

मशीन अनुवाद: वैश्विक सन्दर्भ में विकास एवं विधियाँ

आम्रपाल शेंदरे

पीएच.डी. शोधार्थी,

School of Translation Studies and Training, IGNOU, नई दिल्ली-110068

Email – shendre.amrapals.amrapal@gmail.com Mob no. 9766956741

मशीन अनुवाद का इतिहास-

मशीन अनुवाद या कंप्यूटर साधित अनुवाद, अनुवाद की उस प्रक्रिया को कहते हैं जिसमें अनुवाद के लिए कंप्यूटर या मशीन का प्रयोग किया जाता है। कंप्यूटर आविष्कार के साथ ही इससे अनुवाद की प्रक्रिया की जाए ऐसा विचार कुछ कंप्यूटर विदों के मन में आया जिसकी परिणति बाद में मशीन अनुवाद के रूप में हुई। इसके इतिहास को वैश्विक सन्दर्भ में देखते हैं।

वैश्विक स्तर पर- वैसे तो कहाँ जाता है कि वैश्विक स्तर पर मशीन अनुवाद की शुरुवात द्वितीय विश्वयुद्ध के दौरान मानी जाती है। परन्तु इसकी नीव इससे भी पहले रखी गयी थी। डॉ। कैलाशचंद्र भाटिया ने कंप्यूटर अनुवाद का ऐतिहासिक परिदृश्य प्रस्तुत करते हुए इस तरह के प्रयास और प्रदर्शन की शुरुवात 1939 से स्वीकार की है। विकिपेडिया से मशीन अनुवाद के इतिहास के बारे में पता करने पर मालूम होता है कि सन 1930 के मध्य में जॉर्ज ऐस्ट्रोनी द्वारा पेपर टेप का प्रयोग करके ऑटोमैटिक बायलिंगुअल डिक्शनरी के लिए “ट्रांसलेटिंग मशीन्स” के नाम से पहला पेटेंट देखने को मिलता है, इसी क्रम में रशिया के विद्वान पीटर ट्रोयोनोस्की ने और विस्तार से बायलिंगुअल डिक्शनरी और द्विभाषिक व्याकरणिक महत्त्व को बताने का प्रयास किया था। जैसा कि कंप्यूटर के आविष्कार में गणितज्ञ और इंजिनियर चार्ल्स बेबेज को उसका जनक माना जाता है। मशीन अनुवाद के सन्दर्भ में कंप्यूटर-विज्ञान के संस्थापक नार्वे बाइनेर ने 1947 में कहाँ था “जहाँ तक मशीन अनुवाद का प्रश्न है, मुझे यह करने में हिचक नहीं होती कि विभिन्न भाषाओं के शब्दों की सीमाएं इतनी अधिक स्पष्ट हैं और इनके भावात्मक तथा सर्वादेशिक अर्थ इतने व्यापक हैं कि अर्ध-यांत्रिक अनुवाद की कोई व्यवस्था अधिक संभव नहीं जान पड़ती।” परन्तु इसके कुछ सालों बाद ही मशीन अनुवाद का पहला प्रोटोटाइप सामने आया। कंप्यूटर का आविष्कार मुख्यतः बड़ी- बड़ी सांख्यिकीय गणना के लिए किया गया था। द्वितीय विश्वयुद्ध के दौरान सर्वप्रथम इसका प्रयोग जर्मन सेना के गुप्त संदेशों को जानने के लिए किया गया था। और एटम बम से संबंधित गणना के लिए भी किया गया था। परन्तु उसी युग में प्रसिद्ध गणितज्ञ वारेन वीवर यह महसूस करने लगे थे कि कंप्यूटर की तार्किक प्रणाली का उपयोग भाषा संबंधित कार्यों

के लिए भी किया जाए। वस्तुतः कंप्यूटर अनुवाद का प्रारंभ सुविख्यात गणितज्ञ वारेन वीवर के 1947 में प्रकाशित 'ऑन ट्रांसलेशन' शीर्षक लेख से माना जाता है। वीवर के समय तक कंप्यूटर की तार्किक प्रणाली का उपयोग अंकीय गणना के लिए किया जाना शुरू हो चुका था। आंतर्राष्ट्रीय स्तर पर मशीन अनुवाद के विकास को मुख्य रूप से चार चरणों में रखा जाता है।

