

भारत में मशीन अनुवाद का विकास

आम्रपाल शेंद्रे

shendre.amrapals.amrapal@gmail.com

असिस्टेंट प्रोफेसर, महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा

भारत एक बहुभाषी देश है जिसमें 22 आधिकारिक भाषाएँ हैं। वर्ल्ड एथ्नोलोग के अनुसार भारत में 448 भाषाएँ बोली जाती हैं। पीपल लिंग्विस्टिक्स सर्वे ऑफ़ इंडिया के अनुसार 780 भाषाएँ भारत में बोली जाती हैं। ऐसे में भारत में इतनी सारी भाषाएँ हैं, जिनको अनुवाद के माध्यम से आपस में जोड़ने की जरूरत है। यदि यह काम डिजिटली हो, तो इन भाषाओं को और सुरक्षित किया जा सकता है। वैसे भारत में डिजिटल अनुवाद या मशीन अनुवाद की शुरुआत पश्चिमी देशों की तुलना में बहुत बाद में हुई।

सन 1985 के आस-पास भारत में मशीन अनुवाद की शुरुआत मानी जाती है। यानि यह शुरुआत भारत जैसे बहुभाषी देश के लिए बिलकुल नई है। शायद यही वजह हो सकती है आज भी ऐसी बहुसंख्य भाषाएँ हैं जिनमें मशीन अनुवाद का कार्य नहीं हुआ है। बहरहाल, भारत में कंप्यूटर का प्रयोग ही 1980 के दशक में होने लगा था इसीलिए हो सकता है यहाँ मशीन अनुवाद की संकल्पना भी बाद में आई हो। यदि मशीन अनुवाद के तीसरे और चौथे वैश्विक दौर को देखें; तो लगभग यही से भारत में भी मशीन अनुवाद का कार्य शुरू हो चुका था। भारत में विकसित मशीन अनुवाद प्रणालियाँ और उनपर चल रहे कार्य को विस्तार से देखते हैं।

- 1. अनुसारका- 1995**—अनुसारका परियोजना की शुरुआत आई. आई. टी.कानपुर में हुई थी और बाद में यह आई. आई. आई. टी. हैदराबाद में स्थलांतरित किया गया। इस परियोजना का मुख्य उद्देश्य था कि भारतीय भाषाओंसे भारतीय भाषाओंमें अनुवाद का कार्य किया जाए। यह परियोजना भारत सरकार के सूचना प्रौद्योगिकी मंत्रालय के टी.डी.आई.एल. (Technology Development for Indian Languages-TDIL) विभाग और सत्यम कंप्यूटर प्रायवेट लिमिटेड के द्वारा वित्त पोषित है। इसमें स्रोत भाषा के रूप में तेलुगु, कन्नड़, मराठी, बांग्ला और पंजाबी से लक्ष्य भाषा हिंदी में अनुवाद की प्रक्रिया होती थी। यह प्रणाली सामान्य थी और मुख्य रूप से बच्चों की कहानियों का अनुवाद करती थी। इस प्रणाली का विकास सूचनाओं का संचय करने के लिए किया गया था। अनुसारका मशीन अनुवाद का मुख्य उद्देश्य मशीन अनुवाद का कार्य नहीं अपितु भारतीय भाषाओं तक पहुँचना है।
- 2. मंत्रा (MANTRA)- Machine Assisted Translation-** सन 1999 में मंत्रा मशीन अनुवाद प्रणाली का विकास सी- डैक पुणे में किया गया। इसके विकास में मुख्य रूप से डॉ. हेमंत दरबारी और डॉ. महेंद्र कुमार सी. पाण्डेय इनका महत्वपूर्ण योगदान है। इसका उपयोग भारत सरकार के राजभाषा विभाग के लिए किया जाता है, जो 'राजभाषा मंत्रा' नाम से अँग्रेजी से हिंदी मशीन अनुवाद प्रणाली है। इसमें सरकारी कामकाज, नियुक्तियां तथा तबादले, सरकारी आदेश और परिपत्रों के अँग्रेजी से हिंदी अनुवाद की व्यवस्था की गयी है। साथ ही कार्यालयी भाषा को हिंदी में अनूदित करने के लिए भी इसका प्रयोग हो रहा है, इसका प्रयोग भारत सरकार केराज्यसभा के लिए भी किया गया है। इस मशीन अनुवाद प्रणाली में पेंसिल्वेनिया विश्वविद्यालय द्वारा विकसित टैग (Tree Adjoining Grammar-TAG) और एलटैग (Lexicalize Tree Adjoining Grammar- LTAG) का प्रयोग कर अनुवाद का कार्य किया जाता है। इसके व्याकरण का स्वरूप विशेष रूप से कार्यालयी क्षेत्र के वाक्यों की संरचनाओं का विश्लेषण कर जेनरेट करने के लिए तैयार किया गया है। साथ ही क्षेत्र विशेष के अर्थों के आधार पर इसके शब्दकोश को निर्धारित किया गया है। फ़िलहाल यह प्रणाली अँग्रेजी, हिंदी के अलावा और भी भारतीय भाषाओंमें

