



मशीनी अनुवाद : इतिहास, वर्तमान एवं भविष्य

करुणा निधि

karunanidhi155@gmail.com

भाषा-प्रौद्योगिकी में पी-एच. डी., फिलहाल इंग्लैंड के एक संस्थान में अध्यापनरत

वर्तमान मशीनी युग समाज को तार्किक और वैज्ञानिक दृष्टिकोण प्रदान कर रहा है। यद्यपि मशीनीकरण ने संपूर्ण ज्ञान-जगत को चिंतन की नई धारा प्रदान की तथापि ज्ञान की किसी भी विधा का मशीनीकृत रूप इसी प्रस्थान बिंदु से देखना उचित होगा। संसाधन और समय की कम लागत में विपुल ज्ञान को आत्मसात करना एवं उसे विभिन्न समुदायों में प्रेषित करने की होड़, संपूर्ण मानव-सभ्यता के विकास में विस्तार करती जा रही है, जिसमें एक रोचक भूमिका अनुवाद की भी है, जो विश्व बाजार को एकीकृत करने में अग्रसर है। इस कार्य को सार्थक रूप देने का प्रयास पिछले कई दशकों से हो रहा है, जिसमें सफलता भी मिली है, जिसका उद्देश्य भारत जैसे बहुभाषी और बहुसांस्कृतिक देश में भाषिक बंधनों को तोड़कर संप्रेषण व्यवस्था स्थापित करना है। अनुवाद एक भाषा के वाचिक और लिखित पाठ को अन्य भाषा के वाचिक और लिखित पाठ में ढालने या रूपांतरित करने की प्रक्रिया है, जिसे मानव अनुवादक लक्ष्य और स्रोत भाषाओं की संरचना, अभिव्यक्ति, संस्कृति, शैली और व्याकरण के नियमों द्वारा सींचता है। ऐसे में मशीन का आगमन एक असहज परंतु अपेक्षित घटना कही जा सकती है। मानवानुवाद के साथ मशीनी अनुवाद की नूतन संकल्पना ने संपूर्ण सूचना और संप्रेषण जगत में अभूतपूर्व क्रांति ला दी है, जिसका परिणाम आज अनेक भाषिक अनुप्रयोगों में दृष्टिगोचर है। परंतु मशीनी अनुवाद की नूतन संकल्पना का उद्गम स्रोत अत्यंत पुराना है।

मशीनी अनुवाद का इतिहास

मशीनी अनुवाद की संकल्पना को 'अभिव्यक्ति के मशीनीकरण' की दृष्टि से समझना अधिक प्रासंगिक होगा, जो लेखन पद्धति, छपाई मशीन, पर्सनल कंप्यूटर जैसे मार्गों से गुजरकर आज इंटरनेट की दुनिया में प्रवेश कर चुकी है। 17वीं शताब्दी के प्रारंभ में रेने देस्कर्टेस (René Descartes) और लैब्रिज़ (Leibniz) नामक दर्शनशास्त्री ने भिन्न भाषिक संरचनाओं की भिन्न भाषिक अभिव्यक्ति को एक ही भाषिक प्रतीक द्वारा प्रस्तुत करने के लिए संख्यात्मक और अभियांत्रिकी दृष्टि से एक सार्वभौमिक भाषा का प्रस्ताव रखा। 22 जुलाई 1933 को जॉर्ज आर्ट्सरौनि (Georges Artsrouni), फ्रेंच विद्वान द्वारा पेपर टेप का प्रयोग कर 'मेकैनिकल ब्रेन' नामक अनुवाद यंत्र के लिए एक स्वचालित द्विभाषी कोश बनाया गया। ट्रोयंस्किइ (Troyanskii) और स्मिर्नोव (Smirnov) नामक रूसी विद्वानों ने पहली बार मशीन से अनुवाद करने वाले यंत्र का प्रस्ताव दिया, जो शब्द धातु शृंखला को अन्य भाषाओं में रूपांतरित करता है, जिसका प्रदर्शन 31 जुलाई 1944 को मास्को में कर दिया गया। इस यंत्र की क्षमता मात्र एक हजार शब्द की थी। इसी समय ट्रोयंस्किइ ने एक और प्रस्ताव दिया, जिसमें द्विभाषी कोश के साथ-साथ भाषा (एसपरंतो आधारित) में व्याकरणिक भूमिकाओं को व्यवहृत करने की विधि भी सम्मिलित थी। इन्होंने अनुवाद प्रक्रिया के तीन चरणों का उल्लेख किया, जहाँ मशीन की भूमिका केवल दूसरे चरण पर थी। 1939 में बेल लैब द्वारा न्यूयार्क विश्व मेला (Fair) में प्रथम इलेक्ट्रॉनिक वाक्-संश्लेषण यंत्र का प्रदर्शन हुआ। जुलाई 1949, न्यूयार्क में एन्ड्रेव. डी. बूथ (Andrew D. Booth) और वारेन वीवर (Warren Weaver) के बीच एक सहमति बनी, जिसमें इलेक्ट्रॉनिक कंप्यूटर का अनुवाद के लिए प्रयोग होने के लिए सुझाव दिये गये। इस पत्रक (Memorandum) के अनुसार सूचना के सिद्धांतों के आधार पर प्राकृतिक भाषा में निहित सार्वभौम तत्वों पर विचार किया गया। 1951 में बार हिलेल (Bar-Hillel) ने इस ज्ञापन के तुरंत बाद एक सर्वेक्षण प्रस्तुत किया। 7 से 20 जून 1952 तक एम. आई. टी. (Massachusetts Institute of Technology : MIT) में बार हिलेल द्वारा मशीनी अनुवाद पर एक सम्मेलन करवाया गया, जिसमें 18 शोधार्थियों ने भागीदारी की। इसी समय एम.आई.टी. ने 'कोमिट' नामक प्रोग्रामिंग भाषा का विकास किया। 7 जनवरी 1954 को आई.बी.एम. (International Business Machine : IBM), न्यूयार्क के मुख्य कार्यालय में जॉर्जटाउन-आई.बी.एम. द्वारा सिस्ट्रॉन (SYSTRAN) नामक

मशीनी अनुवाद प्रणाली का प्रदर्शन किया गया, जिसकी क्षमता 250 शब्दों और 6 व्याकरणिक नियमों का प्रयोग कर रसायनशास्त्र (Chemistry) के 49 रूसी वाक्यों को अँग्रेजी में अनुवाद करना था। शब्दानुवाद की प्रमुखता के कारण यह किसी खिलौने के समान प्रतीत होने लगा। इस प्रवृत्ति के कारण कालांतर में इसे 'Toy' सिस्टम कहा जाने लगा। आई.बी.एम. के सुप्रसिद्ध गणितज्ञ डॉ. वाटेन ने यह अनुभव किया कि कंप्यूटर की तार्किक प्रणाली का उपयोग भाषिक विश्लेषण, संश्लेषण और संसाधन के लिए किया जा सकता है। शुरू में यह कार्य अक्षरों और शब्दों तक ही सीमित रहा। संभावनाओं को देखते हुए सरकार द्वारा इस क्षेत्र में अपार धनराशि का प्रावधान किया गया, जिससे 1955 में संयुक्त राष्ट्र, रूस और पश्चिमी यूरोप में पर्याप्त कार्य हुए। 1958 में 50 संस्थानों से 340 प्रतिभागियों की उपस्थिति में मशीनी अनुवाद पर प्रथम सोवियत सम्मेलन का आयोजन हुआ। 1959 में यू.एस.एफ. के लिए आई.बी.एम. प्रणाली की शुरुआत हुई और सिस्ट्रॉन प्रणाली का प्रथम लोक प्रदर्शन हुआ। 1960 में कैलिफोर्निया विश्वविद्यालय द्वारा लॉस एंजलिस (Los Angeles) में मशीनी अनुवाद पर एक राष्ट्रीय परिचर्चा (Symposium) का आयोजन हुआ, जिसमें बार हिलेल द्वारा अनुवाद की गुणवत्ता पर टिप्पणी की गई। 1963 में सिस्ट्रॉन प्रणाली का संस्थापन कार्य हुआ। इन सफलताओं के कारण इस क्षेत्र से अपेक्षाएँ बढ़ने लगी, पर करोड़ों रुपये के व्यय के बावजूद अनुवाद की गुणवत्ता में कमी के कारण इस पर आक्षेप लगने लगे। फलतः इसकी प्रगति का लेखा-जोखा लेने के लिए 1964 में संयुक्त राष्ट्र सरकार द्वारा ALPAC (Automatic Language Processing Advisory Committee) नामक सात वैज्ञानिकों की एक समिति का गठन हुआ, जिसने अपने 1966 के प्रतिवेदन में मुख्य रूप से यह कहा कि वर्तमान स्थिति में मशीनी अनुवाद संभव नहीं। इसमें आधारभूत भाषा-विश्लेषण सिद्धांतों का अभाव है, इसलिए मशीनी अनुवाद गुणवत्ता की दृष्टि से चिंताजनक है। परिणामतः शोध एवं विकास समूह ने अमेरिका और अन्य संलग्न देशों में वित्तीय सहायता देने पर रोक लगा दी। धन की अनुपलब्धता में लगभग दस वर्षों तक इस क्षेत्र में विकास की गति धीमी हो गई। इसलिए दस वर्षों का यह समय (1966-76) अंधकार युग के नाम से जाना जाता है। फिर भी कई देशों में न्यूनाधिक कार्य चलते रहे। इसी समय एलिजा (ELIZA) नामक पहली बातचीत करने वाली प्रणाली का विकास हुआ, जो मानव के बातचीत का अनुकरण करती थी। 1967 में एल. ई. बॉम (L.E.

