



विश्वविद्यालय अनुदान आयोग
University Grants Commission
Quality higher education for all

मशीनी अनुवाद: परिचय

डॉ. अभिज्ञान द्विवेदी

मशीनी अनुवादःपरिचय

डॉ.अभिज्ञान द्विवेदी

विषय सूची

1. भूमिका

- 1.1 मशीनी अनुवाद का प्रारंभ
- 1.2 प्राथमिक मशीनी अनुवाद सिस्टम
- 1.3 भारत में मशीनी अनुवाद का प्रारंभ
- 1.4 भाषा अनुवाद के प्रमुख पहलू
- 1.5 मशीनी अनुवाद (MT) की सीमाएँ

2. मशीनी अनुवाद के प्रकार

- 2.1 नियम-आधारित मशीन अनुवाद (RBMT)
- 2.2 सांख्यिकीय मशीन अनुवाद (SMT)
- 2.3 न्यूरल मशीनी अनुवाद (NMT)
- 2.4 वाक्यांश-आधारित मशीन अनुवाद (PBMT)
- 2.5 उदाहरण-आधारित मशीनी अनुवाद (EBMT)

3. मशीनी अनुवाद के सोपान/चरण

- 3.1 शब्दावली और शब्दार्थ की सम्बद्धता
- 3.2 वाक्य संरचना का विश्लेषण

3.3 भाषा योग्यता और सांद्रता का आकलन

3.4 मशीनी प्रशिक्षण और सुधार

3.5 स्रोत भाषा का विश्लेषण

3.6 कोश एवं व्याकरण नियमों का अवलोकन एवं क्रियान्वयन

3.7 लक्ष्य भाषा का पुनर्गठन

4 मशीनी अनुवाद की समस्याएं

5. मशीनी अनुवाद सिस्टम

5.1 विश्व के प्रमुख मशीनी अनुवाद सिस्टम

5.2 भारत के प्रमुख मशीनी अनुवाद सिस्टम

5.3 भारतीय भाषाओं में मशीनी अनुवाद सिस्टम की
आवश्यकता

6. भारत में एमटी रिसर्च

सन्दर्भ सूची

मशीनी अनुवाद/एमटी (Machine Translation/MT)

1.1 भूमिका

मशीनी अनुवाद भाषा विज्ञान और कम्प्यूटर विज्ञान अथवा संगणकीय भाषाविज्ञान का एक क्षेत्र है जिसका मुख्य उद्देश्य है- एक भाषा से दूसरी भाषा में लिखे गए पाठ को स्वतंत्र और समुचित रूप से अनुवाद करना। यह प्रक्रिया कम्प्यूटर प्रोग्रामों और तकनीकी एल्गोरिदम की मदद से संभव होती है।

1.2 मशीनी अनुवाद का प्रारंभ

अनुवाद के मशीनीकरण का पहला प्रस्ताव **1940** के दशक के आसपास एक रूसी स्मिरनोव-ट्राँयंस्की और एक फ्रांसीसी आर्टरूनी द्वारा किया गया था। इलेक्ट्रॉनिक कम्प्यूटरों की उपलब्धता के बाद ही इसे गंभीरता से लिया जा सका। ब्रिटेन में, **1946** से अस्थायी प्रयोग शुरू किए गए थे। **1949** में वॉरेन वीवर ने अपने **200** परिचितों को एक ज्ञापन भेजा और मशीनी अनुवाद को एक वैज्ञानिक उद्यम के रूप में लॉन्च किया। अमेरिकी सरकारी एजेंसियों को मशीनी अनुवाद के अनुसंधान पर विशाल धन का निवेश करने के लिए प्रोत्साहित किया गया था। **1965** तक यह अनुमान लगाया गया था कि मशीनी अनुवाद में अनुसंधान के लिए **17** संस्थानों में लगभग **20** मिलियन डॉलर का सहयोग दिया गया था। लेकिन वे बहुत उत्साही थे और '**5** साल के भीतर' वाणिज्यिक मशीनी अनुवाद सिस्टम उपलब्ध कराने के

लिए शीघ्रता में वादा किया गया था। इसलिए, यह विफल रहा क्योंकि उत्पादित अनुवाद अधूरे थे और लगभग अपठनीय थे। **1964** में, संयुक्त राज्य अमेरिका में राष्ट्रीय विज्ञान फाउंडेशन ने **ALPAC** (स्वचालित भाषा प्रसंस्करण सलाहकार समिति) नामक एक समिति की स्थापना की। इस समिति को उस समय अमेरिका में चल रहे एमटी अनुसंधान के प्रदर्शन पर रिपोर्ट देनी थी। **1966** में, **ALPAC** ने बताया कि मशीनी अनुवाद, मानव अनुवाद की तुलना में धीमा, कम सटीक और दोगुना महंगा था और 'उपयोगी मशीनी अनुवाद की कोई तत्काल में कोई संभावना नहीं थी। इसके बजाय, समिति ने मशीनी सहायता प्राप्त अनुवादकों के विकास की सिफारिश की, जैसे कि स्वचालित शब्दकोश। इसने संगणकीय भाषाविज्ञान के सामान्य क्षेत्र में आर्थिक सहयोग जारी रखने की भी सिफारिश की। मशीनी अनुवाद के शोधकर्ताओं ने इस सिफारिश का विरोध किया और इसे शीघ्रता में लिए गए निर्णय बताते हुए इसकी निंदा की। इसके बाद, एमटी अनुसंधान को अमेरिका और हर जगह आर्थिक कमी और सहयोग में नुकसान का सामना करना पड़ा। बाद में, **1967** के आसपास एमटी अनुसंधान को कम स्तर पर जारी रखा गया और इसकी उपलब्धियों को मान्यता नहीं मिली। उसी समय, एमटी सिस्टम सीमित क्षेत्रों के भीतर उचित स्वीकार्य अनुवाद के साथ परिचालन के लिए उपलब्ध थे। **1970** के दशक में यूरोप, जापान और कनाडा में विशिष्ट क्षेत्र या डोमेन के स्तर पर अनुसंधान जारी रहा। **1980** के दशक में, कनाडा में, **METEO**

मशीनी अनुवाद सिस्टम या प्रणाली अंग्रेजी से फ्रेंच में मौसम के पूर्वानुमान का अनुवाद करने के लिए **1976** में विकसित किया गया था। **I** जापान में, पांचवीं पीढ़ी की कंप्यूटर प्रणाली ने एमटी अनुसंधान को गति दी। सांख्यिकीय दृष्टिकोण में विभिन्न प्रयोग करने योग्य प्रणालियां विकसित की गईं। इस अवधि के दौरान कृत्रिम बुद्धि (एआई) और ज्ञान आधारित दृष्टिकोण सहित अंतर्भाषिक(इंटरलिंगुअल) एमटी भी विकसित किए गए थे। **1990** के दशक में, कई अनुभवजन्य कॉर्पोरेट आधारित दृष्टिकोण विकसित किए गए।

प्राथमिक मशीनी अनुवाद सिस्टम को सरल और संक्षिप्त भाषा संरचना को अनुवाद करने के लिए विकसित किया गया था। इसमें विभिन्न वाक्य संरचनाओं, शब्दों और वाक्यांशों को संबद्धता और अनुवाद के लिए प्रयोग किया गया। इस प्रकार के सिस्टम द्वारा छोटे और साधारण पाठों का अनुवाद किया जा सकता है, लेकिन यह उच्च स्तर की भाषा और भाषा के विविध पहलुओं को समझने और संबद्धता को सही ढंग से अनुवाद करने में समर्थ नहीं होते।