अनुवाद करती है। जिनमें अँग्रेजी- बांग्ला, अँग्रेजी- तेलुगु, अँग्रेजी- गुजराती और हिंदी- अँग्रेजी शामिल है, साथ ही यह भारतीय भाषाओं से भारतीय भाषाओं जैसे हिंदी- बांग्ला और हिंदी – मराठी में भी अनुवाद का कार्य करती है।

3. **मैट (MAT-2002)** –इस मशीन सहायक अनुवाद प्रणाली का विकास के. मूर्ति ने सन 2002 में किया। यह प्रणाली अँग्रेजी से कन्नड़ भाषा में अनुवाद का कार्य करती है। जो कन्नड़ भाषा के लिए रूपिमिक विश्लेषक और संस्लेषक का प्रयोग करती है। इसमें अँग्रेजी वाक्यों को पार्स करने के लिए यूनिवर्सल क्लाज स्ट्रक्चर ग्रामर (UCSG) का उपयोग किया गया है और द्विभाषिक कोश से क्लाज के रिलेशनशिप उसके प्रकार और संख्या का प्रयोग करके उसे कन्नड़ भाषा के क्लाज के साथ मिलाकर आउटपुट के रूप में प्रस्तुत किया जाता है। अनूदित किए हुए पाठ को एडिट करने के लिए पोस्ट एडिटिंग टूल का प्रयोग किया जाता है। मैट अनुवाद प्रणाली के पूरी तरह से स्वयंचलित अनुवाद की शुद्धता 40-60 प्रतिशत है। यह प्रणाली एक क्षेत्र विशेष सरकारी सर्कुलर का अनुवाद करती है।
4. **शक्ति- (2003)**-‘शक्ति’ मशीन अनुवाद प्रणाली का विकास आर. मूना, पी. रेड्डी, डी. एम. शर्मा और आर. संगल के संयुक्त प्रयासों का प्रतिफल है। यह अनुवाद प्रणाली अँग्रेजी से किसी भी भारतीय भाषा में अनुवाद करती थी। इसमें नियम आधारित उपागम और सांख्यिकीय उपागम दोनों का प्रयोग किया गया था। इस प्रणाली में अलग- अलग तरह के 69 मॉड्यूल्स थे। जिसमें 9 मॉड्यूल्स को स्रोत भाषा (अँग्रेजी) का विश्लेषण करने के लिए इस्तेमाल किया जाता था और 24 मॉड्यूल्स को द्विभाषिक कार्य करने के लिए प्रयोग में लाया जाता था। बाकी मॉड्यूल्स का उपयोग लक्ष्य (भारतीय) भाषा को जेनरेट करने के लिए किया जाता था।
5. **अँग्रेजी- तेलुगु मशीन अनुवाद प्रणाली- (2004)**- अँग्रेजी –तेलुगु मशीन अनुवाद प्रणाली का विकास शिवाजी बंदोपाध्याय इनके नेतृत्व में हुआ। इस प्रणाली में अँग्रेजी- तेलुगु के 42000

शब्दों का समावेश किया गया है। यह प्रणाली अँग्रेजी भाषा के जटिल वाक्यों को बखूबी अनूदित करती थी।

