

खंड

4

मशीनी अनुवाद

इकाई 12	
मशीनी अनुवाद : स्वरूप एवं विकास	189

इकाई 13	
मशीनी अनुवाद : पद्धतियाँ और प्रकार	205

इकाई 14	
मशीनी अनुवाद : प्रक्रिया एवं चुनौतियाँ	228

खंड परिचय 4

‘अनुवाद प्रक्रिया’ अनुवाद अध्ययन में एम.ए. कार्यक्रम का ग्यारहवाँ पाठ्यक्रम है। ‘मशीनी अनुवाद’ इस पाठ्यक्रम का चौथा खंड है। इस खंड का उद्देश्य शिक्षार्थियों को मशीनी अनुवाद की संकल्पना, मशीनी अनुवाद के आरंभिक विकास और वर्तमान स्वरूप से परिचित कराना है। इस खंड की विभिन्न इकाइयों में मशीनी अनुवाद का परिचय एवं क्षेत्र तथा विकास, मशीनी अनुवाद के प्रकार एवं प्रविधियों तथा वर्तमान में मशीनी अनुवाद की सीमाओं तथा उसके भविष्य पर चर्चा की जाएगी। इसके अध्ययन के उपरांत शिक्षार्थी मशीनी अनुवाद के प्रयोजन क्षेत्रों तथा उसकी उपयोगिता और व्यावहारिक समस्याओं को भी जान पाएँगे। शिक्षार्थी, यह भी समझ सकेंगे कि मशीनी अनुवाद में भाषाविज्ञानियों, अनुवादकों और कंप्यूटर विज्ञानियों के सामूहिक प्रयास से ही व्यावहारिक सफलता प्राप्त की जा सकती है। मशीनी अनुवाद में मानवीय भूमिका को भी शिक्षार्थी इस इकाई के अध्ययन क्रम में ठीक-ठीक समझ सकेंगे। कुल मिलाकर यह इकाई मशीनी अनुवाद को वर्तमान परिप्रेक्ष्य में समझने तथा उसके योगदान को सही परिप्रेक्ष्य में मूल्यांकित करने में सहायक होगी। यह खंड एक प्रकार से मशीनी अनुवाद पर निबंधात्मक संग्रह है।

इकाई 11 ‘मशीनी अनुवाद : स्वरूप एवं विकास’ : इस खंड की प्रथम इकाई में मशीनी अनुवाद के स्वरूप की व्याख्या कर उसकी आवश्यकता पर प्रकाश डाला गया है। मशीनी अनुवाद प्रक्रिया का संकेत करते हुए इस इकाई में मशीनी अनुवाद के प्रकारों, मशीनी अनुवाद की अब तक की विकास यात्रा और भारत में मशीनी अनुवाद में हुई प्रगति का एक जायजा आप इस इकाई में पाएँगे। विशेष उल्लेख के रूप में भारतीय भाषाओं के संदर्भ में मशीनी अनुवाद पर भी आप जानकारी हासिल करेंगे।

इकाई 12 ‘मशीनी अनुवाद : पद्धतियाँ और प्रकार’ : प्रथम इकाई में मशीनी अनुवाद का प्रारंभिक परिचय प्राप्त करने के उपरांत शिक्षार्थी खंड की इस दूसरी इकाई (इकाई 12) में मशीनी अनुवाद की विभिन्न पद्धतियों के बारे में जानेंगे। मशीनी अनुवाद की संभावनाओं का भी जायजा इस इकाई में प्रस्तुत किया गया है। मशीनी अनुवाद के संदर्भ में पूर्ण मशीनी अनुवाद तथा अर्ध मशीनी अनुवाद जैसे पक्ष भी चर्चा में सम्मिलित किए गए हैं। मशीनी अनुवाद के संदर्भ में अनुवादकों की श्रेणियों पर भी कुछ जानकारी आप इस इकाई में पाएँगे।

इकाई 13 ‘मशीनी अनुवाद : प्रक्रिया और चुनौतियाँ’ : पिछली दो इकाइयों में मशीनी अनुवाद के प्रारंभिक परिचय तथा मशीनी अनुवाद की पद्धतियों और मशीनी अनुवाद की पूर्णता संबंधी प्रश्नों पर विचार के उपरांत प्रस्तुत इकाई में अनुवाद प्रक्रिया पर विस्तार से चर्चा की जाएगी। साथ ही कुछ नई उभरती हुई पद्धतियों पर भी इस इकाई में प्रकाश डाला जाएगा। शिक्षार्थी इस इकाई के अध्ययनोपरांत यह भी जान पाएँगे कि मशीनी अनुवाद प्रणाली एक विकासोन्मुख प्रक्रिया है तथा इसमें लगातार नए-नए आयाम जुड़ते जा रहे हैं जिनमें अनेक हित-पक्ष सहभागी हैं। मशीनी अनुवाद संबंधी बहु-आयामी चर्चा के उपरांत मशीनी अनुवाद की पूर्णता, विकास तथा भविष्य संबंधी अनेक प्रश्नों का शिक्षार्थियों के मन में उठना स्वभाविक है। अतः खंड की इस अंतिम इकाई में मशीनी अनुवाद से जुड़े मुद्दों, मशीनी प्रणाली निर्माण की भाषिक और तकनीकी जटिलताओं और मशीनी अनुवाद की संरचनाओं पर भी विमर्श किया गया है। मशीनी अनुवाद में आने वाली जटिलताओं के लिए संभावित समाधान खोजने संबंधी प्रयासों पर भी आप यहाँ चर्चा करेंगे। बहुभाषी संसार में मशीनी अनुवाद की अपार संभावनाओं और संप्रेषण के रूप में महत्वपूर्ण साधन स्वरूप इस प्रणाली के महत्व पर भी चर्चा की गई है। यह भी समझने का प्रयास किया गया कि मशीनी अनुवाद प्रणाली के विकास में मानवीय भूमिका अनेक स्तरों पर केंद्रीय हो जाती है।

यह खंड एक प्रकार से मशीनी अनुवाद से जुड़े मुद्दों पर एक निबंध-संग्रह जैसा ही है, जिसमें कुछ मामलों में एकाधिक मत दिखाई देंगे; परंतु वर्तमान मशीनी अनुवाद परिदृश्य में ये सब पक्ष हितभागी हैं। यह संकेत करते हैं कि मशीनी अनुवाद एक संपूर्ण अनुवाद प्रणाली के रूप में स्थापित हो सकेगी या नहीं इसका उत्तर भविष्य के गर्भ में है तथापि इस दिशा में गंभीर प्रयास जारी है। अध्ययन के उपरांत आप यह अनुभव करेंगे कि मशीनी अनुवाद प्रणाली वर्तमान समाज की आवश्यकता है और इसमें मशीन एवं मानव दोनों की सहभागिता अपेक्षित है।

इकाई-12 मशीनी अनुवाद : स्वरूप एवं विकास

इकाई की रूपरेखा

- 12.0 उद्देश्य
- 12.1 प्रस्तावना
- 12.2 मशीनी अनुवाद : अर्थ एवं परिभाषा
 - 12.2.1 मशीनी अनुवाद का अर्थ
 - 12.2.2 मशीनी अनुवाद की परिभाषा
 - 12.2.3 मशीनी अनुवाद की आवश्यकता
 - 12.2.4 मशीनी अनुवाद के प्रकार
 - 12.2.5 मशीनी अनुवाद की प्रतिक्रियाएँ
- 12.3 मशीनी अनुवाद का विकास : पश्चिम में
 - 12.3.1 प्रारंभिक तथा अत्युत्साह का युग
 - 12.3.2 शून्य काल
 - 12.3.3 पुनरुत्साह का युग
 - 12.3.4 कुछ मशीनी अनुवाद तंत्र
- 12.4 मशीनी अनुवाद का विकास : भारत में
 - 12.4.1 भारतीय भाषाओं में अनुवाद
 - 12.4.2 अंग्रेजी-हिंदी मशीनी अनुवाद
- 12.5 इंटरनेट और मशीनी अनुवाद
- 12.6 सारांश
- 12.7 अभ्यास के लिए प्रश्ना
- 12.8 शब्दावली
- 12.9 कुछ उपयोगी पुस्तकें

12.0 उद्देश्य

प्रस्तुत इकाई को पढ़ने के बाद आप :

- मशीनी अनुवाद क्या है, यह समझ सकेंगे,
 - मशीनी अनुवाद की आवश्यकता और प्रक्रिया के बारे में बता सकेंगे,
 - मशीनी अनुवाद के प्रकारों के बारे में समझ सकेंगे,
 - मशीनी अनुवाद की पश्चिम में विकास यात्रा के बारे में चर्चा कर सकेंगे; तथा
 - भारत में मशीनी अनुवाद के क्षेत्र में क्या-क्या कार्य हुए हैं उनके बारे में जान सकेंगे।
-

12.1 प्रस्तावना

पिछले खंडों की विभिन्न इकाइयों में आप अनुवाद के स्वरूप, महत्व एवं परिभाषा के बारे में पढ़ चुके हैं। आप यह जानते हैं कि अनुवाद एक भाषिक प्रक्रिया है जिसमें स्रोतभाषा में दिए गए पाठ या मूलपाठ की सूचना को

लक्ष्यभाषा में रूपांतरित किया जाता है। ऐसा करने में अनुवाद की तीन प्रक्रियाओं का सहारा लिया जाता है 'ये प्रक्रियाएँ-विश्लेषण, अंतरण और पुनर्गठन हैं। मूलपाठ को समझने और उसमें कथ्य को जानने में विश्लेषण की प्रक्रिया हमारी सहायता करती है। यह कार्य स्रोतभाषा के व्याकरणिक एवं प्रयोग के नियमों के आधार पर किया जाता है। इस प्रक्रिया के दौरान मानव अनुवादक स्रोतभाषा के पाठ में निहित संदेश अथवा सूचना को समझता है और 'कथ्य' तक पहुँचने की चेष्टा करता है। मानव अनुवाद में यह कार्य मानव कंप्यूटर अर्थात् मानव अनुवादक का मस्तिष्क करता है और मशीनी अनुवाद में यह कार्य मशीन करती है। अंतरण की प्रक्रिया के अंतर्गत कथ्य का स्रोतभाषा से लक्ष्यभाषा में द्विभाषी कोश की सहायता से अनुवाद किया जाता है अर्थात् अंतरण कोश की सहायता से अनुवाद तत्वों के लक्ष्यभाषा के पर्याय निश्चित किए जाते हैं। ये पर्याय शब्दों तथा बहुशाब्दिक अभिव्यक्तियों के स्तर के हो सकते हैं। यहाँ बहुशाब्दिक अभिव्यक्तियों से तात्पर्य पदबंध, प्रयोग, मुहावरे, लोकोक्तियाँ आदि से है।

'समायोजन' या 'संश्लेषण' की प्रक्रिया के अंतर्गत अंतरित तत्वों को लक्ष्यभाषा की भाषिक प्रकृति के अनुसार समायोजित किया जाता है। यह इसलिए भी आवश्यक हो जाता है कि कई बार स्रोतभाषा और लक्ष्यभाषा की भाषा संरचना में अंतर होता है; जैसे-अंग्रेजी में शब्द-क्रम (Word order) पदबंधों की संरचना तथा उनके विन्यास नियमों में हिंदी के पदक्रम, पदबंधों की संरचना आदि की अपेक्षा अंतर होता है।

12.2 मशीनी अनुवाद का अर्थ और परिभाषा

12.2.1 मशीनी अनुवाद का अर्थ

मशीनी अनुवाद का अर्थ है-मशीन या कंप्यूटर प्रणाली के माध्यम से एक भाषा से दूसरी भाषा में अनुवाद करना। अनुवाद की तीन प्रक्रियाएँ सर्वस्वीकृत हैं-विश्लेषण, अंतरण एवं पुनर्गठन। मशीनी अनुवाद में ये तीनों प्रक्रियाएँ कंप्यूटर द्वारा संपन्न होती हैं। ये प्रक्रियाएँ हैं (पाठ का) विश्लेषण, अंतरवर्ती प्रक्रिया या अंतरण और प्रजनन (generation) या समायोजना। ऐसा माना जाता है कि हम अनुवाद की जाने वाली सामग्री को कंप्यूटर में इनपुट करेंगे और कंप्यूटर प्रणाली में पहले से ही तैयार विश्लेषण प्रक्रिया या पार्सिंग प्रक्रिया द्वारा विश्लेषित कर उसे अनुवाद तत्वों में विखंडित किया जाएगा। फिर अंतरण प्रक्रिया से संबद्ध अंतरण कोश के माध्यम से उसके समतुल्य अनूदित शब्दों, पदबंधों के रूप में प्रस्तुत किया जाएगा। फिर समायोजन या प्रजनन (जेनरेशन) की प्रक्रिया से गुजरते हुए उन अनूदित तत्वों को लक्ष्यभाषा की व्याकरणिक संरचना के अनुसार समायोजित कर लक्ष्यभाषा में अनुवाद के रूप में प्रस्तुत किया जाएगा।

12.2.2 मशीनी अनुवाद की परिभाषा

मशीनी अनुवाद की परिभाषा और मानव अनुवाद की परिभाषा में बिल्कुल साम्य है। दोनों अनुवाद प्रकारों (मानव अनुवाद एवं मशीनी अनुवाद) के अनुसार अनुवाद दो भिन्न भाषाओं के बीच होने वाली एक भाषिक प्रक्रिया है। लक्ष्यभाषा के पाठ को अनूदित पाठ कहा जाता है। स्रोतभाषा और लक्ष्यभाषा के पाठ का मुख्य आधार उसका निहित संदेश होता है। दूसरे शब्दों में, दोनों पाठ निहित संदेश या अर्थ को व्यक्त करते हैं। मशीनी अनुवाद की सार्थकता इसी में है कि मशीन के लिए अनुवाद कार्यक्रम इस प्रकार से तैयार किया जाए कि उससे मशीन सरलतापूर्वक और अच्छी गुणवत्ता (क्वालिटी) वाला अनुवाद कर सके।