Baum) और उनके साथियों द्वारा रक्षा विश्लेषण संस्थान (Institute of Defense Analysis : IDA) प्रिंसटन (Princeton), न्यूजर्सी (New Jersey) में हिडेन मार्कोव मॉडल (Hidden Markov Model: HMM) का विकास किया गया, जो सतत वाक् अभिज्ञान (Continuous Speech Recognition) की रीढ़ की हड्डी साबित हुई। 1968 में जार्जटाउन विश्वविद्यालय के संस्थापक पेटर टोमा (Peter Toma) (भाषाविज्ञानी) ने पहली मशीनी अनुवाद कंपनी 'लैंग्वेज ऑटोमेटेड ट्रांसलेशन सिस्टम एण्ड इलेक्ट्रॉनिक कम्यूनिकेशन' (Language Automated Translation System and Electronic Communications : LATSEC) की शुरुआत की। रैंड कॉरपोरेशन, कैलीफोर्निया में प्रो. डेविड डी. हेज के द्वारा डिपेंडेंसी विश्लेषक (Dependency Analyzer) नामक एक प्रोग्राम तैयार किया गया, जो पद व्याख्या करने में सफल हुआ। 1969 में न्यूयार्क के मिडलटाउन (Middletown) में चार्ल्स बीर्न (Charles Byrne) और बर्नार्ड स्कॉट (Bernard Scott) ने मशीनी अनुवाद प्रणाली लोगोस (LOGOS) का विकास किया। 1973 में ल्यूरनर (LUNAR) नामक एक प्रश्न विशेषज्ञ प्रणाली (Question Expert System) विकसित की गई। 1978 में आर्पा (Advance Research Project Agency: ARPA) के नेटवर्क स्पीच कंप्रेशन (Network Speech Compression : NSC) परियोजना ने इंटरनेट पर प्रथम वाचिक शब्द का प्रसारण किया। इस काल में और भी कार्य हुए, जैसे- नियम आधारित अभिगम : अंतरभाषिक और अंतरण, मौसम संबंधी सूचनाओं के प्रसारण के लिए 'टॉममेटो' (TAUMMATEO) प्रणाली का विकास, जो 30 मिलियन शब्दों का अनुवाद सहजता से करता था। 1980 में पैन अमेरिकन स्वास्थ्य संगठन प्रणाली (Pan American Health Organization : PAHO) द्वारा स्पेनैम (Spanish-English Machine Translation : SPANAM) एवं इंगस्पेन (English- Spanish : ENGSPAN) प्रणाली का संचालन किया गया। इसके समानांतर भाषाविज्ञान के क्षेत्र में भी कई व्याकरणिक मॉडलों का विकास हुआ, जैसे- रूपांतरण प्रजनक व्याकरण, व्यवस्थापरक व्याकरण, स्तरीकृत व्याकरण, कारक व्याकरण, संबंधपरक व्याकरण आदि। इन व्याकरणों के कारण समय-समय पर इन शोधों में लगे विशेषज्ञों को उत्साह मिलता रहा। 1982 में जेम्स के. बेकर (James K. Baker) द्वारा वाक् अभिज्ञान प्रौद्योगिकी के व्यावसायीकरण के लिए ड्रेगन



प्रणाली का विकास किया गया। 1983 में माइक्रोकंप्यूटर के लिए प्रथम मशीनी अनुवाद सॉफ्टवेयर, 'स्वचालित भाषा संसाधन प्रणाली' (Automated Language Processing System) बनाया गया। 1985 में दारपा (Defense Advanced Research Projects Agency : DARPA) ने अपने वाक् अभिज्ञान प्रोग्राम की शुरुआत की। 1986 में जापान ने बहुभाषिक वाक् अनुवाद के अध्ययन के लिए ए.टी.आर. निर्वचन टेलीकम्यूनिकेशन शोध प्रयोगशाला, (ATR Interpreting Telecommunication Research Laboratories) का निर्माण किया। 1997 में कार्नेगी मेलॉन विश्वविद्यालय (Carnegie Mellon University) में सामान्य प्रायोजन श्रुतलेख (General purpose dictation) के लिए ड्रेगन डिक्टेट (Dragon Dictate) नामक प्रथम वाक् से पाठ प्रणाली का विकास हुआ। इसी समय दारपा (DARPA) द्वारा मानव-मशीन अंतरक्रिया के लिए वाक् भाषिक प्रणाली (Spoken Language System) का विकास किया गया। 1991 में आई.बी.एम. अनुवाद प्रबंधक (Translation Manager) का निर्माण हुआ। 1992 में AIR- ITL ने 'वाक् अनुवाद उन्नत अनुसंधान संघ' (Consortium for Speech Translation Advanced Research : C- STAR) की स्थापना की, जिसने अँग्रेजी, जर्मन और जापानी भाषा में हुए फोन अनुवाद का पहला प्रदर्शन किया। 1994 में चयनित कॉम्प्यूसर्व चैट मंच (select CompuServe chat forums) के अंतर्गत निःशुल्क सिस्ट्रॉन मशीनी अनुवाद (Free Systran Machine Translation) उपलब्ध हुआ। इस काल में होने वाले अन्य कार्य हैं- (ALPS), वूडनर (Weidner), पी.सी. अनुवादक (PC-Translator), ग्लोबल्लिंक (Globalink) नामक प्रथम पीसी प्रणाली का विकास, एटलस (ATLAS), हाईकैट्स (HICATS), एसट्रांसैक (ASTRANSAC), डूएट (DUET), टोवना (TOVNA) जैसी व्यावसायिक (Commercial) प्रणाली का विकास, अंतरण आधारित अभिगम पर ग्रोनोब्ल विश्वविद्यालय में एरिएन (ARIANE) का विकास, टैक्सास में मेटल (METAL), सारब्रुकेन में ससी (SUSY) प्रणाली का विकास, न्यूमैक्सिको में ज्ञान आधारित प्रणाली का विकास, रासेटा (ROSSETA) नामक अंतरभाषिक प्रणाली का विकास, टर्मबैंक (Termbank) और शब्दावली प्रबंधन जैसे अनुवाद उपकरणों का विकास। 1980 के दशक में जापान में पिवोट (PIVOT) तंत्र का विकास हुआ, जो अँग्रेजी, जापानी, कोरियन, फ्रेंच

तथा स्पेनिश में अनुवाद के लिए बनाया गया। आई.बी.एम. के जर्मनी, स्पेन, इजराइल तथा अमेरिका में विभिन्न शोध केंद्रों में एल.एम.टी. परियोजना (1985-86) प्रारंभ की गई, जिसका उद्देश्य अंग्रेजी-जर्मन, जर्मन-अंग्रेजी, अंग्रेजी-स्पेनिश अनुवाद तंत्रों का प्रोटोटाइप प्रणाली तैयार करना था। 1989 में यूरोपीय आयोग द्वारा यूरोट्रा (EUROTRA) नामक मशीनी अनुवाद परियोजना की स्थापना की गई, जिसका उद्देश्य यूरोपीय समुदाय (European Community) के नौ कार्यालयी भाषाओं के लिए मशीनी अनुवाद प्रणाली का निर्माण करना था। यह परियोजना यूरोपीय आयोग द्वारा वित्तपोषित है। इस क्षेत्र में सीमित और आवश्यक शोधकार्यों के संयोजन एवं विकास के लिए 'अंतरराष्ट्रीय मशीनी अनुवाद परिषद्' (I.A.M.T) की स्थापना हुई। इसके अलावा कार्पस आधारित अभिगम, उदाहरण आधारित अभिगम, अनुवाद स्मृति, इंटरनेट के लिए प्रणाली विकास, स्थानीयकरण, कोशीय संसाधन, अनुवाद कार्यस्थल (Work Station) का विकास हुआ। इस क्षेत्र में जापान में हुए मशीनी अनुवाद विषयक शोध के संबंध में प्रस्तुत जेटेक (JTEC) नामक रिपोर्ट (1992) से अमेरिकी नीति निर्माताओं के लिए अमेरिका में वित्तीय सहयोग का नया दौर शुरू किया गया। आज जापान में सी.आई.सी.सी. (Centre for the International Cooperation of Computerization : CICC) द्वारा जापानी, चीनी, कोरियन आदि भाषाओं में परस्पर अनुवाद की अंतरराष्ट्रीय परियोजना पर कार्य चल रहा है। इस समय भाषाविज्ञान के क्षेत्र में कई मॉडलों का विकास हुआ, जैसे- कॉन्सट्रेंट आधारित रूपवाद, कोशीय-प्रकार्यात्मक व्याकरण (LFG), तर्क प्रोग्रामिंग (Q- systems, Prolog), डेफिनिट क्लाउज व्याकरण, स्लॉट व्याकरण, सामान्यीकृत पदबंध संरचना व्याकरण, हेड-ड्रिवेन पदबंध संरचना व्याकरण, कोटिकृत व्याकरण, अधिकार और अनुबंध व्याकरण, डिपेंडेंसी व्याकरण, मिनींग टेक्सट मॉडल, मॉन्टेग व्याकरण, सूचना सिद्धांत, प्रिंसिपल एवं पैरामीटर, वृक्ष संलग्न व्याकरण, न्यूरल नेटवर्क। 2000 में एम.आई.टी. (Massachusetts Institute of Technology: MIT) के लिंकलिन (Lincoln) प्रयोगशाला में यंग-शुक (Young Suk) और क्लिफोर्ड वीन्सटाइन (Clifford Weinstein) ने एक अत्याधुनिक कोरियन-अंग्रेजी वाक् से वाक् अनुवाद यंत्र के प्रोटोटाइप प्रणाली का प्रदर्शन किया। 2001 में चीली देश में बोली जाने वाली अल्प भाषा जैसे क्रोशियन (Croatian) के लिए कार्नेगी मेलॉन विश्वविद्यालय

