का टेबल है।

मराठी शब्द - तालिका (नाम) (समान अर्थ वाले शब्द –टेबल,

सारणी,सूची,तक्ता,खानेसुमारी,नावनिशी)

मराठी अनुवाद - ही मुलांच्या नावांची तालिका आहे.

2.हिंदी शब्द - टेबल(नाम) मेज़(नाम)

उदाहरण - यह किताब टेबल पर रख दो।

मराठी शब्द - टेबल (नाम) (समान अर्थ वाले शब्द –मेज, तिवई,तपाई,तिकटे,घडवंची,चौकटी)

मराठी अनुवाद - हे पुस्तक टेबल वरती ठेव.

1.हिंदी शब्द - ट्रस्ट(नाम) विश्वास (नाम)

उदाहरण - मुझ पर ट्रस्ट करो मैं कल पैसा लौटा दूंगा।

मराठी शब्द - विश्वास (नाम)(समान अर्थ वाले शब्द -भरवसा, भरोसा,खात्री, निश्चिती,

हमी,ग्वाही,आशा,विश्रंभ,उमेद,हमी)

मराठी अनुवाद - माझ्यावरती विश्वास ठेव मी उद्या पैसे परत करेन।

2.हिंदी शब्द - ट्रस्ट(नाम) संस्था(नाम)

उदाहरण - निसंतान दंपत्ति ने अपनी सारी संपत्ति एक ट्रस्ट को दान कर दी।

मराठी शब्द - संस्था (नाम) (समान अर्थ वाले शब्द -संघटना,प्रणाली, मंडळ, प्रणाली, समूह,

महामंडळ, प्रणाली, निगम ,केंद्र, मंडळी)

मराठी अनुवाद - नि:संतान दाम्पत्याने ने आपली सर्व संपत्ती एका संस्थेस दान केली.

1.हिंदी शब्द - नेट (नाम) मुनाफा(नाम)

उदाहरण - नेट मिलता हो तो दलाल काम करते हैं।

मराठी शब्द - मिळकत (नाम) (समान अर्थ वाले शब्द -उत्पन्न, प्राप्ती,मत्ता, आवक, फायदा,

फळ, जोड, उपज, अवांतर)

मराठी अनुवाद - मिळकत मिळत असेल तर दलाल काम करतात.

2.हिंदी शब्द - नेट (नाम) जाल(नाम)

उदाहरण - वह नेट पर क्रिकेट का अभ्यास करता है।

मराठी शब्द - जाळी (नाम) (समान अर्थ वाले शब्द - जाल,जाली,अवकुंठन,कुछ,चाळण,पाश)

मराठी अनुवाद - तो क्रिकेटचा अभ्यास जाळी मध्ये करतो.

1.हिंदी शब्द - नोट(नाम) टिप्पणी(नाम)

उदाहरण - इस ग्रंथ के गूढ़ वाक्यों को समझने के लिए जगह-जगह नोट दी गयी है।

मराठी शब्द - टिप्पण (नाम) (समान अर्थ वाले शब्द -टिप्पणी, टिपणी,भाष्य, विवरण,परामर्श, स्पष्टीकरण,खूलासा,टीप,तळटीप)

मराठी अनुवाद - या पुस्तकाच्या गुढ वाक्यांना समजन्यासाठी जागो-जागी टिप्पण दिलं गेलं आहे.

2.हिंदी शब्द - नोट(नाम) · कागज़ी मुद्रा (नाम)

उदाहरण - वह मुझे सौ का नोट दिखा रहा था।

मराठी शब्द - नोट (नाम) (समान अर्थ वाले शब्द -कागदी चलन)

मराठी अनुवाद - तो मला शंभचे नोट दाखवत होता।

1.हिंदी शब्द -पार्टी(नाम) दल (नाम)

उदाहरण - भारत में राजनीतिक दल कुकुरमुत्तों की तरह उग आये हैं।

मराठी शब्द - पक्ष (नाम) (समान अर्थ वाले शब्द -पार्टी,बाजू,पंखपाख, पर)

मराठी अनुवाद - भारता मध्ये राजकारणी दल कुत्र्याच्या छत्री प्रमाणे उगवले आहेत.

2.हिंदी शब्द - पार्टी(नाम) दावत(नाम)

उदाहरण - उसने आज सबको अपने यहाँ पार्टी पर बुलाया है।

मराठी शब्द - मेजवानी (नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - त्याने आज सर्वांना त्याच्या कडे मेजवानी साठी बोलवलं आहे.

1.हिंदी शब्द - पार्ट(नाम) भूमिका(नाम)

उदाहरण - इस नाटक में रमेश का पार्ट अच्छा है।

मराठी शब्द - भूमिका (नाम) (समान अर्थ वाले शब्द -पात्र)

मराठी अनुवाद - या नाटका मध्ये रमेशची भूमिका चांगली आहे.

2. हिंदी शब्द - पार्ट(नाम) भाग(नाम)

उदाहरण - गाडी का यह पार्ट खराब है।

मराठी शब्द - हिस्सा (नाम) (समान अर्थ वाले शब्द - अंश,हिस्सा, भाग)

मराठी अनुवाद - गाडीचा हा भाग खराब आहे.

1.हिंदी शब्द - पावर(नाम) ताकत(नाम)

उदाहरण - इस पहलवान में काफी पावर है।

मराठी शब्द - ताकद (नाम) (समान अर्थ वाले शब्द -समर्थता, क्षमता, सामर्थ्य)

मराठी अनुवाद - या पहिलवानत खूप ताकद आहे.

2.हिंदी शब्द - पावर (बिजली)

उदाहरण - पानी से भी बिजली उत्पन्न की जाती है।

मराठी शब्द - विज (नाम) (समान अर्थ वाले शब्द -विद्युल्लता, सौदामिनी, चपला, दामिनी, तडित, बिजली,विद्युत, करंट)

मराठी अनुवाद - पाण्याने ही वीज निर्माण केली जाते.

1.हिंदी शब्द - पास(नाम) प्रवेश पत्र(नाम)

उदाहरण - राजू ने रेल यात्रा के लिए एक महीने का पास निकाला।

मराठी शब्द - प्रवेश-पत्र (नाम) (समान अर्थ वाले शब्द –पास, प्रमाणपत्र,ओळखपत्र,

अधिकारपत्र, पारपत्र)

मराठी अनुवाद - राजू ने रेल्वे प्रवासासाठी एक महिन्याचे प्रवेश-पत्र बनवले आहे.

2.हिंदी शब्द - पास (विशेषण) मंजूर (विशेषण)

उदाहरण - लोकसभा में पास हुआ विधेयक जल्दी ही लागू हो जाएगा।

मराठी शब्द - मंजूर (विशेषण) (समान अर्थ वाले शब्द –मंजूरीप्राप्त, मान्यताप्राप्त,अभिमत, अनुमत, मान्य, स्वीकृत, पारित, कबूल)

मराठी अनुवाद - लोकसभेमध्ये मंजूर झालेले विधेयक लवकरच लागू होईल.

3.हिंदी शब्द - पास(क्रियाविशेषण) नजदीक(क्रियाविशेषण)

उदाहरण - श्याम के घर के पास एक विद्यालय है।

मराठी शब्द -जवळ(क्रियाविशेषण) (समान अर्थ वाले शब्द -समीप, निकट, आजूबाजूस,भोवताली, आवतीभोवती,जवळपास,सान्निध्यात,लगत,नजीक,बाजूला)

मराठी अनुवाद - श्यामच्या घरा जवळ एक विद्यालय आहे.

1.हिंदी शब्द - पेपर(नाम) कागज़(नाम)

उदाहरण - एक सफेद पेपर ले आओ।

मराठी शब्द - कागद(नाम) (समान अर्थ वाले शब्द - कमलपत्र, भूर्जपत्र, ताडपत्र, पुष्टिपत्र)

मराठी अनुवाद - एक पाढंरा काग घेऊन ये.

2.हिंदी शब्द - पेपर(नाम) प्रश्नपत्र(नाम)

उदाहरण - इस प्रश्नपत्र में कुल आठ प्रश्न हैं।

मराठी शब्द - प्रश्नपत्र(नाम) (समान अर्थ वाले शब्द -प्रश्नपत्रिका)

मराठी अनुवाद - या प्रश्नपत्रिकेत एकून आठ प्रश्न आहेत.

3.हिंदी शब्द - पेपर(नाम) समाचार पत्र(नाम)

उदाहरण - वह सायंकालीन प्रकाशित समाचार पत्र पढ़ रहा है।

मराठी शब्द - वृत्तपत्र(नाम) (समान अर्थ वाले शब्द -दैनिक, वृत्तपत्र, वर्तमानपत्र, समाचार पत्र, माध्यम, बातमीपत्र)

मराठी अनुवाद - तो सायंकाळचे प्रकाशित वृतपत्र वाचत आहे.

1.हिंदी शब्द - प्वाइंट(नाम) मुद्दा(नाम)

उदाहरण - प्वाइंट यह है कि कौन पहले जाएगा।

मराठी शब्द - मुद्दा (समान अर्थ वाले शब्द -बिंदू, तात्पर्य)

मराठी अनुवाद - मुद्दा हा आहे की कोण पहिल्यांदा जणार.

2.हिंदी शब्द - प्वाइंट(नाम) नोक(नाम)

उदाहरण - पेन का प्वाइंट खराब हो गया है।

मराठी शब्द - टोक (समान अर्थ वाले शब्द - टोक,अग्र, शिखर,शेंडा, माथा, शिरोभाग)

मराठी अनुवाद - पेनचे टोक खराब झाले आहे.

1.हिंदी शब्द - फ़ाइन(नाम) जुर्माना(नाम)

उदाहरण - पुलिस ने तुम पर कितना फ़ाइन किया।

मराठी शब्द - दंड(नाम) (समान अर्थ वाले शब्द -जोर,दंड)

मराठी अनुवाद - पोलिसांनी तुमच्या वरती किती दंड केला.

2.हिंदी शब्द - फ़ाइन(विशेषण) अच्छा(विशेषण)

उदाहरण - मैं अच्छा हूँ, तुम कैसे हो ?

मराठी शब्द - चांगला(विशेषण) (समान अर्थ वाले शब्द - भला,सरस,नीट,सही, छान, योग्य, सरस, बेहतर, भला, उत्तम, उत्कृष्ट)

मराठी अनुवाद - मी ठीक आहे, तू कसा आहेस?

1.हिंदी शब्द - बटन(नाम) घुंडी(नाम)

उदाहरण - आपके कुर्ते का एक बटन टूट गया है।

मराठी शब्द - बटन(नाम) (समान अर्थ वाले शब्द -गुंडी,बटण,गुंडाळा, लड)

मराठी अनुवाद - तुमच्या कुर्त्याचे एक बटन तुटले आहे.

2.हिंदी शब्द - बटन(नाम) स्विच(नाम)

उदाहरण - उसने मशीन चालू करने के लिए बटन दबाया।

मराठी शब्द - बटन (समान अर्थ वाले शब्द -स्विच)

मराठी अनुवाद - त्याने मशीन चालू करण्यासाठी बटन दाबले.

1.हिंदी शब्द - बस(क्रियाविशेषण) नियंत्रित(क्रियाविशेषण)

उदाहरण - बस कीजिए ज़्यादा खाना मत परोसिए।

मराठी शब्द - बस (समान अर्थ वाले शब्द – नियंत्रण)

मराठी अनुवाद - जास्त जेवण वाढू नका बस करा.

2.हिंदी शब्द - बस(नाम) मोटर लॉरी (नाम)

उदाहरण - बस, ट्रेन आदि आम जनता के यातायात के सर्वोत्तम साधन हैं।

मराठी शब्द - बस(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - बस,रेल्वे इत्यादी सामान्य जनतेच्या प्रवासाचे सर्वोत्तम साधन आहेत.

3.हिंदी शब्द - बस(विशेषण) केवल (विशेषण)

उदाहरण - इस समय बस भगवान ही उसकी सहायता कर सकते हैं।

मराठी शब्द - केवळ(विशेषण) (समान अर्थ वाले शब्द -निव्वळ, निवळ, नुसता,फक्त)

मराठी अनुवाद - या वेळी केवळ देवच त्यांची सहायता करु शकेल.

1.हिंदी शब्द - बिल(नाम) छिद्र(नाम)

उदाहरण - साँप अपने बिल में घुस गया।

मराठी शब्द - बीळ (नाम) (समान अर्थ वाले शब्द - होल,भोक,विवर,कपार,खबदाड,बळद,कोटर,ढोली)

मराठी अनुवाद - साप त्याच्या बीळात घुसला.

2.हिंदी शब्द - बिल(नाम) विधेयक(नाम)

उदाहरण - इस विधेयक को लेकर विधानसभा में ख़ूब हंगामा हुआ।

मराठी शब्द - विधेयक(नाम) (समान अर्थ वाले शब्द - कायदा,कायदेकानून,घटना, फलक, तक्ता, सूचनाफलक, फळी)

मराठी अनुवाद - या विधेयकाला घेऊन विधानसभेमध्ये खूप गोंधळ झाला.

1.हिंदी शब्द - बोर्ड(नाम) मंडल(नाम)

उदाहरण - परिषद का निर्णय सर्वमान्य होगा।

मराठी शब्द - मंडळ (समान अर्थ वाले शब्द -चमू, गट, संघ, दल)

मराठी अनुवाद - परिषदेचा निर्णय सर्व मान्य असेल.

2.हिंदी शब्द - बोर्ड(नाम) फलक(नाम)

उदाहरण - गुरुजी बोर्ड पर गणित समझा रहे हैं।

मराठी शब्द - फळा(नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - गुरुजी फळ्यावरती गणित समजावत आहेत.

1.हिंदी शब्द - रीडर(नाम) सहायक प्रोफेसर(नाम)

उदाहरण - हिंदी विश्वविद्याल में रीडर का पद खाली है।

मराठी शब्द - सहायक प्रोफेसर (नाम) (समान अर्थ वाले शब्द -)

मराठी अनुवाद - हिंदी विश्वविद्यालयात सहायक प्रोफेसरचे पद रिक्त आहे.

2.हिंदी शब्द - रीडर(नाम) पाठक(नाम)

उदाहरण - इस पुस्तकालय में रोज दो सौ रीडर आते हैं।

मराठी शब्द - पाठक(नाम) (समान अर्थ वाले शब्द –वाचक,वाचनारा,पठन करनारा)

मराठी अनुवाद - या ग्रंथालयात रोज दोनशे पाठक येतात.

1.हिंदी शब्द - सूप(नाम) तरी(नाम)

उदाहरण - आज मैंने खाने से पहले सूप पिया।

मराठी शब्द - तरी (समान अर्थ वाले शब्द - रस्सा,कडी)

मराठी अनुवाद - आज मी जेवणा अगोदर तरी पीली।

2.हिंदी शब्द - सूप(नाम) फटकनी(नाम)

उदाहरण - वह सूप से चावल फटक रही है।

मराठी शब्द - सूप (समान अर्थ वाले शब्द - फटकणी,पाकडणी)

मराठी अनुवाद - ती सूपाने तांदूळ पाकडत आहे.

1.हिंदी शब्द - मिल(नाम) कारखाना(नाम)

उदाहरण - इस कारखाने के श्रमिकों ने हड़ताल कर दी है।

मराठी शब्द - मिल (समान अर्थ वाले शब्द –कारखाना, उत्पदनगृह,टाकसाळ,गिरणी,यंत्रशाळा)

मराठी अनुवाद - या कारखान्याच्या कामगारांनी संप केला आहे.