6. मात्रा (MaTra) 2004- इस अनुवाद प्रणाली का विकास आर. अनन्तकृष्णन, एम. कविता, जे.जे. हेडगे, चंद्रशेखर, रितेश शाह, सावनी बड़े और एम. शशिकुमार इनके नेतृत्व में हुआ है। यह प्रणाली अंतरण उपागम का प्रयोग करती थी। एन.सी.एस.टी. मुंबई समाचार कथा- विवरणों के अँग्रेजी – हिंदी अनुवाद के लिए इस मशीन अनुवाद प्रणाली का विकास किया है। यह प्रणाली वाक्यों के अर्थपरक विश्लेषण पर आधारित है। औरकभी- कभी संदिग्धता को दूर करने के लिए स्वतः घोषित पद्धति (Heuristic Method) का भी प्रयोग किया जाता था। यह प्रणाली मुख्य रूप से समाचारपत्र और तकनीकी लेखों का अनुवाद करती थी। इस प्रणाली में पाठ को अलग-अलग वर्गों में बाँटने के लिए पाठ वर्गीकरण (Text categorization) का इस्तेमाल किया जाता था, जो स्रोत भाषा के पाठ पर प्रक्रिया करने से पहले पाठ का वर्गीकरण करता था- जैसे समाचार के प्रकार (आर्थिक, राजकीय, आपराधिक आदि)। इस प्रणाली में अलग-अलग क्षेत्र के लिए अलग-अलग शब्दकोशों का प्रयोग किया जाता था, जैसे राजकीय, आर्थिक आदि। समाचार के प्रकार के आधार पर उसके लिए योग्य शब्दकोश का चयन किया जाता था। इस प्रणाली में अँग्रेजी के जटिल वाक्यों को बड़ी ही आसानी से सरल वाक्यों में बदल दिया जाता था। इन सामान्य वाक्यों का संस्लेषण कर आगे लक्ष्य भाषा हिंदी के वाक्य जेनरेट किए जाते थे।

7. पंजाबी से हिंदी मशीन अनुवाद प्रणाली- 2007-2008-जी. एस. जोसन औरजी. एस. लेहल, इन्होंने इस मशीन अनुवाद प्रणाली का विकास किया जो प्रत्यक्ष मशीन अनुवाद प्रणाली पर आधारित है। इसमें जिन मॉड्यूल्स का प्रयोग किया गया था, उनमें पूर्व-संसाधन, पंजाबी- हिंदी शब्दकोश का प्रयोग करके शब्द- से शब्द स्तर पर अनुवाद, रूपिमिक विश्लेषण, शब्द स्तर पर संदिग्धता निवारण (Word Sense Disambiguation), लिप्यंतरण और पश्चसंसाधन, इस प्रणाली से अनुवाद की शुद्धता 90.67 प्रतिशत थी। सन 2010 में इसी प्रणाली का संस्करण वेब आधारित किया है।