मशीन अनुवाद का शुरुवाती दौर 1948 से 60 तक-

मशीन अनुवाद की शुरुवात वारेन वीवर के प्रकाशित लेख से मानी जाती है, इसके बाद मशीन अनुवाद को लेकर सर्वप्रथम अंतर्राष्ट्रीय संगोष्ठी का आयोजन सन 1952 में येहोशुआ बार हिलेल के नेतृत्व में MIT(Massachusetts Institute of Technology) में किया गया। जिसमें मशीन अनुवाद की संभावना और उसकी संरचना पर विचार किया गया। इसके दो वर्ष बाद ही जनवरी 1954 में अमेरिका के जॉर्जटाउन विश्वविद्यालय और आई.बी.एम. कंपनी के संयुक्त कार्य से अंग्रेजी से रशियन भाषा में मशीन अनुवाद प्रणाली का विकास किया गया और इसे लोगो के सामने प्रदर्शित किया गया। जिसमें मशीन की सहायता से 49 रशियन वाक्यों को अंग्रेजी से अनुवादित किया गया था। जिसमें 250 शब्द और 6 व्याकरणिक नियमों का उपयोग किया गया था। जो अच्छा अनुवाद करता था। इसका मुख्य उद्देश्य था मशीन अनुवाद के कार्य को प्रोत्साहन मिले तथा इस प्रोजेक्ट के लिए अमेरिकी सरकार के द्वारा बड़ी मात्रा में अनुदान भी मिल सके तथा विश्व में और भी देश मशीन अनुवाद की ओर आकर्षित हो। इसी दौरान Victor Yngve (विक्टर यन्वे) ने मशीन ट्रांसलेशन पर पहला जर्नल "मेकैनिकल ट्रांसलेशन" नाम से प्रकाशित किया। शुरू में रूसी-अंग्रेजी मशीन अनुवाद के क्षेत्र में जॉर्जटाउन विश्वविद्यालय ने 1964 में GAT नाम से रूसी- अंग्रेजी मशीन अनुवाद प्रणाली का विकास किया। इसी मशीन अनुवाद प्रणाली को आगे चलकर 1976 में सिस्ट्रान(SYSTRAN) नाम से जाना जाने लगा।

मशीन अनुवाद का दूसरा दौर – 1961 से 1980 तक-

इस दौर तक आते-आते मशीन अनुवाद के सन्दर्भ में काफी कुछ विकास हो चुका था। कंप्यूटर से अनुवाद करने के लिए प्रोग्रामिंग भाषा का विकास करना आवश्यक होता है। इस दिशा में MIT ने पहला प्रयास किया। इस कंपनी ने 'कोमिट' नाम से प्रोग्रामिंग भाषा का विकास किया इस दिशा में एक और प्रयास 'नॅशनल ब्यूरो ऑफ़ स्टैण्डर्ड' नामक केंद्र ने किया था। वाशिंगटन की महिला गणितज्ञ श्रीमती आइडा रोड्स ने एक प्रोग्रामिंग भाषा विकसित की। गत वर्षों में मशीन अनुवाद करने वाले कंप्यूटरों का निर्माण किया गया जिसमें मुख्य रूप से रूसी- से अंग्रेजी और अंग्रेजी से रूसी में अनुवाद करने वाले कंप्यूटर मुख्य रूप से शामिल है। सन 1958 तक पूरे विश्व में तकनीकी लेखों का अनुवाद करने वाले केवल तीन कंप्यूटर थे। इनमें भी सबसे प्रभावशाली कंप्यूटर सोवियत रशिया ने विकसित कर लिया था। जिसकी स्टोरेज क्षमता 952 अंग्रेजी शब्दों और 1973 रूसी शब्दों की थी। सोवियत संघ में फ्रांसिसी- रूसी कंप्यूटर अनुवाद के 17 प्रोग्राम तैयार हुए जिनमें 8500 सूचनाएं थीं। इसी क्रम में 1961 में डेविड जी. हेज के साप्ताहिक व्याख्यानो के कारण