2.हिंदी शब्द - मिल(नाम) बत्तख(नाम)

उदाहरण - बत्तख मटमैले रंग की होती है।

मराठी शब्द - बदक (समान अर्थ वाले शब्द -)

मराठी अनुवाद - बदक मळकट रंगाचा असतो.

3.हिंदी शब्द - मील(नाम) किलोमीटर(नाम)

उदाहरण - एक मील बराबर एक दशमलव छः किलोमीटर होता है।

मराठी शब्द - मैल (समान अर्थ वाले शब्द -)

मराठी अनुवाद - एक मैल बरोबर एक दशमलव सहा किलोमीटर.

1.हिंदी शब्द - सील(नाम) बंद(नाम)

उदाहरण - कर न भरने से दुकान सील कर दी गयी।

मराठी शब्द - बंद (समान अर्थ वाले शब्द - ठप्प, थांबलेला, स्थगित,संस्थागित, प्रतिबंध, कुलूपबंद, कुंठित, खंडित)

मराठी अनुवाद - कर न भरल्याणे दूकान बंद करण्यात आले.

2.हिंदी शब्द - सिल(नाम) पत्थर की पटिया(नाम)

उदाहरण - गीता सिल पर भीगी दाल पीस रही है।

मराठी शब्द - पाटा(नाम) (समान अर्थ वाले शब्द -नामफलक,बुरडी, हारा, डाल डालगे,टोपला)

मराठी अनुवाद - गीता पाट्या वरती दाळ बारीक करत आहे.

अंग्रेजी से हिंदी में आए शब्द हिंदी में अनेकार्थकता निर्माण करने का काम करते हैं। जिससे केवल भाषिक संप्रेषण में ही नहीं बल्कि मशीनी अनुवाद के कार्य में भी समस्याऐं उत्पन्न होती हैं। इस समस्या का निराकरण शब्दों की व्याकरणिक कोटियों के अधार पर करना अधिक सरल होगा। इन व्याकरणिक कोटियों को देने से पूर्व एक और समस्या दिखाई देती है, इन शब्दों की व्याकरणिक कोटि अंग्रेजी के अनुसार देनी चाहिए या हिंदी के अनुसार? यह स्पष्ट नहीं हो रहा है। कुछ शब्दों को अंग्रेजी और हिंदी दोनों शब्दों की व्याकरणिक कोटियों में विभाजित करके देखा जाए तो भिन्नताऐं दिखाई देती हैं चूंकि यह शब्द अंग्रेजी से आये हुए हैं, इसलिए इन्हें अंग्रेजी की व्याकरणिक कोटियों में रखना ही उचित होगा। कुछ विद्वानों का मत है कि इन शब्दों को हिंदी ने अपनाया है इसलिए इन शब्दों की व्यारणिक कोटि हिंदी की व्याकरणिक कोटियों के अनुसार ही होनी चाहिए। इस प्रकार से हम यह देख सकते हैं कि मशीन को जिस व्याकरणिक कोटि के आधार पर हिंदी में निर्माण होने वाली अनेकार्थकता की समस्या को स्पष्ट करने में अधिक सहायता होगी हमें उसी व्याकरणिक कोटि का उपयोग करना चाहिए। कुछ जगहों पर हम दोनों सूत्रों का

भी उपयोग कर सकते हैं। कुछ जगह हम इस विषय पर भी विचार कर सकते हैं "A similar treatment may be made of all cases of so-called motion, ie.e.of the "transition of the word from one grammatical category to another"[25] "यह निराकरण उन सभी कोटियों के शब्दों के लिए हो सकता है जो भाषा में गतिशीलता निर्माण करते हों। इस प्रकार के शब्दों को एक व्याकरणिक कोटि से दूसरे व्याकरणिक कोटि में परिवर्तित किया जाए।" जिससे व्याकरणिक कोटियों के भेद के कारण मशीन को अनुवाद करने में सहायता हो। यह कार्य व्याकरणिक नियमों के विरोध में जाता है। इस सूत्र निर्माण के कारण शब्दों को व्याकरणिक कोटि प्रदान करने का काम उत्पन हो जाएगा, ऐसे समय यह काम आसानी से नहीं हो सकता। चीनी भाषा मे निर्माण होने वाले अनेकार्थक शब्दों के निराकरण के लिए इस सूत्र का उपयोग किया जाता है। लेकिन भारतीय भाषाओं के लिए यह सूत्र कितना उपयोगी सिद्ध होगा यह निश्चित नहीं कहा जा सकता। चीनी भाषा की लिपि चित्र लिपि है, हिंदी एक ध्वनि अधारित भाषा है। हिंदी में स्वर और व्यंजनों के आधार से अक्षरों का निर्माण होता है इसलिए इस कार्य के लिए अन्य विद्वान सहमत नहीं होंगे। हिंदी-मराठी में कई समानताऐं हैं जैसे लिपि की समानता, संकृति की समानता, और समान भाषा परिवार। हिंदी और मराठी भाषा में मिलने वाली इन समानताओं के कारण अन्य भाषाओं के मशीनी अनुवाद में मिलने वाली समस्याओं से हिंदी से मराठी मशीनी अनुवाद में मिलने वाली मशीनी अनुवाद की समस्याऐं भिन्न होंगी। जैसे:- लिपि, संस्कृति और भाषा परिवार भिन्न होने वाली दो भाषाऐं। एक समस्या को निम्न रुप से ले सकते हैं। स्रोत भाषा के समान उच्चारण वाले शब्दों का निराकरण लक्ष्य भाषा में अनुदित करने के उपरांत स्पष्ट होता है। लक्ष्य भाषा में भी अनेकार्थक शब्द होने के कारण अनुवाद करते समय यह ध्यान रखा जाना चाहिए कि शब्द चयन के समय ऐसे शब्दों का चयन न हो जिससे अनुदित पाठ में अस्पष्टता का निर्माण हो।

1. माल-हार(noun)

2. माल-सामान(noun)

---

[25] Manual of Lexicography, Homonymy – IGUSTA (Le xical meaning) – page No 83

माल शब्द का प्रथम अर्थ हार दिया गया है परंतु "हार" शब्द भी अपने आपमें अनेकार्थक शब्द है जिससे द्विअर्थकता उत्पन्न होती है। 'माल' शब्द के दो अर्थ हैं। 1.वस्तु,सामान 2.पुष्पहार वाक्य में शब्द का अर्थ 'फूलों की माला' के संबंध में आ जाए तो शब्द चयन करते समय यह ध्यान रखा जाए की शब्द चयन के समय 'माल' शब्द का अनुवाद 'पुष्पहार' ही किया जाए जिससे लक्ष्य पाठ में द्विअर्थकता निर्माण न हो सके। इस कार्य से लक्ष्य पाठ का अनुवाद अधिक सुंदर हो जाएगा। जिससे मशीनी अनुवाद का स्तर बढ़ाने में मदद होगी। इस प्रक्रिया को पूर्ण रुप देने के लिए मशीन के पास लक्ष्य भाषा के कार्पस एवं लक्ष्य भाषा के अनेकार्थक शब्दों का विशिष्ट कोश भी होना चाहिए जो लक्ष्य भाषा के अनेकार्थक शब्दों को स्पष्ट कर सके।

हिंदी के अनेकार्थक शब्दों के अनुवाद में एक और समस्या दिखाई देती है। दूसरी भाषाओं से आये शब्द और हिंदी में होने वाले शब्दों के टकराव के कारण निर्माण होने वाली अनेकार्थकता की समस्या। जैसे - सिल(क्रिया) सिना(क्रिया), सिल पत्थर की पटिया(नाम), सिल(नाम) बंद(नाम) इन शब्दों पर विचार किया जाए तो समान लिपि लेकिन भिन्न अर्थ वाले यह अंग्रेजी और हिंदी के शब्द दो भिन्न भाषाओं से आये हैं। इसलिए अनेकार्थता की समस्या के निराकरण के लिए शब्द उत्पत्ती के उपर भी ध्यान दिया जाना चाहिए।

3.हिंदी शब्द - सिल(क्रिया) सिना(क्रिया)

उदाहरण - कपड़े को सिल दो नहीं तो फट जाएगा

मराठी शब्द - शिलाई(क्रिया) (समान अर्थ वाले शब्द -शिवणे)

मराठी अनुवाद - कपड्याची शिलाई कर नाही तर फटेल.

इस प्रकार से हम देखते हैं कि 'अनेकार्थकता की समस्या मूलत: शब्द पहचान की समस्या है और शब्द पहचान की समस्या किसी भी कार्य में सबसे महत्वपूर्ण एवं कठिन समस्या है भले ही इसके लिए कई दार्शनिक आधार हों।" "The" preceding discussion must not unduly dispirit the lexicographer. After all, homonymy is a problem of the identity of the word; and problems of identity are frequently the most and difficult ones in any

investigation, irrespective of the philosophical basis of the approach.[26] शाब्दिक अस्पष्टता की यह समस्या किसी भाषा के निराकरण के लिए शब्द पहचान के निराकरण के साथ ही समाप्त हो सकती है। शब्द पहचान की इस समस्या को स्पष्ट करने के लिए डेटाबेस में कई प्रकार के उपाय किए जा सकते हैं जिसमें शब्दों को शब्द संख्या के क्रमांक के आधार पर दिए जा सकता हैं। एक शब्द के दो या दो से अधिक अर्थ हो सकते हैं ऐसे समय अर्थ के आधार पर पुन: शब्द संख्या के आधार पर क्रमांक दिए जा सकते हैं। इसी आधार पर इस अनुसंधान के डेटाबेस में शब्द पहचान की समस्या को दूर करने का प्रयास किया गया है। इसी के साथ व्याकरण की कोटियों को संक्षिप्त अक्षरों एवं अनेकार्थक शब्दों की कोटियों से भी दर्शाया गया है।

अनुसंधान का यह अध्याय अनेकार्थक शब्दों का हिंदी से मराठी में अनुवाद कर के किया गया है। यह कार्य मात्र एक संकलन नहीं है, इन शब्दों के अतिरिक्त अन्य कई शब्द हैं जो हिंदी में अनेकार्थकता की समस्या निर्माण करते हैं। अनेकार्थक शब्दों की यह विशेषता होती है कि यह शब्द वाक्य में आने वाले अन्य किसी भी शब्दों की सहायता लिए बगैर अनेकार्थकता निर्माण कर सकते हैं। जब इन शब्दों की अनेकार्थकता के निराकरण पर कार्य होता है तब उस समय वाक्य में आने वाले शब्द भी महत्वपूर्ण हो जाते है। इस प्रकार देखा जाए तो एक शब्द दो प्रकार की भूमिकाओं में दिखाई देता है। सामान्य शब्दकोशों में भी इन शब्दों के अर्थ व्याकरणिक दृष्टि से दोनों कोटियों में स्पष्ट किए जाते हैं। इसी प्रकार मशीनी अनुवाद में भी इन शब्दों को स्पष्ट करना आवश्यक है।

इन शब्दों के संकलन के साथ ही 306 जोडे यानी कुल 612 शब्दों का संकलन तैयार किया गया है। इन शब्दों के मराठी एवं हिंदी समानार्थी शब्दों को भी दिया गया है। इन शब्दों के माध्यम से अनेकार्थक शब्दों के मूल अर्थ तक पहुँचा जा सकता है जिसके आधार पर अनुवाद करने में सहायता होगी। इन शब्दों के आधार पर ही मशीन काम करेगी। यह कार्य किस प्रकार होगा यह संक्षिप्त रुप में दिया जा रहा है। जब मशीन को हिंदी टेक्स्ट दें तो पहले मशीन यह जाँच करेगी की दिए गए टेक्स्ट में कितने अनेकार्थक शब्द हैं।मशीन जब (output)

[26] Manual of Lexicography, Homonymy – IGUSTA (Le xical meaning) – page No 78

लक्ष्य पाठ दे तो मशीन यह स्पष्ट करे की अनेकार्थक शब्दों का अनुवाद क्या किया गया है, जैसे – मूल टेक्स्ट में आने वाले अनेकार्थक शब्दों के किए गए अनूदित शब्दों का रंग बदल जाएगा।

यह कार्य पाठ में निर्माण होने वाले अनेकार्थक शब्दों का मशीन के माध्यम से पूर्ण रुप से निराकरण नहीं करता। हिंदी में निर्माण होने वाले अनेकार्थक शब्दों को हिंदी-से-मराठी अनुवाद के समय अनेकार्थक शब्दों का अनुवाद सही रुप बतलाता है। अनुवाद के लिए सही एवं उपयुक्त पर्याय दिखलाता है। अनेकार्थकता की समस्या ने भाषा में काफी विस्तृत रुप ले लिया है। जिसके निराकरण के लिए कई संसाधनों का एक साथ होना आवश्यक है। हिंदी-मराठी मशीनी अनुवाद का विकास अभी बाल्यअवस्था का ही रुप ले पाया है। अत: हिंदी-मराठी मशीनी अनुवाद यंत्र के अन्य संसाधनों की सहायता न मिल पाने से यह कार्य अनेकार्थकता का संपूर्ण निराकरण नहीं कर सकता। यह हिंदी में उत्पन्न होने वाले विशिष्ट शब्दों के माध्यम से शाब्दिक अस्पष्टता निर्माण करने वाले शब्दों को वाक्यों से परिचित करवाता है। अनुवाद के समय पूर्ण रुप से सही अनुवाद न होने पर समान अर्थ वाले शब्द को देगा जिसका उपयोग कर उपयोग कर्ता(user) शब्द का निरसण कर सकता है।

चौथा अध्याय

समान उच्चारण वाले शब्दों की समस्याओं का संगणकीय निराकरण

चौथे अध्याय का सार

1. समान उच्चारण वाले शब्द की अन्य विशेषतायें

2. संगणक कलनविधि

3. उपयोग

    1. भारतीय भाषा से भारतीय भाषा के संगणकीय कोशों के उपयोग के लिए

    2. भारतीय भाषाओं से भारतीय भाषाओं के मशीनी अनुवाद में उपयोग के लिए

चौथे अध्याय का सार

यह अध्याय मुख्य रुप से 'कंम्प्यूटर प्रोग्रामिंग' के लिए रखा गया है। अनेकार्थक शब्दों पर किये गये अनुसंधान की सहायता से मशीन के द्वारा अनुवाद करते समय मशीन को किन अनेकार्थक शब्दों पर अनुवाद करते समय समस्या उत्पन्न हो सकती है, इसकी प्रोग्रामिंग की गई है। प्रोग्रामिंग निर्देशानुसार अनेकार्थक शब्दों (Homonymy) पर विशिष्ट सहायक लघु शब्दकोश की सहायता से मशीनी अनुवाद के शब्द निर्देशन का काम करेगी। इसी के साथ कुछ व्याकरणिक नियमों की सहायता से भी कलनविधि संपन्न की गई है। प्रत्येक शब्दों का व्याकरणिक नियमों के माध्यम से निराकरण करना अति कठिन कार्य होने के कारण केवल उदाहरण स्वरुप ही कुछ शब्दों पर व्याकरणिक नियमों के मध्याम से निराकरण किया गया है। अन्य शब्दकोशों को देखा जाए तो अधिकतर मशीनें लक्ष्य भाषा में होने वाले शब्दों को ही प्रस्तुत करती हैं। यह अनुसंधान अन्य कार्यों से इसलिए भिन्न है कि इस अनुसंधान में मशीन को दिया गया स्रोत पाठ ही लक्ष्य भाषा के शब्द रुपों में प्रस्तुत न होकर लक्ष्य भाषा के शब्दों के अर्थ को प्रस्तुत करने का काम करेगा। संभवत: यह कार्य अभी संपूर्ण रुप से नहीं हो पाया है लेकिन इस कार्य के द्वारा मशीन को भाषिक समझ देने का काम किया जा रहा है।