8. **अँग्रेजी – कन्नड़ मशीन आधारित अनुवाद प्रणाली- (2009)** के. नारायण मूर्ति ने अँग्रेजी- कन्नड़ मशीन अनुवाद प्रणाली का विकास किया। इस प्रणाली का विकास हैदराबाद विश्वविद्यालय में किया गया। यह प्रणाली अंतरण विधि पर आधारित थी और सरकारी सर्कुलर का अनुवाद करती थी। इसमें युनिवर्सल क्लाज स्ट्रक्चर ग्रामर का प्रयोग किया गया था। यह प्रणाली कर्नाटक सरकार द्वारा वित्तपोषित थी।
9. **संपर्क- भारतीय भाषाओंमें अनुवाद प्रणाली-(2009)** भारत के प्रसिद्ध 11 संस्थानों ने मिलकर भारतीय भाषा से भारतीय भाषाओंमें अनुवाद (Indian language to Indian language machine translation) करने के लिए इस प्रणाली का विकास किया। यह प्रोग्राम भारत सरकार के इलेक्ट्रॉनिक्स एवं सूचना प्रौद्योगिकी मंत्रालय के विभाग टी.डी,आई.एल. द्वारा वित्तपोषित था। इसमें पाणिनीयन कंप्यूटेश्रल व्याकरण का प्रयोग कर भाषा का विश्लेषण किया जाता था और उसे मशीन लर्निंग के साथ जोड़ा जाता था। इन संस्थाओं ने 9 भारतीय भाषाओंके लिए टेक्नोलॉजी का विकास किया था जिससे 18 भाषा के युग्मों में अनुवाद का कार्य होता था।
10. **आंग्लभारती (ANGLABHARTI-2001)** – आई. आई. टी. कानपुर ने इस मशीन अनुवाद प्रणाली का विकास किया है। जिसमें मुख्य भूमिका के रूप में प्रो. आर. एम. के. सिन्हा का महत्वपूर्ण योगदान है। यह प्रणाली स्वास्थ्य सेवाओं में प्रयोग की जाने वाली भाषा का अनुवाद करती है। इसमें उदाहरण-आधारित अनुवाद प्रणाली का उपयोग किया गया है। इस प्रणाली का वेब संस्करण भी उपलब्ध है। यह अँग्रेजी से भारतीय भाषाओंमें अनुवाद का कार्य करती थी। इसमें स्यूडो- अंतराभाषाउपागम का प्रयोग किया गया है। अंतराभाषा की वजह से अँग्रेजी से एक से अधिक भारतीय भाषाओंमें अनुवाद करना आसान हो गया है और अँग्रेजी से भारतीय भाषाओं में अनुवाद करने के लिए प्रत्येक भाषा के लिए अलग से अनुवाद प्रणाली का विकास करने की जरूरत नहीं पड़ती है। स्यूडो अंतराभाषा के नियम सभी भाषाओं के लिए लागू होते हैं। इसमें अँग्रेजी भाषा (स्रोत भाषा) का एक बार ही विश्लेषण किया जाता है। इसके बाद उसे प्रत्येक भारतीय भाषाओं के लिए स्यूडो मध्यवर्ती नियमों (Pseudo Lingua for Indian

Languages-PLIL)के द्वारा एक संरचना तैयार की जाती है। बाद में पी.एल.आई.एल. उन संरचनाओं को भारतीय भाषाओं में बदलकर पाठ जेनरेशन की प्रक्रिया द्वारा जेनरेट किया जाता है। पी.एल.आई.एल. जेनरेट करने के लिए 70% और पाठजेनरेट करने के लिए 30% प्रयास करना पड़ता है। इस प्रणाली में 90% कार्य मशीन से होता है और 10% पश्च संपादन का कार्य मनुष्य के द्वारा किया जाता है। इस मशीन अनुवाद प्रणाली का प्रयोग स्वास्थ्य सेवाओं के क्षेत्र में किया जाता है।

11. **यू.एन.एल.- आधारित अँग्रेजी- हिंदी मशीन अनुवाद प्रणाली- (2001)** इस प्रणाली के विकास में पुष्पक भट्टाचार्य, एस. देव, जे. पारीख इनका महत्त्वपूर्ण योगदान रहा है। यह प्रणाली आई. आई. टी. मुंबई में विकसित की गयी। यह प्रणाली अंतराभाषा की संरचना में यू.एन.एल.(Universal Networking Language) का प्रयोग करती थी। यू.एन.एल. एक अंतर्राष्ट्रीय परियोजना है जिसका उद्देश्य मनुष्य द्वारा बोली जाने वाली सभी भाषाओंके लिए एक अंतराभाषा संरचना का निर्माण करें। यू.एन.एल. परियोजना में आई. आई. टी. मुंबई शामिल है, जो अँग्रेजी- हिंदी, हिंदी- यू.एन.एल., यू.एन.एल.- हिंदी, अँग्रेजी-मराठी और अँग्रेजी- बांग्ला आदि। भाषा को यू.एन.एल. तकनीक का प्रयोग करके विकास किया जा रहा है।
12. **आंग्लहिंदी (AnglaHindi -2003)-** इस अनुवाद प्रणाली को आंग्ल भारती से लिया गया है। इसके विकास में मुख्य रूप से डॉ. आर. एम. के. सिन्हा और ए. जैन इनका सहयोग है। इसका विकास अँग्रेजी से भारतीय भाषाओंमें अनुवाद के लिए किया गया था। यह प्रणाली स्यूडो-अंतराभाषा नियम- आधारित प्रणाली थी। आंग्ल भारती के सारे मॉड्यूल्स का प्रयोग इस मशीन अनुवाद प्रणाली में होता था। इस प्रणाली से किए गए अनुवाद की शुद्धता 90% थी।
13. **अनुभारती- (1995, 2004)** अनुभारती मशीन अनुवाद प्रणाली का विकास संकर उपागम का प्रयोग करके किया गया। इसमें उदाहरण आधारित और कॉर्पस आधारित उपागम का प्रयोग किया गया। साथ ही कुछ बुनियादी व्याकरणिक विश्लेषण का उपयोग करके इसको बनाया गया।