12.2.3 मशीनी अनुवाद की आवश्यकता

अनुवाद के महत्व और उसकी प्रासंगिकता के बारे में आप पिछली इकाइयों में पढ़ चुके हैं। अनुवाद की परंपरा बहुत पुरानी है और उसका महत्व भी स्वयंसिद्ध है। सामाजिक, सांस्कृतिक, राजनैतिक, व्यापारिक एवं वाणिज्यिक उद्देश्यों के लिए हमें अनुवाद की सहायता की आवश्यकता होती है।

एक समय था जब अनुवाद को वैयक्तिक रुचि के आधार पर किया जाता है। ऐसा अनुवाद 'स्वांतःसुखाय' होता था। इसके अंतर्गत धार्मिक ग्रंथों का अनुवाद आता है। आधुनिक युग में अनुवाद एक संगठित व्यवसाय के रूप में उभर कर आया है। यह आधुनिक युग की सामाजिक, राजनैतिक और आर्थिक आवश्यकताओं के द्वारा प्रेरित

है। आज अनुवाद की आवश्यकता इतनी अधिक बढ़ गई है कि केवल मानव अनुवादकों से यह संभव नहीं रह गया है।

प्रत्येक देश ज्ञान-विज्ञान के क्षेत्र में हुए अनुसंधान एवं विकास से दूसरी भाषाओं की ऐतिहासिक, सांस्कृतिक एवं वैज्ञानिक धरोहर की गहनता और व्यापकता से अपनी भाषाओं को संपन्न करना चाहता है। मानव अनुवाद एक धीमी प्रक्रिया है। मानव अनुवादक सीमित मात्रा में ही अनुवाद कर सकता है जबकि मशीन तीव्र गति से सैकड़ों पृष्ठों का अनुवाद कुछ ही समय में कर सकती है। इस प्रकार मशीनी अनुवाद प्रणाली हमें बड़ी मात्रा में समरूपी (uniformed) अनुवाद करने का त्वरित साधन उपलब्ध कराती है।

आज हम सूचना प्रौद्योगिकी के युग में जी रहे हैं। नई-नई सूचनाओं एवं ज्ञान-विज्ञान का विकास प्रति क्षण हो रहा है। उन्हें अपनी भाषा में लाने का तीव्रतम उपाय मशीनी अनुवाद ही है। भूमंडलीकरण के इस युग में जहाँ परस्पर दूरियाँ कम हो रही हैं, वहीं एक दूसरे की आवश्यकताओं की पूर्ति के लिए अनुवाद की आवश्यकता भी काफी बढ़ गई है। बहुराष्ट्रीय कंपनियाँ अपने उत्पाद उपभोक्ताओं तक उनकी भाषा में लाना चाहती हैं। मानव क्षमताओं की सीमितता और उपलब्धता के कारण उनकी यह आवश्यकता मशीनी अनुवाद से ही पूरी की जा सकती है।

अब अनुवाद विभिन्न राष्ट्रों के राजनैतिक-सांस्कृतिक-आर्थिक संबंधों का आधार बन रहा है। विभिन्न प्रकार के राजनयिक दस्तावेज दोनों देशों की भाषाओं में तैयार किए जाते हैं। ऐसा केवल अनुवाद के कारण संभव हो पाता है। संयुक्त राष्ट्र संघ में जो भी वक्तव्य दिए जाते हैं, उन्हें संयुक्त राष्ट्रसंघ की सभी भाषाओं में मशीन अनुवाद तंत्र द्वारा अनूदित कर उपलब्ध कराया जाता है। इस प्रकार विश्व मंच पर भी अनुवाद की आवश्यकता अनुभव की जाती है। भारतीय संसद में भी इस प्रकार के प्रावधान किए जा रहे हैं। इस प्रकार आप यह निष्कर्ष निकाल सकते हैं कि इन सभी कारणों से अनुवाद की जाने वाली सामग्री की मात्रा इतनी अधिक है कि मानव अनुवादक के सहारे इस विशाल सामग्री का अनुवाद निर्धारित समय-सीमा में कर पाना कठिन है, ऐसी स्थिति में एक ही उपाय रह जाता है कि मशीनी अनुवाद का सहारा लिया जाए और उसे निरंतर शोध एवं विकास प्रक्रिया से विश्वसनीय बनाया जाए।

12.2.4 मशीनी अनुवाद के प्रकार

मानव तथा मशीन के अंतःसंबंध को दृष्टि में रखते हुए अनुवाद के सामान्यतया तीन प्रकार माने जाते हैं-

- मशीन साधित मानव अनुवाद (Machine Aided Human Translation)
- मानव साधित मशीनी अनुवाद (Human Aided Machine Translation)
- मशीनी अनुवाद (Machine Translation)

मशीन साधित मानव अनुवाद में मशीन मानव की सहायता करती है। इसी क्रम में अनुवाद स्मृतियों (Translation Memories) का विकास शामिल है, जिसके अंतर्गत कंप्यूटरीकृत शब्दकोशों, पदबंध कोशों, बहुशाब्दिक अभिव्यक्तियों का निर्माण आता है, जिससे अधिक समय नष्ट किए बिना मानव अनुवादक मशीन की सहायता लेकर अनुवाद कार्य को तीव्र गति से कर सकता है।

दूसरा प्रकार मानव साधित मशीनी अनुवाद है। इसमें विश्लेषण करते हुए जेनेरेशन प्रक्रिया के लिए कई कार्यक्रम विकसित किए जाते हैं। इनमें टैगिंग (व्याकरणिक कोटियों की पहचान), एनोटेशन (क्रिया के साथ वाक्य के अन्य घटकों (कर्ता, कर्म, कारक संबंधों आदि की पहचान), मिश्र वाक्यों के अंतर्गत गौण उपवाक्यों का प्रधान उपवाक्य से संबंध आदि कार्यक्रम आते हैं। इनकी सहायता से मशीन को स्रोतभाषा के पाठ विश्लेषण (पार्सिंग) में सहायता मिलती है।

अंतिम प्रकार मशीनी अनुवाद है जिसमें विश्लेषण, अंतरण तथा समायोजन (जेनेरेशन) की सारी प्रक्रियाएँ मशीन करती हैं। तेरहवीं इकाई में इन प्रकारों की विस्तृत जानकारी दी जाएगी।

12.2.5 मशीनी अनुवाद की प्रक्रियाएँ

मशीनी अनुवाद से तात्पर्य है कि अनुवाद की उपर्युक्त तीनों प्रक्रियाएँ मशीन यानी कंप्यूटर संपन्न करे। हम कंप्यूटर में स्रोतभाषा के पाठ की सामग्री का इनपुट करें और आउटपुट के रूप में उससे लक्ष्यभाषा में अनूदित सामग्री प्राप्त कर लें। कंप्यूटर में स्रोतभाषा की व्यवस्था के आधार पर विश्लेषण प्रक्रिया को संपन्न करने की क्षमता हो, जिसके आधार पर मशीन स्रोतभाषा पाठ को पढ़कर उसे समझ सके और उसमें से अनुवाद तत्व निकाल सके। मशीनी अनुवाद में विश्लेषण की प्रक्रिया को 'पार्सिंग' कहा जाता है।

अंतरण के अंतर्गत स्रोतभाषा से लक्ष्यभाषा के शब्दों, अभिव्यक्तियों आदि के आधार पर अंतरण कोश तैयार किया जाता है, जिसके आधार पर स्रोतभाषा की संरचनात्मक सामग्री का लक्ष्यभाषा की समरूपी संरचना में अंतरण किया जाता है।

समायोजन की प्रक्रिया के अंतर्गत स्रोतभाषा की अंतरिम सामग्री का लक्ष्यभाषा की भाषिक संरचना के आधार पर समायोजन किया जाता है। इसे मशीनी अनुवाद की भाषा में 'जेनरेशन' कहते हैं।

इस प्रकार हम देखते हैं कि मानव अनुवाद तथा मशीनी अनुवाद की प्रक्रियाओं में समानता है-फर्क इतना है कि मशीनी अनुवाद में ये तीनों प्रक्रियाएँ पूर्णतया मशीन द्वारा संपन्न होती हैं। मशीनी अनुवाद के ये प्रक्रियात्मक चरण लंबे अनुसंधान और श्रम से साधित हुए हैं। अतः मशीनी अनुवाद के इतिहास को जानना आपको रुचिकर लगेगा।

12.3 मशीनी अनुवाद का विकास : पश्चिम में

यदि मशीनी अनुवाद के विकास से संबंधित साहित्य पर नजर डाली जाए तो ज्ञात होता है कि इस दिशा में पिछले सात-आठ दशकों से व्यवस्थित प्रयास किए जा रहे हैं। जैसे मशीनी अनुवाद के विचार के बीज 17वीं शताब्दी में, सन् 1629 में, रेने देकार्त और लाइबनिज के प्रस्तावों में मिल जाते हैं। बाद में सन् 1933 में फ्रांसीसी वैज्ञानिक जार्ज आर्त्सूनी ने स्वचालित द्विभाषी कोश और 1939 में रूसी वैज्ञानिक पीटर त्रशेयांस्की ने द्विभाषी कोश तथा एस्परान्तो पर आधारित विभिन्न भाषाओं के बीच व्याकरणिक संबंधों के समाधान की विधि प्रस्तुत कर इस दिशा में पहल की थी। त्रोयांस्की अनुवाद प्रक्रिया को तीन चरणों में विभाजित करने वाले पहले व्यक्ति थे। इसके अतिरिक्त माध्यमिक भाषा की संकल्पना के भी जनक माने जाते हैं। उनके मतानुसार सभी भाषाओं की तार्किक संरचना समान होती है, जबकि उनकी कोशीय और व्याकरणिक संरचना अलग-अलग होती है। यही समान तार्किक संरचना सार्वभाषिक तार्किक माध्यमिक भाषा के माध्यम से एक भाषा से दूसरी भाषा में अनुवाद करने में सहायक होती है। त्रोयांस्की के अनुवाद तंत्र का जुलाई 1944 में मास्को में इस संबंध में एक सफल प्रदर्शन भी किया गया था।

प्राकृत भाषाओं के अनुवाद के क्षेत्र में अंकीय कंप्यूटर के प्रयोग का प्रस्ताव 1946 में ए.डी. बूथ एवं अन्य वैज्ञानिकों ने दे दिया था। बूथ ने ही कंप्यूटरीकृत शब्दकोश तैयार करने का विचार प्रस्तुत किया था। इसी युग में पूव-संपादन और पश्च-संपादन की संकल्पनाओं की स्थापना हो चुकी थी। मशीनी अनुवाद का प्रारंभ वास्तव में सन् 1947 में वारेन वीवर के लेख 'ऑन ट्रांसलेशन' से माना जाता है। इसमें बूथ एवं रिशेंस के कार्यों की समीक्षा करते हुए मशीनी अनुवाद की तुलना 'कोड ब्रेकिंग' से की गई थी। अंकीय कंप्यूटर दूसरे विश्वयुद्ध में सफलतापूर्वक काम कर चुके थे। वीवर ने शब्दों की अनेकार्थता की समस्या को सुलझाने के लिए सांख्यिकी-आर्थी अध्ययनों का सुझाव दिया। इसी समय आई.बी.एम तथा जार्जटाउन विश्वविद्यालय के एक दल ने मिलकर एक खिलौना मशीनी अनुवाद तंत्र का नमूना प्रदर्शित किया। इसमें मात्र 250 शब्दों पर आधारित सावधानी पूर्वक 59 रूसी वाक्यों का अंग्रेजी में अनुवाद प्रदर्शित किया गया था। इस प्रदर्शन ने मशीनी अनुवाद के क्षेत्र में एक प्रकार से चिंगारी का काम किया।

सन् 1950 में एक चीनी भाषाविद् ई. राइफलर ने पहली बार मशीनी अनुवाद के क्षेत्र में अनुवाद की प्रक्रियाओं के मशीनीकरण का प्रस्ताव प्रस्तुत किया। मशीनी अनुवाद के क्षेत्र में एक अन्य उत्साही शोधकर्ता तर्कशास्त्री बार-हिलेल थे, जिन्होंने सन् 1954 में एम.आई.टी (Massachusetts Institute of Technology) में मशीनी

अनुवाद पर पहली संगोष्ठी का आयोजन कर कंप्यूटिंग में वाक्यविन्यासीय विश्लेषण के मशीनीकरण का विचार प्रस्तुत किया।

पश्चिम में मशीनी अनुवाद का व्यवस्थित विकास सन् 1954 जार्ज टाउन विश्वविद्यालय, वाशिंगटन, अमेरिका में 'मशीन ट्रांसलेशन तथा लैंग्वेज प्रोजेक्ट' से माना जाता है। इसके अंतर्गत 60 रूसी वाक्यों का अंग्रेजी में मशीनी अनुवाद सफलतापूर्वक करके दिखाया गया था। विर्कबेक कालेज (लंदन विश्वविद्यालय) में सन् 1954 में अंग्रेजी से फ्रेंच में मशीनी अनुवाद का प्रदर्शन किया गया। यह अनुभव इतना सफल माना गया कि मशीनी अनुवाद के क्षेत्र में अनुसंधान एवं विकास के लिए अपार अनुदान राशि प्राप्त होनी शुरू हो गई। ऐसा लगने लगा था कि मशीनी अनुवाद की समस्या का सफल समाधान अगले तीन से पाँच वर्षों के बीच प्राप्त हो जाएगा।

12.3.1 प्रारंभिक युग

वीवर से लेकर सन् 1954 तक मशीनी अनुवाद के युग को प्रारंभिक युग की संज्ञा दी जाती है। मशीनी अनुवाद का प्रारंभ रूसी-अंग्रेजी एवं अंग्रेजी-रूसी सामग्री के त्वरित अनुवाद की सामरिक आवश्यकता पर आधारित था। सन् 1950-60 का समय अंतरिक्ष में जाने की अमेरिका तथा सोवियत संघ की होड़ का समय था। दोनों देश एक दूसरे की सूचनाओं का अपनी-अपनी भाषा में शीघ्रातिशीघ्र अनुवाद करना चाहते थे और उनका उपयोग अपने-अपने अंतरिक्ष विकास कार्यक्रम में करना चाहते थे। इसी उद्देश्य की पूर्ति के लिए मशीनी अनुवाद के विकास के लिए असीमित अनुदान का रास्ता खुल गया। परिणामस्वरूप, यूरोप तथा अमेरिका में कई विश्वविद्यालयों तथा संस्थानों में मशीनी अनुवाद की कई शोध परियोजनाएँ शुरू की गईं। इस युग को प्रारंभिक युग या अत्युत्साह युग की संज्ञा दी जाती है।