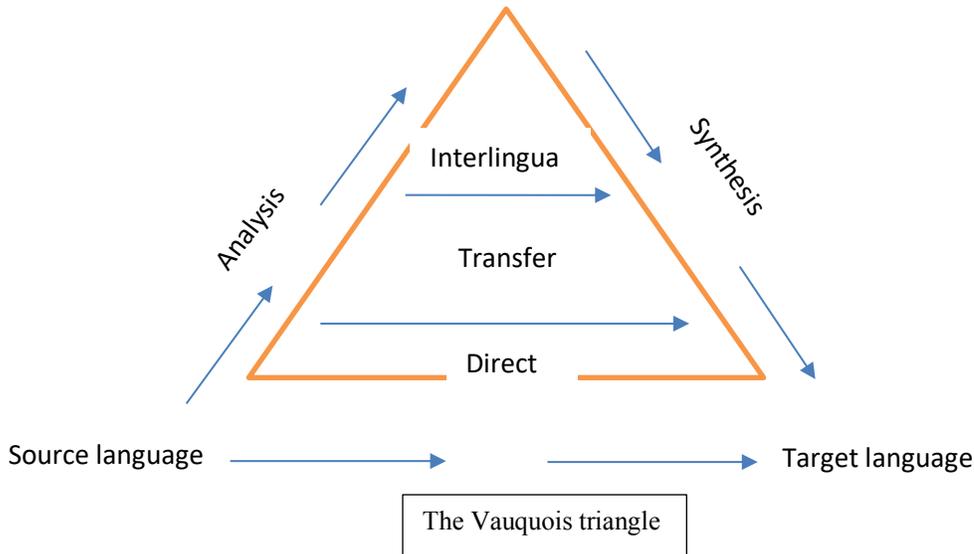
कंप्यूटेश्रल भाषाविज्ञान विषय का उदय हुआ। जिनका आयोजन लॉस-एंजेलेस के रैंड कार्पोरेशन में होता था। इसी साल मशीन ट्रांसलेशन पर पहले अंतर्राष्ट्रीय संगोष्ठी का आयोजन भी किया गया था। जिसमें भाषा विज्ञानी और कंप्यूटर विज्ञानी शामिल थे। जिसमें मुख्य रूप से चार्ल्स हॉकेट, सिडनी एम. लैम्ब, बर्नार्ड वॉकवॉइस आदि विद्वान शामिल थे। सन 1963 के बाद ऐसे रूसी- अंग्रेजी अनुवाद करने वाले कंप्यूटर बन गए थे, जो एक दिन में तकनीकी लेखों के एक लाख शब्दों का अनुवाद करने में समर्थ थे। जापान विश्व में तकनीकी अन्वेषण के लिए विख्यात है, मशीन अनुवाद के विकास के क्षेत्र में जापान ने नए कीर्तिमान स्थापित किए हैं। टोकियो की एक फार्म ने जापानी- अंग्रेजी अनुवाद का एक कंप्यूटर यंत्र तैयार किया है। यह यंत्र एक घंटे में 3000 शब्दों का अनुवाद प्रस्तुत करता है। परन्तु इस तरह किया गया अनुवाद एकदम सुनिश्चित नहीं होता, उसका संपादन करना होता है। जापान ने यू. एन. 700 नामक मिनी कंप्यूटर और भी विकसित बनाया है। तोशिबा कार्पोरेशन द्वारा निर्मित इस अनुवादक कंप्यूटर से अंग्रेजी से जापानी में अनुवाद करना अधिक शुद्ध व तीव्र गति से संभव है। माना जाता है कि इस कंप्यूटर द्वारा 90 प्रतिशत से ज्यादा शुद्ध अनुवाद हो सकता है। और यह कंप्यूटर एक घंटे में लगभग 5000 शब्दों का अनुवाद करने में समर्थ है। इस कंप्यूटर का निर्माण पेटेंट दस्तावेजों तथा वैज्ञानिक एवं तकनीकी लेखों का सही अनुवाद करने के लिए किया गया है। इसके सॉफ्टवेयर में 30000 बुनियादी शब्दों का समावेश किया गया है, साथ ही विशेष विषयों के 50000 अतिरिक्त शब्दों की व्यवस्था भी की गयी है। मशीन अनुवाद के चल रहे इन कार्यों को सही दिशा देने के लिए सन 1964 में अमेरिका द्वारा एक कमिटी का गठन किया गया जिसका नाम ALPAC(Automatic Language Processing Advisory committee) कमिटी था। इसका कार्य था मशीन अनुवाद का दृष्टिकोण और सही अनुवाद की संभावनाओं का अध्ययन कर एक रिपोर्ट तैयार करना था। इस समिति ने दो वर्षों तक काम किया और सन 1966 में अपनी रिपोर्ट अमेरिका सरकार के समक्ष प्रस्तुत की, जिसमें मशीन अनुवाद की भावी संभावनाओं को लेकर कुछ तथ्य सामने रखे गए। जिसमें कहा गया था कि मशीन अनुवाद का कार्य मतलब समय और पैसा दोनों की बर्बादी है। जितना भी मशीन ट्रांसलेशन का काम हो रहा है उसकी गुणवत्ता बहुत ही खराब है, जिसमें भाषावैज्ञानिक नियमों का अभाव है। इस रिपोर्ट का मशीन अनुवाद के कार्य पर बहुत ही बुरा असर हुआ और अगले दस सालों के लिए मशीन अनुवाद का कार्य पूरी तरह से बंद हो गया। सन 1966 से लेकर 1977 के इस कालखंड को मशीन अनुवाद के इतिहास में काले युग के रूप में जाना जाता है। इस युग में मशीन अनुवाद के कार्य को सरकारी सहायता और अनुदान पूरी तरह से बंद हो गया था। ALPAC समिति ने मशीन अनुवाद का मुल्यांकन करने के लिए कुछ मापदंड दिए थे। जिससे कि मशीन अनुवाद की गुणवत्ता में सुधार आ सके। अल्पैक समिति के इन सुझावों को अमेरिकी सरकार ने स्वीकार कर इससे संबंधित शोध कार्यों को प्रशासनिक सहायता देनी बंद कर दी। इस तरह मूलभूत शोध कार्य किए बिना मशीन अनुवाद प्रणाली का विकास करने के प्रयासों पर मुख्य रूप से अमेरिका और आंशिक रूप से अन्य देशों में शोध कार्य के लिए सरकारी सहायता मिलना बंद हो गयी। बौद्धिक जगत में इस रिपोर्ट का काफी विरोध होने के बावजूद भी अमेरिका में मशीन अनुवाद का काम ठप हो गया। मशीन अनुवाद के इस कठिन समय के बावजूद भी कुछ देश ऐसे भी थे जिन्होंने इस दिशा में शोध कार्य किया। जिनमें इंग्लैण्ड, फ्रांस, जर्मनी, रशिया आदि देश शामिल थे।