(Carnegie Mellon University) के भाषा-प्रौद्योगिकी संस्थान के जेमी कारबोनेल (Jaime Carbonell) ने वाक् से वाक् अनुवाद प्रणाली का निर्माण किया। यू.एस.सी. (USC) के जैव चिकित्सक अभियंता थियोडोर बर्जर (Theodor Berger) और जिम शिह (Jim Shih) ने एक नये बर्जर-लिअव (Berger-Liaw) तंत्रिकीय संजाल वाक् अभिज्ञान प्रणाली (Neural Network Speech Recognition System : SRS) का विकास किया है, जो मानव की अपेक्षा वाचिक भाषा को समझने में अधिक सक्षम हैं। 2002 में एक एजेंट आधारित न्यूज़रीडर प्रणाली (Agent-Based News Reader Device) का विकास हुआ, जो आलेखों का अनुवाद कर उसे एमपी3 श्रव्य फाइल के रूप में परिवर्तित करता था। 2006 में नासा के निर्देशक रफु संजली (Rafu Sanjali) ने पृथ्वी से एक रोबोट नियंत्रित यान द्वारा मंगल ग्रह पर होने वाले चौथे आपदा को 99.999 प्रतिशत परिशुद्धता के साथ मशीनी अनुवाद प्रौद्योगिकी का प्रयोग कर नाकाम कर दिया। 2007 में माइक्रोसॉफ्ट ने “What do you want to think today?” अभियान (Campaign) के माध्यम से एक विचार अभिज्ञान अंतरापृष्ठ (TRI) का प्रदर्शन किया। 2008 में एल एण्ड एच (L&H) के ट्रैवल सनग्लासेज द्वारा सनग्लासेज धारक की मातृभाषा में रोड चिह्न, ट्रैफिक चिह्नों को तत्काल अनुवाद करने की सुविधा प्रदान की गई। 2009 में जापानी से अँग्रेजी डाक्यूमेंटेशन अनुवाद प्रोग्राम की अंतिम कॉपी बनाई गई है, जो मानव संपादन की जरूरत को कृत्रिम बुद्धि आधारित अर्थीय संजाल का प्रयोग कर कम करता है।

भारत में मशीनी अनुवाद सर्वथा नवीन है। इस क्षेत्र में हो रहे शोध और विकास को आज लगभग दो दशक बीत चुके हैं। मशीनी अनुवाद के क्षेत्र में भारत के प्रयास को निम्न लिखित अनुप्रयोगों में देखा जा सकता है-

आंग्लभारती एवं अनुभारती

अँग्रेजी से भारतीय भाषाओं में अनुवाद करने वाली मशीन साधित अनुवाद प्रणाली हेतु 1991 में भारतीय प्रौद्योगिकी संस्थान, कानपुर के प्रो. आर. एम. के० सिन्हा द्वारा आंग्लभारती परियोजना आरंभ की गई। भारत में यंत्रानुवाद क्षेत्र में किया जाने वाला यह पहला कार्य था। इसका पहला प्रोटोटाइप अँग्रेजी से तमिल के लिए बनाया गया और बाद में इसे अँग्रेजी से हिंदी के लिए भी प्रयोग लिया जाने लगा। नियम

आधारित अभिगम और संदर्भ निरपेक्ष व्याकरण का प्रयोग कर यह प्रणाली कुछ विशेष क्षेत्रों, जैसे- स्वास्थ्य अभियान, नियमित कार्यालयी पत्राचार आदि में प्रयोग होती है। बाद में इसमें उदाहरण आधारित अभिगम के साथ पश्च संपादन की व्यवस्था भी जोड़ दी गयी। 1995 में प्रो. सिन्हा ने अनुभारती प्रविधि का आरंभ किया। 2004 में चरण-2 के अंतर्गत अनुभारती- II और आंग्ल भारती- II विकसित की गई। अनुभारती- II का ढाँचा सामान्यीकृत पदानुक्रमिक उदाहरण आधारित एवं टेम्पलेट आधारित है। उदाहरण आधारित अभिगम के रूप में यह भूतकालिक अनुभव से ज्ञान को भंडारित कर भविष्य में इसका प्रयोग करने के लिए मानव अभिगम प्रक्रिया को अपनाती है। मशीन साधित अनुवाद के लिए विकसित यह अनुभारती प्रौद्योगिकी उदाहरण व कॉर्पस आधारित अभिगम एवं प्रारंभिक व्याकरणिक विश्लेषण के संयोजन की, एक संकरित उदाहरण आधारित मशीनी अनुवाद अभिगम है। यह अनुवाद प्रणाली हिंदी से अन्यज भाषाओं में अनुवाद करने के लिए विकसित की गई है। आंग्लभारती- II संकरण (Hybridisation) के लिए कच्चा उदाहरण आधार (Raw Example Base : REB) के अतिरिक्त सामान्यीकृत उदाहरण आधार (Generalised Example Base : GEB) का प्रयोग करती है।

अनुसारका

प्रो. राजीव संगल के निर्देशन में प्रारंभ हुई यह परियोजना सूचना प्रौद्योगिकी विभाग, संप्रेषण और सूचना प्रौद्योगिकी मंत्रालय, भारत सरकार, के भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (Technology Development for Indian Languages : TDIL) परियोजना के अंतर्गत वित्तपोषित है, जो तेलुगु, कन्नड़, पंजाबी, मराठी और बांग्ला से हिंदी में अनुवाद प्रणाली के विकास हेतु बनाई गई है। पाणिनि व्याकरण के सिद्धांतों का प्रयोग करते हुए यह स्रोत भाषा से लक्ष्य भाषा में लोकल वर्ड ग्रुप का प्रतिचित्रण कर, बच्चों की कहानियों का अनुवाद करती है। यह अनुवाद प्रणाली लैंग्वेज एक्सेसर कहलाती है, क्योंकि यह प्रमुख रूप से भारतीय भाषाओं के बीच भाषा को एक्सेस करती है। इसका उद्गम भारतीय प्रौद्योगिकी संस्थान, कानपुर, 1995 में हुआ, फिर बाद में स्कूल ऑफ ह्यूमेनिटी (School of Humanity), हैदराबाद विश्वविद्यालय और अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान, हैदराबाद के भाषा-प्रौद्योगिकी अनुसंधान केंद्र (Language Technology Research Centre : LTRC) में स्थानांतरित हो गया, जो



सत्यम कंप्यूटर प्राइवेट लिमिटेड (Satyam computer Pvt. Ltd.) द्वारा वित्तपोषित है। यह प्रणाली संस्था की वेबसाइट http://www.iiit.net/ltrc/Anusaaraka/anu_home.html पर उपलब्ध है।

शक्ति

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।

मंत्रा

मंत्रा राजभाषा (MAchiNE assisted TRAnslation tool : MANTRA Rajbhasha) नामक परियोजना प्रगत संगणन विकास केंद्र (C-DAC) पुणे के मंत्रा प्रौद्योगिकी के अंतर्गत राजभाषा विभाग, गृह मंत्रालय, भारत सरकार द्वारा वित्तपोषित है, जिसे भारतीय संसद के उच्च सदन राज्यसभा सचिवालय के लिए 1999 में डॉ. हेमन्त दरबारी और डॉ. महेन्द्र कुमार सी. पाण्डेय के निर्देशन में विकसित किया गया। यह प्रणाली प्रशासनिक दस्तावेजों, ज्ञापन, कार्यालय अध्यादेश, परिपत्र, अधिसूचना, प्रतिवेदन, प्रारूपण आदि को अंग्रेजी से हिंदी में अनुवाद करती है। वर्तमान में यह प्रणाली अंग्रेजी-बांग्ला, अंग्रेजी-तेलुगु, अंग्रेजी-गुजराती, हिंदी-बांगाली और हिंदी-अंग्रेजी भाषा-युग्मों के लिए कार्य कर रही है। अंग्रेजी और हिंदी व्याकरण को प्रस्तुत करने के लिए कोशीयकृत वृक्ष संलग्न व्याकरण (Lexicalized Tree Adjoining Grammar : LTAG, Bandyopadhyay, 2004) का प्रयोग किया जाता है तथा पार्सिंग व प्रजनन के लिए वृक्ष संलग्न व्याकरण (प्रो. अरविंद जोशी, पेन्सिलवेनिया विश्वविद्यालय) का प्रयोग होता है। बाद में यह स्मिथसोनियन इस्टिट्यूट के नेशनल म्यूजियम ऑफ अमेरिकन हिस्ट्री (Smithsonian Institution's National Museum of American History), वासिंगटन डीसी संयुक्त राष्ट्र अमेरिका (Washington DC, USA) में “द 1999 इनोवेशन कलेक्शन (The 1999 innovation collection)” का हिस्सा बना। यह प्रणाली संस्था के वेबसाइट <http://cdac.in/html/aai/mantra.asp> पर उपलब्ध है।

यू.एन.एल. आधारित मशीनी अनुवाद प्रणाली

यूनिवर्सल नेटवर्किंग लैंग्वेज (UNL) आधारित यह प्रणाली भारतीय प्रौद्योगिकी संस्थान, मुंबई में प्रो. पुष्पक भट्टाचार्या के निर्देशन में विकसित की गई है, जो यूनिवर्सल नेटवर्किंग लैंग्वेज रूपवाद (Formalism) को अंग्रेजी-हिंदी और बंगाली अनुवाद के लिए अंतरभाषा के रूप में प्रयोग करती है। यह संयुक्त राष्ट्र विश्वविद्यालय की अंतरराष्ट्रीय परियोजना है, जिसका लक्ष्य सभी प्रमुख मानव भाषाओं के लिए एक अंतरभाषा का निर्माण करना है। अन्य दो प्रणालियों यू.एन.एल. से हिंदी और हिंदी से यू.एन.एल. का प्रदर्शन इसकी साइट पर किया जा चुका है।