इसी युग में कई अनुवाद तंत्र विकसित किए गए। संयुक्त राज्य अमेरिका में वायु सेना तथा परमाणु ऊर्जा आयोग आई.बी.एम एवं वाशिंगटन विश्वविद्यालय द्वारा विकसित अनुवाद तंत्र का प्रयोग कर रहे थे। इटली में यूराटोम भी जार्जटाउन विश्वविद्यालय द्वारा विकसित अनुवाद तंत्र का प्रयोग कर रहा था। हालाँकि इन तंत्रों की क्षमता बहुत ही कम थी फिर भी अनुवाद गति की दृष्टि से ये अपने प्रयोगकर्ताओं की आवश्यकताओं को पूरा करते थे। पचास के दशक के अंत में अमेरिकी सरकार ने बार हिलेल को उच्च गुणवत्ता वाले मशीनी अनुवाद की संभाव्यता के बारे में अनुसंधान करने के लिए कहा। उन्होंने आर्थी स्तर पर अनेकार्थकता को इसका मुख्य कारण माना।

साठ के दशक में सोवियत संघ तथा अमेरिका में रूसी-अंग्रेजी भाषा वर्ग पर काम किया गया। यह अनुवाद तकनीकी पत्रिकाओं में प्रकाशित वैज्ञानिक तथा तकनीकी दस्तावेजों तक सीमित था। इस अनुवाद से लेख में निहित सूचना को समझने में मदद मिलती थी।

इस युग में मशीन से सामान्य उद्देश्यीय एवं उच्च गुणवत्ता वाले मशीनी अनुवाद (General Purpose High Quality Machine Translation) की अपेक्षा की जा रही थी। हम सभी जानते हैं कि उस समय कंप्यूटर की शक्ति और क्षमता आज की तुलना में बहुत सीमित थी। इस क्षेत्र में अपार धनराशि नियोजित किए जाने के बावजूद प्राप्त मशीनी अनुवाद की गुणवत्ता मानव अनुवाद की अपेक्षा बहुत ही कम थी। यहाँ यह भी संकेत करना असमीचीन न होगा कि इस युग में मशीनी अनुवाद को मूलरूप से एक शुद्ध इंजीनियरी प्रक्रिया के रूप में देखा गया था तथा उसके व्याकरणिक पक्ष की उपेक्षा की गई थी। अपेक्षित परिणाम प्राप्त न होने का एक यह भी कारण था।

12.3.2 शून्यकाल

- सन् 1954 से 1964 तक मशीनी अनुवाद के क्षेत्र में किए गए कार्य तथा दावों के परिप्रेक्ष्य में अमेरिका में एक सात सदस्यीय समिति का गठन किया गया जिसे एल्पाक (ALPAC-Automatic Language Processing Advisory Committee) समिति के नाम से जाना जाता है। चूँकि अपार धनराशि व्यय करने पर भी तत्कालीन मशीनी अनुवाद की गुणवत्ता मानव अनुवाद की तुलना में बहुत ही कम थी, अतः इस संबंध में तत्कालीन मशीनी अनुवाद की स्थिति का अध्ययन करने और उसके संबंध में आवश्यक सुझाव देने का कार्य इस समिति को सौंपा गया।

समिति ने लगभग दो वर्ष अध्ययन करने के बाद सन् 1966 में अपनी रिपोर्ट प्रस्तुत की, जिसकी मुख्य सिफारिशें इस प्रकार थीं-

- उपलब्ध मशीनी अनुवाद की गुणवत्ता बहुत की कम और चिंताजनक है।
- आज की (तत्कालीन) स्थिति में मशीनी अनुवाद संभव नहीं है।
- मशीनी अनुवाद के लिए आधारभूत भाषा विश्लेषण सिद्धांतों का अभाव है। भाषा विश्लेषण माडलों के विकास पर बल देना आवश्यक है।
- समिति ने अनुवादकों की सहायता के लिए स्वचालित कोशों के विकास की सिफारिश की।

एल्फैक समिति की रिपोर्ट का प्रभाव अमेरिका में पूर्ण रूप से तथा सोवियत संघ और इंग्लैंड में आंशिक रूप से हुआ। इसके बाद अमेरिका में मशीनी अनुवाद के विकास के लिए सहायता बिलकुल बंद कर दी गई सिस्ट्रान को छोड़कर अमेरिका में अन्य मशीनी अनुवाद संबंधी काम बंद कर दिए गए। कनाडा, फ्रांस तथा जर्मनी में इस दिशा में शोध कार्य चलता रहा। व्याकरणिक विश्लेषण के सिद्धांतों के विकास पर बल दिया जाने लगा। सन् 1966 से संगणक भाषाविज्ञान के क्षेत्र में कई भाषा विश्लेषण मॉडलों का विकास हुआ जिसने मशीनी अनुवाद के भावी विकास की नींव रखी। किसी विशेष सरकारी एवं निजी प्रयास की अनुपस्थिति में उदासीनता के कारण अन्यमनस्कता की स्थिति होने से 1966-76 का समय मशीनी अनुवाद के इतिहास में 'शून्यकाल' के नाम से जाना जाता है।

12.3.3 पुनरुत्साह का युग

एल्फैक समिति की रिपोर्ट आने के बाद अमेरिका में तो इस क्षेत्र में काम बिलकुल ठप्प हो गया किंतु यूरोप तथा अन्य देशों में छिट-पुट काम होते रहे। भाषा विश्लेषण सिद्धांतों के विकास के लिए धन उपलब्ध होना शुरू हो गया। फलस्वरूप, कई भाषा विश्लेषण मॉडलों का विकास हुआ जिसने आगे चलकर इस कार्य के लिए अत्यावश्यक भाषा-विश्लेषण के लिए आधार प्रस्तुत किया और भाषा विश्लेषकों के विकास का मार्ग प्रशस्त किया।

सन् 1976 में कनाडा में कनाडा ब्राडकस्टिंग कार्पोरेशन (सी.बी.सी) ने टॉम-मैटो (Taum-Meteo) के अनुवाद तंत्र का विकास तथा उसका सफल प्रदर्शन करके मशीनी अनुवाद के क्षेत्र में पुनरुत्साह के युग की शुरुआत की। इस अनुवाद तंत्र में मौसम विज्ञान संबंधी सूचनाओं का अंग्रेजी से फ्रेंच तथा फ्रेंच से अंग्रेजी में मशीनी द्वारा अनुवाद किया गया। यह तंत्र अत्यधिक सफल रहा और इसने मशीनी अनुवाद के मृतप्राय विषय में जान फूँक दी। यह अनुवाद तंत्र मौसम की सूचना संबंधी अत्यंत सीमित प्रयुक्ति पर आधारित था। इसने प्रयुक्तिपरक अनुवाद तंत्रों के विकास का मार्ग प्रशस्त कर दिया। अब 'फोकस' सामान्य उद्देश्य के अनुवाद तंत्रों के विकास पर न होकर सीमित प्रयोग-क्षेत्र के अनुवाद तंत्रों पर हो गया। इससे मशीनी अनुवाद के क्षेत्र में पुनरुत्साह की लहर दौड़ गई और विश्व के अनेक देशों, विशेषकर जापान में इस क्षेत्र में काफी काम होने लगा। वहाँ सरकार ने अपने स्तर पर जापानी-अंग्रेजी कंप्यूटरीकृत कोश पर काम कराकर विभिन्न निजी कार्पोरेशन (जैसे सोनी, मात्सीविशु, तोशीबा) को सौंप दिया। इन संस्थाओं ने उसमें अपने उपयोग में आने वाली शब्दावली को जोड़कर अपने लिए सीमित (इन-हाउस) अनुवाद तंत्रों के विकास का रास्ता खोल दिया। इन कंपनियों को अपने उत्पादों के मैनुअलों को विभिन्न भाषाओं में अनूदित करने की आवश्यकता तो थी ही, इससे उनके सीमित उद्देश्यीय अनुवाद तंत्रों के विकास का रास्ता और भी उपयोगी उद्देश्यीय हेतु खुल गया।

सन् 1980 के बाद मशीनी अनुवाद की इन विशेषताओं को स्वीकृति मिली :

- सामान्य उद्देश्यीय अनुवाद के स्थान पर विशिष्ट एवं सीमित उद्देश्यीय अनुवाद पर बल दिया जाने लगा।
- मशीनी अनुवाद की गुणवत्ता मानव अनुवाद के समकक्ष नहीं हो सकती। अतः मशीनी अनुवाद में पश्च-संपादन की आवश्यकता को स्वीकार कर लिया गया।
- मशीनी अनुवाद का उद्देश्य मानव अनुवादकों का स्थान लेना न होकर उनकी सहायता करना हो गया।

विश्व में छिट-पुट मशीनी अनुवाद प्रयासों की आंशिक सफलता के बावजूद इसे यूरोप में विधिवत अनुसंधान के लिए अब स्वीकार कर लिया गया है तथा समन्वित प्रयास प्रारंभ हो गए हैं। इन प्रयासों के माध्यम से कुछ विशिष्ट अनुवाद तंत्र विकसित हुए।

12.3.4 कुछ मशीनी अनुवाद तंत्र

कनाडा, अमेरिका तथा जापान में हुए प्रयासों के उपरांत आइए, अब पश्चिम में विकसित कुछ मशीनी अनुवाद तंत्रों के बारे में जानें। सबसे पहला अनुवाद तंत्र रूसी से अंग्रेजी तथा अंग्रेजी से रूसी में विकसित किया गया, इसका नाम 'सिस्ट्रान' (Systran) था। इसका विकास आगे चलकर अन्य यूरोपीय भाषाओं के लिए भी किया गया।

- विभिन्न विश्वविद्यालयों तथा वाणिज्यिक संस्थाओं ने एक के बाद एक अनुवाद तंत्रों के विकास के लिए कार्य करना शुरू कर दिया। ग्रेनोब्ल विश्वविद्यालयों में 'आरियान (ARIANE) टैक्सास विश्वविद्यालय में 'मेटल' (METAL), सारब्रुकेन विश्वविद्यालय में 'सूसी' (SUSY), क्योटो (जापान) विश्वविद्यालय में 'एम.यू.' (MU) अनुवाद तंत्र विकसित किए गए। ये चारों अनुवाद तंत्र कार्यक्षमता की दृष्टि से प्रतिनिधि माने जाते हैं। इसी समय यूरोपीय समुदायों ने 'यूरोट्रा' (Eurotra), नामक बहुभाषिक अनुवाद तंत्र के विकास पर काम शुरू किया गया है। अब मशीनी अनुवाद तंत्र के विकास पर काम शुरू किया गया है तथा मशीनी अनुवाद 'प्रयोगशाला' से निकलकर बाजार में पहुँचना शुरू हो गया है। सन् 1960 के आसपास एल्प्स (ALPS), वाइडनर (Weidner) तथा लोगोस (Logos) कार्पोरेशनों ने विभिन्न प्रकार के अनुवाद तंत्र उपलब्ध कराने शुरू कर दिए। पाहो (Paho) के 'स्पैनम' (Spanam) और इंगस्पैन (Engspan) अनुवाद तंत्रों ने सफलतापूर्वक काम करना शुरू कर दिया था।

- इस युग में सीमित अनुवाद तंत्रों के विकास पर बल दिया गया। मशीनी अनुवाद के क्षेत्र में शोध कार्यों के संयोजन एवं विकास के लिए मशीनी अनुवाद की अंतरराष्ट्रीय परिषद की स्थापना हुई इसका उद्देश्य इस दिशा में किए जा रहे प्रयोगों को अंतरराष्ट्रीय मंच पर प्रस्तुत करना था। जापान में हुए मशीनी अनुवाद संबंधी शोध से तथा जीईक (JIEC) नाम रिपोर्ट से इस क्षेत्र में सरकारी वित्तपोषण का नया दौर शुरू हुआ।

सन् 1989 में मशीनी अनुवाद के क्षेत्र में नए युग का सूत्रपात होने से अनुवाद-सहायक तंत्रों के विकास पर अधिक बल दिया जाने लगा। अब मशीनी अनुवाद को मानव अनुवाद के समान समकक्ष बनाने के प्रयासों को छोड़कर मानव अनुवाद के सहायक रूप में ग्रहण किया गया है।

इस युग में यूरोपीय समुदाय द्वारा वित्तपोषित 'यूरोलांग' (Eurolang) परियोजना की शुरुआत की गई इसके अंतर्गत अंग्रेजी से फ्रेंच, जर्मन, स्पेनिश, इटैलियन तथा इन भाषाओं से अंग्रेजी में दस भाषायी जोड़ों में अनुवाद क्षमता का विकास किया जाना था।

जापान में इस दिशा में कई परियोजनाओं का संचालन शुरू हुआ। परिणामस्वरूप 'पिवट' (Pivot) अनुवाद परियोजना का विकास हुआ जिसका अंग्रेजी, जापानी, कोरियन, फ्रेंच तथा स्पेनिश में अनुवाद के लिए सफल प्रद न किया गया। जापान में CICC द्वारा जापानी, चीनी, कोरियन, थाई भाषा, इंडोनेशियाई, भासा और मलय में परस्पर अनुवाद की अंतरराष्ट्रीय परियोजना पर काम शुरू हो गया है।

- विभिन्न व्याकरण सिद्धांतों के विकास ओर उनके अनुप्रयोग से कई अनुवाद तंत्रों का विकास हुआ। इन अनुवाद तंत्रों के अतिरिक्त एक महत्वपूर्ण अनुवाद तंत्र ज्ञान आधारित मशीनी अनुवाद KBMT कार्नेजी मेलन विश्वविद्यालय, अमेरिका में शुरू किया गया था। इसके अंतर्गत यह माना गया कि अनुवाद के लिए केवल भाषावैज्ञानिक सूचनाएँ ही पर्याप्त नहीं हैं। इस तंत्र के विकास में सामान्य वाक्यविन्यासीय विश्लेषण और प्रजनन (Generation) के अतिरिक्त भाषिक संरचनाओं को आर्थी प्रतिरूप में बदलने की आवश्यकता होती है। इसकी सफलता से प्रेरित होकर न्यू मैक्सिको स्टेट विश्वविद्यालय में दूसरी अमेरिकी परियोजना 'अल्ट्रा' (Ultra) की शुरुआत हुई जिसके अंतर्गत बहुमाध्यमिक भाषी तंत्र (अंग्रेजी, चीनी, जर्मन, स्पेनिश) का विकास शुरू हुआ।