अनुसंधान में संकलित किए गए शब्दों का एक डेटाबेस तैयार किया गया है। इसी के साथ अनुसंधान के सैद्धांतिक पक्ष के रुप में लिए गए शब्द भी हैं। इन दोनों संकलित शब्दों का एक विशिष्ट रुप से डेटाबेस बनाया गया है। इस संकलन का उपयोग कर अनेकार्थक शब्दों के संकलन को प्रोग्राम के माध्यम से जोडा गया है। हिंदी में मिलने वाले कुल 640 अनेकार्थक शब्दों को मराठी अनुवाद एवं व्याकरणिक कोटियों के साथ दिया गया है। जिससे इस अध्याय का महत्व दिखाई पड़ता है। इस अध्याय को अंतिम रुप देकर जो कार्य सामने आयेगा उसे http://sanskrit.jnu.ac.in/hhnym की वेब साईट पर रखा गया है। जिसका उपयोग देश के किसी भी कोने से किया जा सकेगा जो इस कार्य की एक महत्वपूर्ण उपलब्धि है।

1. समान उच्चारण वाले शब्द की अन्य विशेषतायें

किसी भी तकनीकी समस्या का अंतिम पड़ाव कलनविधि के द्वारा ही सम्पन्न होता है। जिसके द्वारा अभ्यासक प्राकृतिक भाषा को मशीन की कृत्रिम भाषा की सहायता लेकर समस्या का समाधान खोजने का प्रयास करता है। प्राकृतिक भाषा की कई समस्याओं को कलनविधि के द्वारा सम्पन्न किया गया है। अर्थ निष्पादन की समस्या के समाधान के लिए अन्य ऑन लाईन शब्द कोशों का भी उपयोग किया गया है। वर्डनेट में मशीन द्वारा संचयित शब्द कोश को जितनी आसानी के साथ अंजाम दिया जाता है उसकी तुलना में शब्द और वाक्य के अर्थ का निष्पादन करना सबसे कठिन काम है जिसके कारण मशीन मानव बुद्धि से आगे नहीं जा रही है। इसलिए किसी भी कलनविधि का सम्पन्न होना मात्र उस अनुसंधान कार्य का समापन न मानकर उस अनुसंधान पर पुन: विचार कर अगले अनुसंधान की असफल परस्थितियों का अवलोकन कराना है। भाषा में उत्पन्न होने वाली अनेकार्थकता की समस्या के लिए भी यही सूत्र अपनाया जा रहा है। किसी भी भाषा में अनेकार्थकता आम बात है। यह अनेकार्थकता कितनी होनी चाहिए या किस आधार पर होनी चाहिए इसका कोई पैमाना निर्धारित नहीं किया गया है। प्रमुख रुप से भाषा में अनेकार्थकता का अनुशीलन भाषा के विकास और भाषा की प्रकृति पर निर्भर करता है। इसलिए यह प्रश्न उठना स्वाभाविक है कि एक भाषा में एक शब्द के कितने अर्थ हो सकते हैं? क्या इन शब्दों का विभाजन हो सकता है? क्या इन शब्दों की संख्या समान होती है? जब हम किसी शब्द का निराकरण करने जा रहे हों तो यह जानना आवश्यक हो जाता है कि एक शब्द के कितने अर्थ हो सकते हैं। जिसके आधार पर हम मशीन को यह निर्देश दे सकते हैं कि सामान्य रुप से एक भाषा में एक शब्द के कितने अर्थ हो सकते हैं। मशीन पूर्णत: लॉजिक के आधार पर काम करती है। अत: मशीन को यह जानना आवश्यक होता है कि किस लॉजिक के आधार पर काम करना है। एक सर्वेक्षण के अनुसार एक भाषा में अनेकार्थक शब्द के बारह अर्थ बताए गए हैं लेकिन यह सभी भाषाओं में समान रुप से लागू नहीं है। केट्य पेर्ट्सोव (Katya Pertsova)ने अपने अनुसंधान में सामान्य रुप से एक भाषा में एक शब्द के कितने अनेकार्थक उदारहण मिल सकते हैं इसका प्रमाण प्रस्तुत किया है जो निम्न टेबल में दिया जा रहा है। इस टेबल में स्पष्ट रुप से दिया गया है कि एक भाषा में कम से कम बारह अर्थ हो सकते हैं। टेबल में कुल तीस भाषाओं का सर्वेक्षण किया गया है

जिसमें सबसे अधिक अर्थ देने वाली भाषा चौदह नंबर की भाषा 'डागा' भाषा है, जिसमें सर्वाधिक एक शब्द के बारह अर्थ दिए गए हैं। इस टेबल में हिंदी में एक अनेकार्थक शब्द के पांच अर्थ बताए गए हैं।

	Language	Family	Region	Num. of paradigms
1.	Aleut	Eskimo-Aleut	North America	1
2.	Amahuaca	Panoan	South America	3
3.	Amele	Trans New-Guinea	New-Guinea	4
4.	Atakapa	Gulf	North America	1
5.	Bagirmi	Nilo Saharan	North Africa	3
6.	Beja	Afro-Asiatic	North Africa	1
7.	Bulgarian	Indo-European	East Europe	2
8.	Burarra	Australian	Australia	1
9.	Burushaski	Isolate	South Asia	9
10.	Canelo-Cnaho	Macro-Ge	South America	2
11.	Carib	Carib	South America	2
12.	Cayuvava	Isolate	South America	2
13.	Chinantec	Oto-Manguean	Central America	7
14.	Daga	Trans New-Guinea	New-Guinea	12
15.	Dargwa (Icari)	North Caucasian	South Russia	4
16.	Diola-Fogni	Niger-Kongo	West-Africa	2
17.	French	Indo-European	Europe	5
18.	Harar Oromo	Afro-Asiatic	North Africa	3
29.	Hebrew	Afro-Asiatic	Middle East	3
20.	Hayu	Sino-Tibetan	South-Asia	2
21.	Hindi	Indo-European	South Asia	5
22.	Ibibio	Niger-Kongo	West-Africa	4
23.	Ket	Yenisean	Asia	3
24.	Kiwai	Trans New-Guinea	New Guinea	1
25.	Krongo	Nilo-Saharan	North Africa	1
26.	Kwamera	Austronesian	South-East Asia	1
27.	Ngarinjin	Australian	Australia	1
28.	Olo	Torricelli	New-Guinea	1
29.	Rongpo	Sino-Tibetan	South Asia	6
30.	Teribe	Chibchan	Central America	1

एक भाषा में मिलने वाले अर्थ निर्देशक शब्द : टेबल क्र. 5[1]

हर समाज में प्रत्येक व्यक्ति का अलग-अलग महत्व होता है। इस आधार पर प्रत्येक व्यक्ति की भाषा को विश्लेषित कर देखा जाए तो प्रत्येक व्यक्ति के साथ उसकी भाषा में अनेकार्थक शब्दों का समन्वयन दिखाई देता है। प्रत्येक व्यक्ति अपने-अपने भाषिक ज्ञान के अनुसार अनेकार्थक शब्दों की पहचान करता है। यह निर्धारण प्राकृतिक भाषा और अन्य रुपों से आये शब्द वर्ग के आधार पर होता है। संप्रेषण के अंतराल में व्यक्ति किस प्रकार के सामाजिक परिवेश वाले व्यक्तियों के साथ संप्रेषण कर रहा है इसके आधार पर भी अनेकार्थक शब्दों की दिशा निर्धारित होती है। अनेकार्थक शब्द पहचान (word identity) का निर्धारण

[1] Katya, Pertsova, **2007,** *Learning Form-Meaning Mappings in Presence of Homonymy: a linguistically motivated model of learning inflection* (A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in) Linguistics, University of California, Los Angeles,page No - 84

संप्रेषण की किस अवस्था में हो रहा है इसके आधार पर भी होता है। जैसे संप्रेषण प्रथम पुरुष, द्वितीय पुरुष या तृतीय पुरुष में से किस कोटि में हो रहा है इस पर भी निर्भर करता है। केट्य पेर्ट्सोव (Katya Pertsova) के भाषिक सर्वेक्षण के आधार पर हिंदी में (2=3) दो-तीन के आधार पर अनेकार्थकता का निर्धारण होता है लेकिन कई बार हिंदी में शब्द (1=2=3) यानी एक शब्द, एक अर्थ, दो अर्थ या तीन अर्थ के साथ भी आते हैं। सामान्यत: हिंदी में अनेकार्थक शब्द पहली बार आते हैं तो वह दो अर्थों के साथ आते हैं और विशेष बल दिए जाने पर ही तीन अर्थों के साथ आते हैं। प्रत्येक भाषा में इस अवस्था की संख्या भिन्न-भिन्न होती है। अधिकतर भाषाओं में अनेकार्थक शब्द दो अर्थों के साथ ही आते हैं। कुछ भाषाओं में तीसरे अर्थ की संकल्पना नहीं की गई है,तो कुछ भाषाओं में अनेकार्थक शब्द तीन अर्थों के साथ ही आते हैं। व्यक्ति समन्वयात्मक अर्थ निर्देशक की यह संकल्पना भाषा की विशेषता को भी प्रकट करती है। मशीन द्वारा इस सूत्र का उपयोग व्यक्ति एवं भाषा के अनुसार अर्थ देने के लिए सहायक हो सकता है। व्यक्ति समन्वयात्मक अर्थ निर्देशक शब्दों को हम निम्न टेबल के माध्यम से देख सकते हैं।

form ambiguity	sg	du	pl	number independent
1=2	Amele, Daga, Burushaski		Bagirmi, Rongpo	Amahuaca
2=3	Hindi, Atakapa, Hayu, Bulgarian, French,Rongpo, Amele, Ibibio	Hayu,Amele	Amele,Oromo, Carib,Chinantec, Burushaski, Ibibio	Chinantec, Amahuaca, Kiwai
1inc.=1excl	-		Cayuvava	
1=2=3	French,Dargwa		Burushaski, Rongpo	Hindi,Rongpo Chinantec
1incl=2=3		Hayu		
1incl=3		Kwamera		
1excl=3				Canela-Craho
1excl=2			Burarra	

व्यक्ति समन्वयात्मक अर्थ निर्देशक शब्द : टेबल क्र. 6.[2]

अनेकार्थक शब्दों में उत्पन्न होनेवाली संख्या और उसके विभाजन पर ध्यान देना भी आवश्यक है। शब्दों को अधिक से अधिक सूक्ष्म रुप से विभाजित करके देखा जाए तो शब्दों के अर्थों को समझने में अधिक आसानी होगी। प्रत्येक भाषा में प्राकृतिक शब्द वर्ग एवं अन्य रुपों से आए शब्दों का वर्ग इन दो वर्गों में विभाजित कर अनेकार्थक शब्दों पर विचार होता है। भाषा में इस शब्द विभाजन का अन्यन्य साधारण महत्व है। जो शब्द स्पष्ट रुप से भाषा में अर्थ को स्पष्ट करते हैं ऐसे शब्दों को प्राकृतिक शब्द वर्ग की कोटि

[2] वही पृष्ठ क्र.- 83

में रखा जाता है। हिंदी में **प्राकृतिक एवं अन्य रुपों से आए अनेकार्थक शब्द** यह दो अनेकार्थक शब्द वर्ग मिलते हैं। इन शब्दों का महत्व सामान्य शब्दों की तुलना में अनेकार्थक शब्दों के अनुवाद में अधिक महत्वपूर्ण होता है इसलिए अनेकार्थक शब्दों के अनुवाद के लिए अनुवादक को यह जान लेना आवश्यक है कि अनेकार्थक शब्दों में एक शब्द के पांच अर्थ हों तो इन पांच अर्थों में से कितने अर्थ प्राकृतिक शब्द वर्ग के अनुसार आए हैं और कितने अर्थ अन्य रुपों से आए शब्द वर्ग के अर्थों के अनुसार आते हैं। एक भाषा में प्राकृतिक शब्द वर्ग से आने वाले शब्दों की संख्या अन्य रुपों से आने वाले शब्दों की तुलना में कम हो सकती है। जैसे – बुरुशकी भाषा में एक अनेकार्थक शब्द के सात अर्थ हैं। इन अर्थों को विभाजित करके देखा जाए तो प्राकृतिक शब्द वर्ग से आने वाले शब्दों की संख्या दो है और अन्य रुपों से आने वाले शब्द वर्ग के शब्दों की संख्या पांच है। कुछ ऐसी भी भाषाऐं हैं जिनके सभी अनेकार्थक शब्द अन्य रुपों से आने वाले शब्द वर्ग से आते हैं जैसे – बेगीरमी, बेजा, क्रोन्गो, क्वामेरा, ओलो आदि। तो कुछ ऐसी भी भाषायें हैं जिनके सभी अनेकार्थक शब्द प्राकृतिक शब्द वर्ग की कोटि में आते है। जैसे – अमाहुआका, अमेली, बुल्गेरीयन, फ्रेंच, किवाई आदि। "हिंदी में अनेकार्थक शब्द प्राकृतिक शब्द वर्ग और अन्य रुपों से आए शब्दों का वर्ग इन दोनों रुपों से आते हैं। जिनकी संख्या चार और एक है।"[3] हिंदी में अनेकार्थक शब्दों की संख्या केट्य पेर्ट्सोव (Katya Pertsova) के अनुसार पांच है। जिसे हम 4(प्राकृतिक शब्द वर्ग)+1(अन्य रुपों से आए शब्दों का वर्ग)=5 के रुप में देख सकते हैं। इससे यह भी स्पष्ट हो जाता है कि हिंदी में अनेकार्थक शब्द विभाजन में प्राकृतिक शब्द वर्ग से अर्थ देने वाले शब्दों की संख्या अन्य रुपों से आने वाले शब्द वर्ग की तुलना में अधिक है। केट्य पेर्ट्सोव (Katya Pertsova) के सर्वेक्षण के अनुसार निम्न रुप से अनेकार्थक शब्दों को प्राकृतिक शब्द वर्ग और अन्य रुपों से आए शब्दों के वर्ग से आने वाले शब्दों की संख्या को हम निम्न टेबल में देख सकते हैं।

[3] वही पृष्ठ क्र.- 84

Language	# of parad. with only natural class syncretism	# of parad. with natural class syncretism and other types of syncretism
Amahuaca	2	0
Amele	3	0
Atakapa	1	0
Bagirmi	0	1
Beja	0	1
Bulgarian	1	0
Burushaki	2	5
Carib	1	0
Cayuvava	1	0
Chinantec	5	0
Daga	5	4
Dargwa	2	1
French	2	0
Harar Oromo	2	1
Hayu	1	1
Hebrew	2	1
Hindi	4	1
Ibibio	1	0
Ket	1	1
Kiwai	1	0
Krongo	0	1
Kwamera	0	1
Ngarinjin	1	0
Olo	0	1
Rongpo	3	1

प्राकृतिक वर्ग के अनुसार मिलने वाले उदाहरण एवं शून्य अनेकार्थक शब्द टेबल क्र.7.[4]