सन 1970 में रशियन रिसर्चर्स के द्वारा REVERSO नाम से एक प्रोजेक्ट की शुरुवात की गयी। और इसी दौर में पीटर टोमा जो एक समय में जॉर्ज टाउन में रिसर्चर ग्रुप के सदस्य थे, द्वारा (SYSTRAN-1) सिस्ट्रान-1 रशियन-अंग्रेजी मशीन अनुवाद प्रणाली का विकास किया गया। सन 1970 के दशक में मशीन अनुवाद के क्षेत्र में तकनीकी और वाणिज्यिक सामग्री के अनुवाद एवं कम लागत वाले अनुवाद- तंत्र विकसित करने की मांग की जाने लगी। भूमंडलीकरण की वजह से वैश्विक स्तर पर अनुवाद की बढ़ती मांग ने मशीन अनुवाद प्रणाली की मांग को और बढ़ा दिया। सन 1973 में अमेरिका ने 'लोगोस'(LOGOS) नाम से मशीन अनुवाद सिस्टम का विकास कर लिया था। इसका विकास अमेरिकी वायुसेना के उपकरणों से संबंधित मैनुअलों को अंग्रेजी से वियतनामी में अनुवाद करने के लिए किया गया था। सन 1960 के दशक में टेक्सास विश्वविद्यालय के भाषाविज्ञान केंद्र (Linguistics Research Centre) में जर्मनी से अंग्रेजी में अनुवाद के लिए मेटल (METAL) अनुवाद प्रणाली का विकास किया गया। यह अनुवाद प्रणाली भाषाई दृष्टि से उन्नत थी। इसका विकास फोर्ट्रॉन (FORTRON) में किया गया था। जिसे बाद में सन 1970 के दशक में लिस्प (LISP) में विकसित किया गया। सन 1976 प्रोजेक्ट (TAUM METEO) टॉम मेटो के अंतर्गत मौसम संबंधित जानकारी का अनुवाद करने के लिए विकास किया गया। यह प्रणाली मॉंट्रियल विश्वविद्यालय में विकसित की गयी। यह अनुवाद प्रणाली रोज 80,000 शब्द और प्रति वर्ष लगभग 3 करोड़ शब्दों का अनुवाद करती थी। इसी क्रम में सन 1978 में जापान की एक संस्था फुजित्सु (FUJITSU) ने एक मशीन अनुवाद प्रणाली का विकास किया जिसका नाम एटलस2 (ATLAS2) था। यह एक नियम आधारित प्रणाली थी जो कोरियाई भाषा से जापानी भाषा में अनुवाद करती थी।