मात्रा

कविता मोहनराज के निर्देशन में अंग्रेजी-हिंदी में अनुवाद करने वाली मानव साधित अंतरण आधारित मात्रा (MaTra) अनुवाद प्रणाली का विकास 2004 में ज्ञान आधारित कंप्यूटर अनुभाग के प्राकृतिक भाषा समूह द्वारा राष्ट्रीय सॉफ्टवेयर प्रौद्योगिकी केंद्र (National Centre for Software Technology : NCST) में किया गया था, जो अब प्रगत संगणन विकास केंद्र (C-DAC) मुंबई नाम से जाना जाता है। समाचारों, वार्षिक प्रतिवेदनों और तकनीकी पदबंधों के क्षेत्र में कार्य करने वाली यह प्रणाली अंग्रेजी पूर्वसर्गों का हिंदी परसर्गों में प्रतिचित्रण करती है। यह भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (Technology Development for Indian Languages : TDIL) परियोजना द्वारा वित्तपोषित है। यह प्रणाली संस्था के साइट <http://www.ncst.ernet.in/matra/> पर उपलब्ध है।

अंग्रेजी-कन्नड़ मशीनी अनुवाद प्रणाली

कर्नाटक सरकार द्वारा वित्तपोषित अंग्रेजी-कन्नड़ मशीन साधित अनुवाद प्रणाली का विकास प्रो. कवि नारायण मूर्ति द्वारा कंप्यूटर और सूचना विज्ञान विभाग के भारतीय भाषा-प्रौद्योगिकी समाधान के लिए स्रोत केंद्र (Resource Centre for Indian Language Technology Solution : RC-ILTS) <http://www.iitg.ernet.in/rcilts> हैदराबाद विश्वविद्यालय में अंतरण आधारित अभिगम का प्रयोग कर किया गया है। यह प्रणाली यूनिवर्सल क्लाउज संरचना व्याकरण (Universal Clause Structure Grammar: UCSG) का प्रयोग कर सरकारी परिपत्रों को अंग्रेजी से कन्नड़ में अनुवाद करती है। यह



प्रणाली संस्था की वेबसाइट <http://www.languagetechnologies.ac.in/lerc/mat/mat.htm> पर उपलब्ध है।

तमिल-हिंदी मैट प्रणाली

प्रो. सी. एन. कृष्णन के द्वारा अन्ना विश्वविद्यालय के के. बी. चंद्रशेखर अनुसंधान केंद्र (Anna University- KB Chandrashekhar : AU-KBC), चेन्नई में इस प्रणाली का विकास हुआ है। अनुसारका मशीनी अनुवाद प्रणाली पर आधारित इस प्रणाली के लिए तमिल रूपप्रक्रियात्मक विश्लेषक और तमिल-हिंदी द्विभाषी कोश तैयार किया गया है। अनुसारका मशीनी अनुवाद प्रणाली पर आधारित होने के कारण इसे 'तमिल अनुसारका' के नाम से भी जाना जाता है। यह प्रणाली संस्था की वेबसाइट http://www.au-kbc.org/research_areas/nlp/demo/mat/ पर उपलब्ध है। अनुवादक सुपर इंफोसॉफ्ट प्राइवेट लिमिटेड (Super Infosoft Pvt.Ltd.), दिल्ली द्वारा श्रीमती अंजली रावचौधरी के निर्देशन में विकसित अंग्रेजी से हिंदी अनुवाद करने वाला अनुवादक http://www.au-kbc.org/research_areas/nlp/demo/mat/ पर उपलब्ध है।

अनुवादक

सुपर इंफोसॉफ्ट प्राइवेट लिमिटेड (Super Infosoft Pvt.Ltd.), दिल्ली द्वारा श्रीमती अंजली रावचौधरी के निर्देशन में विकसित अंग्रेजी से हिंदी अनुवाद करने वाला अनुवादक 5.0 पश्च संपादन में सहायक है। इसके अंदर प्रशासनिक, कार्यालयी, भाषावैज्ञानिक, तकनीकी क्षेत्रों से संबंधित कोश रखे गए हैं। इस सॉफ्टवेयर को विंडो के किसी भी ऑपरेटिंग प्रणाली पर चलाया जा सकता है। यह प्रणाली इस लिंक

<http://www.mysmartschool.com/pls/portal/portal.MSSStatic.ProductAnuvaada> k पर उपलब्ध है।

अनुवाद हाईब्रिड मशीनी अनुवाद

इस प्रणाली को कंप्यूटर विज्ञान इंजीनियरिंग (Computer science Engineering : CSE) विभाग, जाधवपुर विश्वविद्यालय (Jadavpur University) द्वारा वर्ष 2004 में बंदोपाध्याय के निर्देशन में विकसित किया गया। यह प्रणाली उदाहरण आधारित मशीनी अनुवाद अभिगम का प्रयोग कर अंग्रेजी समाचारों के शीर्षक को बंगाली में अनुवाद करती है। वर्तमान में यह प्रणाली वाक्य स्तर तक कार्य कर रही है। यह प्रणाली संस्था की वेबसाइट <http://www.jadavpur.edu/> पर उपलब्ध है।

पंजाबी से हिंदी मशीनी अनुवाद प्रणाली

यह प्रणाली वर्ष 2008 में जोसन और लेहल (Josan and Lehal) के निर्देशन में पंजाब विश्वविद्यालय, पटियाला में विकसित की गई है। 92.8% परिशुद्धता के साथ शब्द से शब्द अनुवाद अभिगम पर आधारित यह प्रणाली विभिन्न स्रोतों का प्रयोग करता है, जैसे- धातु कोश, रूपप्रक्रियात्मक कोश, संदिग्धार्थक शब्दकोश आदि। अनुवाद मॉड्यूल वाक्य के विभिन्न वाक्यीय संरचनाओं, जैसे- नामीय पद अभिज्ञान, शब्द आशय विसंदिग्धक (Word Sense Disambiguator), पुनरावृत्ति, लिप्यंतरण आदि का समाधान करते हैं। इसी संस्था में वर्ष 2009 के अंतर्गत गोयल और लेहल (Goyal and Lehal) के निर्देशन में हिंदी से पंजाबी मशीनी अनुवाद प्रणाली विकसित की गई है। 95% परिशुद्धता के साथ शब्द प्रति शब्द अनुवाद अभिगम पर आधारित इस प्रणाली के लिए भी हिंदी-पंजाबी कोश, रूपप्रक्रियात्मक विश्लेषक शब्द आशय विसंदिग्धक (Word Sense Disambiguator) जैसे मॉड्यूल बनाए गए हैं।

संस्कृत सुबंत अभिज्ञानक और विश्लेषक

जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली के संस्कृत अध्ययन केंद्र के एम. फिल. डिग्री के अंतर्गत शोधार्थी सुभाष चंद्रा द्वारा यह ऑनलाइन उपकरण तैयार किया गया है, जिसे इसकी वेबसाइट <http://sanskrit.jnu.ac.in/subanta/rsubanta.jsp> से प्राप्त किया जा सकता है।

अंग्रेजी से (हिंदी, कन्नड़, तमिल) और तमिल-कन्नड़ भाषा-युग्मों का उदाहरण आधारित मशीनी अनुवाद प्रणाली

इस प्रणाली को वर्ष 2006 में बालाजापल्ली (Balajapalli) द्वारा विकसित किया गया था। यह मशीनी अनुवाद प्रणाली वाक्य, पदबंध, शब्द और स्वनिक कोश जैसे द्विभाषी कोश का प्रयोग कर स्रोत भाषा अँग्रेजी से तीन लक्ष्य. भाषाओं हिंदी, तमिल और कन्नड़ में अनुवाद करती है।

संपर्क

‘संपर्क’ नामक परियोजना के अंतर्गत अंतरराष्ट्रीय सूचना-प्रौद्योगिकी संस्थान, हैदराबाद द्वारा एक मशीनी अनुवाद प्रणाली का विकास किया गया है, जो चार भाषा युग्मों हिंदी से पंजाबी, पंजाबी से हिंदी, उर्दू से हिंदी और तेलुगु से तमिल में अनुवाद करता है। भारत के पूर्व राष्ट्रपति डॉ. ए.पी.जे. अब्दुल कलाम द्वारा 30 मार्च 2011 को इसके प्रथम सेट का उद्घाटन हुआ है। वर्ष 2009 में इस परियोजना के अंतर्गत भारतीय भाषाओं के मध्य विकसित होने वाली अनुवाद प्रणालियों के लिए संस्थाओं के संघ निर्मित किए गए हैं। इनके नाम हैं- अंतरराष्ट्रीय सूचना-प्रौद्योगिकी संस्थान, हैदराबाद, हैदराबाद विश्वविद्यालय, प्रगत संगणन विकास केंद्र (नोएडा, पुणे), भारतीय प्रौद्योगिकी संस्थान, कानपुर, भारतीय प्रौद्योगिकी संस्थान, खड़गपुर, अन्ना विश्वविद्यालय – के.बी.सी. अनुसंधान केंद्र (AU-KBC Research Lab), चेन्नई, भारतीय विज्ञान संस्थान, बंगलौर, भारतीय सूचना प्रौद्योगिकी संस्थान, इलाहाबाद, जाधवपुर विश्वविद्यालय, तमिल विश्वविद्यालय।

इसके अलावा उत्कल विश्वविद्यालय, भुवनेश्वर में प्रो. संघमित्रा मोहन्ती द्वारा ओमट्रांस (OMTrans) नामक अँग्रेजी-उड़िया मशीनी अनुवाद प्रणाली का विकास किया गया है। इसे <http://www.its-utkal.org/omtrans.htm> वेबसाइट से प्राप्त किया जा सकता है। गणित विभाग, भारतीय प्रौद्योगिकी संस्थान दिल्ली द्वारा उदाहरण आधारित अँग्रेजी-हिंदी अनुवाद प्रणाली का विकास और आई.बी.एम. इंडिया अनुसंधान प्रयोगशाला दिल्ली द्वारा अँग्रेजी से अन्य भारतीय भाषाओं के लिए स्टैट मशीनी अनुवाद (StatMT) नामक सांख्यिकीय मशीनी अनुवाद प्रणाली का विकास किया गया है। इसे इस वेबसाइट से <http://www.research.ibm.com/irl/projects/translation.html> प्राप्त किया जा सकता है। मशीनी अनुवाद के लिए वर्ष 1991-1995 के दौरान भारतीय भाषाओं में कॉरपोरा विकास हेतु भारतीय प्रौद्योगिकी संस्थान, नई दिल्ली द्वारा हिंदी, अँग्रेजी, पंजाबी, भारतीय भाषा