"शब्द वर्ग पहचान(Word Identification) की योजना के लिए और शब्दों के उच्चारण को देखने के लिए शब्दकोश सबसे महत्वपूर्ण भूमिका निभाता है। कोशों में निहित शब्दों की व्याख्याओं द्वारा शब्दों के अर्थ को समझने में सहायता होती है। शब्दकोशों में प्रविष्ट शब्दों का वाक्य में उपयोग किस प्रकार किया जाए इसकी भी जानकारी होती है। अत: यह स्पष्ट है कि उच्चारण, अर्थ और शब्दों की व्याख्याओं के द्वारा शब्द वर्ग पहचान करना आसान है।"[5] अनेकार्थक शब्दों के संदर्भ में यह कार्य भी उपयोगी है लेकिन यह नहीं कहा जा सकता कि यह कार्य अनुसंधान का अंतिम उपाय है। जैसे अनेकार्थक शब्दों के एकाधिक अर्थ होने के कारण एक शब्द की अनेक व्याख्याओं को स्पष्ट करना होगा। एक शब्द का उच्चारण समान होते हुए भी अर्थ भिन्न होते हैं, कुछ शब्दों की वर्तनी समान होने के बावजूद उच्चारण भिन्न होता है। इस प्रकार शब्द पहचान के कुछ अन्य उपाय इस प्रकार बताए गए हैं। 1.शब्द को वाक्य में उपयोग कर पढ़ना(Read word in the sentence) 2.उपसर्ग को पहचानना (Look for a prefix) 3.प्रत्यय को पहचानना (look for a suffix) 4.शब्दों के मूलरुप को देखना (Look for stem) 5.अन्य शब्दों के उपयोग के साथ अनेकार्थक शब्दों को

[4] वही पृष्ठ क्र.- 84

[5] http://www.how-to-study.com/study-skills/en/language-arts/16/a-word-identification-strategy/

स्पष्ट करना भी एक उपाय है। इस प्रकार के कुछ उपायों को ध्यान में रखते हुए अनेकार्थक शब्द पहचान के कुछ उपाय कलनविधि के माध्यम से करने की कोशिश की जा रही है।

इस अध्याय में किए गए कलनविधि का प्रमुख उद्देश्य हिंदी में मिलने वाले अनेकार्थक शब्दों को दिखाना है। साथ ही यह भी स्पष्ट करना कि हिंदी से मराठी अनुवाद करते समय अनेकार्थक शब्दों का अनुवाद क्या होगा। इस प्रोग्राम के माध्यम से हम केवल अनेकार्थक शब्दों को ही देख पाऐंगे अन्य शब्दों को इस शब्द संचयन में नहीं दिया गया है। इसलिए यह शब्द संचयन काफी छोटा होते हुए भी महत्वपूर्ण है। कलनविधि इस अनुसंधान के प्रमुख कार्यों में से एक है। इसलिए कलनविधि को विशेषज्ञ की सहायता से पूर्ण किया गया है, जो निम्न रुप से है।

हिंदी में संगणकीय भाषा (programming Language) के किए "क्रमादेशन भाषा" शब्द का उपयोग किया जाता है। जिसके माध्यम से संगणक को निर्देश देने का काम किया जाता है। संगणक मानव द्वारा दिए गए पाठ को क्रमादेशन के साथ जोड़कर अंतिम लक्ष्य पाठ का निर्वाहन करता है। इसलिए किसी भी संगणकीय कार्य में क्रमादेशन का सर्वाधिक महत्व होता है। क्रमादेशन कई प्रकार की भाषाओं में संपन्न होता है। जैसे – सी, सी++, जावा डॉट.नेट, विजुअल बेसिक, एच.टी.एम.एल. आदि। इस अनुसंधान के क्रमादेशन के लिए "जावा" भाषा का उपयोग किया जा रहा है। यह क्रमादेशन अनेकार्थक शब्दों के संकलन पर किया जा रहा है। इन अनेकार्थक शब्दों की संख्या 644 है।

2. संगणक कलनविधि

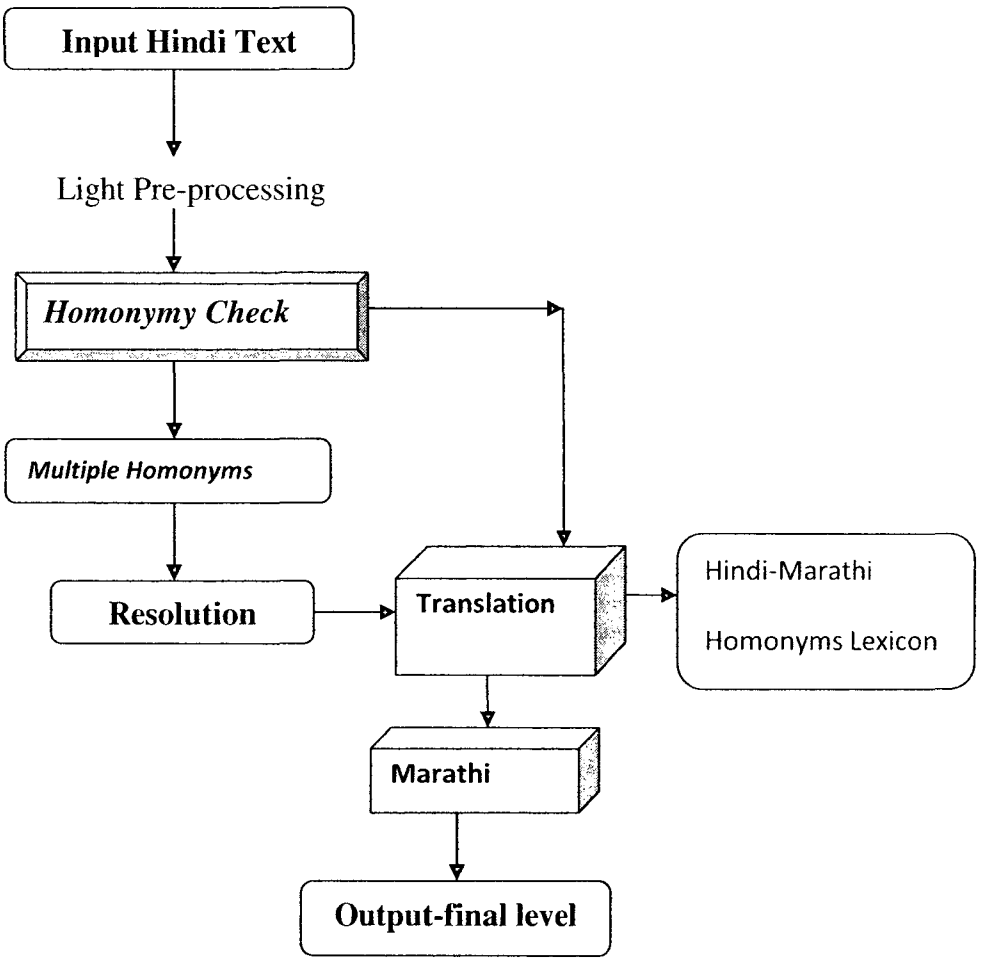
pre-processing –

1.वाक्य में यदि परसर्ग मिल रहा हो तो मशीन परसर्गों की जांच करेगी। एवं उनको शब्दकोश को रुप में परिवर्तीत करेगी।

2.वर्तनी जाँच - अगर दिए गए वाक्य में वर्तनी गलत दी गई हो तो मशीन उसकी भी जाँच करगी।

3.POS टैगर के माध्यम से सभी शब्दों की व्याकरणिक कोटियों का देखा जाएगा।

processing -



(Fig No.3 : process of Homonymy word Translation Tool(Hindi to Marathi))[6]

इस प्रक्रिया का क्रमादेशन निम्न रुप से किया गया है।

(Development Platform)

'जावा भाषा' - जावा एक क्रमादेशन भाषा है जिसे मूल रुप से सन माईक्रोसिस्टम ने बनाया है। इसमा रचनाक्रम काफी हद तक 'सी' एवं सी++ के समान होती है। इसका आब्जेट मॉडल अन्य भाषाओं से तुलनात्मक रुप में सरल होता है। जावा के अनुप्रयोग को कम्पाइल करने पर बाईटकोड प्राप्त होता है। जिसे किसी भी जावा वर्चुअल मशीन पर चलाना संभव है। इस अनुसंधान में जावा भाषा का उपयोग किया गया है। इसी के साथ जे.एस.पी. सर्वर एवं एच.टी.एम.एल. का उपयोग कर वेब पेज का निर्माण किया गया है। एच.टी.एम.एल.एक संगणणिकीय क्रमादेशन

[6] http://sanskrit.jnu.ac.in/hhm/index.jsp

भाषा है। जिसका प्रमुख उपयोग इंटरनेट के वेबपेज(पन्ने) बनाने के लिए उपयोग में लाया जाता है। इस भाषा का भी इसमें उपयोग किया गया है।

Apache Tomcat:- अपेक सॉफ्टवेयर फाउंडेशन(ASF) के द्वारा विकसित किया गया है। अपेक टॉम कॅट जावा में किए गए क्रमादेशन को चलाने के लिए सहायता प्रदान करता है।

Java objects

इस संसाधन का नाम अनेकार्थक शब्द विश्लेषक दिया गया है। जो अनेकार्थक शब्दों के विशिष्ट शब्दकोश की सहायता काम करेगा।

HomonymMarker

LexiconReader

Data files

---

il-posts-tagged-hindi-data.txt (IL-POSTS tagged Hindi data 2.2 MB)

hindi-marathi-homonyms.txt (Hindi-marathi homonym lexicon 94 KB)

The JSP file

---

index.jsp (the main jsp file which calls the java object)

code snippet -

```
<%@ page
        language="java"
        pageEncoding="utf-8"
        contentType="text/html; charset=utf-8"
        import="java.util.*"
%>
```

```
<%@ page import="HomonymMarker"  %>

<%

        request.setCharacterEncoding("UTF-8");

        String otext = request.getParameter("itext");

        if (otext==null)
                otext = "";

        HomonymMarker lbt = new HomonymMarker();

%>

<jsp:include page="../js/include.html" />

<% if (otext.length()>0) { %>
        <%=lbt.tagText(otext) %>
<% } %>
```

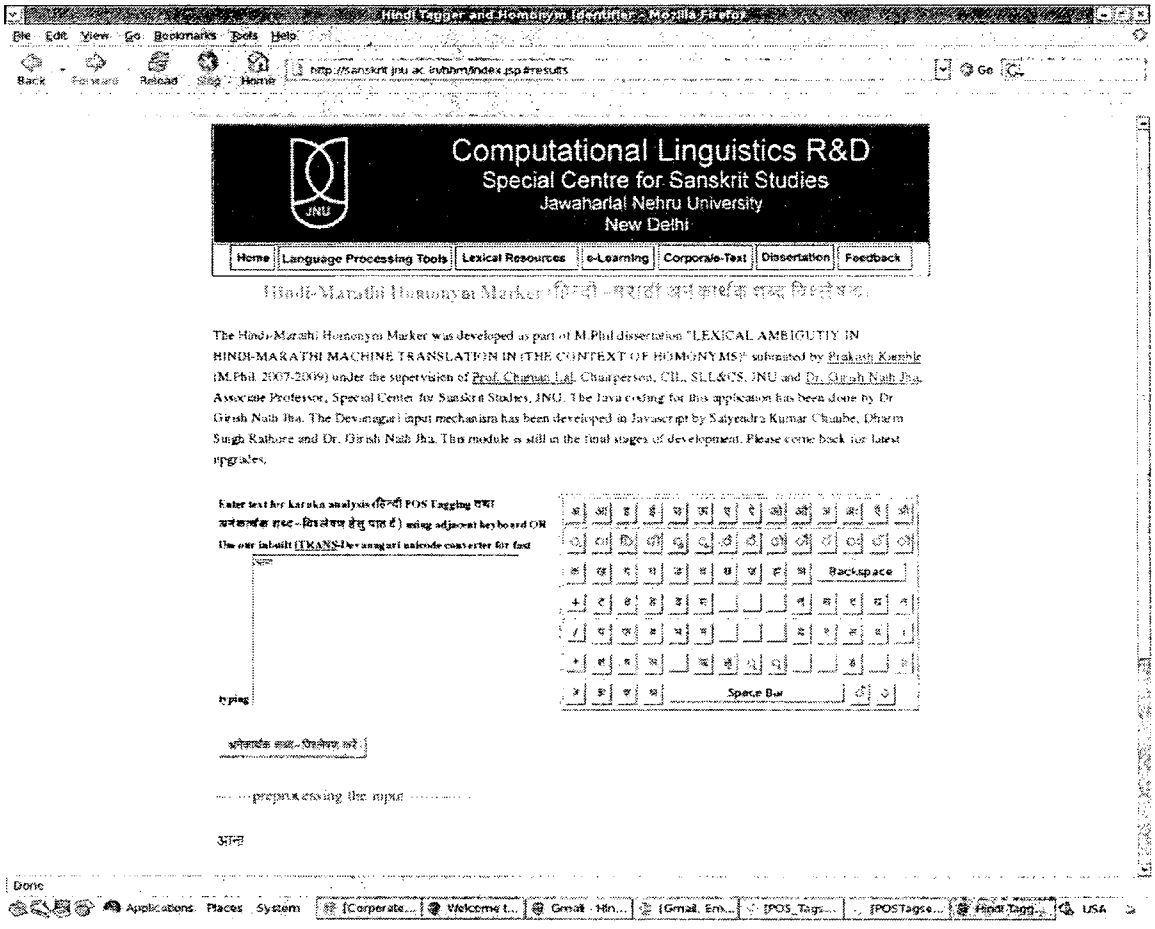
Java code snippet

```
public class HomonymMarker{
```

```
//this function loads the lexicon

        public HomonymMarker(){
                lr = new LexiconReader(LEXFILE);
        }

//this function tags the text as per IL-POSTS convention//(इस प्रक्रिया में मशीन POS टैगर का उपयोग कर व्याकरणिक कोटियों को प्रस्तुत करेगी।)

        public String tagText(String itext){

        }

//this function checks homonyms and translates them//( इस प्रक्रिया में मशीन अनेकार्थक शब्दों को प्रस्तुत करेगी)

        public String checkHomonymyAndTranslate(String itext){
        }
}
```

इस संपूर्ण प्रक्रिया के बाद मिलने वाला वाले वेबपेज एवं भाषिक संसाधन को हम निम्न रुप से देख सकते है। जिसका उपयोग कर अनेकार्थक शब्द का विश्लेषण किया जा सकते है। मशीनी अनुवाद को ध्यान में रखते हुए इस संसाधन में लक्ष्य भाषा के रुप में मराठी भाषा का उपयोग

किया गया है। इस निम्न को और विकसित करने का काम भी किया जा रहा है। जिसका उपयोग अधिक से अधिक जगह हो सके।



चित्र संख्या - 5 : हिंदी मराठी अनेकार्थ शब्द विश्लेषक[7]

Step I: Preprocessing[8]

1. Input की प्रथम प्रक्रिया में हिंदी में शब्द दिए जाएँ.
2. अगर Input वाक्य के रुप में दिया जा रहा हो तो प्रथम मशीन अनेकार्थक शब्दों में जाकर यह देखे की दिए गए वाक्य में कितने अनेकार्थक शब्द हैं।

[7] http://sanskrit.jnu.ac.in/hhm/index.jsp
[8] http://sanskrit.jnu.ac.in/hhm/index.jsp