मशीनी अनुवाद के लिए प्रयुक्त विधियों के आधार पर मशीनी अनुवाद के क्रमिक विकास को तीन पीढ़ियों में विभाजित किया जा सकता है।

- **प्रत्यक्ष विधि** : मशीनी अनुवाद की प्रत्यक्ष विधि पहली पीढ़ी है जो साठ के दशक के मध्य तक चलती है। इस पीढ़ी में मशीनी अनुवाद के लिए प्रत्यक्ष विधि अपनाई गई इसमें किसी पार्सर का प्रयोग नहीं किया जाता है। यह विधि स्रोत और लक्ष्यभाषा के बीच परस्पर अनुवाद के लिए अपनाई जाती रही है। इसमें स्रोतभाषा के शब्दों के स्थान पर लक्ष्यभाषा के शब्दों को रखकर बाद में लक्ष्यभाषा के वाक्य-विन्यास के अनुरूप शब्दों का संयोजन किया जाता था। इस विधि के अनुवाद के विकसित प्रमुख अनुवाद तंत्र हैं - सिस्ट्रान, चीनी-अंग्रेजी अनुवाद के लिए कल्ट (CULT) (1968), जर्मन, स्पेनिश और फ्रेंच से अंग्रेजी के बीच परस्पर अनुवाद के लिए टाइटस (TITUS) और अंग्रेजी से वियतनामी में अनुवाद के लिए लोगोस (LOGOS) (1978).
- **अंतरण विधि** : मशीनी अनुवाद की दूसरी पीढ़ी में विश्लेषण और संश्लेषण के अतिरिक्त अंतरण की भी आवश्यकता महसूस की गई, जिसमें पद-निरूपण की परिकल्पना प्रस्तुत की गई इसमें स्रोतभाषा के प्रत्येक शब्द को व्याकरणिक कोटियों के आधार पर निरूपित कर उनका एक आंतरिक ढाँचा तैयार कर लिया जाता था। ऐसे फिर अंतरण शब्द और संरचना दोनों स्तरों पर किया जाता था। इसमें स्रोत तथा लक्ष्य भाषाओं के एकभाषी कोश और अंतरण के लिए द्विभाषी कोश की आवश्यकता होती थी।
- **माध्यमिक भाषा विधि** : इसके विकास से विभिन्न भाषाओं के बीच परस्पर अनुवाद की व्यवस्था संभव हुई इस विधि के अंतर्गत स्रोतभाषा के अर्थ का प्रतिरूपण किया जाता है, इसी के आधार पर लक्ष्यभाषा में जेनरेशन किया जाता है। साथ ही, कृत्रिम बुद्धि के अंतर्गत फ्रेम, स्क्रिप्ट आदि ज्ञान-प्रतिरूपण की अनेक तकनीकों का उपयोग किया जाता है। विभिन्न यूरोपीय भाषाओं के लिए विकसित यूरोपीय समुदाय का येट्रोत्रा (EUROTRA) तथा विभिन्न एशियाई भाषाओं के लिए विकसित एम.यू. (MU) अनुवाद तंत्र इसी विधि के नमूने हैं।

यूरोपीय और अन्य पश्चिमी तथा एशियाई देशों में मशीन द्वारा अनुवाद के मिश्रित परिणाम हासिल होने के उपरांत भारत में भी इस दिशा में कुछ प्रयत्न प्रारंभ हुए हैं।

12.4 मशीनी अनुवाद का विकास : भारतीय परिदृश्य

भारत के मशीनी अनुवाद के क्षेत्र में रूसी-तमिल के अनुवाद के प्रयास उल्लेखनीय हैं। रूसी-तमिल के क्षेत्र में यह कार्य के.सी. चेलामुत्तु द्वारा तंजावुर विश्वविद्यालय में किया गया था किंतु धन के अभाव में यह कार्य आगे न बढ़ सका। अंग्रेजी-हिंदी के क्षेत्र में आर. चंद्रशेखर का प्रयास भी उल्लेखनीय है।

भारत में प्राकृत भाषा संसाधन (NLP) के अंतर्गत मुख्य रूप से मशीनी अनुवाद के क्षेत्र में कार्य सन् अस्सी के दशक में शुरू किया गया था।

12.4.1 भारतीय भाषाओं में अनुवाद

सन् 1987 में भारतीय प्रौद्योगिकी संस्थान कानपुर में 'अक्षरभारती' वर्ग के द्वारा भारतीय भाषाओं में परस्पर अनुवाद की एक महत्वाकांक्षी परियोजना पर कार्य प्रारंभ हुआ। इसके अंतर्गत तेलुगु से हिंदी के अनुवाद का लक्ष्य निर्धारित किया गया। इस परियोजना से संबंधित साहित्य को देखने से पता चलता है कि इसमें 'माध्यमिक भाषा' की संकल्पना को आधार बनाकर अनुवाद कार्य किया गया। इसमें पाणिनि के कारक व्याकरण सिद्धांत के आधार पर माध्यमिक भाषा के विकास पर बल दिया गया। इसी परियोजना में हिंदी का पार्सर-जेनरेटर तथा तेलुगु भाषा के पार्सर के विकास तथा तेलुगु-हिंदी द्विभाषी कोश का कार्य हाथ में लिया गया।

सन् 1992 में भारतीय प्रौद्योगिकी संस्थान (IIT) कानपुर में संपन्न एशियाई भाषाओं में कंप्यूटर संसाधन (Computer Processing in Asian Language) विषयक अंतरराष्ट्रीय संगोष्ठी के बाद यह मान लिया गया कि कंप्यूटर से पूर्ण अनुवाद कराना दुष्कर एवं असंभव कार्य है। अतः पूर्ण अनुवाद तंत्र विकसित करने की अपेक्षा

अनुसारक (सहायक अनुवाद तंत्र) विकसित करने पर ध्यान दिया जाए। परिणामस्वरूप, इस वर्ग ने अनुसारकों के निर्माण की दिशा में 1995 में कार्य प्रारंभ किया जो इस वर्ग के हैदराबाद विश्वविद्यालय में स्थानांतरित हो जाने के बाद भी चलता रहा। केंद्रीय विश्वविद्यालय, हैदराबाद में पांच अनुसारकों पर कार्य हुआ - तेलुगु-हिंदी, कन्नड़-हिंदी, पंजाबी-हिंदी, बंगाली-हिंदी और मराठी-हिंदी अनुसारक। ये अब संचार और सूचना प्रौद्योगिकी विभाग (Technological Development in Indian Language-TDIL) की वेबसाइट पर उपलब्ध हैं। इन अनुसारकों की क्षमता 65 से 90 प्रतिशत तक आँकी गई है। ये अनुसारक 'लिनक्स प्लेटफार्म' पर तैयार किए गए हैं - इस कारण प्रयोगकर्ता इनका लाभ नहीं उठा पा रहे हैं। सूचना प्रौद्योगिकी मंत्रालय के सूचना प्रौद्योगिकी विभाग के वर्ष 1998 में अंग्रेजी-हिंदी अनुवाद को प्राथमिकता देने के कारण इन अनुसारकों को अनुवादकों के रूप में विकसित होने का कार्य अधूरा रह गया।

इसी दिशा में अन्ना विश्वविद्यालय में के.बी. चंद्रन वर्ग ने तमिल-हिंदी अनुसारक पर कार्य किया है। यह भी ज्ञात हुआ है कि अनुसारक प्रौद्योगिकी को अपनाते हुए अंग्रेजी-हिंदी में अनुसारक निर्माण का कार्य भी अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान (आई.आई.टी.) हैदराबाद में किया जा रहा है। अभी तक इसके भी मिश्रित परिणाम ही हासिल हो रहे हैं, क्योंकि कोई अंतिम नतीजा अभी प्राप्त नहीं हुआ है।

12.4.2 अंग्रेजी-हिंदी मशीनी अनुवाद

सन् 1992 के बाद अंग्रेजी-हिंदी मशीनी अनुवाद के कार्य को प्राथमिकता दिए जाने के बाद संचार एवं सूचना प्रौद्योगिकी मंत्रालय भारत सरकार द्वारा कई अंग्रेजी-हिंदी अनुवाद परियोजनाओं को शुरू किया गया। इनके अंतर्गत विकसित मुख्य अनुवाद-तंत्र निम्नलिखित हैं।

मशीन साधित अनुवाद (Machine Aided Translation)

भारतीय प्रौद्योगिकी संस्थान (आई.आई.टी.), कानपुर में 'आंग्ल भारती' प्रविधि के अंतर्गत उदाहरण आधारित मशीनी अनुवाद तंत्र के विकास की दिशा में कार्पोरा आधारित मशीनी अनुवाद तंत्र विकसित करने का कार्य प्रारंभ हुआ। इसमें स्वास्थ्य मंत्रालय के अंग्रेजी मैनुयूलों के हिंदी अनुवाद का लक्ष्य निर्धारित किया गया। सन् 1992-95 के दौरान 'लिनक्स' प्लेटफार्म पर इस तंत्र का प्रोटोटाइप विकसित किया गया जिसे नोएडा स्थित एर्जेसी प्रगत अभिकलन विकास केंद्र अर्थात् सी-डैक (C-DAC) को समुचित रूप से व्यवस्थित करने और फीडबैक देने के लिए दे दिया गया। उन्होंने पहले तो उसको 'लिनक्स' से 'विंडो' प्लेटफार्म में परिवर्तित किया तथा उसके परीक्षण के लिए अपेक्षित कार्य किया। आगे चलकर इस परियोजना में सूचना प्रौद्योगिकी की प्रयुक्ति को जोड़ते हुए वर्ष 2000 में Test Bed in Translation नामक एक महत्वपूर्ण परियोजना को स्वीकृति दी गई जिससे इस अनुवाद तंत्र को विकसित किया जा सके। इस अनुवाद तंत्र के विकास के लिए निर्धारित उद्देश्यों से ज्ञात होता है कि यह एक प्रयुक्तिपरक (Domain specific) अनुवाद तंत्र के रूप में विकसित करने का उद्देश्य लेकर चला था कि किंतु वर्ष 2002 में इसे 'इंटरनेट' पर प्रस्तुत कर दिया गया ताकि प्रयोगकर्ताओं के सुझावों के आधार पर इसे और आगे विकसित किया जा सके।

पहले यह अनुवाद तंत्र एक वाक्य के एकाधिक अनुवाद प्रस्तुत करता था। पहले यह निर्धारित किया गया था कि उपयोगकर्ता (अनुवादक) उनमें से उपयुक्त अनुवाद (समुचित संप्रेषक वाक्य) का चयन कर ले। यह प्रक्रिया व्यवहार की दृष्टि से अधिक क्लिष्ट प्रतीत होती थी - बाद में अनुवादों की संख्या को घटाने का प्रयास किया गया।

यह तंत्र मुख्यतः 'सरल वाक्यों' के अनुवाद तक सीमित है। दूसरी बात यह है कि प्रयुक्तिपरक अनुवाद के क्षेत्र में इस तंत्र से काफी उम्मीद लगाई जा रही थी। अब इस प्रविधि का प्रयोग अंग्रेजी से अन्य भारतीय भाषाओं में अनुवाद तंत्र विकसित करने के लिए किया जा रहा है।

मात्रा (Matra), सी.डैक. मुंबई

यह अनुवाद तंत्र प्रगत अभिकलन विकास केंद्र (C-DAC) द्वारा विकसित हुआ है। इसका विकास समाचारों (News stories) के अनुवाद के लिए हुआ है। इस तंत्र पर वर्ष 1995 से 2000 के दौरान कार्य किया गया।

उपलब्ध जानकारी के आधार पर इस तंत्र की विशेषता यह है कि इसमें इंटरएक्टिव संरचना एडिटर (Interactive structure editor) कार्य करता है, जिसके माध्यम से अंग्रेजी इनपुट वाक्यों का संरचनात्मक प्रतिरूपण प्राप्त होता था। इसमें प्रयोक्ता की सहायता भी ली जाती थी। यह तंत्र सरल वाक्यों का ही अनुवाद करता था। इसके अंतर्गत मात्रा - लाइट और मात्रा-प्रो वर्जन तैयार किए गए। मात्रा-लाइट को सूचना प्रौद्योगिकी मंत्रालय की वेबसाइट पर प्रस्तुत किया गया है। मात्रा-प्रो के लिए प्रयोगकर्ताओं को लाइसेंस लेना अपेक्षित है। इसके विकासकर्ताओं का यह मानना है कि इस विधि का उपयोग किसी भी प्रयुक्तिपरक अनुवाद तंत्र (Domain based Translation System) के विकास में किया जा सकता है।