उदाहरण आधारित उपागम का इस्तेमाल मनुष्य के सीखने की प्रक्रिया के ज्ञान को भूतकाल के अनुभव से संग्रहित करना और भविष्य में उसका उपयोग करने के लिए किया जाता था। इस प्रणाली में पुराने उदाहरण आधारित मशीन अनुवाद (Example Based Machine Translation) उपागम को सुधारा गया जिससे कि बड़े उदाहरणों के ऊपर की निर्भरता को कम किया जा सके। वाक्यों के समूहों को कच्चे उदाहरणों से पहचान कर उन्हें स्रोत भाषा की संरचना के आधार पर मैप किया जाता था। यह कार्य स्रोत भाषा के वाक्य कोटियाँ और आर्थी टैग के आधार पर किया जाता था। आंग्ल भारती और अनुभारती दोनों की संरचना अपने शुरुआती दौर से काफी बदल गयी थी। सन 2004 में इनका नाम अनुभारती-II और आंग्ल भारती-II किया गया।

14. **बांग्ला- हिंदी मशीन अनुवाद प्रणाली- (2009)**- एस. चटर्जी, डी. सरकार, एस. रॉय और ए. बासु ने इस मशीन अनुवाद प्रणाली के विकास में महत्वपूर्ण योगदान दिया है। इनके ही नेतृत्व में इस प्रणाली का विकास हुआ है। इस प्रणाली में सांख्यिकीय मशीन अनुवाद और नियम आधारित मशीन अनुवाद इन उपागमों का प्रयोग किया गया था। इस संकर प्राणाली का प्रदर्शन बहुत अच्छा था और इसका ब्लू स्कोर 0.2275 भी काफी अच्छा था।

15. **अनुवाद ANUBAAD (2000, 2004)**-इस प्रणाली का विकास शिवाजी बंदोपाध्याय इनके नेतृत्व में हुआ। यह प्रणाली अँग्रेजी से बांग्ला भाषा में अनुवाद करती थी, जो उदाहरण आधारित उपागम पर आधारित है। इसमें इनपुट के रूप में अँग्रेजी समाचार के शीर्षकों को दिया जाता है। और उसके सटीक मेल को उदाहरण आधारित उपागम में ढूँढा जाता है। अगर मेल नहीं मिलता है तो शीर्षक को टैग किया जाता है और उस टैग शीर्षक को जनरलाइज टैग उदाहरण आधारित डाटा में खोजा जाता है। टैग डाटा में मेल होने के बाद संस्लेषण कर बांग्ला भाषा में पाठ को जेनरेट किया जाता है। अगर टैग डाटा में भी मेल नहीं मिला तो पदबंधीय उदाहरण आधारित डाटा का उपयोग करके लक्ष्य भाषा के वाक्य को जेनरेट किया जाता है। यदि अभी भी शीर्षक का अनुवाद नहीं होता है, तो शब्द के स्तर पर जा कर स्वानुभाविक नीति का प्रयोग किया जाता है जिसमें शब्द के क्रम के

अनुसार लक्ष्य भाषा में अनुवाद किया जाता है। समाचार शीर्षकों के अनुवाद के लिए संबंधित शब्दकोशों का सहारा लिया जाता है।