3. जो शब्द अनेकार्थक हों उस शब्द का रंग बदल दिया जाएगा। और व्याकरणिक कोटियाँ भी दी जाएगी।
4. दिए गए अर्थ के अनुसार उसी शब्द का मराठी अनुवाद मशीन प्रस्तुत करेगी।
5. अगर input केवल शब्द स्तर का हो तो केवल सर्च प्रोग्राम के माध्यम से डेटाबेस में मिलने वाले अनेकार्थक शब्दों का अर्थ प्रस्तुत किया जाएगा जैसे शब्द कोशों में किया जाता है। लेकिन साथ में दिए गए हिंदी एवं मराठी दोनों उदाहरणों को भी साथ-साथ दिया जाएगा। जिससे शब्द के कारण निर्माण होने वाली अनेकार्थकता की समस्या शब्द स्तर पर स्पष्ट हो जाएगी।

Step II: Database

1. डेटा बेस का उपयोग मशीनी अनुवाद के रुप में किया जा रहा हो तब मशीन प्रमुखत: व्याकरण की कोटियों पर ध्यान देगी। डेटाबेस में ऐसे कई शब्द हैं जिनकी व्याकरणिक कोटियाँ समान हैं लेकिन इन शब्दों की sub-grammatical Category भिन्न होगी। ऐसे समय इन्हें sub-grammatical Category के माध्यम से दूर किया जा सकता है। (डेटाबेस में sub-grammatical Category नहीं दे पाया हूँ।)
2. दूसरे स्तर पर व्याकरण के नियमों का उपयोग होगा।

   जैसे – कल शब्द के दो अर्थ है एक मराठी - उद्या (tomorrow), काल(yesterday)

   'कल' शब्द का उपयोग किसी भी वाक्य में हो अंत में था, थी, थे या गा, गी, गे,/ हैं, है के साथ ही आता है।

   जब वाक्य के अंत में था, थी, थे का प्रयोग किया जाएगा उस समय मशीन केवल मराठी अनुवाद के लिए 'काल' शब्द का ही प्रयोग करेगी। जिस वाक्य के अंत में गा, गी, गे,/ हैं, है का उपयोग किया जाएगा उस समय मराठी अनुवाद के लिए – 'उद्या' का ही उपयोग करें।

Step III: Output level-1

समान उच्चारण वाले शब्दों को अन्य ऑनलाईन शब्दकोशों का उपयोग कर भी तुलना कर सकते।

Step V: Output - final level[9]

अंतिम output के रुप में user तीन रुपों को देख सकता है।

1. हिंदी में दिए गए वाक्यों में कितने शब्द अनेकार्थक हैं इसकी निष्पक्ष रुप से मशीनी अनुवाद के लिए जांच करेगा। जो user को यह बताएगी की वाक्य में किस शब्द में अनेकार्थकता है।
2. दिए गए अनेकार्थक हिंदी शब्द के लिए मराठी अनुवाद किस प्रकार हो सकता है।
3. मशीन के द्वारा किस प्रकार अनेकार्थक शब्दों का अनुवाद किया जा सकता है।
4. डेटाबेस के रुप में हिंदी में मिलने वाले संपूर्ण अनेकार्थक शब्दों की सूची।
5. अनेकार्थक शब्दों का इस संसाधन के द्वारा विश्लेषण और समाधान।

3.उपयोग

1.भारतीय भाषाओं के शब्दकोशीय संसाधनों के उपयोग के लिए।

अभिकलनात्मक शब्दकोश रचना का निर्माण अभिकलनात्मक भाषाविज्ञान का ही एक उपकार्यक्षेत्र है, इसलिए शब्दकोश संसाधनों के कार्यान्वयन से संबंधित कार्य के रुप में अभिकलनात्मक शब्दकोश रचना की व्याख्या एन.एल.पी 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 भूमीका निभाता है। पहले तो कोशकार को संपादन, शब्द प्रतिनिधित्व और शब्दकोशीय आधार सामग्री संसाधन की प्रक्रिया में सहायता करता है तो दूसरी ओर कोश के उपयोग कर्ता(user) को इलेक्ट्रानिक रुप में शब्द को खोजने में मदद करता है। इस संपूर्ण प्रक्रिया को पूर्ण रुप प्रदान करने के लिए एन.एल.पी. में वाचिक और लिखित दोनों रुपों में शब्दकोशीय संसाधनों की सूचना/आधार सामग्री की आवश्यकता होती है। अकादमिक एवं उद्योग जगत में अधिकतर कार्य एन.एल.पी. के संसाधनों पर ही निर्भर है, इसलिए शब्द के आधार सामग्री (lexical resources) की अधिकतर आवश्यकता इस क्षेत्र में है। यह निर्माण प्रक्रिया शब्दकोश संपादन और

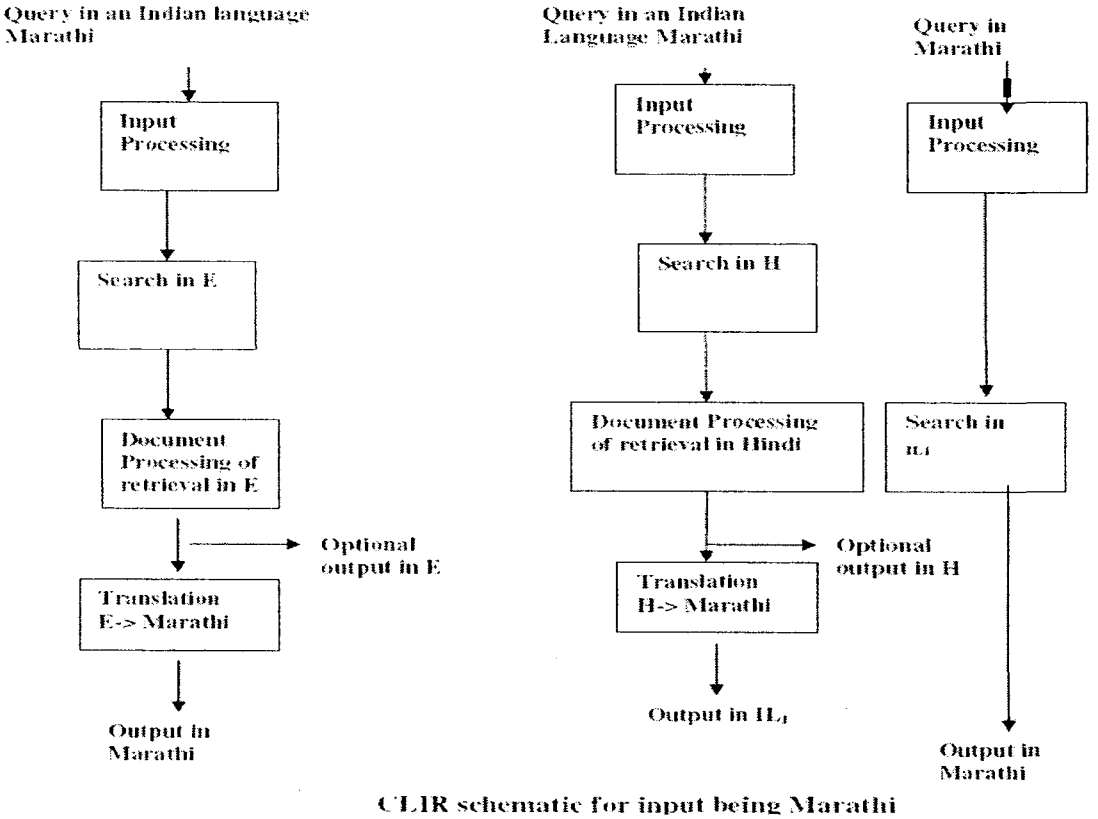
[9] http://sanskrit.jnu.ac.in/subanta/rsubanta.jsp

शब्दकोशों पर हो रहे अनुसंधानों के आधार सामग्री और अभिकलनात्मक भाषाविज्ञान के लिए वरदान साबित हो सकती है। अभिकलनात्मक भाषाविज्ञान की दोनों पद्धतियों में (भाषाई और सांख्यिकीय) भाषा प्रयोग के विशाल नमूने कॉर्पोरा और डेटाबेस के रूप में उपलब्ध हैं। इस कार्य में किसी पुराने शब्दकोश को परिष्कृत करना हो या उसमें विकास करना भी अभिकलनात्मक कार्य की सूक्ष्मता दिखाता है। यह अनुसंधान भी कार्पोरा में मिलने वाले अस्पष्ट शब्दों को स्पष्ट करने में महत्वपूर्ण भूमिका निभाएगा। इस कार्य से न केवल शब्दकोशों को बल मिलेगा वरन यह कार्पोरा की भी सहयता करेगा। इस अनुसंधान में हिंदी में मिलने वाले लगभग सभी अनेकार्थक शब्दों(Homonyms) को दिया गया है। इसी के साथ इन शब्दों का वाक्य में किस प्रकार उपयोग होगा यह भी दिया गया है। भाषाई अभिकलनात्मक संसाधनों में अभिकलनात्मक शब्द आधार सामग्री पर विचार किया जाए तो कई भाषिक संसाधनों के निर्माण में इस कार्य की भूमिका स्पष्ट हो सकती है जैसे – रुपविश्लेषक, चंकर, शब्दविश्लेषक, वर्तनी जाँचकर्ता आदि। अनेक भाषाई अभिकलनात्मक संसाधनों में शब्द आधार सामग्री महत्वपूर्ण भूमिका निभाती है। इस कार्य का संबंध भी इन संसाधनों से जोड़ा जा सकता है। जिस प्रकार भाषा में प्रत्येक शब्द का अन्यन्य महत्व होता है उसी प्रकार छोटे से छोटे प्रारुप में अनुसंधान के द्वारा तैयार किए गए आधार सामग्री का भी भाषाई संसाधनों में महत्व होता है। मशीनी अनुवाद के संबंध में छोटे-छोटे ऐसे कई भाषाई संसाधन है जो महत्वपूर्ण भूमिका निभा रहे हैं। यह संसाधन केवल मशीनी अनुवाद के कार्य को पूर्ण करने के लिए ही उपयोग में लाए गए बल्कि मशीनी अनुवाद में अनुवाद प्रारुप और सामान्य सिद्धांतों का पालन करने के लिए भी उपयोगी सिद्ध हुए हैं। जैसे - "अनुवाद प्रक्रिया में व्यवस्था लाने के लिए कई अनुवादक और विचारक सामने आए और उन्होंने अनुवाद के कुछ सिद्धांत भी विकसित किए। यह अनुवाद सिद्धांत मुलत: अनुवाद में आनेवाली कठिनाईयों व अर्थ संप्रेषण को केंद्र में रखकर विकसित लिए गए।"[10] इस नियम के आधार पर ही मशीनी अनुवाद के निर्माताओं ने मशीनी अनुवाद में अर्थ को महत्व देते हुए कार्य करना आरंभ किया। अनुवाद और मशीनी अनुवाद की प्रक्रिया में अनुवाद सिद्धांत यह संबंध न केवल एक प्रक्रिया को दर्शाता है बल्कि अनुवाद की दोनों विधाओं में लक्ष्य भाषा पाठ में अर्थ के महत्व को भी स्पष्ट करता है। अनुवाद प्रक्रिया में अर्थ स्पष्ट तभी हो सकता है जब मूल पाठ में अनेकार्थक शब्द न हो। अनेकार्थक शब्दों का मानव अनुवाद करते समय 90% अनुवाद कर सकता है परंतु मशीन के लिए यह काम अधिक कठीन होता है। इसलिए मशीनी अनुवाद यंत्रों से प्राप्त उदाहरणों से यह स्पष्ट हो जाता है कि मशीनी अनुवाद में अनेकार्थक शब्दों के सही अनुवाद के लिए विशेष कार्य की आवश्यकता है। यह कार्य प्रथमत: स्त्रोत पाठ में

---

[10] ओझा, डॉ.त्रिभुवन,1994, *हिंदी में अनेकार्थकता का अनुशीलन,* विश्वविद्यालय प्रकाशन, वाराणसी पृष्ठ संख्या-198-199

अनेकार्थक शब्दों को चिन्हित करेगी और उन शब्दों का निराकरण करने के लिए मशीन को कहेगी। किसी भी आधार सामग्री का उपयोग केवल शब्दकोश के लिए ही नहीं होता बल्कि उस आधार सामग्री के माध्यम से अन्य कार्यों को भी अंजाम दिया जाता है। विशिष्ट शब्द संसाधनों पर किए गए कार्य के आधार पर सी.एल.आई.आर.यंत्र निर्माण प्रक्रिया में इसी प्रकार का कार्य सी.एल.आई.आर.यंत्र में सम्मलित मराठी, हिंदी और अंग्रेजी के अर्थ निष्पादन से संबधित कार्य में हुआ है।[11] इस कार्य में मराठी स्रोत भाषा के रुप में है।



CLIR Schematic for input being Marathi[12]

1. सी.एल.आई.आर में मराठी, हिंदी एवं अंग्रेजी भाषाओं में आप शब्द की सूचना प्राप्त कर सकते है।
2. स्रोत भाषा हिंदी या अंग्रेजी न होते हुए भी स्रोत भाषा और लक्ष्य भाषाओं की उचित सूचना इस यंत्र के माध्यम से प्राप्त कर सकते हैं।

---

[11] Dr. Bhattacharyya, Pushpak. *White Paper on Cross Lingual Search and Machine Translation* Department of Computer Science and Engineering, IIT Bombay

[12] Dr. Bhattacharyya, Pushpak, White Paper on Cross Lingual Search and Machine Translation, Department of Computer Science and Engineering, IIT Bombay

3. सही उच्चारण की सूचना प्राप्त कर सकते हैं।
4. भाषावैज्ञानिक रुप से परिष्कृत न हुए शब्दों की सूचना।

इन उपयोगों के साथ ही मशीनी अनुवाद यंत्र एवं शब्दकोशों में भी सी.एल.आई.आर शब्द संकलन का उपयोग प्रमुख रुप से किया गया है। इसी प्रकार अनेकार्थक शब्दों पर किए गए कार्य का भी उपयोग इन सभी कार्यों के लिए किया जा सकता है। कार्य में केवल दो भाषाओं को ही रखा गया है लेकिन इस कार्य को अन्य समृद्ध शब्दकोशों के साथ जोड़ा जाएगा इसलिए दो से अधिक भाषाओं में भी सूचना प्राप्त की जा सकती हैं। किसी भाषा में जितने भी अनेकार्थक शब्द हों उनकी संख्या शब्दकोशों की तुलना में नाममात्र ही होती है। इन शब्दों की संख्या कम होने के कारण इन शब्दों के निराकारण के लिए संसाधन निर्माण करना हो तो संपूर्ण शब्दकोश का डेटाबेस तैयार करने की आवश्यकता नहीं होनी चाहिए। अब सामान्य रुप से किसी भी भाषा में शब्दों का छोटे डेटाबेस के आधार पर भाषा में उत्पन्न होने वाली समस्याओं को दूर करना ही आज प्रौद्योगिकी क्षेत्र का व्यवस्थित कार्य माना जा रहा है। जिस प्रकार सभी बोलचाल की भाषाओं में "नानार्थक (अनेकार्थक) शब्द होते हुए भी उस भाषा में लोकव्यवहार, साहित्य-लेखन, वैचारिक विवेचन और वैज्ञानिक विश्लेषण सुचारु रुप से चल रहा हों। इससे यह सिद्ध होता है कि अनेकार्थकता का निराकरण किसी न किसी रुप में हो सकता है।" यह अनुसंधान मात्र इस कार्य को आगे की ओर ले जा रहा है। यह अनुसंधान भाषा में अनेकार्थकता उत्पन्न करने वाले शब्दों को बतलाएगी। साथ ही इन शब्दों के निरकरण के ऐच्छिक शब्दों को भी बतलाएगी।