संस्थान, मैसूर द्वारा कन्नड़, मलयालम, तमिल, तेलुगु, डेक्कन महाविद्यालय, पुणे विश्वविद्यालय द्वारा मराठी, गुजराती, भारतीय अनुप्रयोग भाषाविज्ञान संस्थान (Indian Institute of Applied Language Science), भुवनेश्वर द्वारा उड़िया, असमी, बांग्ला, संपूर्णानन्द संस्कृत विश्वविद्यालय द्वारा संस्कृत भाषाओं के लिए कॉरपोरा का निर्माण किया गया है। वर्ष 1992-1995 में अलीगढ़ मुस्लिम विश्वविद्यालय द्वारा उर्दू, सिंधी और कश्मीरी भाषाओं के लिए कॉरपोरा निर्माण हो चुका है।

मशीनी अनुवाद की वर्तमान स्थिति

इस क्षेत्र में विश्व के प्रत्येक देशों में भिन्न-भिन्न तरीके से प्रगति हुई है। मशीनी अनुवाद प्रणाली प्राकृतिक भाषा संसाधन के भिन्नक अनुप्रयोगों में से एक है। इसके विकास हेतु वाक् और पाठ दोनों स्तरों पर अनेक प्रकार के यंत्रानुवाद से इतर मॉड्यूल, उपकरण और सॉफ्टवेयर की आवश्यकता होती है, जिसका निर्माण कार्य अनेक संस्थाओं में चल रहा है एवं विभिन्न परियोजनाओं के अंतर्गत ये वित्तपोषित हो रही हैं। भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (TDIL) परियोजना के अंतर्गत मशीनी अनुवाद के लिए कई उपकरण निर्मित किए गए हैं। इनमें प्रमुख रूप से ज्ञान स्रोत (समानांतर : Parallel) कॉरपोरा, बहुभाषिक लाइब्रेरी कोश, कोशीय स्रोत), ज्ञान उपकरण (भाषा संसाधन टूल्स, अनुवाद स्मृतति टूल्स), अनुवाद सहायक प्रणाली (मशीनी अनुवाद, बहुभाषिक सूचना एक्सेस, सूचना प्रत्यानयन), मानव-मशीन अंतरापृष्ठ प्रणाली (संप्रतीक अभिज्ञान प्रणाली, वाणी अभिज्ञान प्रणाली, पाठ से वाक् प्रणाली), स्थानीयकरण (Localization), भाषा-प्रौद्योगिकी मानव स्रोत विकास (प्राकृतिक भाषा संसाधन और कंप्यूटेशनल भाषाविज्ञान में मानवशक्ति का विकास) (Language Technology Human Resource Development), मानकीकरण के क्षेत्र में विकास किया जा चुका है।

ज्ञान स्रोत विकास के अंतर्गत भारतीय भाषा संस्थान, मैसूर द्वारा मशीन पठित रूप में टैग किए हुए पाठ कॉरपोरा को विकसित किया जा चुका है। पाठगत कॉरपोरा के संकलन हेतु भारतीय भाषाओं में कॉरपोरा विकास कार्य निम्न लिखित संस्थाओं में चल रहे हैं- भारतीय भाषा संस्थान, मैसूर द्वारा 'अनुकृति' नामक डाटाबेस का निर्माण, कथा-भारती नामक भारतीय क्लासिक अनुवाद और भाषा-भारती नामक लाइब्रेरी

स्रोतों का डिजीटलीकरण, प्रगत संगणन विकास केंद्र (C-DAC) नोएडा द्वारा 'ज्ञान निधि' नामक समानान्तर पाठगत कॉरपोरा का निर्माण, ई.एम.एल.ई. (EMLE- Enabling Minority Language Engineering) द्वारा लिखित एवं वाचिक डाटा का संग्रह किया जा रहा है। वाचिक कॉरपोरा के संकलन हेतु भारतीय भाषाओं में कॉरपोरा विकास कार्य निम्नलिखित संस्थाओं में चल रहे हैं- केंद्रीय इलेक्ट्रॉनिक अभियांत्रिकी अनुसंधान संस्थान (CEERI), नई दिल्ली द्वारा संचालित रेलवे पूछ-ताछ प्रणाली और अक्षर-विभाजन (Syllabification) के लिए हिंदी-बंगाली पाठ से वाक् पद-विच्छेदन नियमों का निर्माण, प्रगत संगणन विकास केंद्र (C-DAC) नोएडा द्वारा 'ज्ञाननिधि' कॉरपोरा का प्रयोग कर हिंदी, मराठी और पंजाबी के लिए 'विश्लेषिका' नामक सांख्यिकीय पाठ विश्लेषक उपकरण का निर्माण, प्रगत संगणन विकास केंद्र (C-DAC) कोलकाता द्वारा असमी और मणिपुरी भाषा में वाक्-संश्लेषक और स्व चालित वाक्-अभिज्ञान प्रणाली का निर्माण, एच०पी० लैब (HP Labs) द्वारा पाठ से वाक् प्रणाली के लिए हिंदी और अँग्रेजी डाटाबेस का निर्माण, स्वरचालित वाक्-अभिज्ञान प्रणाली के लिए असमी और भारतीय अँग्रेजी डाटाबेस का निर्माण, वर्तमान में अंतरराष्ट्रीय सूचना-प्रौद्योगिकी संस्थान, हैदराबाद के संयुक्त तत्वावधान में मराठी, तमिल और तेलुगु कि लिए संग्रह कार्य जारी है। उत्कल विश्वविद्यालय, भुवनेश्वर द्वारा टेलीफोनी प्रणाली में प्रयोग हेतु शोर रहित वातावरण और वाक् अभिज्ञान एलगोरिदम का निर्माण, भारतीय प्रौद्योगिकी संस्थान, मद्रास में हिंदी और तेलुगु समाचार बुलेटिन का निर्माण, भारतीय प्रौद्योगिकी संस्थान, मुंबई में 'वाणी' नामक पाठ से वाक् प्रणाली के प्रोटोटाईप का निर्माण, सी.एफ.एस.एल चंडीगढ़ द्वारा अँग्रेजी और हिंदी के लिए वक्ता अभिज्ञान डाटाबेस का निर्माण, भारती विद्यापीठ, पुणे द्वारा मराठी आधारित वाक्-संश्लेषक का निर्माण, आई.सी.एस. हैदराबाद द्वारा संवादात्मक वाणी प्रतिक्रिया (Interactive Voice Response) प्रणाली का विकास किया जा रहा है। टाटा अनुसंधान संस्थान (Tata Institute of Fundamental Research), मुंबई द्वारा वाक्-अभिज्ञानक, वाक्-संश्लेषक, भाषा-मॉडलिंग और वाक्-डाटाबेस का निर्माण, प्रोलॉगिक्स सॉफ्टवेयर (Prologix Software), लखनऊ द्वारा हिंदी वाक्-संश्लेषक का निर्माण, भ्रिगस सॉफ्टवेयर लिमिटेड (Bhrigus Software Limited), हैदराबाद द्वारा हिंदी, तेलुगु वाक्-अभिज्ञानक, वाक्-संश्लेषक का निर्माण, वेबल

मीडियाट्रॉनिक्स (Webel Mediatronic), कोलकाता द्वारा हिंदी, बंगाली वाक्-संश्लेषक का निर्माण किया जा रहा है। प्रगत संगणन विकास केंद्र (C-DAC) नोएडा एवं वैज्ञानिक एवं तकनीकी शब्दावली आयोग द्वारा हिंदी विश्वकोश, भारत भाषा शब्दकोश, पैन-इंडियन (Pan-Indian) शब्दकोश, अंग्रेजी-हिंदी शब्दकोश, संस्कृत-हिंदी इलेक्ट्रॉनिक द्विभाषी शब्दकोश निर्माण कार्य प्रारंभ होने वाला है। नागरी प्रचारणी सभा, वाराणसी द्वारा प्रकाशित हिंदी विश्वकोश का ऑनलाइन प्रदर्शन केंद्रीय हिंदी संस्थान, आगरा एवं प्रगत संगणन विकास केंद्र (C-DAC) की संयुक्त परियोजना के अंतर्गत किया जा चुका है। जवाहरलाल नेहरू विश्वविद्यालय नई दिल्ली में सूचना-प्रौद्योगिकी विभाग की परियोजना हेतु भारतीय भाषा कॉरपोरा उपक्रम (Indian Language Corpora Initiative) के अंतर्गत आई.एल.सी.आई. (ILCI) टूल विकसित किया गया है, जो इसकी वेबसाइट <http://sanskrit.jnu.ac.in/ilciann/index.jsp> पर उपलब्ध है। भारतीय प्रौद्योगिकी संस्थान, मुंबई संयुक्त राष्ट्र संघ की वित्तपोषित परियोजना हिंदी शब्द संजाल के लिए कार्य कर रही है। इसके अलावा लिनक्स मंच के लिए देवनागरी की एच.टी.एम.एल. दस्तावेजों के अनुक्रमण और खोज के लिए हिंदी खोज इंजन का विकास किया जा चुका है और हिंदी बुलेटिन बोर्ड प्रणाली का कार्य विकासाधीन है।