मंत्रा (Mantra) सी.डैक. पुणे

सरकारी अधिसूचनाओं, संकल्पों, आदेशों आदि कार्यालयी साहित्य का अंग्रेजी से हिंदी में अनुवाद करने के लिए 'मंत्रा' अनुवाद तंत्र का कार्य सन् 1997 में प्रारंभ किया गया। इसके लिए भारत सरकार के राजभाषा विभाग ने वित्तपोषण का भार लिया था। इस अनुवाद तंत्र का प्रोटोटाइप वर्ष 1998-99 में पूरा कर लिया था जो केवल सरल वाक्यों का ही अनुवाद करता था। इसमें व्याकरणिक विश्लेषण सिद्धांत के रूप में (पेंसिलवेनिया विश्वविद्यालय अमेरिका) सुप्रसिद्ध कंप्यूटेशनल भाषाविज्ञानी प्रो. अरविंद जोशी के 'ट्री एडज्वानिंग ग्रामर' को आधार बनाया गया था। इस तंत्र में निविष्ट अंग्रेजी वाक्यों का पार्स-ट्री के रूप में विश्लेषण होता है। इसके अनुसार वृक्ष-आरेख बनता है और उनके अनुरूप हिंदी में अनुवाद किया जाता है। वर्ष 1999-2000 में इसे भारत सरकार के पाँच मंत्रालयों में परीक्षण के तौर पर स्थापित किया गया था किंतु अनुकूल परीक्षण रिपोर्ट प्राप्त न होने के कारण इसे सी.डैक, पुणे को संशोधन/परिवर्तन के लिए दे दिया गया। इधर सी.डैक., पुणे ने इसे संशोधित कर दिया है और इसे राजभाषा विभाग की वेबसाइट पर उपलब्ध करा दिया गया है। प्रशासन के अतिरिक्त इसे लेखा संबंधी मामलों के लिए भी सक्षम बनाया गया है। इसकी अनुवाद क्षमता सीमित होने के कारण भारत सरकार के कार्यालयों में इसका उपयोग बहुत कम किया जा रहा है। सी.डैक. पुणे मंत्रा में प्रयुक्त प्रविधि का प्रयोग अन्य विभागों के लिए अनुवाद तंत्र विकसित करने में भी कर रहा है।

आंग्ल भारती - आई.आई.टी. कानपुर और सी-डैक, नोएडा

आंग्ल भारती का विकास भारतीय प्रौद्योगिकी संस्थान, (आई.आई.टी) कानपुर और प्रगत अभिकलन विकास केंद्र (सी-डैक) नोएडा के संयुक्त प्रयास से सन् 1994 में हुआ है। इसमें नियम आधारित प्रणाली और उदाहरण आधारित प्रणाली दोनों का संयुक्त रूप है। यह अंग्रेजी से हिन्दी में अनुवाद करता है। इसकी विषय-सामग्री लोक स्वास्थ्य सेवा, कार्यालयी पत्राचार और तकनीकी संदर्भिका है। इसके दो संस्करण हैं। दूसरे संस्करण में काफी संशोधन और परिवर्द्धन हुआ है। आंग्ल भारती के मॉडल के आधार पर अंग्रेजी-पंजाबी, अंग्रेजी-उर्दू, अंग्रेजी-बंगला, अंग्रेजी-मलयालम, अंग्रेजी-तेलुगु, अंग्रेजी-नेपाली आदि पर कार्य हो रहा है।

अनुवादक - (सुपर इंफोसाफ्ट, नई दिल्ली)

मैट, मात्रा तथा मंत्र - ये तीनों अनुवाद तंत्र सरकारी वित्तपोषण के आधार पर विकसित किए गए हैं। इसके विपरीत अनुवाद तंत्र 'अनुवादक' एक निजी कंपनी (सुपर इंफोसाफ्ट, नई दिल्ली) द्वारा विकसित किया गया है। इसका विकास कार्य सन् 1998-99 में शुरू किया गया था। प्राप्त जानकारी के आधार पर, यह अनुवाद तंत्र सरकारी, सार्वजनिक तथा निजी सैक्टर के कार्यालयों, बैंकों आदि के अंग्रेजी सूचनात्मक साहित्य का अनुवाद करता है। इससे वार्षिक रिपोर्टें, समाचारिकाओं (बुलेटिन), परिपत्रों, अधिसूचनाओं आदि का अनुवाद किया जा सकता है।

अनुवादक एक ऐसा अनुवाद तंत्र है जो कृत्रिम बुद्धिमत्ता वाले अनुवाद इंजिन की मदद से अंग्रेजी दस्तावेजों का हिंदी भाषा और व्याकरण के अनुसार अनुवाद करता है। इसमें अनुवाद से पहले तथा बाद में संबंधित भाषा (स्रोत एवं लक्ष्य) के पाठ की संबंधित भाषा के व्याकरण शोधक द्वारा जांच की जाती है। इसमें प्रशासन, रेल, बैंक, कृषि चिकित्सा तथा तकनीकी शब्दकोश समाहित हैं। अंतरण कोश में प्रविष्टियों की संख्या चार लाख से भी अधिक

है। साथ ही एक अच्छा पदबंध कोश भी विकसित किया गया है। इसमें कोशों के संशोधन-परिवर्तन की सुविधा भी उपलब्ध है। सटीक अनुवाद के लिए मूल अंग्रेजी शब्द के कई पर्याय उपलब्ध हैं, जिससे उपयुक्त शब्द के चयन में सुविधा होती है। गुणवत्ता की दृष्टि से इसका अनुवाद अन्य अनुवाद तंत्रों की अपेक्षा बेहतर है, हालाँकि सामान्य प्रयोक्ताओं तक अनुवादक की पहुँच नहीं हो पाई है। यदि इसका प्रचार-प्रसार समुचित रूप से किया जाए तो इसमें परस्पर फीडबैक (प्रतिपुष्टि) से और भी सुधार हो सकता है।

युनिवर्सल नेटवर्किंग लैंग्वेज (UNL) - भारतीय प्रौद्योगिकी संस्थान (IIT), मुंबई

हिंदी को संयुक्त राष्ट्र संघ की भाषा के रूप में प्रतिष्ठित करने के लिए जहाँ कई प्रकार की भाषा-नियोजन संबंधी तैयारियाँ अपेक्षित हैं, वहीं संयुक्त राष्ट्र संघ की भाषाओं से हिंदी तथा हिंदी से उन भाषाओं में अनुवाद की क्षमता विकसित करने के लिए प्राथमिकता के आधार पर कार्य करना भी अपेक्षित है। इस संबंध में संयुक्त राष्ट्र संघ की एक महत्वपूर्ण संस्था यू.एन.डी.एल. (United Nations Digital Library) ने सन् 1996 से कार्य करना आरंभ कर दिया था-वह है एक माध्यमिक भाषा के रूप में युनिवर्सल नेटवर्किंग लैंग्वेज (यू.एन.एल.) का विकास। इस पर सन् 1996 से भारतीय प्रौद्योगिकी संस्थान, मुंबई में कार्य किया जा रहा है। संयुक्त राष्ट्र की मौजूदा छह भाषाओं में इसके माध्यम से अनुवाद तंत्र विकसित कर लिया गया है। आज विश्व की अनेक भाषाओं में यू.एन.एल. पर कार्य हो रहा है।

इसके विकासकर्ताओं के अनुसार यू.एन.एल. के अंतर्गत अंग्रेजी के माध्यम से लगभग दो लाख शब्दों का युनिवर्सल कोश तैयार हो चुका है। हिंदी से यू.एन.एल. में अनुवाद करने के लिए एनकन्वर्टर (Enconverter) और यू.एन.एल. से हिंदी में अनुवाद करने के लिए डिक्न्वर्टर (Deconverter) के विकास का काम चल रहा है। यदि इस कार्य को प्राथमिकता के आधार पर, समयबद्ध रूप से, पूरा किया जा सके तो संयुक्त राष्ट्र संघ की भाषाओं तथा हिंदी में परस्पर अनुवाद की क्षमता का विकास जल्दी हो जाएगा।

अन्य प्रयास

(i) 'शिव' और 'शक्ति' अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान (IIIT), हैदराबाद

अभी तक जिन अनुवाद तंत्रों पर चर्चा हुई है उनके परिणाम अंतिम रूप से व्यवहार में उपयोगी साबित नहीं हुए हैं तथापि जिन पाँच अनुवाद तंत्रों पर हमने ऊपर चर्चा की, उनके अतिरिक्त अन्य अंग्रेजी-हिंदी अनुवाद तंत्र 'शिव' और 'शक्ति' का भी विकास किया जा रहा है। पहले अनुवाद तंत्र - उदाहरण आधारित मशीनी अनुवाद तंत्र (Example Based Machine Translation System-EBMT) 'शिव' के विकास का काम अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान, हैदराबाद, भारतीय विज्ञान संस्थान, बंगलौर तथा कार्नेजी मेलन विश्वविद्यालय अमेरिका के सहयोग से प्रारंभ किया गया था। साथ ही, इंटरनेट आधारित अनुवाद तंत्र 'शक्ति' का विकास अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान, हैदराबाद में सन् 2003 में प्रारंभ हुआ। इन दोनों अनुवाद तंत्रों के लिए साझा कोशीय संसाधनों के विकास का कार्य प्रारंभ किया गया था। 'शिव' का विकास उदाहरण आधारित प्रविधि के आधार पर तथा 'शक्ति' का विकास मशीन अधिगम प्रविधि के आधार पर करना निश्चित किया गया था।

कुछ समय बाद यह पाया गया कि उदाहरण आधारित प्रविधि के परिणाम मशीन अधिगम प्रविधि की तुलना में अधिक उत्साहजनक नहीं हैं। इस कारण सन् 2005 में 'शिव' अनुवाद तंत्र के विकास का कार्य रोक दिया गया। 'शक्ति' अनुवाद तंत्र पर कार्य चलता रहा। इसके लिए अपेक्षित मात्रा में कोशीय संसाधनों का विकास किया गया है, जिससे अंग्रेजी भाषा में उपलब्ध इंटरनेट सामग्री का हिंदी रूपांतरण प्राप्त किया जा सके। यह संप्रेषण सहायक उपकरण के तौर पर कार्य करेगा। इसके लिए कोशीय संसाधनों का विकास ब्रिटिश नेशनल कार्पस के आधार पर किया गया है।

(ii) अंग्रेजी-हिंदी अनुसारक

विभिन्न भारतीय भाषाओं में अनुसारक निर्माण करने वाली अक्षर भारती टीम के प्रमुख सदस्य डॉ. विनीत चैतन्य के द्वारा 'चिन्मय मिशन' के सहयोग से अंग्रेजी-हिंदी अनुसारक की महत्वाकांक्षी परियोजना पर काम प्रारंभ किया

गया है। इसमें अंग्रेजी भाषा के हिंदी में रूपांतरण नियमों पर आधारित अंतरण व्याकरण के आधार पर मशीनी अनुवाद तंत्र विकसित करने का लक्ष्य रखा गया है। इसमें चैंबर अंग्रेजी कोश को आधार बनाकर अंतरण व्याकरण के नियम तैयार करने के लिए विभिन्न शोध वर्गों द्वारा कार्य किया जा रहा है।

12.5 इंटरनेट और मशीनी अनुवाद

इंटरनेट का विकास बीसवीं शताब्दी की एक महत्वपूर्ण कंप्यूटेशनल घटना है। इससे मशीनी अनुवाद को वेब अनुवाद के रूप में विकसित होने का अवसर मिला है। सर्वाधिक प्राचीन अनुवाद तंत्र 'सिस्ट्रान' पर आधारित 'आल्ताविस्ता' के माध्यम से विभिन्न यूरोपीय भाषाओं में परस्पर अनुवाद की सुविधा प्राप्त हुई है। इसके अंतर्गत सीमित पाठ का किसी भी यूरोपीय भाषा में अनुवाद प्राप्त किया जा सकता है। वेब आधारित अनुवाद के विकास में मशीनी अनुवाद के उद्देश्यों में महत्वपूर्ण परिवर्तन हुआ है। पहले इस प्रकार के अनुवाद को 'अनुवाद उपकरण' के रूप में माना जाता था किंतु अब इसे 'संप्रेषण उपकरण' के रूप में माना जाने लगा है। इसमें अनुवाद की शुद्धता पर बल न देकर संप्रेषणीयता पर अधिक बल दिया जाता है। जैसा कि आपने ऊपर की गई चर्चा में पाया कि मशीनी अनुवाद की आवश्यकता आज के सूचना विस्फोट के युग में बढ़ती जा रही है और अब मशीनी अनुवाद सहायता के रूप में न होकर संप्रेषण सहायता के रूप में समझा जाने लगा है।

आज विश्व की महत्वपूर्ण भाषाओं में वेब अनुवाद की सुविधा प्राप्त है। विभिन्न सर्विस प्रोवाइडर (याहू, हॉटमेल, गूगल आदि) सर्च इंजिन का कार्य तो करते ही हैं, साथ ही अनुवाद की सुविधा के भी लिंक उपलब्ध कराते हैं। अन्य विश्वसनीय वेब अनुवाद तंत्र "www.altavista.com", "www.freetranslation.com" आदि के वेबसाइट पर भी उपलब्ध हैं जिनसे विभिन्न यूरोपीय भाषाओं तथा अंग्रेजी में परस्पर अनुवाद प्राप्त किया जा सकता है। ऐसे अनेक उपलब्ध अनुवाद तंत्रों की सहायता से प्राप्त अनुवाद का पुनरीक्षण कर प्रकाशन योग्य बनाया जा सकता है।

12.6 सारांश

ऊपर हम देख चुके हैं कि मशीनी अनुवाद तंत्रों के विकास के संबंध में पिछले पांच-छह दशकों में काफी कार्य हुआ है, साथ ही मशीन की प्रोसेसिंग तथा स्मृति क्षमता में भी अपार वृद्धि हुई है। भाषा विश्लेषण मॉडलों का भी विकास किया गया है।

प्रारंभिक युग में जहां मशीनी अनुवाद को एक शुद्ध इंजीनियरी कार्य माना जाता था, अब इसे अंतरविषयी भाषा इंजीनियरी का कार्य माना जाता है। यह क्षेत्र कई विषयों के विशेषज्ञों के एक साथ मिलकर कार्य करने की अपेक्षा रखता है। मशीनी अनुवाद के सफल तंत्रों के विकास में कंप्यूटर विज्ञानियों के साथ-साथ भाषाविज्ञानियों (संगणक भाषाविज्ञानी, कोश विज्ञानी एवं अनुवाद विशेषज्ञों) का सहयोग अपेक्षित है।

इस प्रकार के तंत्रों के विकास के लिए एक और आवश्यकता है - कोशीय संसाधनों के विकास की। भारत में विकसित अनुवाद तंत्रों के लिए उपयुक्त मात्रा में कोशीय संसाधनों का अभाव स्पष्ट दिखाई देता है। अंतरराष्ट्रीय सूचना प्रौद्योगिकी संस्थान (आई.आई.टी.), हैदराबाद में विकसित किए जा रहे, 'शक्ति' अनुवाद तंत्र के लिए कोशीय संसाधन विकसित करने का कार्य किया गया। इसमें अंग्रेजी-हिंदी अंतरण व्याकरण के नियम विकसित किए गए। इससे अनुवाद की गुणवत्ता में काफी सुधार दिखाई दिया। इधर सी.डैक. नोएडा में भी कोशीय संसाधनों के विकास का कार्य प्रगति पर है।