मशीन अनुवाद का तीसरा दौर सन 1981 से 2000 तक- 1980 का दशक आते-आते मशीन अनुवाद के क्षेत्र में काफी तरक्की हो चुकी थी अन्य देशों के साथ-साथ जापान में भी मशीन अनुवाद की प्रणालियों का विकास विभिन्न संस्थाओं द्वारा किया गया। सन 1981 में जापान की एक संस्था शार्प (SHARP) ने ऑटोमेटिक ट्रांसलेटर का विकास किया जिसका नाम 'ड्युएट' (DUET) था, जो अंग्रेजी से जापानी भाषा में अनुवाद करता था और नियम आधारित था। सन 1983 में जापान की संस्था NEC ने अल्गोरिथम पर आधारित एक प्रणाली का विकास किया। जिसे PIVOT नाम दिया गया, जो इंटरलिंग्वा पद्धति पर आधारित था। सन 1986 में जापान की ही एक कंपनी OKI3 ने जापानी से अंग्रेजी में अनुवाद करने के लिए एक प्रणाली का विकास किया जिसका नाम (PENSEE) पेन्सी था। यह मुख्य रूप से नियम आधारित प्रणाली थी। इसी समय जापान की ही एक और कंपनी (HITACHI) हिताची में जापानी से अंग्रेजी में अनुवाद करने के लिए एक प्रणाली विकसित की जिसका नाम (HICATS) हिकैट्स(Hitachi Computer Aided Translation System) था। सन 1993 में C-STAR (Consortium for Speech Translation Advanced Research) नाम से एक प्रोजेक्ट की शुरुवात की गयी जो वेब आधारित मशीन अनुवाद प्रणाली के निर्माण में कार्यरत थी। यह त्रय भाषीय मशीन अनुवाद प्रणाली थी जिसमें अंग्रेजी, जर्मनी और जापानी भाषाओं को शामिल किया गया था। यह प्रणाली पर्यटन क्षेत्र के लिए विकसित की गयी थी। सन 1998 में सॉफ्टीसीमो द्वारा रिवर्सो (REVERSO) प्रोजेक्ट को पूरा किया गया। सन 2000 में जापान की लेबोरेटरी ATR ने बहुभाषी मशीन अनुवाद प्रणाली ALPH का निर्माण किया जो जापानी से अंग्रेजी और चीनी से अंग्रेजी भाषा में अनुवाद करती थी। यह उदारण आधारित मशीन अनुवाद प्रणाली थी।

मशीन अनुवाद का चौथा दौर सन 2001 से अब तक- सन 2000 से आगे का दौर वेब प्रौद्योगिकी का युग माना जाता है, इस समय तक www (World wide web) का निर्माण हो चुका था। सन 2005 में गूगल ने मशीन ट्रांसलेशन के लिए पहली वेबसाइट लॉन्च की। सन 2007 में METIS-II नाम से हायब्रिड मशीन अनुवाद प्रणाली का विकास किया गया, जो SMT(Statistical Machine Translation), EBMT(Example Based Machine Translation) और RBMT (Rule Based Machine Translation) मशीन अनुवाद प्रणाली का प्रयोग करके अनुवाद का कार्य करती थी। सन 2008 तक 23% इंटरनेट प्रयोगकर्ता मशीन ट्रांसलेशन का प्रयोग करने लगे थे। और 40% प्रयोगकर्ता इसका प्रयोग करने के बारे में सोच रहे थे। सन 2009 में 30% प्रोफेशनल मशीन ट्रांसलेशन का प्रयोग करने लगे थे और 18% प्रयोक्ता प्रूफ रीडिंग का कार्य करते थे।