1.2 प्राथमिक मशीनी अनुवाद सिस्टम

प्रारंभिक चरण में विकसित विभिन्न प्रयोगात्मक और परिचालित कुछ प्रमुख एमटी सिस्टम इस प्रकार थे:

मार्कII (MARKII) एमटी सिस्टम - यह अमेरिकी वायु सेना के विदेशी प्रौद्योगिकी विभाग में **1964** में स्थापित सबसे शुरुआती एमटी प्रणालियों में से एक था। यह

रूसी को अंग्रेजी में अनुवाद करने के लिए विकसित किया गया था। यह **1970** तक उपयोग में रहा।

जॉर्जटाउन(GEORGETOWN) एमटी सिस्टम - यह एमटी सिस्टम **1964** में लियोन डोस्टर्ट के नेतृत्व में जॉर्जटाउन विश्वविद्यालय के अनुसंधान समूह द्वारा विकसित किया गया था। यह परमाणु ऊर्जा आयोग के ओक रिज राष्ट्रीय प्रयोगशाला और इटली के इस्प्रा में यूराटॉम में स्थापित किया गया था।

सिसट्रान (Systran) - Systran MT सिस्टम 1970 में Latsec Inc. द्वारा रूसी में अनुवाद करने के लिए विकसित किया गया था। इस प्रणाली ने अमेरिकी वायु सेना के विदेशी प्रौद्योगिकी विभाग में मार्क **II (ऊपर)** को प्रतिस्थापित किया। इसका उपयोग नासा द्वारा अपोलो-सोयुज परियोजना के दौरान भी किया गया था। **1976** में सिस्ट्रान ने **EURATOM** में जॉर्जटाउन सिस्टम की जगह ली। इसके अलावा इसे अंग्रेजी से फ्रेंच अनुवाद के लिए भी विकसित किया गया था और परीक्षण अवधि के लिए यूरोपीय समुदाय आयोग द्वारा लिया गया था।

टॉम मिटियो (TAUM METEO)

टीएयूएम परियोजना में विभिन्न प्रयोगात्मक और परिचालन प्रणालियां विकसित की गई थीं। यह **1968** और **1980** के बीच मॉन्ट्रियल विश्वविद्यालय (टीएयूएम) में एमटी समूह द्वारा विकसित किया गया था। विकसित प्रयोगात्मक अन्य अनुवाद

प्रणालियां थीं- टीएयूएम -71, टीएयूएम -73, टीएयूएम -76 आदि । टीएयूएम-एमईटीओ वर्तमान में परिचालन प्रणालियों के बीच पूरी तरह से स्वचालित उच्च गुणवत्ता वाले अनुवाद के निकटतम अनुवाद का अच्छा उदाहरण है। यह 1974 में कनाडाई पर्यावरण विभाग द्वारा जारी मौसम रिपोर्ट को अंग्रेजी से फ्रेंच में अनुवाद करने के लिए विकसित किया गया था। टीएयूएम एमईटीओ की सबसे उल्लेखनीय खासियत इसकी अति अर्थशीलता और हस्तांतरण मॉड्यूल की कमी है।

इसके रूपांतरण घटक में दो उप-घटक शामिल हैं; शाब्दिक और संरचनात्मक। टीएयूएम प्रणाली में विश्लेषण चरण वाक्यात्मक है, और इसमें कई शब्दार्थ विशेषताएं भी शामिल हैं।

टीएयूएम-एमईटीओ 1977 से चालू है, जो पश्च संपादन के बिना 80% की सफलता की दर से सालाना लगभग पांच मिलियन शब्दों का अनुवाद करता है। यह परिचालन एमटी प्रणालियों के बीच गुणवत्ता का एक विशिष्ट उच्च स्तर है।

1.3 भारत में मशीनी अनुवाद का प्रारंभ

भारत में मशीनी अनुवाद का प्रारंभ सामान्यतः 1960 के दशक में हुआ। उस समय सरकारी संस्थानों और वैज्ञानिक संस्थानों ने मशीनी अनुवाद के अध्ययन और विकास में गहरी रुचि दिखाई। भारतीय भाषाओं के लिए मशीनी अनुवाद सिस्टमों का विकास हुआ, जिनमें विभिन्न भाषाओं के पाठों को अनुवाद करने की कोशिश

की गई। यह प्रयास भारतीय भाषाओं के अनुवाद क्षेत्र में एक महत्वपूर्ण योगदान है, जो भाषाओं के बीच संचार को सुगम बनाने में मदद करता है। भारतीय मशीनी अनुवाद सिस्टमों में बहुभाषिकता और वैश्विक भाषाओं के साथ कार्य करने की क्षमता भी विकसित की गई है।

1.4 भाषा अनुवाद के प्रमुख पहलू

अनुवाद परियोजना की विशिष्ट आवश्यकताओं और लक्ष्यों के आधार पर अनुवाद को विभिन्न तरीकों से संपादित किया जा सकता है। भाषा अनुवाद के कुछ प्रमुख पहलू इस प्रकार हैं:

- a. स्रोत भाषा और लक्ष्य भाषा:** स्रोत भाषा मूल पाठ की भाषा है जिसका अनुवाद किया जाना है, जबकि लक्ष्य भाषा वह भाषा है जिसमें पाठ का अनुवाद किया जाएगा।
- b. भाषाई क्षमता:** एक अनुवादक के पास स्रोत और लक्ष्य दोनों भाषाओं में मजबूत भाषाई कौशल होना चाहिए। इसमें व्याकरण, शब्दावली, वाक्यविन्यास, मुहावरेदार अभिव्यक्तियों और सांस्कृतिक बारीकियों की गहरी समझ शामिल है।
- c. अनुवाद का अर्थ और संदर्भ:** प्रभावी अनुवाद शब्द-दर-शब्द रूपांतरण से परे है। इसके लिए स्रोत पाठ के अर्थ, आशय और संदर्भ को पकड़ने और लक्ष्य भाषा में सटीक रूप से संप्रेषित करने की आवश्यकता होती है। इसमें दोनों भाषाओं के सांस्कृतिक और सामाजिक संदर्भ को समझना शामिल है।

- d. सांस्कृतिक अनुकूलन:** अनुवाद के लिए अक्सर सामग्री को लक्षित दर्शकों की सांस्कृतिक मानदंडों, प्राथमिकताओं और संवेदनशीलता के अनुकूल बनाने की आवश्यकता होती है। इसमें मुहावरों, संदर्भों, हास्य या रूपकों को समायोजित करना शामिल हो सकता है ताकि यह सुनिश्चित किया जा सके कि वे लक्ष्य भाषा और संस्कृति में प्रासंगिक और समझने योग्य हैं।
- e. विशिष्ट ज्ञान:** कुछ अनुवाद परियोजनाओं के लिए विशिष्ट डोमेन जैसे कानूनी, चिकित्सा, तकनीकी या साहित्यिक क्षेत्रों में विशेषज्ञता की आवश्यकता हो सकती है। इन क्षेत्रों में विशेष ज्ञान रखने वाले अनुवादक सटीक और प्रासंगिक रूप से उपयुक्त अनुवाद प्रदान कर सकते हैं।
- f. उपकरण और प्रौद्योगिकी:** अनुवाद उपकरण और प्रौद्योगिकियां, जैसे कंप्यूटर-सहायता प्राप्त अनुवाद (सीएटी) उपकरण, मशीन अनुवाद (एमटी), और शब्दावली प्रबंधन प्रणाली, दक्षता, स्थिरता और सटीकता में सुधार करने में अनुवादकों की सहायता कर सकती हैं।
- g. पुनरीक्षण और गुणवत्ता आश्वासन:** अनुवाद सामग्री की सटीकता, निरंतरता और समग्र गुणवत्ता सुनिश्चित करने के लिए अनुवाद में संपूर्ण संशोधन और गुणवत्ता आश्वासन प्रक्रियाएं महत्वपूर्ण हैं। इसमें अनुभवी भाषाविदों या विषय वस्तु विशेषज्ञों द्वारा प्रूफरीडिंग, संपादन और अनुवाद की समीक्षा शामिल हो सकती है।