ज्ञान उपकरण (Knowledge Tool) के विकास के अंतर्गत बिड़ला संस्थान (बीट्स पिलानी) के सहयोग से देवनागरी कंप्यूटर भी विकसित किया जा रहा है। 'अक्षर', 'शब्दमाला', 'शब्दरत्न', 'आलेख', 'भारती', 'मल्टीवर्ड' आदि शब्द संसाधन के अलावा 'जिस्ट', तकनीक (ग्राफिक एंड इंडियन स्क्रिप्ट टर्मिनल) पर कई हार्डवेयर युक्ति का विकास किया जा चुका है। सी-डैक, बंगलौर द्वारा संस्कृत शब्द संसाधक का कार्य निर्माणाधीन है। पाठ-संसाधन के क्षेत्र में संस्कृत विद्वानों के प्रयोग के लिए संस्कृत शब्द संसाधक सहित संस्कृत संलेखन प्रणाली विकसित की जा रही है। संस्कृत भाषा के लिए प्राकृतिक भाषा समझ प्रणाली के रूप में 'देशिका' नामक सॉफ्टवेयर पैकेज विकसित किया गया है, जो प्राचीन भारतीय विज्ञान के सिद्धांतों पर आधारित है। कंप्यूटर भाषाविज्ञान शोध एवं विकास के संस्कृत अध्ययन केंद्र के अंतर्गत, जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली द्वारा डॉ॰ गिरीश नाथ झा के मार्गदर्शन में एम.फिल. शोध के दौरान कोशीय स्रोत के रूप में वर्ष 2011 में 'सुश्रुत संहिता' (Sushruta Samhita) नामक ऑन-लाईन

आयुर्वेद अनुक्रमणी (Online Indexing of Ayurved) निर्मित की गई है। वर्ष 2008 में बच्चोंब के लिए मल्टीमीडिया और ई-शिक्षण सामग्री के निर्माण हेतु संस्कृत कंप्यूटेशनल टूलकिट्स और संस्कृत-हिंदी मशीनी अनुवाद <http://sanskrit.jnu.ac.in/shmt/index.jsp> का विकास किया जा रहा है, जिसके लिए अंबा कुलकर्णी (हैदराबाद विश्वविद्यालय) के नेतृत्व में सात संस्थाओं के संघ निर्मित हुए हैं- जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली, अंतरराष्ट्रीय सूचना-प्रौद्योगिकी संस्थान, हैदराबाद, संस्कृत अकादमी, पूमा प्रजना विद्यापीठ, बंगलौर, जे.आर.आर.एस.यू. (Jagadguru Ramanandacharya Rajasthan Samskrita University), जयपुर और तिरुपति विद्यापीठ। कासटल (Castle) सॉफ्टवेयर ने जिस्ट कार्ड के साथ डॉस प्लेंटफार्म पर संस्कृत शिक्षण और अभिगम परियोजना विकसित की थी, जिसके अंतर्गत संस्कृत स्वनविज्ञान और रूपविज्ञान के संश्लेषण पक्ष को संश्लेषक का निर्माण किया गया है। सी.बी.एस.ई. के माध्यमिक और उच्च माध्यमिक विद्यालयों के सूचना-प्रौद्योगिकी पाठ्यक्रम में भारतीय भाषाओं को स्थान देने में भी सूचना-प्रौद्योगिकी मंत्रालय की अहम भूमिका रही है। बी.ए. और एम.ए. स्तर पर प्रयोजनमूलक हिंदी के लिए भी सूचना-प्रौद्योगिकीय पाठ्यक्रम निर्मित किए जा रहे हैं। वनस्थली विद्यापीठ द्वारा मशीन पठित रूप में डोएक 'ओ' स्तर के लिए हिंदी कंप्यूटर कोर्सवेयर का विकास हो चुका है, जो अब वेब पर भी उपलब्ध होगा। भारतीय प्रौद्योगिकी संस्थान, कानपुर द्वारा भारतीय परंपरा को केंद्र में रखकर 'उपनिषद् और 'भगवद्गीता' के संपूर्ण पाठ को एक विरासत वेबसाइट के माध्यम से प्रस्तुत किया गया है।

मानव-मशीन अंतरापृष्ठ प्रणाली के विकास के अंतर्गत केंद्रीय इलेक्ट्रॉनिक इंजीनियरिंग अनुसंधान संस्थान (CEERI), नई द्वारा सुर, अनुत्तान जैसे स्वनगुण से युक्ति 'हिंदी वाणी' नामक पाठ से वाक् परिवर्तन सॉफ्टवेयर विकसित किया गया है। भारतीय प्रौद्योगिकी संस्थान, मद्रास के वाक्-प्रौद्योगिकी समूह द्वारा भारतीय भाषाओं के लिए प्रौद्योगिकी विकास कार्य जारी है। अमेरिका द्वारा एक ऐसे तंत्र का विकास किया जा रहा है, जो वाचिक भाषा को संकेत में परिवर्तित कर शीघ्रता से अनुवाद करने में सक्षम है।

अनुवाद सहायक प्रणाली के विकास के अंतर्गत 'प्रबंधिका' नामक कॉर्पस प्रबंधक, 'चित्राक्षरिका' नामक हिंदी प्रकाशित संप्रतीक अभिज्ञान (OCR), बहुभाषिक सूचना प्रत्यानयन प्रणाली, प्रति भाषिक सूचना

प्रत्यानयन प्रणाली (Cross Lingual information retrieval system), 'लेखिका' भारतीय भाषा शब्द, संसाधक का विकास हो चुका है। यूनिवर्सल डिजिटल कम्यूनिकेशन रिसर्च इंस्टिट्यूट ने अंकीय कोश की विस्तृत योजना बनाई है। आंग्लभारती मिशन के अंतर्गत मशीन साधित मशीनी अनुवाद के विकास के लिए आंग्ल प्रौद्योगिकी के अंतर्गत अँग्रेजी से भारत की 12 भाषाओं में अनुवाद के लिए इसे 8 अलग-अलग संस्थानों में स्थानांतरित किया गया है। इनमें भारतीय प्रौद्योगिकी संस्थान, मुंबई आंग्ल-मराठी और आंग्ल-कोंकणी पर, भारतीय प्रौद्योगिकी संस्थान, गुवाहाटी आंग्ल-असमी और आंग्ल-मणिपुरी पर, सी-डैक, पुणे आंग्ल-सिंधी, आंग्ल-उर्दू और आंग्ल-कश्मीरी पर, सी-डैक, कोलकाता आंग्ल-बांग्ला पर, सी-डैक, तिरुवनंतपुरम आंग्ल-मलयालम पर, थापर अभियांत्रिकी एवं प्रौद्योगिकी संस्थान, पंजाब आंग्ल-पंजाबी पर, जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली, आंग्ल-संस्कृत पर कार्यरत हैं। माइक्रोसॉफ्ट द्वारा माइक्रोसॉफ्ट शोध-मशीनी अनुवाद (Microsoft Research Machine Translation : MSR-MT) नामक एक डाटा-चालित (Data Driven) अनुवाद प्रणाली विकसित की गई है, जिसकी क्षमता अपने समस्त कोशीय और पदबंधीय ज्ञान को सीधे उपलब्ध डाटा से अनुवाद करने की है। अभी इस प्रणाली के पद-विच्छेदक (Parser) सात भाषाओं अँग्रेजी, फ्रेंच, जर्मन, कोरियन, जापानी, चीनी और स्पेनिश एवं जेनेरेटर (Generator) पांच भाषाओं अँग्रेजी, फ्रेंच, जर्मन, जापानी और स्पेनिश में उपलब्ध है। माइक्रोसॉफ्ट अनुवादक प्रौद्योगिकी द्वारा बिंग अनुवादक (Bing Translator) नामक सेवा उपलब्ध कराई गई है, जो पूरे पाठ या वेबपृष्ठ को विभिन्न भाषाओं में अनुवाद करती है। वर्ष 1988 में इलेक्ट्रॉनिक विभाग, भारत सरकार द्वारा सूचना इंटरचेंज के लिए भारतीय लिपि कोड (Indian Script Code for Information Interchange : ISCII) विकसित हुई थी, जिसे 1991 में भारतीय मानक ब्यूरो द्वारा संशोधित किया गया था। प्रगत संगणन विकास केंद्र (C-DAC) द्वारा भारतीय भाषा दस्तावेजों के लिए सीडी संलेखन उपकरण का विकास किया जा चुका है। प्रगत संगणन विकास केंद्र (C-DAC) मोहाली द्वारा संस्कृत वाक्य की अर्थ और वाक्यात्मक संरचना को विश्लेषित करने के लिए 'शब्दबोध' नामक एक संवादात्मक अनुप्रयोग विकसित किया गया है।

स्थानीयकरण (Localization) हेतु विभिन्न सॉफ्टवेयर एवं प्लेटफॉर्म पर निजी क्षेत्रों और कंपनियों द्वारा प्रयास चल रहे हैं। सी-डैक, मॉड्यूलर, इंफोटेक द्वारा अनेक भारतीय भाषा प्रोसेसिंग सॉफ्टवेयर पैकेज विकसित किए गए हैं। सोनाटा, शिखर, सॉफ्टेक, वेब दुनिया, इंड-लिनक्स, सरल सॉफ्ट, इलैक्ट्रानिकी शोध एवं विकास केंद्र, प्रगत संगणन विकास केंद्र (C-DAC) मुंबई, भारतीय प्रौद्योगिकी संस्थान, कानपुर, राष्ट्रीय इंफॉरमेटिक्स केंद्र (National Informatics Centre : NIC), टीसीएस, आई.बी.एम. इंडिया अनुसंधान लैब, ओरेकल आदि सरकारी, अर्द्ध सरकारी एवं गैर-सरकारी संगठनों द्वारा पता प्रबंधन प्रणाली (Address Management System), भारतीय भाषा-शिक्षा प्रणाली, व्यापार प्रबंधन प्रणाली आदि के रूप में भारतीय भाषाओं में फॉन्ट आधारित बहुभाषी पैकेज, बहुभाषी शब्द-संसाधक, प्रतिलेखन सुविधा, फॉन्ट आधारित भारतीय डी.टी.पी. पैकेज, डेटाबेस पैकेज के लिए सक्षम स्क्रिप्ट, भारतीय लिपि सक्षम पैकेज, डेटा प्रविष्टि पैकेज, ई-मेल प्रणाली, अनुप्रयोग सॉफ्टवेयर पैकेज को विकसित किया जा चुका है। माइक्रोसॉफ्ट द्वारा महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा के सहयोग से वर्ष 2010 में हिंदी में कैप्शन लैंग्वेज अंतरापृष्ठ पैकेज (Caption Language Interface Package) का निर्माण किया गया है।