16. **वासानुवाद- (VAASAANUBAAD), (2002)**- इस प्रणाली का विकास करने में के. विजयानंद, आई. एस. चौधरी और पी. रत्ना इनका विशेष योगदान है। यह प्रणाली बांग्ला से असमी भाषा में ऑटोमेटिक अनुवाद करती थी। इसके अनुवाद का क्षेत्र है, समाचार के पाठ और इस मशीन अनुवाद प्रणाली में उदाहरण आधारित उपागम का प्रयोग किया गया है। यह प्रणाली बांग्ला से असमी में वाक्य स्तर पर अनुवाद का कार्य करती है। इसमें पूर्व-संसाधन और पश्च-संसाधन का कार्य करना पड़ता था। द्विभाषिक कार्पस का निर्माण वास्तविक उदाहरणों को लेकर मनुष्य के द्वारा स्यूडो कोड का प्रयोग करके किए जाते थे। अनुवाद की गुणवत्ता अच्छी आने के लिए बड़े- बड़े वाक्यों को विराम चिन्हों का उपयोग करके तोड़ा जाता था।

17. **शिवा और शक्ति (2003)** – कार्नेगी मेलॉन विश्वविद्यालय (Carnegie Mellon University, USA), भारतीय विज्ञान संस्थान, बंगलौर और अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान, हैदराबाद के सहयोग से विकसित यह अनुवाद प्रणाली अपने वेबसाइट <http://shakti.iiit.net> पर चलती है। इसका नया वर्जन 18 अप्रैल 2005 को आरंभ किया गया, जो तीन लक्ष्य भाषाओं हिंदी, मराठी और तेलुगु के लिए कार्य करता है। इसमें उदाहरण आधारित उपागम, नियम आधारित और सांख्यिकीय उपागम का प्रयोग किया गया। साथ ही इस प्रणाली में कुछ मॉड्यूल्स के द्वारा आर्थी सूचनाएं भी दी गयी थी।

18. **हिंगलिश (Hinglish- 2004)**- इस मशीन अनुवाद प्रणाली का विकास सिन्हा और ठाकुर इनके नेतृत्व में किया गया। यह मशीन अनुवाद शुद्ध हिंदी से शुद्ध अंग्रेजी में अनुवाद करने के लिए किया गया था। इसमें आंग्लभारती-II (अंग्रेजी से हिंदी) और अनुभारती-II (हिंदी से अंग्रेजी) दोनों को सम्मिलित किया गया था। इस अनुवाद प्रणाली की शुद्धता 90% से अधिक है।

19. **अँग्रेजी से भारतीय भाषाओंमें मशीन अनुवाद- (E-ILMT) 2006-** E-ILMT मशीन ट्रांसलेशन सिस्टम पर्यटन और स्वास्थ्य संबंधित क्षेत्र के लिए विकसित किया गया। इसमें अँग्रेजी से 8 भारतीय भाषाओं- हिंदी, मराठी, बंगाली, उर्दू, तमिल, उड़िया, गुजराती और बोडों में अनुवाद होता था। इसके विकास में भारत के 12 संस्थानों, 1. सी-डैक मुंबई, 2. आई.आई.आई.टी हैदराबाद, 3. आई.आई.एस.सी. बैंगलोर, 4. आई.आई.टी. मुंबई, 5. जाधवपुर विश्वविद्यालय कोलकत्ता, 6. आई.आई.आई.टी. अलाहाबाद, 7. उत्कल विश्वविद्यालय- भुवनेश्वर, 8. अमृता विश्वविद्यालय- कोइम्बतूर, 9. वनस्थली विद्यापीठ, 10. उत्तर महाराष्ट्र विद्यापीठ – जलगाँव, 11. धरमसिंह देसाई विश्वविद्यालय-नडियाड, और 12. नॉर्थ इस्टर्न हिल विश्वविद्यालय- शिलॉन्ग के साथ सी-डैक पुणे मुख्य भूमिका में था। यह मशीन अनुवाद प्रणाली संकर विधि पर आधारित है। जिसमें तीन अनुवाद इंजिन (EBMT, SMT और TAG) एक साथ काम करते हैं। (TAG) वृक्ष संलग्न व्याकरण आधारित इंजिन और (SMT) सांख्यिकीय मशीन अनुवाद इंजिन पार्सिंग और जेनरेशन का काम करते हैं। इसके आलावा और भी भाषावैज्ञानिक उपकरणों का प्रयोग इसमें किया जाता है। जैसे- (Input format extraction, morph analyser and synthesize, Named Entity Recognizer, POS tagger, Word sense Disambiguator, Semantic TAG Parser) आदि का भी इसमें उपयोग होता है। इस परियोजना को सूचना प्रौद्योगिकी विभाग भारत सरकार द्वारा वित्तपोष मिलाता था। C-DAC मुंबई ने सांख्यिकीय मशीन ट्रांसलेशन के लिए सांख्यिकीय मॉडल का विकास किया है, जो अँग्रेजी से मराठी, हिंदी और बंगाली के लिए था। इसका ट्रेनिंग कार्पस 5000 वाक्यों का है जिसमें 800 वाक्य टेस्ट और ट्यून करने के लिए है।