हिंदी 'वर्डनेट' का विकास भी इस कार्य में काफी सहायता कर सकता है। आई.आई.टी. मुंबई में इस संबंध में काफी काम किया गया है। 'गूगल' पर अंग्रेजी तथा अन्य विदेशी भाषाओं से हिंदी में अनुवाद की सुविधा उपलब्ध है। यह सांख्यिकीय मशीनी अनुवाद प्रणाली पर आधारित है। इसकी गुणवत्ता अभी बहुत ही कम है।

इस प्रयासों से यह आशा प्रतीत होती है कि निकट भविष्य में मशीनी अनुवाद तंत्रों के विकास में सफलता प्राप्त हो जाएगी।

12.7 अभ्यास के लिए प्रश्न

प्रश्न 1 : रिक्त स्थानों की पूर्ति कीजिए :

- i) अनुवाद की तीन प्रक्रियाएँ होती हैं-
1..... 2. अंतरण 3.
- ii) अंतरण कोश के अंतर्गत शब्द स्तरीय कोश और स्तरीय कोश आते हैं।
- iii) विश्लेषण की प्रक्रिया को मशीनी अनुवाद में कहा जाता है।
- iv) मानव अनुवाद या मशीनी अनुवाद दो भिन्न भाषाओं के बीच होने वाली एक प्रक्रिया होती है।
- v) मशीन अनुवाद अवरोधों को दूर करने का त्वरित साधन है।
- vi) 'टैगिंग' और 'एनोटेशन' साधित मशीनी अनुवाद के अंतर्गत आते हैं।
- vii) जापान में कई कंपनियों के इन हाउस अनुवाद तंत्रों के विकास से..... उद्देश्यीय अनुवाद तंत्रों के विकास का रास्ता खुल गया।
- viii) इंटरनेट पर अनुवाद को उपकरण के रूप में माना जाता है।
- ix) सिस्ट्रान पर आधारित विभिन्न..... भाषाओं में परस्पर अनुवाद की सुविधा का विकास हुआ।
- x) सन् 1980 के आसपास मशीनी अनुवाद को..... अनुवाद के रूप में विकसित किया गया है।

प्रश्न 2 : निम्नलिखित कथनों पर सही या गलत का निशान लगाइए :

- i) मानव अनुवाद और मशीनी अनुवाद की प्रक्रियाएँ अलग-अलग हैं। ()
- ii) मशीनी अनुवाद में अनुवाद की तीनों प्रक्रियाएँ मशीन संपन्न करती हैं। ()
- iii) मशीनी अनुवाद की प्रक्रियाएँ हैं-पार्सिंग, अंतरण और जेनरेशन। ()
- iv) 'पार्सिंग' संबंधी साफ्टवेयर के निर्माण का आधार लक्ष्यभाषा का व्याकरण होता है। ()
- v) मशीनी अनुवाद की अपेक्षा मानव अनुवाद अधिक विश्वसनीय है। ()
- vi) मशीनी अनुवाद का विकास सबसे पहले पश्चिम में हुआ। ()
- vii) मशीनी अनुवाद का व्यवस्थित विकास सन् 1933 से माना जाता है। ()
- viii) अनुवाद के क्षेत्र में अंकीय (डिजिटल) कंप्यूटर के प्रयोग का प्रस्ताव सन् 1946 में ए.डी.बूथ ने दिया था। ()
- ix) प्रारम्भ में मशीनी अनुवाद की गुणवत्ता बहुत अच्छी थी। ()
- x) मशीनी अनुवाद के क्षेत्र में पहले तंत्र का नाम 'सिस्ट्रान' था। ()
- xi) सन् 1966 के बाद भाषा विश्लेषण के कंप्यूटेडानल माडलों का विकास हुआ। ()
- xii) 'टॉउम मेटो' के सीमित तंत्र के विकास ने मशीनी अनुवाद के क्षेत्र में नई जान फूंक दी। ()

(प्रश्न संख्या 1 एवं 2 के लिए संकेत पाठ्यसामग्री में तलाशें)।

प्रश्न 3 : प्रश्नों के उत्तर दीजिए

- i) मानव और मशीन के संदर्भ में अनुवाद के प्रकार बताइए।

- ii) 'पार्सिंग' से आप क्या समझते हैं?
- iii) मानव साधित मशीनी अनुवाद से क्या तात्पर्य है ?
- iv) मशीनी अनुवाद के प्रारंभिक युग को अत्युत्साह के युग की संज्ञा क्यों दी जाती है?
- v) एल्फैक समिति की मुख्य सिफारिशें क्या-क्या थीं? तथा एल्फैक समिति की रिपोर्ट के बाद मशीन अनुवाद का विकास किस प्रकार हुआ?
- vi) मशीन साधित अनुवाद (MAT) अनुवाद तंत्र पर प्रकाश डालिए।
- vii) यूनिवर्सल नेटवर्किंग लैंग्वेज से क्या तात्पर्य है? उससे मशीनी अनुवाद तंत्र का कैसे विकास होता है?

12.8 शब्दावली

संगणक भाषाविज्ञान (Computational linguistics), व्याकरण शोधक (Grammar checker), अंतरण कोश (Transfer grammar), कृत्रिम बुद्धिमत्ता (Artificial intelligence), मशीन अधिगम (Machine learning), कोशीय संसाधन (Lexical Resources), संप्रेषण उपकरण (Communication tool)

12.9 कुछ उपयोगी पुस्तकें

- कृष्ण कुमार गोस्वामी, 2012, अनुवाद विज्ञान की भूमिका, नई दिल्ली : राजकमल प्रकाशन।
- ठाकुर दास, मशीनी अनुवाद : विधियां एवं प्रविधियां 'अनुवाद' अंक 129, नई दिल्ली, भारतीय अनुवाद परिषद।
- कंप्यूटर अनुवाद, (विशेषांक) भारतीय अनुवाद परिषद, नई दिल्ली।
- भाषा का सूचना प्रौद्योगिकी और भारतीय भाषाएं विशेषांक 2002 में प्रकाशित मशीनी अनुवाद संबंधी लेख, केंद्रीय हिंदी निदेशालय, नई दिल्ली, रामकृष्ण पुरम्।
- सूचना प्रौद्योगिकी और हिन्दी अनुवाद (विशेषांक), नई दिल्ली, भारतीय अनुवाद परिषद।
- Gargesh, R. and K.K. Goswami, Translation and Interpreting (Reader & Workbook), New Delhi, Orient Longman, Pvt.Ltd.
- Hutchins, J. (2005) The History of machine translation in a nutshell.
- Hauenschild, Christa and Susanne Heizmann (eds), 1997, Machine Translation and Translation Theory, (Text, Translation, Computational Processing), Berlin-New York, Mouton de Gruyter.
- Sangal, Rajeev, Vineet Chaitanya, Natural Language Processing, New Delhi. Prentic Hall.
- Sergei Nirenburg, Harold L. Somers and Yorick A. Wilks, Readings in Machine Translation, The Brain Science Connectin, MIT COGnet.
- Steffens, Petra, Ed., April 26-28, 1993, Machine Translation and the Lexicon, Germany, Third International EAMT Workshop Heidelberg Proceedings.
- Envisioning Machine Translation in the Information Future, October 10-14, 2000, 4th Conference of the Association for Machine Translation in the Americas, Cuernavaca, Mexico, AMTA.
- www.freetranslation.com, www.altavista.com, विदेशी भाषाओं तथा अंग्रेजी में परस्पर अनुवाद के लिए।
- www.tdil@mit.gov.in अनुसारकों तथा अंग्रेजी-हिंदी अनुवाद के लिए।

इकाई 13 मशीनी अनुवाद : पद्धतियाँ और प्रकार

इकाई की रूपरेखा

- 13.0 उद्देश्य
- 13.1 प्रस्तावना
- 13.2 मशीनी अनुवाद की विभिन्न पद्धतियाँ
 - 13.2.1 प्रत्यक्ष मशीनी अनुवाद-पद्धति
 - 13.2.2 अंतरण मशीनी अनुवाद-पद्धति
 - 13.2.3 अंतरभाषिक मशीनी अनुवाद पद्धति
- 13.3 मशीनी अनुवाद पद्धति विकास का आधुनिक चरण : कृत्रिम बुद्धि
- 13.4 मशीनी अनुवाद के प्रकार
 - 13.4.1 पूर्णतः स्वचालित मशीनी अनुवाद
 - 13.4.2 अर्ध-स्वचालित मशीनी अनुवाद
- 13.5 अर्ध-स्वचालित मशीनी अनुवाद के विविध प्रकार
 - 13.5.1 मानव-साधित मशीनी अनुवाद प्रणाली
 - 13.5.2 मशीन-साधित मानव अनुवाद प्रणाली
- 13.6 कंप्यूटर अनुवाद प्रौद्योगिकी के संदर्भ में अनुवादकों के प्रकार
 - 13.6.1 समूह में काम करने वाले पेशेवर अनुवादक
 - 13.6.2 स्वतंत्र पेशेवर अनुवादक
 - 13.6.3 गैर-पेशेवर अनुवादक
- 13.7 सारांश
- 13.8 अभ्यास के लिए प्रश्न
- 13.9 शब्दावली
- 13.10 कुछ उपयोगी पुस्तकें

13.0 उद्देश्य

इस इकाई का अध्ययन करने के बाद आप यह जान सकेंगे कि :

- मशीनी अनुवाद की कौन-सी पद्धतियाँ हैं,
 - मशीनी अनुवाद कितने प्रकार का होता है और क्या पूर्णतः स्वचालित मशीनी अनुवाद संभव है?
 - अर्ध-स्वचालित मशीनी अनुवाद कितने प्रकार का होता है; और
 - मशीनी अनुवाद के संदर्भ में अनुवादकों को किन-किन श्रेणियों में रखा जा सकता है।
-

13.1 प्रस्तावना

मशीनी अनुवाद से संबंधित खंड की यह दूसरी इकाई है। आप पहले यह जान चुके हैं कि कंप्यूटर अनुवाद पद्धति का विकास करते समय कई दृष्टिकोणों को आधार बनाया गया है। इन दृष्टिकोणों को नियम-आधारित मशीन अनुवाद (Rule-based Machine Translation), उदाहरण-आधारित मशीन अनुवाद (Example-based Machine

Translation) और सांख्यिकी-आधारित मशीन अनुवाद (Statistical-based Machine Translation) के संदर्भ में देखा जाता है। इसी के साथ-साथ आप यह भी जान चुके हैं कि कंप्यूटर अनुवाद तीनों प्रक्रियाओं में संपन्न होता है। इनमें से विश्लेषण संबंधी सॉफ्टवेयर को “पद-निरूपित” (पार्सर), अंतरण संबंधी सॉफ्टवेयर को “अनुवादित” (ट्रांसलेटर) तथा पुनर्गठन संबंधी सॉफ्टवेयर को “जनित्र” (जेनरेटर) की संज्ञा दी जाती है। “अनुवादित” शब्द संचय और अंतरण-व्याकरण की सहायता लेते हुए कार्य करता है। इस इकाई में आपको यह बताया जा रहा है कि मशीनी अनुवाद की विभिन्न पद्धतियाँ कौन-सी हैं। इसके अतिरिक्त यह भी बताया जा रहा है कि क्या पूर्णतः स्वचालित कंप्यूटर अनुवाद संभव है? इससे आप इस निष्कर्ष पर पहुँच सकेंगे कि पूर्णतः स्वचालित कंप्यूटर अनुवाद संभव नहीं है। इसलिए शोधकर्ता इसके अर्ध-स्वचालित को ध्यान में रखकर शोध की दिशा में आगे बढ़े हैं। यह जानकारी देना भी जरूरी है कि अर्ध-स्वचालित मशीनी अनुवाद मशीनी अनुवाद कितने प्रकार का होता है। इसके अतिरिक्त यह भी अध्ययन किया जाएगा कि मशीनी अनुवाद के संदर्भ में अनुवादकों को किस प्रकार वर्गीकृत करके देखा जा सकता है।

13.2 मशीनी अनुवाद की विभिन्न पद्धतियाँ

कंप्यूटर अनुवाद प्रणाली के क्रमिक विकास का अवलोकन मुख्य रूप से (1) प्रारंभिक और (2) आधुनिक चरणों के रूप में किया जा सकता है। प्रारंभिक चरण के अंतर्गत उन पद्धतियों का उल्लेख किया जाता है जिन्हें अनुसंधानकर्ताओं ने क्रमिक रूप में शुरू में अपनाया था। ये हैं - 1. “प्रत्यक्ष मशीनी अनुवाद पद्धति”, 2. “अंतरण मशीनी अनुवाद पद्धति”, और 3. “अंतरभाषिक मशीनी अनुवाद पद्धति”।

13.2.1 प्रत्यक्ष मशीनी अनुवाद-पद्धति (Direct Machine Translation System)