प्रौद्योगिकी संस्थानों एवं विश्वविद्यालयों में शोध कार्यों द्वारा मशीनी अनुवाद की अनेक विधियों, प्रविधियों, उच्चस्तरीय मध्यवर्ती भाषा (High Level Medium Language) व अंतरभाषा एवं द्विभाषी कोशों का विकास किया जा रहा है। अनूदित सामग्री को संपादित कर और अधिक बेहतर बनाने के लिए विभिन्न प्रकार के अंतरापृष्ठ निर्मित किए जा रहे हैं। महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा में भाषा-प्रौद्योगिकी विभाग, कंप्यूटेशनल भाषाविज्ञान विभाग और इन्फॉरमेटिक्स एण्ड लैंग्वेज इंजीनियरिंग विभाग भी हिंदी से अन्य भारतीय भाषाओं एवं अंग्रेजी में अनुवाद के लिए ऐसी प्रणालियों के विकास कार्य में संलग्न है। एम.फिल. और पी-एच.डी. शोध के अंतर्गत हिंदी के कई सहायक उपकरण का विकास किया जा चुका है एवं शोधार्थियों द्वारा ऐसे अनेक उपकरणों पर निर्माण कार्य जारी है।

मशीनी अनुवाद की संभावनाएँ

भविष्य में इस कार्य के विस्तार हेतु विभिन्न मंत्रालयों और उसके विभागों द्वारा अलग-अलग परियोजनाओं की रूपरेखा बनाई जा रही है, जिनमें अनेक संस्थाएँ संलग्न हैं। मशीनी अनुवाद के दीर्घकालिक लक्ष्य की प्राप्ति हेतु भारतीय भाषाओं के लिए प्रौद्योगिकी विकास (TDIL) परियोजना के अंतर्गत वाक् से वाक् अनुवाद प्रणाली और मानव सहयोगी प्रणालियों एवं विभिन्न स्रोत केंद्रों का विकास होना सुनिश्चित हुआ है। इसकी पूर्ति हेतु विश्व भर में विभिन्न परियोजनाओं पर कार्य चल रहे हैं। भारतीय भाषा प्रौद्योगिकी प्रसरण एवं विस्तारण केंद्र (Indian Language Technology Proliferation and Deployment Centre : ILTP-DC) द्वारा अनेक शोध क्षेत्रों को बढ़ावा दिया जा रहा है। इसके अंतर्गत प्रति भाषिक सूचना एक्सेस (Cross-lingual Information Access) का विकास, भारतीय भाषाओं के लिए रोबस्ट दस्तावेज विश्लेषण एवं अभिज्ञान प्रणाली (Robust Document Analysis & Recognition System) का विकास, ऑन-लाइन हस्तलिखित अभिज्ञान प्रणाली (On-line Hand Writing Recognition System) का विकास किया जाना प्रस्तावित है। जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली के संस्कृत अध्ययन केंद्र में कंप्यूटर भाषाविज्ञान शोध एवं विकास के अंतर्गत डॉ॰ गिरीश नाथ झा के मार्गदर्शन में संस्कृत एवं अन्य भारतीय भाषाओं के लिए भाषा-प्रौद्योगिकी के विभिन्न क्षेत्रों में शोध एवं विकास कार्य जारी है। कंप्यूटेशनल भाषाविज्ञान समूह द्वारा जे-टेस (JNU Text Encoding and Search for Sanskrit : J-TESS) नामक परियोजना प्रारंभ करने की संभावना है, जिसके अंतर्गत संस्कृत पाठ के लिए खोजीय डाटाबेस (Searchable Database) का निर्माण किया जाएगा। वर्तमान में यह संस्था संस्कृत-हिंदी अनुवादक (Sanskrit Hindi Translator :SAHIT) के निर्माण हेतु संस्कृत विश्लेषण उपकरण के विकास में संलग्न है। इसी संस्था में सूचना-प्रौद्योगिकी विभाग के परियोजना के लिए भारतीय भाषा कॉरपोरा उपक्रम (Indian Language Corpora Initiative) के अंतर्गत अँग्रेजी सहित 12 भारतीय भाषाओं में पर्यटन और स्वास्थ्य प्रक्षेत्र (Domain) में डॉ॰ गिरीश नाथ झा के नेतृत्व में समानांतर (Parallel) एनोटेटेड कॉरपोरा के निर्माण की संभावना है। इस परियोजना के अंतर्गत कार्यरत संघ की सूची निम्नवत् है- जे.एन.यू. नई दिल्ली, पंजाबी विश्वविद्यालय, पटियाला, भारतीय सांख्यिकीय संस्थान (Indian Statistical Institute) कोलकाता, उत्कल विश्वविद्यालय, भुवनेश्वर, भारतीय

प्रौद्योगिकी संस्थान मुंबई, गुजरात विश्वविद्यालय, गोवा विश्वविद्यालय, तमिल विश्वविद्यालय, तंजावुर, द्रविण विश्वविद्यालय कुप्पम, आई.आई.आई.टी.एम. केरल। सितम्बर 2006 के बाद सूचना-प्रौद्योगिकी विभाग, कंप्यूटर और सूचना-प्रौद्योगिकी मंत्रालय, भारत सरकार द्वारा पर्यटन (Tourism) और स्वास्थ्य प्रक्षेत्र में अंग्रेजी से भारतीय भाषाओं में मशीनी अनुवाद प्रणाली के विकास हेतु एक परियोजना तैयार की गई थी, जिसके लिए इस क्षेत्र से संबंधित निम्नलिखित संस्थाओं के संघ निर्मित हुए हैं, जिनके सदस्य निम्नवत् हैं- अंतरराष्ट्रीय सूचना-प्रौद्योगिकी संस्थान, हैदराबाद, पुणे और मुंबई का प्रगत संगणन विकास केंद्र (C- DAC), भारतीय विज्ञान संस्थान, बंगलौर, भारतीय प्रौद्योगिकी संस्थान, मुंबई, जाधवपुर विश्वविद्यालय, भारतीय सूचना प्रौद्योगिकी संस्थान, इलाहाबाद, उत्कल विश्वविद्यालय, भुवनेश्वर, अमरीता विश्वविद्यालय, कोयम्बटूर और बनस्थाली विद्यापीठ, राजस्थान। इनमें प्रगत संगणन विकास केंद्र (C-DAC) मुंबई अंग्रेजी से भारतीय भाषाओं में सांख्यिकीय मशीनी अनुवाद प्रणाली के विकास हेतु सांख्यिकीय मॉडल और स्रोत निर्मित करेगा। आंग्लभारती मिशन के अंतर्गत उत्कल विश्वविद्यालय, भुवनेश्वर द्वारा आंग्ल-उड़िया मशीनी अनुवाद प्रणाली विकसित करने की संभावना है। अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान, हैदराबाद के 'संपर्क' नामक परियोजना के प्रथम सेट का प्रदर्शन होने के बाद भावी चार-पाँच महीनों में 18 भाषा-युग्मों पर कार्य होने की संभावना बताई जा रही है, जो प्रो. राजीव संगल के निर्देशन में पूर्ण होगा। टाटा कंसलटेंसी सर्विस (Tata Consultancy Service), याहू और गूगल भी मशीनी अनुवाद प्रणाली के विकास कार्य से जुड़े हैं। गूगल द्वारा अप्रैल 2006 में निःशुल्क सांख्यिकीय मशीनी अनुवाद प्रणाली सेवा उपलब्ध कराई गई है और 2011 में इसके 23वें चरण का प्रदर्शन होने वाला है। सूचना-प्रौद्योगिकी मंत्रालय ने भारतीय भाषा-प्रौद्योगिकी के लिए 13 स्रोत केंद्रों (Resource Centres) की स्थापना की है, जिनमें भाषा और उससे संबंधित संगठनों को रखा गया है, जो संघ की 22 कार्यालयी भाषाओं को लेते हुए उसका समाधान प्रस्तुत करेगी। इन स्रोत केंद्रों की सूची निम्नवत् है-

1. असमी और मणिपुरी के लिए भारतीय प्रौद्योगिकी संस्थान, गुवाहाटी।
2. बांग्ला के लिए भारतीय सांख्यिकीय संस्थान, कोलकाता।

3. विदेशी भाषाओं (जापानी, चीनी) और संस्कृत (भाषा अभिगम प्रणाली) के लिए जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली ।
4. गुजराती के लिए एम.एस. विश्वविद्यालय, बड़ौदा ।
5. हिंदी और नेपाली के लिए भारतीय प्रौद्योगिकी संस्थान, कानपुर ।
6. कन्नड़ और संस्कृत (संज्ञानात्मक मॉडल) के लिए भारतीय विज्ञान संस्थान, बंगलौर ।
7. मलयालम के लिए प्रगत संगणन विकास केंद्र (C-DAC) तिरुवनंतपुरम ।
8. मराठी और कोंकणी के लिए भारतीय प्रौद्योगिकी संस्थान, मुंबई ।
9. उड़िया के लिए कंप्यूटर विज्ञान और अनुप्रयोग विभाग, उत्कल विश्वविद्यालय ।
10. पंजाबी के लिए थापर इंजीनियरिंग एवं प्रौद्योगिकी संस्थान, पटियाला ।
11. तमिल के लिए अन्ना विश्वविद्यालय, चेन्नई ।
12. तेलुगु के लिए हैदराबाद विश्वविद्यालय, हैदराबाद ।
13. उर्दू, सिंधी के लिए सी-डैक, पुणे ।