मशीन अनुवाद की भाषावैज्ञानिक विधियाँ-

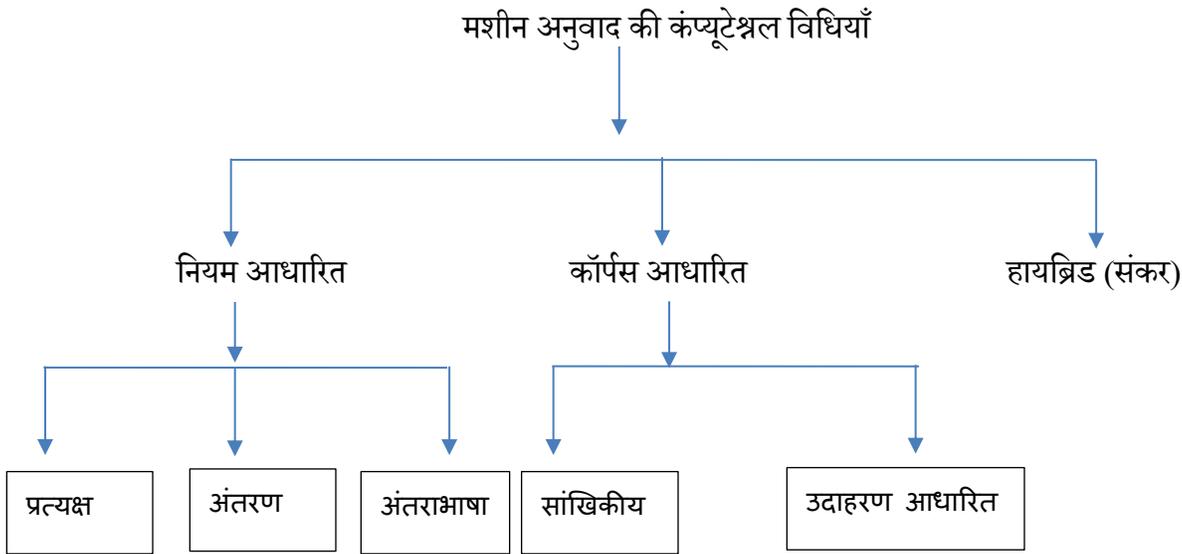


1. **प्रत्यक्ष विधि (Direct Method)** - यह विधि मशीन अनुवाद की सबसे पुरानी विधि है। मशीन अनुवाद के शुरुवाती दौर में इसी विधि का प्रयोग किया जाता था। सबसे पहले इस विधि का प्रयोग सन 1964 जॉर्जटाउन विश्वविद्यालय में देखने को मिलता है। रूसी से अंग्रेजी मशीन अनुवाद के सन्दर्भ में इसका पहली बार प्रयोग किया गया। मुख्यतः ये प्रणाली मूल रूप से दो भाषाओं के शब्दकोश, मोटे नियम और सतही वाक्य सांचों पर आधारित है। इस विधि में ज्यादातर अनुवाद शब्द के स्तर पर होता था। सन 1976 में इस विधि पर आधारित मशीन अनुवाद प्रणाली का विकास किया गया जो SYSTRAN नाम से थी। आज इसमें लगभग 20 से अधिक भाषा युग्मों के अनुवाद की सुविधा है। तथा इसकी शुद्धता 60-70 प्रतिशत के आसपास मानी जाती है। मशीन अनुवाद की पहली पीढ़ी में ज्यादातर अनुवाद का काम प्रत्यक्ष विधि से ही होता था। नासा ने अपोलो – सोयूज सहयोग से संबंधित सामग्री का अनुवाद इसी प्रणाली की सहायता से किया था। सन 1968 में हांगकांग के चीनी विश्वविद्यालय ने गणित और भौतिकी से संबंधित चीनी पत्रिकाओं का अनुवाद करने के लिए इसी विधि पर

आधारित CULT नामक प्रणाली का विकास किया था। इसमें अनुवाद से जुड़ी संदिग्धताओं को दूर करने के लिए प्री-प्रोससिंग और पोस्ट-प्रोसेसिंग का उपयोग किया जाता था।

2. **अंतरण विधि (Transfer Method)** - यह विधि मशीन अनुवाद के दूसरे चरण में प्रयोग की जाने लगी। इसमें शब्द स्तर के अनुवाद के साथ भाषिक नियमों का भी प्रयोग किया गया था। इसमें पार्सिंग(parsing) का प्रयोग किया गया जो वाक्य स्तर पर की जाती है। जिसमें स्रोत भाषा के प्रत्येक शब्द को व्याकरणिक कोटियों के आधार पर निरूपित करके उनका एक आंतरिक ढांचा तैयार किया जाता था और फिर शब्द और संरचना दोनों ही स्तर पर अंतरण किया जाता था। इसमें तीन शब्दकोशों की आवश्यकता पड़ती थी। स्रोत भाषा कोश, लक्ष्य भाषा कोश और द्विभाषी कोश। इस विधि का इस्तेमाल करके सन 1977 में मांट्रियल विश्वविद्यालय द्वारा मौसम संबंधी जानकारी को अंग्रेजी से फ्रेंच में अनुवाद करने के लिए TAUM- METEO प्रणाली का विकास किया था। यह एक स्वचालित प्रणाली थी इसमें केवल 1500 शब्दों का समावेश था। यह दुनिया की सफलतम प्रणालियों में से एक मानी जाती है। अंतरण विधि के अंतर्गत रूसी-जर्मन प्रोटोटाइप सिस्टम पर आधारित SUSY प्रणाली का विकास सन 1984 में किया गया। इसमें अनुवाद की प्रक्रिया तीन चरणों में होती थी। सबसे पहले स्रोत भाषा के वाक्यों का व्याकरणिक विश्लेषण, फिर उसकी सूक्ष्म अमूर्त संरचना निर्माण की जाती थी। दूसरे चरण में स्रोत भाषा की संरचना को लक्ष्य भाषा के समकक्ष संरचना में अंतरित किया जाता था। तीसरे चरण में आउटपुट को लक्ष्य भाषा में पुनर्गठित किया जाता था।
3. **इंटरलिंग्वा- अंतराभाषा (Interlingua Method)** - यह विधि अंतरण विधि से अद्यतन है, अंतरण विधि से एक स्टेप आगे बढ़कर इस इंटरलिंग्वा विधि के बारे में विचार किया गया। इसमें स्रोत भाषा के वाक्यों को पार्सिंग और लक्ष भाषा में जनरेशन के साथ ही इसके अर्थिय पक्ष (Semantic Aspect) का भी उपयोग किया गया। इन अर्थिय पक्ष को आधार मानकर ऐसे कृत्रिम भाषा नियमों को विकसित करने का प्रयास किया गया जो स्रोत भाषा और लक्ष्य भाषा दोनों के लिए काम में आ सके और दोनों के लिए समान हो। इस पद्धति में दोनों भाषाओं के बीच एक अन्य मध्यस्थ भाषा का निर्माण किया जाता है। यह मध्यस्थ भाषा, वास्तविक भाषा न होकर एक ऐसी भाषानिरपेक्ष भाषा होती है, जिसमें सांसारिक ज्ञान को संकल्पनाओं के प्रतिको के रूप में निरूपित किया जाता है। सांसारिक ज्ञान के अंतर्गत विश्व की सभी वस्तुएं, प्राणी, विचार और संबंध आदि संकल्पनाएँ शामिल हैं। ये कोटियाँ कई प्रकार के संबंधों से परस्पर जुड़ी रहती हैं। जैसे अंगांगी संबंध, जाती संबंध, कारण और परिणाम संबंध, पर्यायवाची संबंध आदि। व्याकरण के रूप में विश्वज्ञान का वर्गीकरण और निरूपण करने के लिए एक से अधिक विधियाँ हो सकती हैं परन्तु इसमें हर कोटि को अर्थपरक लक्षणों के रूप में प्रस्तुत किया जाता है। इस विधि का प्रयोग कर भारत में आंग्लभारती, UNL आधारित अंग्रेजी-हिंदी मशीन अनुवाद प्रणाली आदि का विकास किया गया है। आंग्लभारती में (PLIL), (Pseudo Lingua for Indian Languages) का प्रयोग मध्यस्थ भाषा के रूप में भारतीय भाषाओं में अनुवाद करने के लिए किया गया है। तो UNL आधारित अंग्रेजी- हिंदी मशीन अनुवाद प्रणाली में (Universal Networking Language) का प्रयोग मध्यस्थ भाषा के रूप में किया गया है। यह विधि पूर्णतः कृत्रिम बुद्धि के क्षेत्र से ली गयी है। इस विधि में कृत्रिम बुद्धि के अंतर्गत विकसित फ्रेम, स्क्रिप्ट, आदि ज्ञान निरूपण की अनेक तकनीकों का प्रयोग किया गया है।