शब्द कोशों की बढ़ती मांग और अर्थ आधारित शब्दकोशों की बढ़ती आवश्यकता के लिए इस कार्य को अन्य शब्दकोशों के साथ भी जोड़ा गया हैं। जिनके प्रथम पृष्ठों को निम्न प्रकार से देख सकते हैं।

यह शब्दकोश विशिष्ट संस्कृत केंद्र, जे.एन.यू. में किया गया है जिसमें हिंदी लक्ष्य भाषा के रुप में है। साथ ही इस शब्द कोश को युनिकोड के साथ संल्लग्न किया गया है जिससे हिंदी में भी आप शब्दों का शोध ले सकते है। हिंदी में कई शब्द संस्कृत से लिए गए हैं इसलिए यह शब्दकोश बहुत महत्वपूर्ण है।

हिंदी शब्दतंत्र एवं मराठी शब्दबंध आई.आई.टी मुंबई द्वारा बनाए गए हैं। इंटरनेट पर उपलबद्ध भारतीय भाषाओं के शब्दकोशों में यह सबसे बेहतरीन कार्य करने वाले शब्दकोश माने जाते हैं। इस अनुसंधान में इन कोशों को जोड़ने का महत्वपूर्ण उद्देश्य यह है कि हिंदी इस अनुसंधान में स्रोत भाषा के रुप में है और मराठी लक्ष्य भाषा के रुप में अत: अभिकलनात्मक रुप में उपलब्ध शब्दकोशों की सहायता के लिए इन कोशों का होना अत्यंत जरुरी है। लेकिन इन शब्दकोशों में केवल एक ही भाषा के शब्द मिल पाते हैं।

इस अनुसधान में दोनों भाषाओं को एक साथ रखा गया है इसलिए यह कार्य शब्दतंत्र एवं शब्दबंध जैसे शब्दकोशों में दोनों भाषाओं के शब्दों को किस प्रकार एक साथ रखा जा सकता है इसके लिए उपयोगी हो सकता है। केवल कोश के लिए ही नहीं बल्कि हिंदी-मराठी मशीनी अनुवाद के आरंभ के लिए भी। यह कार्य अत्यंत छोटा होने के बावजूद शब्दकोशों के विकास में जरुर सहायता करेगा।

2. भारतीय भाषाओं से भारतीय भाषाओं के मशीनी अनुवाद के उपयोग के लिए।

शब्दकोश आधार सामग्री में किया गया कार्य अधिकतर मशीनी अनुवाद के उद्देश्य से ही किया जा रहा है। जो आधार सामग्री अर्थ की दृष्टी से महत्वपूर्ण है ऐसी आधार सामग्री को और ज्यादा महत्व दिया गया है। मशीनी अनुवाद के उपयोग में लाए जाने वाले प्रत्येक संसाधनों में आधार सामग्री की आवश्यकता होने के कारण एन.एल.पी में इन आधार सामग्री को महत्व प्राप्त हुआ है। भारत में भारत सरकार की भारतीय भाषाओं से भारतीय भाषाओं में मशीनी अनुवाद बनाने की जो महत्वपूर्ण परियोजना बनाई गई है इसमें कार्य कर रही कई संस्थायें शब्दकोशीय आधार सामग्री इकट्ठा करने की प्रक्रिया में कार्य कर रही है। इस परियोजना में 'हिंदी' अधिकर जगह स्रोत भाषा और लक्ष्य भाषा के रुप में महत्वपूर्ण भूमिका निभा रही है। इस स्थिति में हिंदी को स्रोत भाषा के रुप में रखकर किया गया यह कार्य उपयोगी सिद्ध हो सकता है। भा.भा. से भा.भा.में अबतक कोई अनुवाद यंत्र इंटरनेट पर उपलबद्ध नहीं हो सका है। आज के समय भारत में निर्माण किए एवं इंटरनेट पर ऑन लाईन कार्य कर रहें मशीनी अनुवाद यंत्र अंग्रेजी भाषा को स्रोत भाषा के रुप में रखकर काम कर रहे हैं। इन सभी अनुवाद यंत्रों में हिंदी लक्ष्य भाषा के रुप में ली गई है। जिनका निर्गत पाठ उच्च कोटि का नहीं है तथा जिससे मिलने वाले निर्गत पाठ में अनेकार्थक शब्दों की फिर से जांच करने की आवश्यकता है। इस कार्य में यह अनुसंधान मह त्वपूर्ण योगदान दे सकता है। इनकी की सूची एवं इन मशीनी अनुवाद यंत्र के प्रथम पृष्ठ को हम निम्न रुप से देख सकते है। साथ ही अनुवादक मशीनी अनुवाद यंत्र का भी पता दिया गया है।

1.मंत्रा अनुवाद यंत्र : -

☏ http://mantra-rajbhasha.cdac.in/mantrarajbhasha/index.html

2.शक्ति अनुवाद यंत्र

☏ http://shakti.iiit.net/~shakti/cgi-bin/mt/translate.cgi (हिंदी अनुवाद यंत्र)

☏ http://shakti.iiit.net/~shakti_tl/cgi-bin/mt/translate.cgi (तेलुगु अनुवाद)

☏ http://shakti.iiit.net/~shakti_tl/cgi-bin/mt/translate.cgi (मराठी अनुवाद)

3.अनुसारका अनुआद यंत्र : -

☏ http://sanskrit.uohyd.ernet.in/~anusaaraka/english/index.html

4.टी.डी.आई.एल द्वारा किया गया अनुवाद यंत्र

☏ http//translate.ildc.in:8080/jsp-examples/work/back3.jsp.

5.आय.आय.टी मुंबई के अनुवाद यंत्र एवं वर्ड नेट का पता

☏ http://www.cfilt.iitb.ac.in/machine-translation/eng-hindi-mt/

6. अनुवादक मशीनी अनुवाद यंत्र (केवल डेमो देखने का पता)

☏ http://www.mysmartschool.com/Images/Anuvadak/AnuvaadakDemo.swf

7.उर्दू-हिंदी मशीनी अनुवाद यंत्र

☏ sanskrit.uohyd.ernet.in/~anusaaraka/urdu/cgi-bin/urdu/u-n/new4.cgi

8.गुगल मशीनी अनुवाद यंत्र :- ☏ http://translate.google.com/translate_t#en|hi|speaker

9. मात्रा मशीनी अनुवाद यंत्र :- ☏ http://www.cdacmumbai.in/matra/

मशीनी अनुवाद की आज की परिस्थितियों पर विचार किया जाए तो यह स्पष्ट रुप से दिखाई देता है की मशीनी अनुवाद के माध्यम से मिलने वाले निर्गत पाठ की प्रक्रिया करने की आवश्यकता है। प्रमुख रुप से वाक्य संरचना एवं अर्थ स्पष्टता के क्षेत्र में अधिक कार्य करने की आवश्यकता है। इस विषय में अनुसंधान करने के आसार दिखाई देते हैं। इस दृष्टि से भी इस अनुसंधान का उपयोग हो सकता है। निर्गत पाठ में कितने अनेकार्थक शब्द हैं इसकी सूचना प्राप्त करने के उद्देश्य से यह अनुसंधान उपयोगी हो सकता है। आज मशीनी अनुवाद के क्षेत्र में अभिकलनात्मक भाषाविज्ञान के छोटे-छोटे संसाधनों के उपयोग की आवश्यकता को ध्यान में रखते हुए यह कार्य भी सहायक सिद्ध हो ऐसी आशा है।

अनुसंधान के सैद्धांतिक पक्ष में अनेकार्थक शब्दों की विस्तृत चर्चा की गई है जिसका उपयोग हिंदी में मिलने वाले अनेकार्थक शब्दों का व्यतिरेकी एवं भाषाशास्त्रीय दृष्टि से अध्ययन करने के लिए भी यह कार्य उपयोगी होगा। प्रत्येक कार्य की सफलता इसी में होती है कि किया गया कार्य अधिक से अधिक उपयोग में आ जाए। इस दृष्टि से यह अनुसंधान अभिकलनात्मक भाषाविज्ञान, शब्दकोश निर्माण, भाषाई संसाधनों का निर्माण इन क्षेत्रों के साथ जिन क्षेत्रों में भी उपयोग में आए यह इस अनुसंधान की सफलता ही मानी जाएगी।

उपसंहार

अनुवाद कार्य के शुरुआती दौर से लेकर मशीनी अनुवाद की उपस्थिति तक के इतिहास को देखा जाए तो आज मनुष्य ने मशीनी अनुवाद के क्षेत्र में जो अभूतपूर्व काम किया है उसकी तुलना "मिस्र के सम्राट टोल्मी द्वितीय पिला डेल्फस(285-246 ई.पू) के द्वारा लिए गए उस निर्णय से हो सकती है, जिसमें पुस्तकालय अध्यक्ष डेमेट्रियस फालेरयूज के साथ अरिस्टिअस भी उपस्थित थे। सम्राट की इच्छा थी कि वह अपने पुस्तकालय में संसार की सभी अच्छी पुस्तकों का संकलन करे।"[1] इसके उपरांत "पुस्तकायल अध्यक्ष ने सम्राट को जो ज्ञापन भेजा था, उसमें उसने बताया कि यहूदियों की धर्म पुस्तकें हिब्रू भाषा में है और उनका यूनानी भाषा में नये सिरे से अनुवाद होना चाहिए।"[2] एलियाजर ने सम्राट की इच्छानुसार 72 योग्य अनुवादकों की नियुक्ति की। इन अनुवादकों को एक शांत एवं सुहानी जगह पर रखा गया। अनुवादक प्रति दिन अनुवाद आरंभ करने से पूर्व नदी में हाथ धोकर ईश्वर से प्रार्थना करते थे। इस प्रकार 72 अनुवादकों ने 72 दिनों में हिब्रू भाषा से यूनानी भाषा में अनुवाद किया। अनुवाद सभी लोगों को पसंद आया, राजा ने भी इस अनुवाद की मुक्त कंठ से प्रशंसा की। अनुवाद क्षेत्र में सबसे महत्त्वपूर्ण एवं ऐतिहासिक माने जाने वाले इस अनुवाद का महत्त्व केवल इसलिए नहीं है कि यह प्रथम अनुवाद है बल्कि इस अनुवाद का महत्त्व इस लिए भी है कि यह अनुवाद ऐतिहासिक अनुवादों में सबसे अधिक निरपेक्ष रूप से तथा अनुभवी योग्य अनुवादकों के द्वारा किया गया था। जिसे अन्य धार्मिक अनुवाद कार्यों के लिए आदर्श के रुप में भी देखा जाता है। यह मात्र अनुवाद इतिहास की कहानी नहीं है, बल्कि अनुवाद क्षेत्र में डाली गई प्रथम सशक्त नीव भी है। मशीनी अनुवाद के क्षेत्र में आज भी यह स्पष्ट रुप से नहीं कहा जा सकता कि कौनसा मशीनी अनुवाद यंत्र सबसे सफल या व्यवस्थित है। लेकिन "सेंट जेरोम संभवत: पहला व्यवस्थित अनुवादक है। जिसने स्वयं अनुवाद करने के साथ-साथ अनुवाद कार्य की समस्याओं का भी निरुपण किया है। 364 ई. में डेमास ने सेंट जेरोम से न्यू टेस्टामेंट का लैटिन अनुवाद प्रस्तुत करने के लिए कहा था।"[3] अनुवाद के शुरुआती दौर से अनुवाद में कई पड़ाव दिखाई दिए जिसे कई विद्वानों ने नकारा और आलोचना भी की। अनुवाद न केवल दो समकालीन समाजों को ही जोड़ता है, बल्कि पश्चिम में रहने वाली संस्कृतियों को पूर्व में रहने वाली

---

[1] गुप्त, डॉ.गार्गी & गुप्त, विश्वप्रकाश, 1992, *अनुवाद के मूल स्रोत,* भारतीय अनुवाद परिषद, पृष्ठ संख्या-30

[2] वहीं - पृष्ठ संख्या - 31

[3] वहीं - पृष्ठ संख्या - 34

संस्कृतियों से परिचित भी कराता है। अनुवाद के दूसरे पक्ष को देखा जाए तो साहित्यिक कृतियों के अनुवादों में विशेष रुप से काव्यानुवादों को कुछ "पाश्चात्य विद्वानों ने 'उबली हुई स्ट्राबेरी के समान स्वाद हीन(Forest Smith) 'पाप' (Grant Showerman), विवेकहीन और असंगत विचार' (Victor Hugo), छपे हुए कपड़े की उलटी तरफ'(Cervants) आदि। कहकर उसे निम्न कोटि का बताया है। परंतु ये लोग भूल जाते हैं कि अनुवाद की एक महत्वपूर्ण भूमिका यह है कि "जिन रचनाकारों का नाम कई बार उनके देशवासी तक नहीं जानते उन्हें भी यह विश्व साहित्य के मंच पर स्मरणीय बना देता है।"[4] अनुवाद की इस महत्ता को आगे बढ़ाते हुए डॉ.गार्गी गुप्त ने अनुवाद के क्षेत्र को काल क्रम की दृष्टि से तीन कालखंडों में विभाजित किया है।[5] "प्रथम कालखंड जो ईसा से दो-तीन सदी पूर्व यूनानी भाषा में बाइबिल के अनुवाद से आरम्भ हुआ और अठारवीं सदी के अन्त तक चला। इसे हम धार्मिक साहित्य के अनुवाद का युग कह सकते हैं। द्वितीय कालखंड उन्नीसवीं सदी के आरम्भ से द्वितीय विश्वयुद्ध के अन्त तक का है जिसमें (साहित्यिक) विशेषकर काव्यानुवाद की समस्याओं पर विचार हुआ। इसे काव्यानुवाद का युग कहा जा सकता है। विश्वयुद्ध के बाद विद्वानों का ध्यान वैज्ञानिक साहित्य के अनुवाद की ओर गया। अनुवाद के इस तीसरे कालखंड में ही मशीनी अनुवाद (यंत्रानुवाद) की धारणा का विकास हुआ। इसे वैज्ञानिक साहित्य के अनुवाद का युग कह सकते हैं।" गार्गी गुप्त एवं विश्वप्रकाश गुप्त का यह काल विभाजन अनुवाद चिंतन के आधार पर हुआ है। आज अनुवाद विधाओं का भी कालक्रम के आधार पर विभाजन हुआ है। मशीनी अनुवाद विधा भी इससे अछुती नहीं रहीं, जिसका विभाजन चार भागों में किया जा चुका है।[6]

एक सरसरी निगाह से देखा जाए तो "आज से प्राय: 2000-2300 वर्ष पूर्व हिब्रू भाषा से यूनानी भाषा में 'बाइबिल' के अनुवाद की यह जो प्रक्रिया अपनाई गई थी उसका आधुनिक म.अनु.(यंत्रानुवाद) की प्रक्रिया से गहन संबंध है। आधुनिक शब्दावली में कहा जाए तो धार्मिक समाज का उच्चतम पदाधिकारी बाइबिल के अनुवाद के लिए एक आयोग की नियुक्ति करता था। इस आयोग के सदस्य बहुत योग्य व्यक्ति होते थे, जिन्हें शास्त्र-ज्ञान भी