अनुवाद एक जटिल और गतिशील प्रक्रिया है जिसके लिए भाषाई दक्षता, सांस्कृतिक समझ और विषय वस्तु विशेषज्ञता की आवश्यकता होती है। यह

विभिन्न भाषा समुदायों के बीच ज्ञान, सूचना और विचारों के आदान-प्रदान को सक्षम करने, क्रॉस-सांस्कृतिक संचार को सुविधाजनक बनाने में महत्वपूर्ण भूमिका निभाता है।

1.5 मशीनी अनुवाद (MT) की सीमाएँ

मशीनी अनुवाद (MT) के क्षेत्र ने महत्वपूर्ण प्रगति की है, लेकिन यह अभी भी कुछ सीमाओं का सामना करता है जो अनुवाद की सटीकता और गुणवत्ता को प्रभावित कर सकती हैं। यहाँ मशीनी अनुवाद प्रणालियों की कुछ सामान्य सीमाएँ हैं:

1. अस्पष्टता: प्राकृतिक भाषाएँ स्वाभाविक रूप से अस्पष्ट होती हैं, और एक वाक्य के इच्छित अर्थ को समझने के लिए अक्सर प्रासंगिक ज्ञान और पृष्ठभूमि की जानकारी की आवश्यकता होती है। मशीनी अनुवाद प्रणालियाँ पर्याप्त संदर्भ के बिना शब्दों, वाक्यांशों या वाक्यों को स्पष्ट करने के लिए संघर्ष कर सकती हैं, जिससे गलत या अस्पष्ट अनुवाद हो सकते हैं।

2. मुहावरेदार अभिव्यक्तियाँ: मुहावरों, कहावतों और बोलचाल की भाषा सहित मुहावरे की अभिव्यक्तियाँ आमतौर पर भाषाओं में उपयोग की जाती हैं। इन अभिव्यक्तियों के आलंकारिक अर्थ हैं जिन्हें केवल अलग-अलग शब्दों का विश्लेषण करके नहीं समझा जा सकता है। मशीनी अनुवाद प्रणालियाँ अक्सर मुहावरेदार भावों का सटीक अनुवाद करने के लिए संघर्ष करती हैं, जिसके परिणामस्वरूप शाब्दिक या निरर्थक अनुवाद होते हैं।

3. सांस्कृतिक बारीकियाँ: भाषाएँ संस्कृति में गहराई से निहित हैं, और कुछ वाक्यांशों, संदर्भों, या रूपकों में सांस्कृतिक अर्थ हो सकते हैं जो मशीनी अनुवाद प्रणाली द्वारा समझने के लिए चुनौतीपूर्ण हैं। सांस्कृतिक बारीकियों का सही-सही अनुवाद करने के लिए सांस्कृतिक संदर्भ की गहरी समझ की आवश्यकता होती है, जिसे समझना मशीनों के लिए मुश्किल हो सकता है।

4. दुर्लभ और डोमेन-विशिष्ट शब्दावली: मशीनी अनुवाद सिस्टम उपलब्ध प्रशिक्षण डेटा पर भरोसा करते हैं, जिसमें दुर्लभ या डोमेन-विशिष्ट शब्द शामिल नहीं हो सकते हैं। तकनीकी, वैज्ञानिक, या कानूनी शब्दों जैसी विशिष्ट शब्दावली का सही-सही अनुवाद करना **MT** सिस्टम के लिए चुनौतीपूर्ण हो सकता है, जब तक कि उन्हें प्रासंगिक डोमेन-विशिष्ट डेटा पर प्रशिक्षित नहीं किया गया हो।

5. रूपात्मक और वाक्यात्मक विविधताएँ: भाषाएँ आकृति विज्ञान, वाक्य रचना और शब्द क्रम में विविधताएँ प्रदर्शित करती हैं। मशीनी अनुवाद प्रणालियाँ इन विविधताओं को संभालने के लिए संघर्ष कर सकती हैं, विशेष रूप से महत्वपूर्ण भिन्न संरचनाओं वाली भाषाओं के बीच अनुवाद करते समय। एमटी सिस्टम के लिए सही शब्द विभक्ति, क्रिया काल, या संज्ञा घोषणाओं को पकड़ना चुनौतीपूर्ण हो सकता है।

6. प्रासंगिक समझ का अभाव: मशीनी अनुवाद प्रणालियाँ आमतौर पर वाक्य-दर-वाक्य के आधार पर काम करती हैं और इसमें संवाद या दस्तावेज़-स्तर के संदर्भ की व्यापक समझ का अभाव हो सकता है। जब अर्थ पिछले या बाद के वाक्यों की

जानकारी पर निर्भर करता है तो इसके परिणामस्वरूप अनुवाद में विसंगतियां या त्रुटियां हो सकती हैं।

7. सीमित प्रशिक्षण डेटा: मशीनी अनुवाद मॉडल को सटीक अनुवाद पैटर्न सीखने के लिए बड़ी मात्रा में उच्च गुणवत्ता वाले द्विभाषी प्रशिक्षण डेटा की आवश्यकता होती है। हालांकि, कम संसाधन वाली भाषाओं या सीमित उपलब्ध डेटा वाले भाषा जोड़े के लिए, मशीनी अनुवाद प्रणाली का प्रशिक्षण अधिक चुनौतीपूर्ण हो जाता है, जिससे अनुवाद की गुणवत्ता खराब हो जाती है।

8. पश्च संपादन आवश्यकताएं : चूंकि मशीनी अनुवाद एक प्रारंभिक बिंदु प्रदान कर सकता है, आउटपुट में अक्सर त्रुटियों को ठीक करने, प्रवाह में सुधार करने और अंतिम अनुवाद को गुणवत्ता मानकों को पूरा करने के लिए सुनिश्चित करने के लिए मानव अनुवादकों द्वारा पोस्ट-एडिटिंग की आवश्यकता होती है। संपादन के बाद की इस प्रक्रिया में अतिरिक्त समय और प्रयास की आवश्यकता होती है।

इन सीमाओं के बावजूद, अधिक परिष्कृत मॉडल के विकास, प्रशिक्षण तकनीकों में सुधार, बड़े और अधिक विविध प्रशिक्षण डेटासेट की उपलब्धता के कारण मशीन अनुवाद आगे बढ़ना जारी है। शोधकर्ता इन सीमाओं को दूर करने और मशीन अनुवाद प्रणालियों की गुणवत्ता और प्रवाह को बढ़ाने के लिए लगातार काम कर रहे हैं।

2.