मशीन अनुवाद की कंप्यूटेश्रल विधियाँ- इन विधियों में निम्नलिखित विधियों का समावेश किया जाता है। जिनको एक आरेख के माध्यम से रेखांकित किया जा सकता है। इसमें नियम आधारित विधि के अंतर्गत प्रत्यक्ष (Direct), अंतरण (Transfer) और अंतराभाषा (Interlingua) इन विधियों को रखा जाता है। तथा कार्पस आधारित विधि में सांखिकीय (Statistic) और उदाहरण आधारित (Example base) को रखा जाता है। मशीन अनुवाद की संकर (Hybrid) विधि में एक से अधिक विधियों का समावेश होता है। जैसे- नियम आधारित+ उदाहरण आधारित, प्रत्यक्ष + सांखिकीय आदि। नीचे दिए गए आरेख के माध्यम से इन्हें विस्तार से समझाने का प्रयास किया गया है।



1. **नियम आधारित (Rule Based)** – मशीन अनुवाद की इस विधि में प्रत्यक्ष, अंतरण और अंतराभाषा इन तीनों विधियों का प्रयोग होता है। नियम आधारित मशीन अनुवाद में भाषा वैज्ञानिक नियमों के उपयोग को विशेष महत्त्व दिया जाता है। इसमें द्विभाषिक कोश के अलावा रूपिमिक विश्लेषण, टैगिंग, पार्सिंग और जनरेशन के नियम होते हैं। नियम आधारित मशीन अनुवाद भाषाओं का वाक्यपरक व्याकरण है। इसमें भाषा के अलग-अलग स्तर पर भाषावैज्ञानिक नियमों के द्वारा भाषा का विश्लेषण किया जाता है और अनुवाद का कार्य संपन्न होता है। जैसे रूपिमिक स्तर पर रूपिमिक विश्लेषण और संस्लेषण की प्रक्रिया होती है। शब्द के स्तर पर टैगिंग की प्रक्रिया होती है, पद के स्तर पर चंकिंग कार्य करता है। वाक्य के स्तर पर पार्सिंग की प्रक्रिया होती है। इस अनुवाद प्रणाली की प्रभावशालिता इस बात पर आधारित होती है कि प्रणाली के लिए बनाए गए नियमों की गुणवत्ता कैसी है। ALPAC रिपोर्ट के बाद नियम आधारित मशीन अनुवाद प्रणाली के विकास को गति मिली। इसके पहले की अनुवाद प्रणालियाँ शब्दानुवाद के रूप में विकसित हुई थीं। जिनसे अनुवाद की गुणवत्ता काफी हद तक प्रभावित हो रही थी।