प्रत्यक्ष मशीनी अनुवाद पद्धति को “साक्षात् पद्धति” भी कहा जाता है। इसमें पद-निरूपण (पार्सिंग) सिद्धांतों अथवा भाषायी सैद्धांतिक विश्लेषण आदि की कोई खास जरूरत नहीं होती। इसमें वस्तुतः किसी विशिष्ट भाषा वैज्ञानिक सिद्धांत का सहारा न लेकर स्रोतभाषा सामग्री का विश्लेषण करने के बाद उसके शब्दों के स्थान पर लक्ष्यभाषा के वाक्य-विन्यास के अनुरूप शब्दों को संयोजित किया जाता है। इस पद्धति के मूल में यह धारणा निहित है कि वांछित दो भाषाओं (स्रोतभाषा और लक्ष्यभाषा) को कंप्यूटर द्वारा सीधे जोड़कर उनमें परस्पर अनुवाद करने का प्रयत्न किया जाता है। किंतु यदि उन दो भाषाओं से भिन्न भाषाओं के बीच कंप्यूटर के जरिए अनुवाद करने की आवश्यकता है तो तंत्र को उस रूप में परिवर्तित कर दिया जाता है। इस पद्धति को मूलतः डॉ. टोमा ने 1964 में विकसित किया था। इसे जॉर्जटाउन विश्वविद्यालय द्वारा रूसी-अंग्रेजी अनुवाद के संदर्भ में विकसित किया गया था। इसकी गति 3,00,000 शब्द प्रति घंटे की थी। इस पद्धति में मूलतः इन दोनों भाषाओं के शब्दकोशों में उपलब्ध जानकारी, शब्द-रचना के स्थूल नियमों और स्थूल वाक्य-साँचों को आधार बनाया गया था। इसमें व्याकरणिक नियमों का समावेश नहीं था। इसके अलावा, पद-निरूपण (पार्सर) का विधान भी इसमें नहीं था। इसलिए इस विधि से प्राप्त अनुवाद अधिकांशतः मात्रा शाब्दिक अथवा कृत्रिम अनुवाद होता था। शाब्दिक अथवा कृत्रिम अनुवाद के कारण इसमें कई बार अर्थ का अनर्थ होने की संभावना रहती थी। बाद में इस पद्धति का और अधिक विकास हुआ।

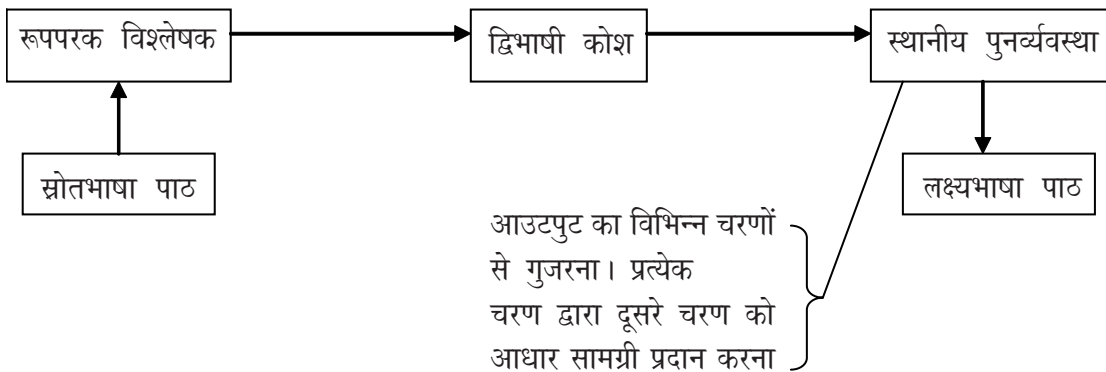
प्रत्यक्ष मशीनी अनुवाद पद्धति के आधार पर 1964 में जॉर्जटाउन विश्वविद्यालय ने रूसी-अंग्रेजी अनुवाद तंत्र GAT का विकास हुआ। इसी अनुवाद-तंत्र का परिष्कृत रूप “सिस्ट्रॉन” (SYSTRAN) पद्धति के रूप में 1976 में विकसित हुआ। इस अनुवाद-तंत्र की सहायता से “नासा” (NASA) ने अपोलो-सोयूज सहयोग से संबंधित सामग्री को अनूदित किया था। यह विश्व का सर्वप्रथम व्यावसायिक मशीनी अनुवाद-तंत्र था। प्रत्यक्ष मशीनी अनुवाद पद्धति के आधार पर हांगकांग के चीनी विश्वविद्यालय ने 1968 में “कल्ट” (CULT) नामक मशीनी अनुवाद-तंत्र को भी विकसित किया। अनुवाद संबंधी संदिग्धताओं के निवारण के लिए इस अनुवाद प्रणाली में पूर्व-संपादन (pre-editing) एवं पश्च-संपादन (post-editing) की व्यवस्था थी। इस अनुवाद-तंत्र की सहायता से गणित एवं भौतिकी विषयक चीनी शोधपत्र का अंग्रेजी अनुवाद किया गया। प्रत्यक्ष मशीनी अनुवाद पद्धति पर आधारित इन दोनों अनुवाद-तंत्रों (सिस्ट्रॉन और कल्ट) के अलावा टाइटस (TITUS) एवं लोगोस (LOGO) का

नाम विशेष तौर पर उल्लेखनीय है। टाइटस (TITUS) जर्मन, स्पेनिश और फ्रांसीसी से अंग्रेजी के बीच परस्पर अनुवाद की प्रणाली है और 'लोगोस' (LOGOS) अंग्रेजी से वियतनामी में अनुवाद के लिए 1973 में विकसित प्रणाली है। इनमें से 'लोगोस' का विकास अमेरिकी वायुसेना के उपकरणों के मैनुअलों के अंग्रेजी से वियतनामी भाषा में अनुवाद के लिए किया गया। इन विभिन्न तंत्रों में अनुवाद प्रक्रिया के प्रत्येक चरण के उपयोग में थोड़ा-बहुत अंतर रहता है।

इस पद्धति में स्रोतभाषा और लक्ष्यभाषा के पर्यायों को शामिल किया जाता है। यह मशीनी अनुवाद प्रणाली मुख्य रूप से तीन कारकों पर आधारित होती है। ये हैं - (1) उच्च-स्तरीय गुणवत्ता वाले कंप्यूटर कोश और रूपपरक विश्लेषण (Morphological analysis) (2) अंतिम पाठ तैयार होने से पूर्व कच्चे अनुवाद का मानव-अनुवादक द्वारा पश्च संपादन और (3) पश्च संपादन और शब्दकोश नियंत्रण के लिए विकसित शब्द और पाठ संसाधक उपकरण (सॉफ्टवेयर)। इस प्रत्यक्ष मशीनी अनुवाद पद्धति में अनूद्य पाठ के प्रत्येक वाक्य को करीब 10 चरणों से गुजारा जाता है। इनमें से प्रत्येक चरण कुछ इस प्रकार का होता है कि पहला चरण दूसरे के लिए आधारभूत सामग्री प्रदान करता है यानी प्रत्येक चरण का 'आउटपुट' अगले चरण का 'इनपुट' होता है। इसे यों भी कह जा सकते हैं कि प्रत्येक चरण का आउटपुट अपने पूर्ववर्ती चरण की प्रक्रिया का परिणाम होता है। इस प्रकार सभी चरण एक-दूसरे पर परस्पर निर्भर होते हैं। कंप्यूटर में विकसित इस शब्दकोश को 'कंप्यूटर आधारित शब्दकोश' भी कहा जाता है। इस पद्धति में सर्वप्रथम अनुवाद किए जाने वाले पाठ वाक्य का कंप्यूटर में प्रविष्टिकरण किया जाता है। उसके बाद प्रविष्ट वाक्य जिन अवस्थाओं से गुजरता है, वे प्रमुख रूप से इस प्रकार हैं

1. स्रोतभाषा पाठ के लिए कोश देखना और उसका रूपपरक विश्लेषण।
2. समरूपों अथवा समलेखों (Homographs) की पहचान और उनका अर्थ निर्धारण।
3. संयुक्त/यौगिक संज्ञाओं (Compound nouns) की पहचान।
4. संज्ञा पदबंध एवं क्रिया पदबंध की पहचान।
5. मुहावरों का विश्लेषण और उनके लिए सही पर्यायों की तलाश।
6. पूर्वसर्गों अथवा संबंधबोधकों (Preposition) के लिए सही पर्यायों की पहचान।
7. उद्देश्य एवं विधेय की पहचान।
8. वाक्यपरक द्विअर्थकता अथवा अनेकार्थता की पहचान।
9. स्रोतभाषा पाठ का लक्ष्यभाषा में संश्लेषण और उसकी रूपपरक संरचना को संसाधित करना।
10. शब्दों और पदबंधों का लक्ष्यभाषा में सही क्रम (वाक्यों के रूप) में संयोजन।

प्रत्यक्ष मशीनी अनुवाद पद्धति के अभिकल्प का आरेखीय निरूपण इस प्रकार है :



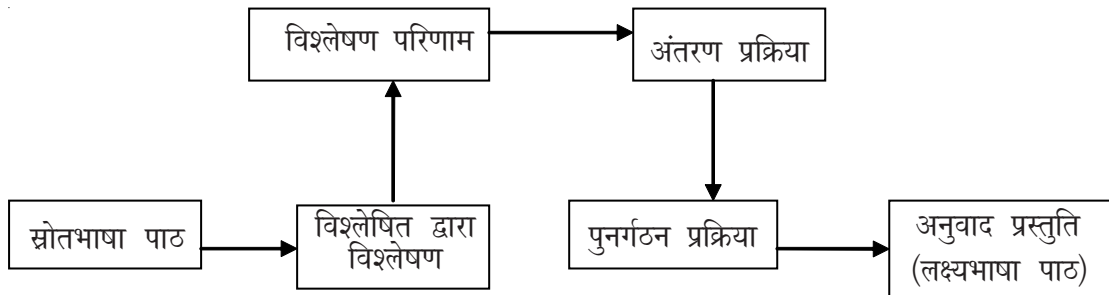
‘सिस्ट्रॉन’ (SYSTRAN), ‘कल्ट’ (CULT), ‘टाइटस’ (TITUS) एवं ‘लोगोस’ (LOGOS) मशीनी अनुवाद-तंत्रों के विकास पर आधारित आज प्रत्यक्ष मशीनी अनुवाद पद्धति में लगभग 20 भाषा-युग्मों में अनुवाद की सुविधा उपलब्ध है। किंतु यह अनुवाद शत-प्रतिशत शुद्ध नहीं होता, इससे केवल 50-60 प्रतिशत तक ही शुद्ध अनुवाद संभव है। इस मशीनी अनुवाद विधि के ‘सिस्ट्रान’ नामक परिष्कृत-संवर्द्धित संस्करण के जरिए इंटरनेट की कुछ वेबसाइटों में ऑन-लाइन अनुवाद किए जाते हैं। ये ऑन-लाइन अनुवाद अंग्रेजी, फ्रांसीसी, जर्मन और स्पेनिश भाषाओं में किए जाते हैं। प्रत्यक्ष मशीनी अनुवाद पद्धति का ऐतिहासिक महत्व अधिक है। आज, वस्तुस्थिति यह है कि इस प्रत्यक्ष मशीनी अनुवाद पद्धतियों में नई कंप्यूटर पद्धतियों का अब बहुत अनुसंधान-विकास नहीं हो रहा है। इस संबंध में कुछ चर्चा इस खंड की प्रथम इकाई में भी की जा चुकी है।

13.2.2 अंतरण मशीनी अनुवाद-पद्धति (Transfer Machine Translation System)

कंप्यूटर-विज्ञानियों ने 19वीं शताब्दी के सत्तर के दशक के अंत में अंतरण मशीनी अनुवाद पद्धति का प्रयोग करना शुरू किया। इस पद्धति के अंतर्गत पद-विच्छेद करते हुए स्रोतभाषा के प्रत्येक शब्द की व्याकरणिक कोटि को निर्धारित करके उसके वाक्यों की आंतरिक संरचना का गठन किया जाता है। इसके बाद दोनों स्तरों -- शब्द और संरचना -- पर अंतरण की प्रक्रिया का आश्रय लिया जाता है। अंतरण मशीनी अनुवाद पद्धति के अंतर्गत अनुवाद की प्रक्रिया तीन चरणों से होकर गुजरती है। ये हैं - (1) विश्लेषण (2) अंतरण और (3) पुनर्गठन। इसके बारे में आप पिछली इकाई में पहले पढ़ चुके हैं। कुछ संदर्भ आप अगली इकाई में भी देखेंगे।

जैसा कि आपको पहले बताया गया है कि मशीनी अनुवाद के पहले चरण में स्रोतभाषा के वाक्य का व्याकरणिक विश्लेषण कर उसकी पद-व्याख्या (पार्सिंग) की जाती है, किंतु वाक्य का यह व्याकरणिक विश्लेषण मूर्त न होकर अमूर्त रूप में होता है। दूसरे चरण के अंतर्गत कंप्यूटर, स्रोतभाषा के वाक्य की इस संरचना और उसके शब्दों को लक्ष्यभाषा की समानधर्मी संरचना और शब्दों में अंतरित (ट्रांसफर) करता है। तीसरे चरण में इस आउटपुट को लक्ष्यभाषा में पुनर्गठित किया जाता है। आउटपुट का लक्ष्यभाषा में यह पुनर्गठन वास्तव में इस पद्धति का अनुवाद आउटपुट होता है। उल्लेखनीय है कि अंतरण पद्धति से अनुवाद के लिए शब्द-सूचियों के रूप में क्रमशः जिन कोशों की आवश्यकता पड़ती है, वे हैं - (1) स्रोतभाषा शब्द-सूची (2) द्विभाषिक अंतरण शब्द-सूची तथा (3) लक्ष्यभाषा शब्द-सूची। अंतरण प्रक्रिया के तीनों चरणों में से पहले और तीसरे चरण में एकभाषिक कोश की तथा दूसरे चरण में द्विभाषिक कोश प्रयुक्त होता है।

अंतरण मशीन अनुवाद पद्धति को निम्नलिखित आरेख की सहायता से सरलता से समझा जा सकता है :



आरेख नं. 13.2

प्रत्यक्ष मशीनी ‘अनुवाद पद्धति’ की तुलना में “अंतरण पद्धति” बेहतर परिणाम उपलब्ध कराती है, क्योंकि प्रत्यक्ष मशीनी अनुवाद पद्धति में अनुद्य पाठ की संरचनात्मक जानकारी का उपयोग नहीं होता। “प्रत्यक्ष मशीनी अनुवाद पद्धति” और “अंतरण मशीनी अनुवाद पद्धति” में मूलभूत अंतर यही है कि पहली विधि में व्याकरण संबंधी सूचनाओं का उपयोग नहीं होता, जबकि दूसरी विधि में उनकी भूमिका अत्यधिक महत्वपूर्ण होती है। प्रत्यक्ष अनुवाद पद्धति में अनुवाद शाब्दिक होता है, क्योंकि इसमें अंतरण एक कोशीय प्रक्रिया है जो विश्लेषण और पुनर्गठन से भिन्न नहीं। विश्लेषण-पुनर्गठन के आधार पर बेहतर प्रस्तुति का प्रयास नहीं होता, जबकि अंतरण विधि में इसका अपना अलग अस्तित्व होता है। यही कारण है कि अंतरण पद्धति काफी लोकप्रिय हुई है।