---

[4] अनुवाद(पत्रिका) अक्तुबर-दिसंबर-१९९४ संपादकीय- डॉ.गार्गी गुप्त, पृष्ठ संख्या - 4

[5] वहीं - पृष्ठ संख्या- 4

[6] वही - पृष्ठ संख्या - 17

होता था और भाषा ज्ञान भी।"[7] मशीनी अनुवाद के निर्माण में इसी आधार पर भाषिक एवं सामाजिक क्षेत्र की कई संस्थाएँ, भाषा-विज्ञान के नामी विद्वान एवं प्रौद्योगिकी के क्षेत्र के महापंडित कार्य कर रहे हैं। जिस प्रकार हिब्रू से यूनानी भाषा में प्रथम अनुवाद के रुप में 'बाइबिल' को एक निरपेक्ष अनुवाद के रुप में देखा जाता है, उसी प्रकार मशीनी अनुवाद के क्षेत्र में काम कर रहे विशेषज्ञ मशीन के द्वारा मिलने वाला प्रत्येक पाठ सफल, सुंदर एवं निरपेक्ष अनुवाद के रुप में मिलने के लिए कार्य कर रहे हैं। जिसमें अनेकार्थक शब्दों के अनुवाद कार्यों पर भी काम हो रहा है। जैसे -'Root' शब्द वनस्पति के लिए 'जड़' अर्थ देगा तो व्याकरण की भाषा में 'धातु'। इसी के साथ प्रयोगधर्मी अर्थ भी अपने अलग-अलग वैकल्पिक समतुल्य लिए होते हैं। इसी प्रकार कोश के निर्माण में अनुवाद की उपयोगिता काफी महत्वपूर्ण है, जबकि अनुवाद में कोश कभी-कभी सफल हो जाता है। इसलिए कोशकार लक्ष्य भाषा के परंपरागत प्रयोगों को बनाए रखने तथा मूल अर्थ की संपुष्टि करने के साथ-साथ अनुवाद समतुल्यता को भी बनाए रक्खे। इसके अतिरिक्त कोशकार को लक्ष्य भाषा का काफी ज्ञान और अनुभव होना चाहिए, उसे मूल भाषा के सूचक से बराबर सहायता लेते रहना चाहिए ताकि अनुचित प्रयोग से बचा जा सके और उसमें अनुवादकीय गंध न आ सके।"[8] इस प्रक्रिया में संगणक सही अनुवाद के लिए अनिवार्य उपर्युक्त सभी पहलुओं से निरपेक्ष होता है, जबकि "अनुवाद में इनके अतिरिक्त कलात्मकता, सृजनात्मकता, शिल्पवत्ता और सहजता भी लानी होती है।"[9]

उन्नीसवीं शताब्दी में अनुवाद के नए प्रतिमान निर्धारित किए गए और अनुवाद में परिशुद्धता पर जोर दिया गया। जिससे मूल पाठ को अधिक महत्व मिला। मूल पाठ को मिले महत्व के कारण यह भी देखा जाने लगा की अनुवाद के लक्ष्य पाठ में आने वाली समस्या केवल अनुवादक की गलतियों के कारण ही होती है या मूल पाठ में आने वाले अति कठिन शब्दों के कारण भी होती है। इसलिए अनुवाद-पूर्व मूल पाठ का परिक्षण करने का प्रकार अनुवाद में प्रचलित हुआ। इस प्रक्रिया में अनुवाद को प्रथम पढ़ा जाने लगा जिसे 'पठन' प्रक्रिया का रुप मिला। मशीनी अनुवाद में भी यह प्रक्रिया मानव साधित मशीनी अनुवाद यंत्र (Human Aided Machine Translation System) के नाम से प्रचलित हुई। इस

---

[7] गुप्त, डॉ.गार्गी & गुप्त, विश्वप्रकाश, 1992, *अनुवाद के मूल स्रोत,* भारतीय अनुवाद परिषद, पृष्ठ संख्या-31

[8] डॉ.कृष्ण कुमार गोस्वामी - राजकमल प्रकाशन नई दिल्ली - पृष्ठ संख्या - 399

[9] चंद्र, रमेश, *राजभाषा हिंदी और तकनीकी अनुवाद*- - कल्याणी शिक्षा परिषद दरयागंज नई दिल्ली - 2 - पृष्ठ संख्या -134

प्रक्रिया में उन शब्दों को देखा जाने लगा जो अनुवाद के लिए कठिन हों। जिसमें समान उच्चारण वाले शब्दों की समस्या एक प्रमुख समस्या के रुप में सामने आयी। मशीनी अनुवाद में यह क्रिया शब्दों को बदल कर दी जाती है। जैसे:- बाल-बच्चा, बाल-केश जिस जगह पर शब्दबोध 'बच्चे' के रुप में होता हो उस जगह 'बाल' शब्द की जगह 'बच्चा' शब्द देकर मशीन को पाठ दिया जाए और जिस जगह 'बाल' शब्द के लिए 'केश' शब्दबोध होता हो उस स्थान पर 'बाल' शब्द की जगह 'केश' शब्द देकर मशीन को पाठ दिया जाए। लेकिन इस प्रकिया में मानव का हस्तक्षेप हो जाएगा। जिसके कारण मशीन पुन: निरपेक्षता की कसौटी से दूर हो जाएगी। इसलिए मशीन ही स्वयं शब्दों के बोध का चयन करे, ऐसी व्यवस्था की जाए। इस संकल्पना के आधार पर ही इस अनुसंधान ने आकार लिया है। जिसके परीक्षण कुछ सीमा तक हिंदी-मराठी मशीनी अनुवाद में सफल होंगे। इस आधार पर यह कार्य अन्य मशीनी अनुवाद यंत्रों को भी सहायता कर सकता है।

भाषा संप्रेषण में शब्द केवल भाव के वाहक होते हैं। वे किसी वस्तुविशेष या भावविशेष को स्पष्ट करने के लिए संकेत मात्र हैं। अनुवाद करते समय उस शब्द के अस्तित्व से अधिक उस भाव का ध्यान रखना चाहिए जो उसमें निहित है। इसी उद्देश्य से हिंदी अनुवाद में अनेकार्थक शब्द संबंधी सामान्य विचारों को पूरा करना, संबंधित जानकारी का सूचीबद्ध रुप से डेटाबेस तैयार करना। इसलिए हिंदी में उपलब्ध अनेकार्थक शब्दों का मराठी अनुवाद में विवेचन, विश्लेषण, वर्गीकरण करना और हिंदी में अनेकार्थक शब्दों के प्रकारों को स्पष्ट करने पर जोर दिया गया है। हिंदी में मिलने वाले अनेकार्थक शब्दों के प्रभावों से अनुवाद के समय आने वाली समस्याओं को संगणक के माध्यम से स्पष्ट करने का प्रयास भी किया गया है। इस कार्य से हिंदी में अनेकार्थक शब्दों की कई बातें प्रकाश में आ जाती हैं। इस सार्थकता के साथ ही हम इस विषय में सीमाओं और संभावनाओं से शब्द संचयन का पुन: उपयोग कर ज्ञान प्राप्त करने की उम्मीद कर सकते हैं। भिन्नार्थकता के अन्वेषण में हम देख सकते हैं कि अनेक विदेशी शब्द हिंदी शब्दों के समरुप हो जाते हैं। इन शब्दों का हिंदीकरण किया जाता है, विशेष रुप से ध्वनि के स्तर पर, लेकिन अर्थ के स्तर पर बहुत ही कम भिन्नता दिखाई देती है।

अनेकार्थकता भाषा-समृद्धि को अधोरेखित करती है। विश्व की सभी समृद्ध भाषाओं में यह विशेषता पायी जाती है। जो भाषा समाज, जाति, राष्ट्र, सभ्यता और संस्कृति जितनी

अधिक उन्नत होती है उस भाषा में उतनी ही अधिक अनेकार्थक अभिव्यक्तियाँ मिलती हैं। हिंदी विश्व की समृद्ध भाषाओं में से एक है। वैश्विक स्तर पर मिली लोकप्रियता के चलते अंग्रेजी एवं यूरोपीय भाषाओं में हिंदी के शब्दों को मिली स्वीकृति के कारण हिंदी के भाषा-सामर्थ्य को और बल मिला है।

किसी भी मशीन की कृत्रिम बुद्धि मनुष्य के कार्य का बौद्धिक अनुकरण ही होती है एवं पूर्ण-संचित नियमों से परिचालित भी होती है। मस्तिष्क में सूक्ष्म और अव्यक्त रुप में विद्यमान इन नियमों और प्रतिबंधों को कलनविधि कहते हैं। इस कार्यक्रम को सूत्रों के माध्यम से शब्द संचयन एवं भाषिक सूत्र को जोड़ कर संगणक की मेमोरी में डाल देना ही संगणक की कृत्रिम बुद्धि कहलाती है। इसीलिए अनुवाद व्याकरण एवं संदर्भ से जुड़ी निरसीम भाषिक अभिव्यक्तियों को नियमबद्ध कर पाना अत्यंत कठिन कार्य है। इस प्रक्रिया के कठिन होने के बावजूद, इस अनुसंधान में इस कार्य पर कलनविधि का प्रयोग कर विचार किया गया है। जिसके अच्छे परिणाम भी हमें दिखाई दे रहे हैं।

भारत एक विकासशील देश के रुप में जिस तेजी के साथ बढ़ रहा है। उसमें प्रौद्योगिकी का सर्वाधिक महत्वपूर्ण योगदान है, खास कर उन क्षेत्रों में जहाँ एकाधिक भाषाओं के कारण समस्यायें उत्पन्न होती हैं। जिसका निराकरण करने के लिए "भारत सरकार के मानव संसाधन विकास मंत्रालय ने टी.डी.आई.एल.जैसा उपक्रम किया है। जो भाषाओं की समस्याओं पर प्रौद्योगिकी के माध्यम से निराकरण पर कार्य कर रही है। जिसके अंतर्गत भारतीय भाषाओं में शब्दकोश, मशीनी अनुवाद, भाषिक संसाधन, फॉन्ट, यूनीकोड परिवर्तक, वर्तनी जाँच कर्ता, आदि कुल १६ भाषिक संसाधनों पर काम किया जा रहा है, जिसकी प्रमुख भाषा हिंदी ही है।"[10] इस कार्यक्रम के अंतर्गत अधिकतर भाषिक संसाधनों पर जोर दिया गया है। लेकिन किसी भी संस्था में अनेकार्थक शब्दों(Homonyms) पर विषेश रुप से कार्य नहीं हुआ है। पाश्चात्य देशों में इन शब्दों का महत्व समझ कर कई पुस्तकें लिखी गई। आधुनिक युग में संगणक एवं इंटरनेट की महत्ता को समझकर कई वेब साईट का निर्माण किया गया है। जिनका उपयोग कर अनेकार्थक शब्दों का निराकरण किया जाता है। लेकिन भारतीय भाषाओं में इस प्रकार की कोई वेब साईट उपलब्ध नहीं है।

---

[10] हिंदी सॉफ़्टवेयर उपकरण सी.डी उपकरण सूची, संचार एवं प्रौद्योगिकी मंत्रालय, भारत सरकार द्वारा प्रमोचित

जब की कई भारतीय प्रशासनिक सेवाओं की परीक्षाओं में अनेकार्थक शब्दों के संबंध में प्रश्न पूछे जाते हैं। इसलिए यह कार्य केवल भाषा की दृष्टि से ही महत्वपूर्ण न होकर परीक्षार्थीयों एवं इन शब्दों के कारण बच्चों के सामने आने वाली समस्याओं के निराकरण के लिए भी लाभदायक है। यह कार्य प्राथमिक अवस्था में ही इन समस्याओं का समाधान कर देगा। इस लिए यह कार्य हिंदी-मराठी मशीनी अनुवाद में मिलने वाले अनेकार्थक शब्दों की समस्याओं पर केंद्रित है।

अन्य भाषाओं से आये शब्दों के कारण इन शब्दों ने हिंदी में भाषिक रुप से अनेकार्थकता उत्पन्न कर काफी उधम मचाया है। खास कर ऐसे समय जब बहुभाषिक भारत में बहुभाषिकता की समस्या का निराकरण अनुवाद कार्य के माध्यम से चल रहा हो, अंग्रेजी से हिंदी में आए शब्द अधिक समस्या उत्पन्न कर रहे हैं। इन शब्दों को जब हिंदी में स्वीकृत किए जाने के बाद जो अनेकार्थकता उत्पन्न होती है, उसपर भी इस अनुसंधान में विचार किया गया है। जिससे कुछ सीमा तक हिंदी-मराठी अनुवाद की समस्याओं का समाधान हो सके। इन शब्दों के कारण समस्या क्यों उत्पन्न होती है, इसपर विचार किया जाए तो मात्र अनेकार्थकता को छोडकर किसी भी भाषा में अनेकार्थ शब्द अन्य शब्दों से भिन्न नहीं होते। इन शब्दों की भी व्याकरणिक कोटियाँ होती है। भाषा में कई जगह सामान्य रुप से इन शब्दों का उपयोग होता है, तो कहीं ये शब्द विवाद का कारण भी बन जाते है। ये शब्द भी अन्य शब्दों की तरह वाक्य में एक दूसरे से संबंध रखते हैं। शब्द संबंध के आधार पर ही कई बार सामान्य शब्द भी अनेकार्थक बन जाते हैं। यह गुण अनेकार्थक शब्दों में भी दिखाई देता है। इन गुणों के उपरांत अनेकार्थक शब्द अपनी विशेषताओं के कारण सामान्य शब्दों से भिन्न होते हैं। जिसका विवरण दूसरे और तीसरे पाठ में किया गया है। अनेकार्थक शब्दों का अधिक संबंध ध्वनि विशेष से होता है लेकिन इस अनुसंधान का उद्देश्य अर्थ स्पष्टता से संबंधित होने के कारण अर्थ को ध्यान में रखकर ही अनुसंधान को आगे बढ़ाया गया है।

यह कार्य अभी पूर्ण रुप से अनेकार्थकता का निराकरण करने का दावा नहीं करता। यह अनुसंधान हिंदी से मराठी मशीनी अनुवाद में अनेकार्थक शब्दों के विषय में पहली बार हो रहा है, इसीलिए इस कार्य को प्राथमिक स्तर से शुरु करना पड़ा जिसके कारण अनेकार्थता को मानव की सहायता लिए बिना मशीन के द्वारा पूर्ण रुप से निराकरण करना संभव नहीं

हो सका है। लेकिन अनेकार्थक शब्दों की समस्याओं के तह तक जाने का प्रयास किया गया है। इसके उपरान्त इस कार्य को आगे बढ़ाने और संपूर्ण रुप प्रदान करने के लिए आगे और अनुसंधान की गुंजाइश है। आज विश्व में जितने भी मशीनी अनुवाद यंत्र विकसित हुए हैं इनमें से अधिकतर मशीनी अनुवाद यंत्र, शब्दों के निकटतम, सरल और समतुल्य रुप देकर शब्दों को एक भाषा से दूसरी भाषा में शब्दों को अंतरित करने का काम करते हैं इसके बावजूद अभी बहुत-सी सीमाएँ और संसाधनों की कमी हैं। "आज की स्थिति में भी हम संगणक से बहुत अच्छे अनुवाद की आकांक्षाएँ नहीं कर सकते। इसलिए जब तक इस दिशा में और अधिक विकास न हो जाए मानव अनुवाद पर ही निर्भर रहना समीचीन होगा।"[11]

[11] रमेश चंद्र, *राजभाषा हिंदी और तकनीकी अनुवाद*- - कल्याणी शिक्षा परिषद दरयागंज नई दिल्ली - 2 - पृष्ठ संख्या 140

ग्रंथानुक्रमणिका

आधार ग्रंथ

संदर्भ ग्रंथ, लेख, कोश एवं पत्रिका

शब्दकोश

इंटरनेट वेब संदर्भ

1. आधार ग्रंथ

1.हिंदी आधार ग्रंथ

1. ओझा, डॉ.त्रिभुवन,1994, *हिंदी में अनेकार्थकता का अनुशीलन,* विश्वविद्यालय प्रकाशन, वाराणसी

2.अंग्रेजी आधार ग्रंथ

1. Katya, Pertsova, **2007**, *Learning Form-Meaning Mappings in Presence of Homonymy: a linguistically motivated model of learning inflection* (A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in) Linguistics, University of California, Los Angeles

2 . संदर्भ ग्रंथ, लेख, कोश एवं पत्रिका

1.अंग्रेजी

1. Akshr Bharati - Dipti Misra-sharma, Amba Kulkarani, Vinit Chaitanya, Anusaraka:an approach to machine Translation

2. Ananthanarayana, H.S. , Jan. 9-11, 1980, *Treatment of homonymy in Páṇini Aṣṭádhyáyi*,

Papers presented at the Second International Conference on South Asian Languages and Linguistics, Dept. of Linguistics, Osmania University, Hyderabad, India page No.- 50

3. B Kumara Shanmugam, 2002, *Machine Translation as related to Tamil,* AU-KBC Research Centre, M.I.T., Anna University,Chennai, India.