मशीनी अनुवाद के प्रकार

विभिन्न विशेषताओं और दृष्टिकोणों के आधार पर, मशीनी अनुवाद (MT) कई प्रकार में वर्गीकृत किया जा सकता है। यहाँ कुछ सामान्य मशीनी अनुवाद के प्रकार हैं:

- 1. नियम-आधारित मशीन अनुवाद (RBMT):** RBMT सिस्टम अनुवाद करने के लिए भाषाई नियम और शब्दकोशों के सेटों पर निर्भर करते हैं। भाषा विशेषज्ञ और भाषाविज्ञानी इन नियमों को मानवीय रूप से बनाते हैं ताकि स्रोत भाषा से लक्ष्य भाषा में वाक्यों को रूपांतरित कर सकें। RBMT सिस्टम इनपुट पाठ की व्याकरणिक संरचना का विश्लेषण करते हैं और उचित नियमों को लागू कर अनुवाद उत्पन्न करते हैं। RBMT विशिष्ट क्षेत्र या डोमेन में भाषाओं को अच्छी तरह से संभाल सकता है, इसमें बहुत अधिक मैन्युअल रूप से नियमों के विकास की आवश्यकता होती है और जटिल भाषिक विशेषताओं से जूझ सकता है।
- 2. सांख्यिकीय मशीन अनुवाद (SMT):** SMT सिस्टम बड़ी द्विभाषीय कॉर्पस से अनुवाद पैटर्न सीखने के लिए सांख्यिकीय मॉडल का उपयोग करते हैं। ये मॉडल स्रोत और लक्ष्य भाषाओं में शब्दों और वाक्यांशों के बीच सह-घटना और सांख्यिकीय संबंधों का विश्लेषण करते हैं। SMT सिस्टम विभिन्न अनुवाद विकल्पों की संभावनाओं की गणना करते हैं और सीखे गए सांख्यिकीय डाटा के आधार पर सबसे अधिक सही संभावित अनुवाद का चयन करते हैं।

3. न्यूरल मशीनी अनुवाद (NMT): NMT मशीनी अनुवाद एक नए और बहुत प्रभावी तरीके को दर्शाता है। यह कृत्रिम न्यूरल नेटवर्क का उपयोग करता है, विशेष रूप से डीप लर्निंग मॉडल्स, जो द्विभाषीय डेटा से अनुवाद मैपिंग को सीधे सीखने के लिए इस्तेमाल किया जाता है। **NMT** मॉडल में एक एनकोडर-डिकोडर आर्किटेक्चर होता है, जहाँ एनकोडर नेटवर्क स्रोत भाषा इनपुट को प्रसंस्कृत करता है और इसे एक निश्चित-लंबाई के प्रतिनिधित्व में एन्कोड करता है। डिकोडर नेटवर्क फिर एनकोड रिप्रेजेंटेशन पर आधारित अनुवादित आउटपुट उत्पन्न करता है। **NMT** ने अनुवाद गुणवत्ता में विशेष सुधार दिखाया है और मशीन अनुवाद में स्टेट ऑफ द आर्ट तकनीक बन गया है।

4. वाक्यांश-आधारित मशीन अनुवाद (PBMT): PBMT एक पुरानी सांख्यिकीय मशीनी अनुवाद तकनीक है जो इनपुट पाठ को वाक्यांशों में तोड़ती है और उन्हें व्यक्तिगत रूप से अनुवादित करती है। द्विभाषी डेटा से सीखे गए सांख्यिकीय पैटर्न के आधार पर वाक्यांशों का अनुवाद निर्धारित किया जाता है। स्रोत और लक्ष्य भाषाओं में वाक्यांशों के बीच पत्राचार की पहचान करने के लिए पीबीएमटी सिस्टम अक्सर वाक्यांश संरेखण मॉडल का उपयोग करते हैं।

5. उदाहरण-आधारित मशीनी अनुवाद (EBMT): EBMT पूर्व में अनुवादित वाक्यों या वाक्यांश युग्मों के डेटाबेस पर निर्भर करता है। एक नए वाक्य का सामना करते समय, सिस्टम डेटाबेस में समान उदाहरणों की खोज करता है और अनुवाद उत्पन्न करने के आधार के रूप में उनका उपयोग करता है।

ईबीएमटी सिस्टम आमतौर पर निकटतम मिलान उदाहरण खोजने के लिए पैटर्न मिलान और पुनर्प्राप्ति तकनीकों का उपयोग करते हैं।

ये कुछ प्रमुख मशीन अनुवाद के प्रकार हैं। हर प्रकार के अपने फायदे और सीमाएं होती हैं, और चल रहे अनुसंधान का ध्यान इन दृष्टियों को जोड़ने और हाइब्रिड मॉडलों का अन्वेषण करने पर होता है ताकि अनुवाद गुणवत्ता और ध्वनि को और अधिक सुधारा जा सके।

3.

मशीनी अनुवाद के सोपान/चरण

मशीनी अनुवाद के सोपान या चरण निम्नलिखित हो सकते हैं:

3.1 शब्दावली और शब्दार्थ की सम्बद्धता : इस चरण में, मशीनी अनुवाद सिस्टम को भाषा की शब्दावली और उनके शब्दार्थ को सम्बद्ध करना होता है। यह शब्दावली और शब्दार्थ का संग्रहण, उनके पर्यायवाची अथवा समानार्थी, विलोम, अनेकार्थता आदि विशेषताओं पर आधारित होता है। उदाहरण के लिए, एक मशीनी अनुवाद सिस्टम को विभिन्न भाषाओं में "नमस्ते" शब्द को संबद्ध करने की क्षमता होनी चाहिए।

3.2 वाक्य संरचना का विश्लेषण: इस चरण में, मशीनी अनुवाद सिस्टम को वाक्य संरचना को समझने और विश्लेषण करने की क्षमता होती है। यह शब्दों के क्रम, वाक्यांशों के निर्माण, भाषा नियमों की जांच, और संवादित वाक्यों की समझ पर आधारित होता है। उदाहरण के लिए, एक मशीनी अनुवाद सिस्टम को वाक्य संरचना को समझने की क्षमता होनी चाहिए, जैसे "मैं एक किताब पढ़ रहा हूँ" और "**I am reading a book**" को विश्लेषित और सम्बद्ध करने की क्षमता होनी चाहिए ।

3.3 भाषा योग्यता और सांद्रता का आकलन: इस चरण में, मशीनी अनुवाद सिस्टम को भाषा की योग्यता और सांद्रता को मापने की क्षमता होती है। यह

शब्दों की श्रेष्ठता, वाक्य संरचना की उपयुक्तता , विविधता की पहचान, वाणिज्यिक उपयोग की योग्यता, और भाषा के सामाजिक मानकों के आधार पर संभव होता है। उदाहरण के लिए, एक मशीनी अनुवाद सिस्टम को सामाजिक संदेशों को सही ढंग से अनुवाद करने की क्षमता होनी चाहिए, जिसमें सटीकता, अर्थवत्ता, और व्यावसायिक भाषा के आदान-प्रदान को ध्यान में रखा जाता है।

3.4 मशीनी प्रशिक्षण और सुधार: इस चरण में, मशीनी अनुवाद सिस्टम को स्वतंत्रता से सीखने और सुधार करने की क्षमता होती है। यह चरण अद्यतित भाषा डेटा का उपयोग करके सिस्टम को सुधारने और नई जानकारी और योग्यताओं को सीखने का मार्ग प्रदान करता है। उदाहरण के लिए, एक मशीनी अनुवाद सिस्टम को नए शब्दों को सीखने और भाषा योग्यता को सुधारने की क्षमता होनी चाहिए, जिससे उसका अनुवाद देने का गुणवत्ता और सटीकता बढ़ती है।

मशीनी अनुवाद के इन सोपानों द्वारा, सिस्टम को भाषा को समझने, वाक्यों को अनुवाद करने, भाषा की योग्यता और सांद्रता को मापने, और स्वतंत्रता से सीखने की क्षमता मिलती है। यह सोपान समग्र मशीनी अनुवाद प्रक्रिया को संभव बनाते हैं और अनुवाद की गुणवत्ता को सुधारते हैं।