इसकी जानकारी इस वेबसाइट http://tdil.mit.gov.in/resource_centre.htm से भी प्राप्त की जा सकती है। ये संस्थाएँ संबंधित भाषाओं के लिए कॉरपोरा और प्रौद्योगिकी विकास हेतु अग्रसर हैं। एशिया में वाक् मशीनी अनुवाद प्रणाली विकसित करने के लिए ए-स्टार (A-STAR) नामक संघ का निर्माण किया गया है, जिसका लक्ष्य एशिया में बड़े पैमाने पर वाचिक कॉरपोरा के विकास को आरंभ करना, संबंधित प्रौद्योगिकी, जैसे- बहुभाषिक वाक् अनुवाद, बहुभाषिक वाक् लिप्यंकन, बहुभाषिक सूचना प्रत्यानयन का विकास एवं एक अंतरराष्ट्रीय शोध समूह की स्थापना एवं वाक् अनुवाद की परीक्षण (Trial) सेवा को प्रारंभ करना है। इस कार्य के लिए ए-स्टार सदस्य के रूप से मुख्यतः जापान, चीन, कोरिया, थाइलैंड, इंडोनेशिया और भारत की अग्रणी भूमिका है। इसके अलावा वाचिक और प्रसारित समाचार को हिंदी और भारतीय अंग्रेजी में लिप्यंकित करने हेतु अंतर- संस्थानात्मक (Inter-institutional) परियोजना के अंतर्गत अनुवाद प्रणाली के विकास एवं बहुभाषिक स्रोत व प्रकाशित संप्रतीक अभिज्ञानक (OCR) पर संभावित कार्य चल रहे हैं। इस कार्य का लाभ यह होगा कि संबंधित

क्षेत्रों विभिन्न भाषिक-समुदाय आपस में सरलता से संपर्क कर सकेंगे। वैज्ञानिक एवं तकनीकी शब्दावली आयोग के संयुक्त तत्वावधान में सूचना-प्रौद्योगिकी के लिए हिंदी मानक शब्दावली का कार्य निर्माणाधीन है। एक सरल बहुभाषी कंप्यूटर (Simple Inexpensive Multi-lingual Computer : SIMPUTER) को इस उद्देश्य से डिजाइन किया गया है कि यह इंटरनेट अनुप्रयोग के लिए स्मार्ट कार्ड, पाठ से वाक्, सूचना मार्कअप लैंग्वेज के रूप में प्रयोग होने में सक्षम हो सके। सी-डैक, मुंबई द्वारा संस्कृत के लिए 'अमरकोश' नामक शब्दकोश, रूपवैज्ञानिक, वाक्यवैज्ञानिक और अर्थवैज्ञानिक विश्लेषक एवं टैगर, संस्कृत पार्सर, जेनरेटर और थिसॉरस का निर्माण होने वाला है। अंतरराष्ट्रीय सूचना-प्रौद्योगिकी संस्थान, हैदराबाद द्वारा हिंदी-तेलुगु कॉरपोरा, भारतीय भाषाओं में रूपवैज्ञानिक अभिगमक (Learner), वक्ता सत्यापन प्रणाली का निर्माण होने की संभावना है। साथ ही इस संस्थान -संघ के नेतृत्व में भारतीय भाषा से भारतीय भाषा में मशीनी अनुवाद प्रणाली (IL-IL MT) के विकास के लिए एक परियोजना प्रस्तावित की गई है, जो भविष्य में भारतीय प्रौद्योगिकी संस्थान, मुंबई में प्रारंभ होने वाली है। भारतीय प्रौद्योगिकी संस्थान, मुंबई द्वारा कोंकणी भाषा में एक शब्दी संजाल विकसित कर कोश निर्मित करने की भी संभावना है। 2012 तक सरुजुनो (Saruzuno) नामक विद्वान द्वारा मेमग्रेन माइक्रोफोन से सुसज्जित स्मार्टकार्ड के लिए एक कोशीय विसंदिग्धक प्रणाली (Lexical Disambiguation System : LDS) का विकास किया जाएगा, जिससे यात्री दर्जनों भाषाओं में अपने अधिकारियों से बात कर सकेंगे। 2020 तक बच्चों के अध्ययन के लिए एक ऐसा मशीन तैयार किया जाएगा, जिससे लिखने और पढ़ने में लगने वाले समय की बचत हो सकेगी। यह संकल्पना सिंगापुर के शिक्षामंत्री की है। 2043 तक न्यूट्रल भाषा के साथ नेक प्रौद्योगिकी (NEC Technology) के जुड़ जाने की संभावना है। 2045 तक यूरोपीय समुदाय द्वारा एक टेलीपैथी प्रणाली विकसित की जाएगी, जो हाई स्पीड बेतार संप्रेषी अभिग्राही (High-Speed Wireless Transceiver) के साथ मशीनी अनुवाद प्रौद्योगिकी और विचार अभिज्ञान (Thought Recognition) से युक्त होगी। 22वीं शताब्दी तक प्रोक्तेर (Procter) और गम्ब्ले (Gamble) द्वारा डिस्ट्रीब्यूटेड तच्योन स्वार्म प्रणाली (Distributed Tachyon Swarm System : DTSS) विकसित करने की संभावना है, जो मेडिकल क्षेत्र के लिए उपयोगी होगी।

निष्कर्षतः कहा जा सकता है कि एक तत्कालीन आवश्यकता की पूर्ति हेतु प्रस्तुत इस अत्याधुनिक संकल्पना का प्रसार आज विश्व के लगभग सभी देशों तक हो चुका है, जिसमें जापान की भूमिका अग्रणी है। भारत जैसे बहुभाषी देशों के लिए यह और अधिक सुखद स्थिति है। प्रस्तुत इतिहास की व्यापकता को देखकर यह अंदाजा लगाया जा सकता है कि इसका भविष्य कितना उज्वल होगा। तत्कालीन प्रगति को देखते हुए ही इसका भविष्य तृतीय विश्व हिंदी सम्मेलन में प्रधानमंत्री इंदिरा गांधी की इस घोषणा से सुनिश्चित हो गया था कि- “भाषा की टेक्नॉलॉजी तेजी से बढ़ रही है। अनुवाद के लिए कंप्यूटर का प्रयोग हो रहा है। हिंदी के वैज्ञानिकों को इस दिशा में समय के साथ ही नहीं, दूर तक सोचना चाहिए, जिससे हिंदी और हमारी दूसरी भाषाएं पिछड़ न जाए।” कंप्यूटर में भाषा संसाधन के लिए विकसित भाषा-विश्लेषण मॉडलों और भाषिक अनुप्रयोगों द्वारा कंप्यूटर और भाषा दोनों क्षेत्रों के वैज्ञानिकों को चिंतन के लिए एक ही वैचारिक पृष्ठभूमि देने का अवसर प्रदान किया जा रहा है। अतः निश्चित रूप से अभिव्यक्ति के माध्यम के रूप में उभरी मशीनी अनुवाद की संकल्पना, वर्तमान में अपनी उपलब्धियों की जड़ सींचते हुए भावी पीढ़ी के लिए एक ज्ञान पोषित समाज का निर्माण कर सकने में सक्षम हो सकेगी।

संदर्भिका

1. झा, गिरीशनाथ. मणि, दिवाकर. मिश्र, दिवाकर. ‘भारत में भाषा प्रौद्योगिकी : एक सर्वेक्षण’, गवेषणा, पृ- 37-47; केंद्रीय हिंदी संस्थान, आगरा।
2. त्रिपाठी, अरिमर्दन कुमार. (2008) ‘हिंदी के भाषा प्रौद्योगिकीय उपकरण’, राष्ट्रभाषा, पृ. 7-11.
3. दास, ठाकुर. (2008) ‘मशीनी अनुवाद : विधियाँ एवं प्रविधियाँ’, गवेषणा, पृ. 635-645; केंद्रीय हिंदी संस्थान, आगरा।
4. भाटिया, कैलाश चन्द्र. (1996) ‘कंप्यूटर अनुवाद की संभावनाएँ’, हिंदी : विकास और संभावनाएँ, पृ. 305-325; सूचना और प्रसारण मंत्रालय, भारत सरकार।
5. विकास, ओम. (2005) ‘हिंदी के विकास में टेक्नोलॉजी का योगदान’, गवेषणा, पृ. 617-641; केंद्रीय हिंदी संस्थान, आगरा।



6. सिंह, सूरजभान. जनवरी-मार्च (2005) सूचना-प्रौद्योगिकी और भाषा क्रांति, राजभाषा भारती, पृ. 7-13;
7. <http://sanskrit.jnu.ac.in/shmt/index.jsp>
8. <http://www.cse.iitb.ac.in/~pb/indtrend2.htm>
9. <http://www.scipub.org/fulltext/jcs/jcs6101082-1087.pdf>
10. http://www.indictrans.in/old/Articles/English/article_src/Indic/ncst2.pdf
11. <http://www.cfilt.iitb.ac.in/Translation-survey/survey.pdf>
12. http://projects ldc.upenn.edu/LDC_Institute/Visitors/Agrawalppt
13. <http://tdil-dc.in/>
14. <http://www.tdil.mit.gov.in/>
15. <http://www.cdacmumbai.in/e-ilmt>
16. http://projects ldc.upenn.edu/LDC_Institute/Visitors/Jha.pdf
17. <http://sanskrit.jnu.ac.in/index.jsp>
18. <http://sanskrit.jnu.ac.in/ilcian>
19. <http://www.iitk.ac.in>
20. <http://www.cse.iitk.ac.in/users/langtech/hist.htm>
21. <http://www.cse.iitk.ac.in/users/langtech/anglabharti.htm>
22. <http://anglahindi.iitk.ac.in>
23. http://prakashblog-google.blogspot.com/2008/04/blog-post_1407.html
24. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.1486&rep=rep1&type=pdf>



Citation: निधि, करुणा (2017). मशीनी अनुवाद : इतिहास, वर्तमान एवं भविष्य, HindiTech: A Blind Double Peer Reviewed Bilingual Web-Research Journal, 8 (1), 1-26. URL: <https://hinditech.in/mashini-anuvad-itih-as-vartman-avan-bhavishya/>