4. Badodekar, Salil, *Translation Resources, Services and Tools for Indian Languages,* Computer Science and Engineering Department, IIT, Mumbai, 400019, India.

5. Bhajantri.S.Y., (M.Phil Dissertation Submitted to JNU - 1996), *A Comparative Analysis of Post-Position in Kannada and Hindi for Machine Translation,* Page - 70

6. Bri.Htte Bl,Ser, Ulrikf, Schwall, 1992, *A Reusable Lexical Database Tool For, Machine Translation,*Institute for Knowledge Based Systems IBM Germany,

7. Cabr, M.Teresa, 1999 *Terminology Theory, Methods and Applications,* John Benjamins Publishing Co, page No.-111

8. Dash , Niladri Sekhar, 2004, *Corpus-based Study of Lexical Polysemy in Bangla for Application in LT,* Computer Vision and Pattern Recognition Unit,ISI, Kolkata

9. Doherty, Martin.j., (Revised 26 May 2004), *Children's difficulty in learning homonyms**, Department of Psychology,University of Stirling

10. Dr. Bhattacharyya, Pushpak, *White Paper on Cross Lingual Search and Machine Translation* Department of Computer Science and Engineering, IIT Bombay

12. Gey, Fredric C., 2002, *Prospects for Machine Translation of the Tamil Language,* University of California, Berkeley

13. Gupta, Deepa, 2005, *Contributions to English to Hindi Machine Translation Using Example-Based Approach, (Ph.D. Thesis)*Department Of Mathematics, IIT, Delhi, India

14. Harold Somers, summer School 8-9 December 2003, *Machine Translation-1- , -2, CCL, UMIST,* Manchester M60, IQD England, Australian Language Technology, Australia

15.Hutchins, W.John, 1988. *Recent developments in Machine Translation a review of the last five years,*University of East Anglia,Norwich

16. Hutchins, John, 13 December 2007. *Machine translation: problems and issues,* panel at conference,

17. Joachim De Beule1, Bart De Vylder1 and Tony, Belpaeme2. *A cross-situational learning algorithm for damping homonymy in the guessing game,* 1Vrije Universiteit Brussel, Belgium, 2University of Plymouth, United Kingdom

18. Kalra,Sachin, June 2004, *Punjabi Text Generation using Interlingua approach in Machine Translation* (A thesis Submitted fulfilment of Master of Engineering) page No.-10

19. Karoly F&bricz, 1986, *Particle Homonymy and Machine Translation,* JATE University of Szeged, Egyetem u. 2.,Hungary I – 6722

20. Marrafa, Palmira,& Mendes, Sara, Vol.19 No1. Jan-Jun 2007, *Using WordNet.PT for Translation: Disambiguation and Lexical Selection Decisions,* International Journal of Translation

21. Mayorcas, Pamela, 1990,*Translating and Computer10,* British Library Catalosuing Publication

22. Michael Thelen and Ellen Rilo, 2002, *A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts,*School of Computing,University of Utah

23. Nikunen, Anja, 11 April 2007, *Different approaches to word sense disambiguation Language technology and applications Essay,* University of Helsinki, Department of Computer Science

24. Paper presented at NATO summer school , July 1963, *Automatic Documentation*

25. S.Raman and Reday , N.Rama Krishna, Vol.No.7.jan-dec.1995, *Parallel MTS for Indian Languages,* Madras, IJT

26. Slocum, Jonathan, 1988, *Machine Translation System*Syndicate of the University of Cambridge

27. Udupa, Raghavendra & Faruquie, Tanveer A , *An Algorithmic Framework for the Decoding Problem in Statistical Machine Translation,*IBM IRL,Block-1A, IIT, Hauz Khas, N.Delhi, Hemanta K Maji, Dept. of Computer Science and Engineering, IIT Kanpur

28. Viegas, Evelyne, Volume 10, Nos 1-2. 17.1995, *Multilingual Computational Semantic Lexicons in Action: The WYSINNWYG Approach to NLP,* New Mexico State University Computing Research Laboratory, USA

29. Vilar et al.& Rimkut`e and Kovalevskat`e, 2007,*Machine translation evaluation, Error analysis, Frameworks – example 1, (English to Lithuanian)*

30. Whithelock, Peter & Kilby, Kieran, 1995 (Second Edition) *Linguistic and Computational Techniques in Machine* Translation System Design, UCL Press Limited

31. विश्वभारत@TDIL (त्रैमासिक पत्रिका), *paradigm Shift Of Language Technology initiatives Under Programme,* सूचना प्रौद्योगिकी मंत्रालय, भारत सरकार, पृष्ठ संख्या - 4

4. शब्दकोश: -

1. Encyclopedia Britannica – vol.No.5.H - page No – 107

2. Langenslheiots, Russian Dicationary - By – E.Wedel

3. The world book encyclopedia volume 9 (H) - page 275-276

4. संपादक- जोशी, श्रीपाद,1994, *श्रीपाद अभिनव शब्दकोश हिंदी-मराठी,* मराठी-हिंदी,शुभदा सारस्वत प्रकाशन,पुणे

5. संपादक - नेने,गो.प.,1994,*बृहत हिंदी-मराठी शब्दकोश,*श्रीपाद जोशी शुभदा प्रकाशन, पुणे

6. संपादक- जोशी, श्रीपाद, कृ.पां.कुलकर्णी, 1993, *मराठी उत्पति कोश,*शुभदा सारस्वत प्रकाशन, पुणे

7. संपादक पं.ग.र.वैशंपायन, 1966,*मराठी-हिंदी शब्द संग्रह,* लोकसंग्रह प्रकाशन, सदाशिव पेठ, पुणे

8. संपादक पं. वैशंपायन, गणेश रघुनाथ, 1967,*हिंदी-मराठी व्यवहार कोश,* साहित्यरत्न (प्रयाग, येवले) अनाथ विद्यार्थी गृह प्रकाशन, पुणे-2

इंटरनेट पर उपलब्द कोश

1. *http://www.cfilt.iitb.ac.in/wordnet/webmwn/wn.php* (ऑन लाईन मराठी वर्डनेट)

2. *http://www.cfilt.iitb.ac.in/wordnet/webhwn/wn.php* (ऑन लाईन हिंदी वर्डनेट)

3. *http://www.cfilt.iitb.ac.in/wordnet/webhwn/wn.php* (ऑन लाईन हिंदी-अंग्रेजी वर्डनेट)

4. *http://www.shabdkosh.com/* (ऑन लाईन हिंदी-अंग्रेजी शब्दकोश)

5. *http://sanskrit.jnu.ac.in/amara/viewdata.jsp* (ऑन लाईन बहुभाषिय शब्द कोश)

2. हिंदी

1. ओम, डॉ.विकास, (संपादक डॉ.नगेंद) 1995, *अनुवाद विज्ञान सिद्धांत और अनुप्रयोग मशीनी अनुवाद की समस्याएँ,* हिंदी कार्यान्वय निदेशालय,दिल्ली विश्वविद्यालय,पृष्ठ सं.३४२

2.कमलेश्वर,2004, *कितने पाकिस्तान,* राजपाल प्रकाशन,नई दिल्ली पृष्ठ संख्या - 12

3.गवेषणा, अक्टूबर-सितंबर - 2007,*भाषा और सूचना पर केंद्रीत विशेषांक, अंक-88* केंद्रीय हिंदी संस्थान, आगरा से प्रकाशित, अनुप्रयुक्त भाषाविज्ञान, भाषाशिक्षण तथा चिंतन की त्रैमासिक पत्रिका

4.गुप्त, डॉ.गार्गी & गुप्त, विश्वप्रकाश, 1992, *अनुवाद के मूल स्रोत,* भारतीय अनुवाद परिषद, पृष्ठ संख्या-31

5.जोगलेकर, न.चि., प्रथम संस्करण-1951,*मराठी का वर्णनात्मक व्याकरण,* प्रकाशक-राष्ट्रभाषा प्रचार समिति, वर्धा

6.डॉ. जैन,दीन दयाल,*हिंदी शब्द रचना,* पृष्ठ संख्या - २८

7. डॉ. नज़ीर, आरिफ़, 2005, *हिंदी में अनुवद की भूमिका और द्विभाषी कम्प्यूटरीकरण,* अनंग प्रकाशन, दिल्ली

8. भाटिया, डॉ. कैलाशचंद्र, डॉ.युगेश्वर,डॉ. कपूर, बदरीनाथ, संपादक-डॉ.भोलानाथ तिवारी, डॉ.हरदेव बाहरी,1989, *कोश विज्ञान सिद्धांत और प्रयोग,* (आचार्य रामचन्द्र वर्मा जन्मशती ग्रंथ)

9.डॉ. भाटिया, कैलाशचन्द्र, 1967, *हिंदी में अंग्रेजी के आगत शब्दों का भाषा-तात्विक अध्ययन,* हिंदुस्तानी एकेडेमी, इलाहाबाद

10.डॉ. राणा, महेंद्र सिंह, *हिंदी व्याकरण-प्रकाश-लेखन,* हर्षा प्रकाशन, आगरा
पृष्ठ सं.-118-216

11.डॉ. शर्मा, प्रमोद, अनुवाद पत्रिका, अप्रैल-जून 2004, *मशीनी अनुवाद की समस्याएँ और संभावनाएं,* अनुवाद पत्रिका(कंप्यूटर विषेशांक-२)

12.प्रेमचंद, *प्रेमचंद की चुनिंदा कहानियाँ,* (नमक का दरोगा, पूस की रात, शतरंज के खिलाड़ी)(सभी कहानियाँ))

13.प्रो. दास, ठाकुर,अप्रैल-जून 2004, *मशीनी अनुवाद:भारतीय परिदृश्य, अनुवाद पत्रिका* (कंप्यूटर विषेशांक-२)

14.प्रो.नारंग, वैष्णा,2006,*समसामायिक भाषा विज्ञान,* यश पब्लिकेशन,दिल्ली
पेज नं. 87

15.प्रो. सिंह, सुरजभान, अप्रैल-जून 2004, *कम्पूटर अनुवाद, सूचना प्रौद्योगिकी, हिंदी और अनुवाद, अनुवाद पत्रिका*(कंप्यूटर विषेशांक-२)

16.सिन्हा, राजेंद्र प्रसाद, *शुध्द हिंदी कैसे सीखें,* भरती भवन प्रकाशन, पटना
पृष्ठ सं.- 431-432

17.मल्होत्रा, विजय कुमार 2007, *कंप्यूटर के भाषिक अनुप्रयोग,* वाणी प्रकाशन, 21-ए दरियागंज, नयी दिल्ली, पृष्ठ संख्या-73

18.वाजपेयी, किशोरीदास, 1985, *हिंदी शब्दानुशासन,* नागरी प्रचारिणी सभा, काशी

19.*हिंदी सॉफ़्टवेयर उपकरण सी.डी उपकरण सूची*, 2006, संचार एवं प्रौद्योगिकी मंत्रालय, भारत सरकार द्वारा प्रमोचित

20. चंद्र, रमेश, *राजभाषा हिंदी और तकनीकी अनुवाद*- - कल्याणी शिक्षा परिषद दरयागंज नई दिल्ली - 2 - पृष्ठ संख्या -134

3. मराठी

1. मूल लेखक-ब्लोक, झुल, अनुवादक-प्रा. परांजपे, वासुदेव गोपाल, *मराठी भाषेचा विकास,* प्रतिभा प्रकाशन शनिवार पेठ, पुणे - २

2. संपादक - डॉ. काले, कल्याण & डॉ. सोमन, अंजली, *आधुनिक भाषा विज्ञान* (मराठी) (संरचनावादी आणि सामान्य)

3.मूल लेखक- प्रेमचंद, अनुवादक – साविरकर, दिनकर, *निवडक प्रेमचंद,* (मिठाचा फौजदार, बुद्धिबळाचे खेळाडू, पौषातील रात्र सभी कहानियाँ)

4. मूल लेखक-कमलेश्वर, मराठी अनुवादक - पद्माकर जोशी, *किती पाकिस्तानं,* राजपाल एंड सन्ज प्रकाशन, दिल्ली, पृष्ठ संख्या - 16

5. कै.वालंबे.मो.रा, 2005, *सुगम मराठी व्याकरण लेखन,* नितिन प्रकाशन, पुणे-३०

3. इंटरनेट वेब संदर्भ

1. http://archiv.tu-chemnitz.de/pub/2001/0043/data/presentation-html/img10.htm
2. http://www.wordreference.com/definition/homonym
3. http://www.all-about-spelling.com/homophones-worksheets.html
4. http://www.all-about-spelling.com/list-of-homophones.html
5. http://www.audienceoftwo.com/mag.php?art_id=598
6. http://www.iloveindia.com/babynames/boy-k.html
7. http://www.magickeys.com/books/riddles/words.html
8. http://www.cooper.com/alan/homonym_list.html
9. http://www.kwiznet.com/p/takeQuiz.php?ChapterID=1905&CurriculumID=14
10. http://www.wordinfo.info/words/index/info/view_unit/3039

11. http://en.wikipedia.org/wiki/Regional_differences_and_dialects_in_Indian_English

12. http://www.diplomacy.edu/language/Translation/machine.htm

13. http://www.todaytranslations.com/index.asp-Q-Page-E-The-Problem-with-Machine-Translation--62999690

14. http://www.ling.helsinki.fi/kit/2001s/ctl310gen/GW-MScThesis/node7.html

15. www.spanicity.com

16. http://en.wikipedia.org/wiki/Homonym

17. http://shakti.iiit.net/~shakti/cgi-bin/mt/translate.cgi

18. http://mantra-rajbhasha.cdac.in/mantrarajbhasha/index.html

19. http://www.cdacmumbai.in/matra/index.jsp

20. http://translate.google.com/translate_t#en|hi|speaker

21. http://www.tribuneindia.com/2000/20000819/windows/roots.htm

22. http://sanskrit.jnu.ac.in/index.jsp