संक्षेप में मशीनी अनुवाद के सोपानों को हम इस प्रकार से देख सकते हैं :

3.5 स्रोत भाषा का विश्लेषण

स्रोत भाषा का विश्लेषण मशीनी अनुवाद में एक महत्वपूर्ण पहलू है। यह स्रोत भाषा के विभिन्न तत्वों की गहराई से विश्लेषण करता है ताकि इसे सुधारने और अच्छे रूप में अनुवाद करने के लिए मशीनी अनुवाद सिस्टम को सही ज्ञान और संदेश प्राप्त हो सके। स्रोत भाषा के विश्लेषण में निम्नलिखित महत्वपूर्ण तत्व शामिल होते हैं:

वर्णमाला और उच्चारण: स्रोत भाषा के वर्णमाला और उच्चारण का विश्लेषण करने से मशीनी अनुवाद सिस्टम को संदेश को सही ढंग से समझने और उसे अनुवाद करने की क्षमता मिलती है। उदाहरण के लिए, जब सिस्टम को पता होता है कि एक वाक्य में कौन से वर्ण हैं और उनका सही उच्चारण क्या होना चाहिए, तो उसे समानार्थक वाक्य को सही प्रकार से अनुवाद करने में मदद मिलती है।

वाक्य संरचना: स्रोत भाषा के वाक्य संरचना का विश्लेषण मशीनी अनुवाद के लिए महत्वपूर्ण है। यह सिस्टम को समझने में मदद करता है कि कौन से शब्द किस प्रकार से एक वाक्य में संयोजित होते हैं और कौन से भाषाई नियमों का पालन किया जाना चाहिए। इसके बिना, संदेश को सही ढंग से अनुवाद करना मुश्किल हो सकता है।

वाक्यांशों का अनुवाद: स्रोत भाषा में वाक्यांशों के विश्लेषण से मशीनी अनुवाद सिस्टम को संदेश को अधिक संगठित और अच्छे रूप से अनुवाद करने में मदद मिलती है। यह सिस्टम को उच्चतम संभावित अर्थ का पता लगाने में मदद करता है और सुनिश्चित करता है कि संदेश का मतलब सही ढंग से संदिग्धता के बिना समझा जा सके।

वाक्यांशों के संबंध: स्रोत भाषा में वाक्यांशों के संबंध का विश्लेषण मशीनी अनुवाद के लिए महत्वपूर्ण है। यह सिस्टम को संदेश के विभिन्न अंगों के बीच के संबंध को समझने में मदद करता है और उसे सही ढंग से अनुवाद करने की क्षमता प्रदान करता है।

इन सभी तत्वों का विश्लेषण मशीनी अनुवाद सिस्टम को स्रोत भाषा को सही ढंग से समझने और उसे सही ढंग से अनुवाद करने के लिए मदद करता है। यह सुनिश्चित करता है कि अंतरभाषा में संदेश की मान्यता, संगठन और उपयुक्तता का पालन होता है।

3.6 कोश एवं व्याकरण नियमों का अवलोकन एवं क्रियान्वयन कोश और व्याकरण नियम भाषा के विभिन्न पहलुओं को संगठित करने और उन्हें समझने में मदद करते हैं। ये नियम भाषा की संरचना, व्याकरणिक नियमों, शब्दावली, अर्थ और प्रयोग के संबंध में होते हैं। इन नियमों का अवलोकन और क्रियान्वयन भाषा विज्ञान के अध्ययन में महत्वपूर्ण है।

कोश नियमों का अवलोकन करता है कि भाषा में शब्दों का उपयोग कैसे होता है और उनका अर्थ क्या होता है। कोश नियम व्याकरण, संधि, विराम चिह्न, शब्दरचना, पर्यायवाची, विलोम शब्द, उपसर्ग और प्रत्यय आदि के बारे में जानकारी प्रदान करते हैं। इन नियमों का क्रियान्वयन करके हम भाषा में सही रूप से शब्दों का उपयोग कर सकते हैं और वाक्यों को संरचित बना सकते हैं। व्याकरण नियमों का अवलोकन करता है कि भाषा में वाक्यों का कैसे रचना की जाती है और उनके भाषाई नियमों क्या होते हैं। व्याकरण नियम संज्ञा, सर्वनाम, क्रिया, कारक, काल, वचन, पुरुष, संबंधबोधक शब्द, वाच्य, समास, अलंकार, रस आदि के बारे में जानकारी प्रदान करते हैं। इन नियमों का क्रियान्वयन करके हम भाषा के वाक्यों को सही रूप से रच सकते हैं और उनका अर्थ समझ सकते हैं। यदि हम इन कोश और व्याकरण नियमों का अवलोकन और क्रियान्वयन नहीं करेंगे, तो हम भाषा में गलतियों कर सकते हैं और संदेश को सही ढंग से समझने में परेशानी हो सकती है। इसलिए, भाषा विज्ञान के माध्यम से हमें कोश और व्याकरण नियमों का अध्ययन करना चाहिए और इन्हें सही ढंग से क्रियान्वयन करना चाहिए ताकि हम सही भाषा का प्रयोग कर सकें।

3.7 लक्ष्य भाषा का पुनर्गठन

लक्ष्य भाषा के पुनर्गठन में लक्ष्य भाषा के व्याकरणिक नियमों के आधार पर अनूदित पाठ को पुनर्गठित किया जाता है I इसके अंतर्गत लक्ष्य भाषा के पदक्रम के अनुसार अनुवादित शब्दों को व्यवस्थित रखा जाता है I उदाहरण के लिए यदि हिंदी से अंग्रेजी अनुवाद किया जाये तो हिंदी वाक्यों का पदक्रम कर्ता-कर्म -क्रिया है जबकि अंग्रेजी में शब्द /पदक्रम कर्ता- क्रिया -कर्म है I इस आधार पर, अनुवाद में लक्ष्य भाषा के पदक्रम के अनुसार वाक्यों का पुनर्गठन किया जाता है I

4.

मशीनी अनुवाद की समस्याएं

मशीनी अनुवाद के बावजूद कुछ सामान्य समस्याएं होती हैं, जो अवधारणात्मक और भाषाई मामलों के कारण हो सकती हैं। यहां कुछ मुख्य समस्याएं हैं:

1. भाषा का अनुवाद: भाषाओं की संरचना, व्याकरण, शब्दावली, और भाषाई प्रतिक्रियाओं के बीच में विशेषताएं होती हैं। मशीनी अनुवाद को इन विशेषताओं को समझना और समान करना मुश्किल हो सकता है, जिसके कारण अनुवाद में अशुद्धियां हो सकती हैं।
2. पाठ का संदर्भ: अनुवाद के दौरान, मशीन को पाठ के संदर्भ को समझना और उसे ठीक से अनुवाद करना होता है। कई बार संदर्भ का अनुवाद गलत हो सकता है और इससे पूरे पाठ का अर्थ बिगड़ सकता है।
3. सांद्रता का अभाव: मशीनी अनुवाद अक्सर भाषा के सांद्रता और परिभाषाओं को समझने में कठिनाई उठाता है। यह समानार्थी शब्दों, उपयोग के संदर्भ पर आधारित अर्थ, और भाषाई अभिव्यक्ति को समझने की क्षमता को प्रभावित कर सकता है।

4. कल्पना और विचार की कमी: मशीन अनुवाद को कल्पना और विचार करने की क्षमता से महत्वपूर्ण अंशों को समझने में कठिनाई होती है। इसके कारण अनुवाद में उच्च स्तर की रचनात्मकता और परिचय की कमी हो सकती है।
5. असंगत विशेषताएं: कई भाषाओं में विशेष व्याकरण नियम और असामान्य विशेषताएं होती हैं जो मशीनी अनुवाद के लिए चुनौतीपूर्ण हो सकती हैं। इसके कारण ऐसी भाषाओं के अनुवाद में अशुद्धियां और गलतियां हो सकती हैं।

ये समस्याएं मशीनी अनुवाद के विकास के साथ-साथ नवीनतम तकनीकों और एल्गोरिदम के द्वारा समाधानित हो रही हैं, लेकिन अभी तक पूर्णता नहीं हासिल की गई है। इसलिए, मशीनी अनुवाद का उपयोग करते समय सावधानी बरतना आवश्यक होता है और प्रतिष्ठानों द्वारा प्रदान की गई अनुवादों की सत्यता की जांच करनी चाहिए।

5.