मशीन अनुवाद के क्षेत्र में उपर्युक्त कार्य भारत में चल रहे हैं।

निष्कर्षतः- यह कह सकते हैं कि मशीन अनुवाद की संकल्पना नई है, यह संकल्पना लगभग 50-60 साल पुरानी है। अभी इस क्षेत्र में बहुत काम होना बाकि है। खास कर भारत के संदर्भ में। क्योंकि भारत में यह संकल्पना पश्चिमी देशों के काफी बाद में यानि सन 1985 के बाद आई इसीलिए यह संकल्पना भारत जैसे देश के लिए नई है। भारत एक बहुभाषी देश है, जहाँ बहुत सी भाषाएँ बोली जाती है, जिनमें कुछ भाषाएँ

लुप्त होने की कगार पर है। यदि इन भाषाओंका जतन और संवर्धन करना है, तो इनको डिजिटल रूप में संगृहीत करने की आवश्यकता है। इन भाषाओंको एक डिजिटल प्लैटफार्म मिल जाएगा साथ ही मशीन अनुवाद के कार्य हेतु इनका आगे उपयोग भी किया जा सकेगा। इस तरह लुप्त हो रही भाषाओंको बचाया जा सकता है। ऐसी बहुत-सी भाषाएँ हैं जिनमें भारतीय भाषाओंसे भारतीय भाषाओंमें अनुवाद अब तक नहीं हुआ है। इस तरह उनका संवर्धन और संगोपन भी हो जाएगा। प्रस्तुत शोध आलेख का उद्देश्य है, मशीन अनुवाद के क्षेत्र को बढ़ाने के लिए सहयोग करना और इस क्षेत्र में रुचि रखने वाले पाठको एवं शोधार्थियों के लिए मशीन अनुवाद की विधियाँ और उसके इतिहास की जानकारी देना है।

संदर्भ ग्रंथ –

1. सेठी, हरीशकुमार, (2013) ई-अनुवाद और हिंदी; किताबघर प्रकाशन, नई दिल्ली.
2. त्रिपाठी, अरिमर्दन कुमार, (2008) 'हिंदी के भाषा प्रौद्योगिकीय उपकरण', राष्ट्रभाषा.
3. दासठाकुर, (2008) 'मशीनी अनुवाद : विधियाँ एवं प्रविधियाँ', गवेषणा, केंद्रीय हिंदी संस्थान, आगरा.
4. भाटिया कैलाश चन्द्र, (1996) 'कंप्यूटर अनुवाद की संभावनाएँ', हिंदी : विकास और संभावनाएँ, सूचना और प्रसारण मंत्रालय, भारत सरकार.
5. विकास, ओम (2005) 'हिंदी के विकास में टेक्नोलॉजी का योगदान', गवेषणा, केंद्रीय हिंदी संस्थान, आगरा.
6. सिंह, सूरजभान, (2005) सूचना-प्रौद्योगिकी और भाषाक्रांति, राजभाषा भारती.
7. V. Garje and G.K. Kharate, Survey of machine translation system in India, International Journal on Natural Language Computing (IJNLC) Vol. 2, No.4, October2013.
8. <http://sanskrit.jnu.ac.in/shmt/index.jsp>
9. <http://www.cse.iitb.ac.in/~pb/indtrend2.htm>
10. <http://www.tdil-dc.in/>



11. <http://www.cdac.in>

Citation: शेंद्रे, आम्रपाल (2017). भारत में मशीन अनुवाद का विकास, HindiTech: A Blind Double Peer Reviewed Bilingual Web-Research Journal, 8 (5), 56-66. URL: <https://hinditech.in/bharat-men-mashin-anuvad-ka-vikas/>