2. **कॉर्पस आधारित (Corpus Based)-** ¹कॉर्पस आधारित मशीन अनुवाद प्रणाली में मुख्य रूप से सांखिकीय विधि और उदाहरण आधारित विधि का प्रयोग किया जाता है। इसमें बड़ी मात्रा में डाटा का संग्रह किया जाता है। यह डाटा द्विभाषिक होता है। इस पूरे डाटा को शब्द के स्तर पर, पदों के स्तर पर और वाक्यों के स्तर पर टैग किया जाता है। और ऐसे टैग डाटा को मॉडल के आधार पर ट्रेन किया जाता है। यहाँ मशीन लर्निंग की प्रक्रिया की जाती है जिसमें मशीन को कॉर्पस उदाहरणों द्वारा सिखाने का काम किया जाता है। फिर वह डाटा किसी क्षेत्र विशेष का भी हो सकता है। टैग किए हुए द्विभाषिक डाटा को कॉर्पोरा कहते और हैं यह कॉर्पोरा बहुत बड़ी मात्रा में होता है। भारत में E-ILMT प्रोजेक्ट जोकि सांखिकीय विधि पर आधारित है इसका उदाहरण माना जाता है।
3. **संकर विधि (Hybrid Method)-** इस विधि में किसी दो या दो से अधिक मशीन अनुवाद विधियों का समावेश किया जाता है। ²जैसे अंतरण विधि और सांखिकीय विधि, सांखिकीय विधि के प्रयोग से मशीन लर्निंग और कॉर्पस ट्रेनिंग की जा सकती है और अंतरण विधि से भाषिक नियमों का विकास किया जा सकता है। इस संकर विधि से मशीन अनुवाद की गुणवत्ता बढ़ाने के लिए प्रयास किए जा रहे हैं।

संदर्भ ग्रंथ

1. सेठी, हरीशकुमार। (2013)ई- अनुवाद और हिंदी: किताब घर प्रकाशन, नई दिल्ली।
2. झा, गिरीशनाथा मणि, दिवाकरा मिश्र, दिवाकरा 'भारत में भाषा प्रौद्योगिकी: एक सर्वेक्षण', गवेषणा, पृ- 37-47; केंद्रीय हिंदी संस्थान, आगरा।
3. त्रिपाठी, अरिमर्दन कुमार। (2008) 'हिंदी के भाषा प्रौद्योगिकीय उपकरण', राष्ट्रभाषा, पृ। 7-11।
4. दास, ठाकुर। (2008) 'मशीनी अनुवाद : विधियाँ एवं प्रविधियाँ', गवेषणा, पृ। 635-645; केंद्रीय हिंदी संस्थान, आगरा।
5. भाटिया, कैलाश चन्द्र। (1996) 'कंप्यूटर अनुवाद की संभावनाएँ', हिंदी : विकास और संभावनाएँ, पृ। 305-325; सूचना और प्रसारण मंत्रालय, भारत सरकार।
6. विकास, ओमा। (2005) 'हिंदी के विकास में टेक्नोलॉजी का योगदान', गवेषणा, पृ। 617-641; केंद्रीय हिंदी संस्थान, आगरा।
7. सिंह, सूरजभान। जनवरी-मार्च (2005) सूचना-प्रौद्योगिकी और भाषा क्रांति, राजभाषा भारती, पृ। 7-13;
8. <http://sanskritjnu.ac.in/shmt/indexl.jsp>
9. <http://www.cseliitb.ac.in/~pb/indtrend2l.htm>
10. <http://www.scipub.org/fulltext/jcs/jcs6101082-1087l.pdf>
11. http://www.indictransl.in/old/Articles/English/article_src/Indic/ncst2l.pdf
12. <http://www.cfiltiitb.ac.in/Translation-survey/surveyl.pdf>
13. http://projects.lldclupenn.edu/LDC_Institute/Visitors/Agrawalppt

¹ Nabeel T. Alsohybe¹, Machine-Translation History and Evolution: Survey for Arabic-English Translations, *Department of Information Technology, Sana'a University, Sana'a, Yemen*, page no.7.

² Mohamed Amine Chérâgui, Theoretical Overview of Machine translation, African University, Adrar, Algeria, page no.165.

14. [http://tdil- dclin/](http://tdil-dclin/)
15. <http://www.tdillmit.gov.in/>
16. <http://www.cdacmumbai.in/e-ilmt>
17. http://projects.ldc.lupenn.edu/LDC_Institute/Visitors/Jha.pdf
18. <http://sanskrit.jnula.ac.in/index.jsp>
19. <http://sanskrit.jnula.ac.in/ilcian>