मशीनी अनुवाद सिस्टम

मशीनी अनुवाद सिस्टम भाषाओं को स्वतंत्रता से संवारते हैं, लेकिन विभिन्न देशों और क्षेत्रों में इनका विकास विभिन्न हो सकता है।

5.1 विश्व के प्रमुख मशीनी अनुवाद सिस्टम:

दुनिया में मशीनी अनुवाद के लिए सबसे महत्वपूर्ण प्रणाली एक ऐसे तंत्र के रूप में कार्य करती है, जिसमें कंप्यूटर या सॉफ्टवेयर की सहायता से लिखित या बोले गए शब्दों, ग्रंथों, वाक्यांशों या संपूर्ण पाठ को एक भाषा से दूसरी भाषा में प्रसारित करने की क्षमता होती है। ये प्रणालियाँ मशीन लर्निंग, आर्टिफिशियल इंटेलिजेंस और भाषा प्रसंस्करण के मूल सिद्धांतों में निहित हैं।

मशीनी अनुवाद प्रणालियों का प्राथमिक उद्देश्य भाषाओं में संचार की प्रक्रिया को आसान बनाना है, जिससे व्यक्ति एक भाषा में पाठ लिख सकें और इसे आसानी से दूसरी भाषा में अनुवाद कर सकें। ये प्रणालियाँ कई उद्देश्यों को पूरा करती हैं, जिनमें वेबसाइट अनुवाद, दस्तावेज़ अनुवाद, संचार और रोज़मर्रा के भाषा आदान-प्रदान शामिल हैं, लेकिन इन्हीं तक सीमित नहीं हैं।

मशीन अनुवाद की प्रमुख प्रणालियों में Google अनुवाद, MS Translate, Amazon Translate, Sapi Translate, दीपल ट्रांसलेशन और यूनिवर्सल ट्रांसलेशन शामिल हैं। ये विशेष प्रणालियाँ अंग्रेजी, स्पेनिश, फ्रेंच, जर्मन, चीनी, जापानी और हिंदी जैसी विभिन्न वैश्विक भाषाओं में सामग्री का अनुवाद करने की क्षमता प्रदर्शित करती हैं।

मशीन अनुवाद प्रणालियाँ यूनिकोड की सहायता से कार्य करती हैं और व्यापार, व्यावहारिकता और सामाजिक मानदंडों के भाषाई नियमों को समझने के लिए शिक्षित हैं। ये प्रणालियाँ पाठ, शब्दावली प्रविष्टियों, अभिव्यक्तियों, वाक्य संगठन और सार्वभौमिक भाषा सिद्धांतों को समझने और उन्हें संदर्भ में रखकर उच्चारण और वाक्यों की व्याख्या करने का प्रयास करती हैं।

हालांकि, ये मशीनी अनुवाद सिस्टम अपूर्णताओं के साथ आते हैं और कई संदर्भों में मानव अनुवाद की गुणवत्ता को पूरी तरह से प्राप्त नहीं कर पाते हैं। भाषा के विषय में जटिलताओं, मूल्यों, और सांस्कृतिक पहलुओं को समझना, सामान्य अभिप्रेतियों को पकड़ना और समानार्थी शब्दों के बीच का विवेक करना मशीनों के लिए कठिनाई का कारण है।

यद्यपि मशीनी अनुवाद सिस्टम बहुत महत्वपूर्ण और उपयोगी हैं, लेकिन इनका उपयोग करते समय हमेशा ध्यान देना चाहिए कि उनके द्वारा उपलब्ध किए जाने वाले अनुवादों में त्रुटियां हो सकती हैं और संदेश की प्रकृति, प्रभाव और अभिप्रेति को पूरी तरह से प्रकट नहीं किया जा सकता है। हालांकि, ये सिस्टम निरंतर विकास कर रहे हैं और आने वाले समय में इसकी क्षमताएं और सुधारें बढ़ेंगी।

1. **Google अनुवाद:** **Google** अनुवाद **Google** द्वारा बनाई गई एक स्वचालित अनुवाद सेवा है, जिसे कई भाषाओं के बीच सहज अनुवाद की सुविधा के लिए डिज़ाइन किया गया है। यह Google Translate वेबसाइट और मोबाइल एप्लिकेशन दोनों के माध्यम से आसानी से उपलब्ध है, और इसमें टेक्स्ट अनुवाद, वेबसाइट अनुवाद, ऑफ़लाइन अनुवाद और वॉइस ट्रांसलेशन सहित कुछ ही नाम हैं।

2. MS Translate (Microsoft Translate): MS Translate, जिसे पहले **Bing Translate** के रूप में मान्यता प्राप्त थी, एक मशीन अनुवाद सुविधा है जिसे **Microsoft** द्वारा तैयार किया गया है। इसका प्राथमिक कार्य ग्रंथों और वाक्यांशों के बीच रूपांतरण करना है। इसके अतिरिक्त, **MS Translate** में **API** के माध्यम से अन्य अनुप्रयोगों और सेवाओं के साथ शामिल करने की क्षमता है। यह अनुवाद सुविधा वेबसाइट, मोबाइल ऐप और वेबसाइट संस्करण के माध्यम से आसानी से उपलब्ध है।

3. Amazon Translate: Amazon Translate एक **AWS (Amazon Web Services)** सेवा है जो ऑटोमेटेड मशीन अनुवाद प्रदान करती है। यह वेबसाइट, ऐप्स, वॉयस टेक्स्ट और डोक्युमेंट्स के बीच अनुवाद करने की क्षमता रखती है। **Amazon Translate** एक सुरक्षित, मैनेज्ड सेवा है जो विश्वसनीयता, स्केलेबिलिटी, और बहुभाषिकता प्रदान करती है। यह वेब और मोबाइल उपयोग के लिए उपलब्ध है और डेवलपर्स को भी **API** के रूप में उपयोग करने की अनुमति देता है।

4. Sapi Translate: **Sapi Translate** एक ऑनलाइन मशीनी अनुवाद सेवा है जिसे **SAPI (Solutions Atlantic Pacific Inc.)** ने विकसित किया है। इस सेवा का उपयोग टेक्स्ट, वेबसाइट, डोक्युमेंट, और ईमेल अनुवाद करने के लिए किया जा सकता है। इसका उपयोगानुक्रम निर्धारित करने के लिए प्लगइन, **API**, और वेब संस्करण उपलब्ध हैं। **Sapi Translate** विभिन्न भाषाओं का समर्थन करता है और उपयोगकर्ताओं को भाषांतर समस्याओं का समाधान प्रदान करने में मदद करता है।

5. दीपल ट्रांसलेशन (DeepL Translation): दीपल ट्रांसलेशन एक ईरूपीय भाषाओं के बीच मशीन अनुवाद प्रदान करने वाली सेवा है। यह वेबसाइट, मोबाइल ऐप, और डेस्कटॉप के रूप में उपलब्ध है। दीपल अनुवाद का उपयोग विज्ञापन, वेबसाइट कंटेंट, और विभिन्न दस्तावेज़ों के लिए किया जाता है। दीपल अनुवाद उन्नत एकल और न्यूरल मशीन अनुवाद (**NMT**) तकनीक का उपयोग करता है और पहले से ही अनुवादित डेटा का उपयोग करता है ताकि उपयोगकर्ताओं को उच्च गुणवत्ता के अनुवाद प्रदान किए जा सकें।

6. यूनिवर्सल ट्रांसलेशन: यूनिवर्सल ट्रांसलेशन (Universal Translation) एक व्यापक मशीनी अनुवाद संगठन की तरह हो सकती है जिसका नाम किसी विशेष सेवा के साथ जुड़ा हो सकता है। इस तरह की सेवाएं विभिन्न भाषाओं के बीच अनुवाद करने की क्षमता प्रदान करती हैं और उपयोगकर्ताओं को एकीकृत और सुविधाजनक अनुवाद समाधान प्रदान करने का प्रयास करती हैं। इन सेवाओं के रूप में अलग-अलग कंपनियों और उपयोगकर्ता संगठनों द्वारा प्रदान की जाने वाली विभिन्न सेवाएं हो सकती हैं। यह संगठन विभिन्न दस्तावेजों, विशेषताओं, और उद्देश्यों के लिए भाषांतर समाधान प्रदान कर सकती हैं।

5.2 भारत के प्रमुख मशीनी अनुवाद सिस्टम:

भारत में मशीन अनुवाद पर शोधकार्य **80** के दशक के अंत और **90** के दशक की शुरुआत से प्रारंभ हुआ। प्रारंभ में यह सरकारी निकायों का प्रयास था, बाद में निजी क्षेत्र द्वारा भी वित्त पोषित हुआ। आजकल, अंग्रेजी को भारतीय भाषाओं में अनुवाद करने के लिए अच्छी संख्या में प्रयोगात्मक और परिचालन में समर्थ एमटी सिस्टम विकसित किए जा रहे हैं। भारतीय भाषाओं के बीच अनुवाद के लिए भी प्रणालियां विकसित की गई हैं। मशीन अनुवाद प्रणालियों के निर्माण में कई विकसित और अविकसित परियोजनाएं हैं और एमटी से संबंधित उपकरणों की

बहुतायत भी है जैसे रूपात्मक विश्लेषक, भाषा टैगर्स के कुछ हिस्से, भाषा पार्सर इत्यादि। भारत में मशीनी अनुवाद अनुसंधान का नेतृत्व करने वाली कुछ मशीन अनुवाद परियोजनाएं इस प्रकार हैं:

आंग्लभारती और अनुभारती एमटी परियोजना-

अंग्रेजी से भारतीय भाषाओं में मशीनी सहायता प्राप्त अनुवाद के लिए **1991** में भारतीय प्रौद्योगिकी संस्थान, कानपुर में आंग्लभारती एमटी परियोजना शुरू की गई थी। यह एक पैटर्न निर्देशित नियम आधारित प्रणाली है जिसमें अंग्रेजी (स्रोत भाषा) के लिए संदर्भ मुक्त व्याकरण जैसी संरचना है। बाद में इसके प्रारंभिक दृष्टिकोण से इसको काफी हद तक संशोधित किया गया है।

अनुभारती को **1995** में हिंदी से अंग्रेजी अनुवाद के लिए विकसित किया गया है। यह उदाहरण आधारित मशीन अनुवाद करता है। इसे इसके अनुभारती-**II** संस्करण में भी संशोधित किया गया है जो विभिन्न प्रतिमानों के मिश्रित विधियों को अपनाता है।

मंत्रा एमटी परियोजना- अंग्रेजी पाठ का हिंदी में अनुवाद करने के लिए यह मशीन अनुवाद प्रणाली सी-डैक, पुणे में विकसित की गई है। यह स्रोत और लक्ष्य भाषा को पार्स और उत्पन्न करने के लिए ट्री एडजोइनिंग ग्रामर का उपयोग करता है। यह प्रशासन, वित्त, कृषि, स्वास्थ्य सेवा, सूचना

प्रौद्योगिकी, शिक्षा और अन्य सामान्य उद्देश्य गतिविधियों के क्षेत्रों सहित विभिन्न सरकारी दस्तावेजों का अनुवाद करने के लिए उपयोग में है।

इस परियोजना को अन्य भाषा जोड़े जैसे अंग्रेजी-बंगाली, अंग्रेजी-तेलुगु, अंग्रेजी-गुजराती, हिंदी-अंग्रेजी और हिंदी-बंगाली और हिंदी-मराठी जैसी भारतीय भाषाओं के बीच विस्तारित किया गया है।

अनुसारक एमटी परियोजना - यह मूल रूप से मशीन अनुवाद प्रणाली के बजाय एक भाषा एक्सेसर है। यह परियोजना तेलुगु, कन्नड़, मराठी, पंजाबी और बंगाली का हिंदी में अनुवाद करने के लिए बनाई गई है। यह भारतीय प्रौद्योगिकी संस्थान, कानपुर में शुरू किया गया था और बाद में हैदराबाद विश्वविद्यालय के मानविकी और सामाजिक विज्ञान विभाग के एप्लाइड भाषाविज्ञान और अनुवाद अध्ययन केंद्र (सीएएलटीएस) में स्थानांतरित कर दिया गया था।

यूएनएल-आधारित एमटी प्रणाली- भारतीय प्रौद्योगिकी संस्थान, मुंबई ने इंटरलिंगुआ के रूप में यूनिवर्सल नेटवर्किंग लैंग्वेज (यूएनएल) का उपयोग करके अंग्रेजी-हिंदी का अनुवाद करने के लिए एक मशीन अनुवाद प्रणाली विकसित की है। अंग्रेजी से मराठी और बंगाली का अनुवाद करने के लिए एक ही तकनीक पर इसी तरह की एमटी प्रणाली विकसित की जा रही है।

मातृ एमटी प्रणाली- यह सी-डैक, मुंबई में विकसित अंग्रेजी के हिंदी में अनुवाद के लिए एक मानव-सहायता प्राप्त हस्तांतरण-आधारित अनुवाद प्रणाली है। इस परियोजना को भारत के सूचना प्रौद्योगिकी मंत्रालय द्वारा वित्त पोषित किया गया है।

मशीन अनुवाद अनुसंधान में इन अग्रणी कार्यों के अलावा, भारत में विभिन्न शैक्षिक और अनुसंधान संस्थानों में विकसित या विकसित की जा रही कई मशीन अनुवाद प्रणालियां हैं। जादवपुर विश्वविद्यालय के कंप्यूटर विज्ञान और इंजीनियरिंग विभाग ने वाक्यांश उदाहरण आधारित मशीनी अनुवाद दृष्टिकोण के आधार पर अंग्रेजी समाचार का बंगाली में अनुवाद करने के लिए एक एमटी सिस्टम (अनुवाद) प्रस्तुत किया है। उत्कल विश्वविद्यालय, भुवनेश्वर अंग्रेजी-उड़िया मशीन अनुवाद प्रणाली पर काम कर रहा है। भारतीय प्रौद्योगिकी संस्थान, दिल्ली के गणित विभाग ने अंग्रेजी-हिंदी के लिए एक उदाहरण आधारित मशीन अनुवाद प्रणाली पर शोध किया है। आईबीएम इंडिया रिसर्च लैब ने अंग्रेजी का भारतीय भाषाओं में अनुवाद करने के लिए सांख्यिकीय मशीनी अनुवाद पर काम शुरू किया है।

Indic Translator: यह सिस्टम भारतीय भाषाओं के अनुवाद के लिए विकसित हुआ है और भारतीय भाषाओं की विशेषताओं को ध्यान में रखता है।

Bharatavani: भारतीय भाषाओं का आधिकारिक मशीनी अनुवाद सिस्टम जो विभिन्न भारतीय भाषाओं के बीच अनुवाद करता है।

5.3 भारतीय भाषाओं में मशीनी अनुवाद सिस्टम की आवश्यकता: भारत एक विविध भाषाओं का देश है, जहां अनेक भाषाएं बोली जाती हैं। मशीनी अनुवाद सिस्टम की आवश्यकता भाषाओं के बीच संचार और समझ में मदद करने के लिए होती है। इससे अलग-अलग भाषाओं के बीच संचार में सुविधा होती है और भाषा सीमाओं को पार किया जा सकता है। इसके साथ ही, व्यापार, प्रशासनिक, और विज्ञान आदि क्षेत्रों में भाषाओं के अनुवाद के लिए भी मशीनी अनुवाद सिस्टम का उपयोग किया जाता है। भारतीय भाषाओं को समर्पित अनेक मशीनी अनुवाद सिस्टम अभी कार्यरत हैं **I** भारतीय बोलियों के अनुवाद के लिए भी अद्यतन मशीनी अनुवाद सिस्टम की आवश्यकता है **I**

6.

भारत में एमटी रिसर्च

भारत में एमटी (मशीन अनुवाद) रिसर्च भाषा प्रसंस्करण, भाषा अनुवाद, और संदर्भ भाषा संसाधनों के विकास और अनुसंधान को समर्पित है। यह रिसर्च क्षेत्र भारतीय संस्थानों और अनुसंधान संगठनों में व्यापक रूप से विकसित हो रहा है।

भारत में कई संस्थान और अनुसंधान संगठन एमटी रिसर्च के क्षेत्र में गतिविधियों को संचालित कर रहे हैं। कुछ प्रमुख संस्थानों और संगठनों में शामिल हैं:

भारत में कई संस्थान और अनुसंधान संगठन एमटी रिसर्च के क्षेत्र में गतिविधियों को संचालित कर रहे हैं। कुछ प्रमुख संस्थानों और संगठनों में शामिल हैं:

1. संगणक विज्ञान और संदर्भ भाषा संस्थान (सीसीएल, आईआईटी बॉम्बे): सीसीएल एमटी रिसर्च के लिए मान्यता प्राप्त एक संस्थान है। यहां विशेषज्ञ अनुसंधानकर्ता मशीन अनुवाद के लिए नवीनतम तकनीकों, एल्गोरिदम्स, और मोडल्स का अध्ययन करते हैं। यह संस्थान मशीन अनुवाद, संदर्भ भाषा संसाधनों, और नैदानिक भाषा संदर्भ से संबंधित अनुसंधान कार्य करता है।
2. संगणक विज्ञान केंद्र (सीएसईआरपी, आईआईटी बॉम्बे): सीएसईआरपी मशीन अनुवाद और न्यूरल मशीन अनुवाद के लिए अनुसंधान कार्य करता है। यहां उन्नत तकनीकों के साथ एमटी के लिए अनुसंधान किया जाता है और विभिन्न भाषाओं के बीच अनुवाद की सुविधा प्रदान की जाती है।
3. भारतीय वैज्ञानिक तथा आण्विक अनुसंधान परिषद (सीएसआरपी): भारतीय वैज्ञानिक तथा आण्विक अनुसंधान परिषद भारतीय अनुसंधान

को बढ़ावा देने के लिए कार्य करता है। यहां मशीन अनुवाद, भाषा प्रसंस्करण, और नेत्रीय भाषा संबंधित अनुसंधान की गई है। इससे भारतीय भाषाओं के लिए उच्च-गुणवत्ता के अनुवाद संबंधी साधन विकसित हो रहे हैं।

इनके अलावा भारतीय प्रौद्योगिकी संस्थान (आईआईटी) और अन्य शोध संस्थानों में भी एमटी रिसर्च की गतिविधियां होती हैं। ये संस्थान विभिन्न एमटी तकनीकों, अनुशासन, और अनुसंधान के लिए सामग्री विकसित करते हैं जो मशीन अनुवाद क्षेत्र में उपयोग हो सकती हैं।

ये संस्थान और संगठनों के अलावा भारत में अन्य अकादमिक संस्थान और विश्वविद्यालय भी एमटी रिसर्च के क्षेत्र में गतिविधियों में सक्रिय हैं। इन संस्थानों में अनुसंधानकर्ता, शोधार्थी, और छात्र एमटी के विभिन्न पहलुओं पर काम करते हैं और नवीनतम तकनीकों और मोडल्स का अध्ययन करते हैं ताकि अनुवाद प्रक्रिया में सुधार किया जा सके।

सन्दर्भ सूची

1. Bandhyopadhyay Sivaji,2001. An Example Based MT System in News Item Domain from English to Indian Languages. Machine Translation Review 12.pp.7-10.
2. Bandyopadhyay,Sivaji. 2000. State and Role of Machine Translation in India. Machine Translation Review 11:25-27.
3. BharatiAkshar et al .1997. Anusaarak: Machine Translation in stages. A quarterly in Artificial Intelligence 10(3):22-25.July.
4. Bharati A,Chaitanya V and Sangal R.1995.Natural Language Processing: A Paninian Perspective. Prentice Hall of India.
5. Brown,P. et al.1990. A statistical approach to language Translation. Computational Linguistics. 16, 79-85.
6. Brown,Ralf D.1990,Example-based Machine Translation in the Pangloss System. In the proceedings of the COLING-96.
7. Hutchins,John. 1978. Machine Translation and Machine-aided Translation. Journal of Documentation. 34(2):119-159.

8. _____.1993.Latest Development in Machine Translation Technology:Beginning a New Era in MT Research.MT Summit.Kobe. Japan.
9. _____.2005.Towards a definition of Example Based Machine Translation.
10. Sato S. and Nagao M. 1990. Towards a Memory Based Translation. In COLING.Vol.3, pp.247-252.
11. Slocum, Jonathan.1984. Machine Translation: its history, current status and future prospects. COLING-84.
12. Somers, H.L.1990. Current Research in Machine Translation.
13. Tucker, Allen B.1987. Current Strategies in machine translation research and development. In Nirenberg's.(ed.).Machine Translation: Theoretical and methodological issues. Cambridge University Press.
14. Mitkov,Ruslin. 2002. The Oxford Handbook of Computational Linguistics. Oxford University Press.

मौलिकता/अनापत्ति प्रमाण-पत्र

यह प्रमाणित किया जाता है कि मेरे द्वारा लिखित पुस्तक मशीनी अनुवाद:परिचय(पुस्तक का नाम) मेरी मौलिक एवं अप्रकाशित कृति है। मैं इसके मुद्रण/प्रकाशन की सहर्ष अनुमति देता हूँ।

यदि डॉक्टर हरीसिंह गौर विश्वविद्यालय, सागर/विश्वविद्यालय अनुदान आयोग/शिक्षा मंत्रालय, भारत सरकार इस पुस् तक को मुद्रित/प्रकाशित करवाता है तो मुझे/ कोई आपत्ति नहीं होगी।

स्थान : सागर

लेखक के हस्ताक्षर एवं : ..डॉ.अभिज्ञान द्विवेदी
नाम

दिनांक :

15/07/23