अंतरण मशीनी अनुवाद पद्धति के आधार पर 1977 में टॉम मेटो (TAUM METEO) नामक एक मशीनी तंत्र विकसित किया गया। इस मशीनी अनुवाद तंत्र को मौसम संबंधी जानकारी का अंग्रेजी से फ्रांसीसी में अनुवाद करने के लिए कनाडा के मांट्रियल विश्वविद्यालय में विकसित किया था। यह तंत्र आज भी कार्यरत है। इसलिए इस अनुवाद पद्धति को आज भी विश्व की अत्यंत सफल पद्धतियों में माना जाता है। यह पूर्णतः स्वचालित मशीनी अनुवाद तंत्र है। इस तंत्र की आधारभूत शब्दावली बहुत सीमित है। इसमें लगभग 1500 शब्दों की प्रविष्टियाँ हैं। इन शब्दों में से भी आधे शब्द वास्तव में स्थानों के नाम हैं। वाक्य संरचनाएँ सरल और सीमित हैं। इसमें अनेकार्थकता की संभावना बिल्कुल नहीं है यानी प्रत्येक शब्द का अर्थ सुनिश्चित है। सीमित परिवेश, शब्दावली और वाक्य-विन्यास के कारण ही संभवतः यह पूर्णतः स्वचालित मशीनी अनुवाद पद्धति के रूप में सफल सिद्ध हुई है और इसे एक सफल अनुवाद तंत्र माना जाता है। इस सीमित किंतु स्वचालित मशीनी अनुवाद तंत्र की सफलता से प्रोत्साहित होकर एक व्यापक विषय “एविएशन” को लेकर भी अनुवाद तंत्र ‘TAUM AVIATION’ विकसित करने का प्रयास किया गया, किंतु विषय की व्यापकता के कारण यह मशीनी अनुवाद तंत्र सफल नहीं हो पाया। इसके अलावा, मांट्रियल विश्वविद्यालय ने टॉम-71, टॉम-73, टॉम-76 नामक प्रायोगिक पद्धतियाँ भी विकसित की गई थीं।

इस अंतरण मशीनी अनुवाद पद्धति में और भी नई-नई कंप्यूटर पद्धतियाँ विकसित हुई हैं। इन पद्धतियों में ‘सूसी’ (Saarbücker über setzungs system - SUSY), ‘गेटा’ (Groupe d’Etudes pour la Traduction Automatique - GETA) आदि विशेष रूप से उल्लेखनीय हैं। 1984 में विकसित और रूसी-जर्मनी प्रोटोटाइप सिस्टम पर आधारित ‘सूसी’ मशीनी अनुवाद तंत्र में वाक्य स्तर पर जर्मन, रूसी, फ्रांसीसी, अंग्रेजी और एस्पेरंटो भाषाओं के परस्पर अनुवाद की व्यवस्था थी। ‘गेटा’ प्रणाली को फ्रांस के गेनोबल विश्वविद्यालय ने 1985 में विकसित किया था।

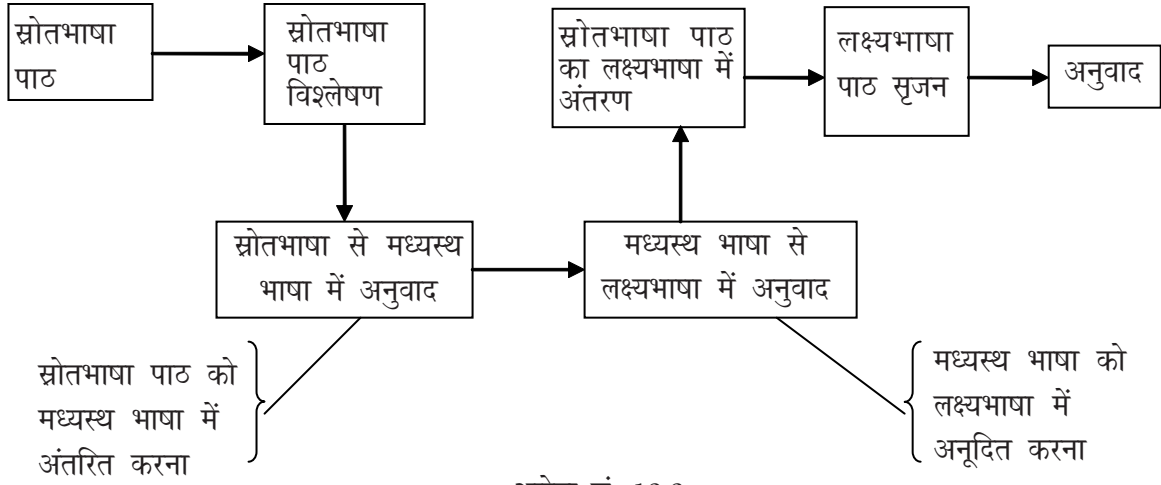
13.2.3 अंतरभाषिक मशीनी अनुवाद पद्धति (Interlingua Machine Translation System)

अंतरभाषिक मशीनी अनुवाद पद्धति का ‘अंतरण मशीनी अनुवाद पद्धति’ के विकल्प के रूप में विकास हुआ है। ‘मध्यस्थ भाषा’ इस पद्धति का सर्वाधिक महत्वपूर्ण अंग है। इस पद्धति में स्रोतभाषा और लक्ष्यभाषा के बीच एक अन्य ‘मध्यस्थ भाषा’ (Interlingua Language) का सृजन किया जाता है। इस मध्यस्थ भाषा को ‘सार्वभौम भाषा’, ‘इंटरलिंग्वा’ अथवा ‘अंतरभाषा’ भी कहा जाता है। नियम आधारित मशीन अनुवाद प्रतिमान के अंतर्गत यह अंतरभाषिक मशीनी अनुवाद पद्धति, प्रत्यक्ष मशीनी अनुवाद और अंतरण मशीनी अनुवाद पद्धति का विकल्प है।

अंतरभाषिक मशीनी अनुवाद पद्धति के अनुसार स्रोतभाषा के पाठ या वाक्य को लक्ष्यभाषा के स्थान पर पहले मध्यस्थ भाषा से जोड़ा जाता है यानी स्रोतभाषा के पाठ या वाक्य को विश्लेषित करके मध्यस्थ भाषा में प्रस्तुत किया जाता है। यह मध्यस्थ भाषा सामान्य मानक भाषा न होकर औपचारिक गणितीय भाषा होती है। मध्यस्थ भाषा सभी भाषाओं के बीच आम मध्यस्थ का काम करती है। इसके लिए पहले प्रत्येक भाषा की समझ जरूरी है। उसके बाद ही उस समझ को मध्यस्थ भाषा में प्रस्तुत किया जाता है। भाषा-सर्जक मध्यस्थ भाषा के प्रस्तुतीकरण को लक्ष्यभाषा के पाठ या वाक्य में बदल देता है। इस तरह की व्यवस्था में भाषा-विश्लेषक और भाषा-सर्जक दो भिन्न-भिन्न स्थितियों से जुड़े होते हैं। कहने का अभिप्राय यह है कि इस अनुवाद पद्धति में स्रोतभाषा के अर्थ का मध्य भाषा में निरूपण करते हुए इस निरूपित अर्थ को लक्ष्यभाषा की अभिव्यक्तियों के रूप में संयोजित-संश्लेषित किया जाता है। यह मध्यस्थ भाषा इस प्रकार की होनी चाहिए कि किसी भी स्रोतभाषा की उपलब्ध सामग्री के अर्थ का पूरी तरह से और सही-सही अर्थ निरूपित कर सके।

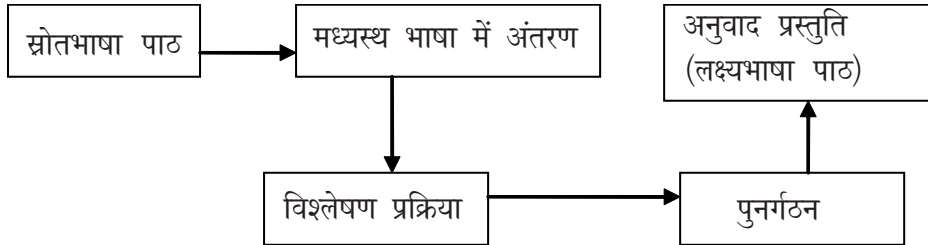
यहाँ यह संकेत करना उपयुक्त प्रतीत होता है कि मध्यस्थ भाषा कोई भाषा-विशेष न होकर एक किसी भाषा-निरपेक्ष भाषा होती है जिसमें विश्व-ज्ञान को अवधारणाओं को प्रतीकों के रूप में निरूपित किया जाता है। इन अवधारणाओं की कई कोटियाँ होती हैं और उनका अपना एक निश्चित पदक्रम (hierarchy) होता है। ये कोटियाँ जाति-सदस्य संबंध, अंग-अंगी संबंध आदि कई प्रकार के संबंधों से परस्पर जुड़ी रहती हैं। उदाहरण के लिए, हाथ और शरीर का, सँड और हाथी का, जड़ और पेड़ का अंग-अंगी संबंध। विश्व-ज्ञान का निरूपण और कोटिकरण करने के संबंध में एकाधिक विधियाँ हो सकती हैं, किंतु प्रत्येक कोटि को अर्थपरक लक्षणों के रूप में प्रस्तुत किया जाता है।

वाक्यात्मक दृष्टिकोण वाला यह अनुवाद-तंत्र 'वाक्य विश्लेषण तकनीक' पर आधारित है। अंतरभाषिक मशीनी अनुवाद की स्थिति का आरेखीय निरूपण निम्न प्रकार से है :



आरेख सं. 13.3

अंतरभाषिक मशीनी अनुवाद पद्धति में भी अनुवाद प्रक्रिया के तीन चरण शामिल हैं। पहले चरण में स्रोतभाषा पाठ को समझते हुए उसके वाक्य की व्याख्या-विश्लेषण करके उसे भाषा-निरपेक्ष संकल्पनात्मक या ज्ञान-प्रतीकों के रूप में निरूपित किया जाता है। दूसरे चरण में इस आउटपुट को अनुमिति (inference) तथा पाठ में निहित संदर्भों से जोड़ते हुए उनके आधार पर समृद्ध एवं संवर्धित किया जाता है। समृद्धि और संवर्धन की इस प्रक्रिया में, आउटपुट में विश्व-ज्ञान का पुट शामिल होता है। तीसरे चरण में इसका भाषा-सापेक्ष संरचना में प्रतिचित्रण (मैपिंग) होता है। इस प्रतिचित्रण कार्य में प्राकृतिक भाषा जनित्र (Natural Language Generator) सहायक सिद्ध होता है। जेनरेटर भाषा-निरपेक्ष प्रस्तुति के विभिन्न भागों को लक्ष्यभाषा में व्यवस्थित करते हुए आउटपुट के रूप में पुनर्गठित करता है और इस तरह हमें लक्ष्यभाषा में अनुवाद प्राप्त हो जाता है। अंतरभाषिक मशीनी अनुवाद की प्रक्रिया को निम्नलिखित आरेख की सहायता से भी स्पष्ट किया जा सकता है :

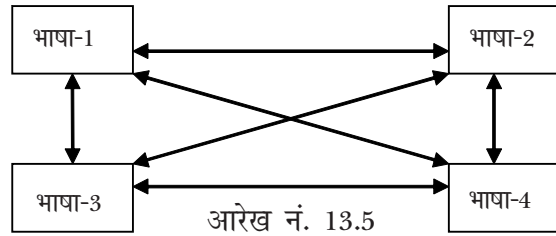


आरेख सं. 13.4

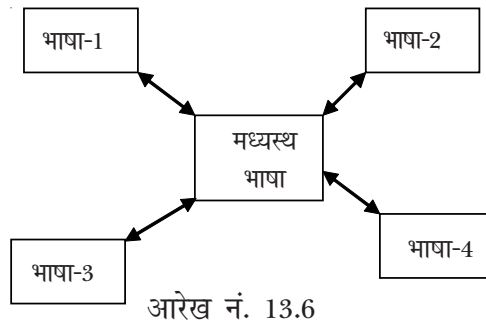
अंतरभाषिक मशीनी अनुवाद प्रणाली को विकसित करने के लिए जिन संसाधनों की अनिवार्य रूप से आवश्यकता होती है उनमें से सर्वप्रमुख है - भाषा का विश्लेषण और पुनर्गठन करने के लिए शब्दभंडार। इसके अलावा क्षेत्र-विशेष (domain specific) का संकल्पनात्मक शब्दभंडार भी जरूरी होता है। यह संकल्पनात्मक शब्दभंडार क्षेत्र-विशेष के लिए ज्ञान-आधार (knowledge-base) होता है। शब्दभंडार के अलावा अनुवाद प्रणाली को क्षेत्र-विशेष और भाषाओं के प्रक्षेपण नियमों से जोड़ना भी होता है। भाषा-विशेष के विश्लेषण और पुनर्गठन के लिए व्याकरणों का होना भी जरूरी होता है तभी इस अनुवाद पद्धति को विकसित कर पाना संभव होता है।

स्पष्ट है कि सभी भाषाओं के बीच एक आम मध्यस्थ भाषा इस पद्धति का आधार है। इसलिए इस पद्धति के प्रयोक्ताओं का यह मानना है कि स्रोतभाषा से लक्ष्यभाषा में अनुवाद के बीच यह मध्यस्थ भाषा होने की वजह से ऐसा लगता है कि मानो काम दोगुना बढ़ गया है। लेकिन इस मध्यस्थ भाषा की अवधारणा का व्यावहारिक पक्ष यह है कि यह प्रणाली ज्यादा से ज्यादा भाषाओं के बीच अनुवाद की दृष्टि से महत्वपूर्ण है। उदाहरण के लिए, अगर नौ भाषाओं के बीच अनुवाद किया जाए तो 72 अनुवादकों की आवश्यकता पड़ेगी जबकि मध्यस्थ भाषा पद्धति से अनुवाद करने पर केवल 18 अनुवादकों की।

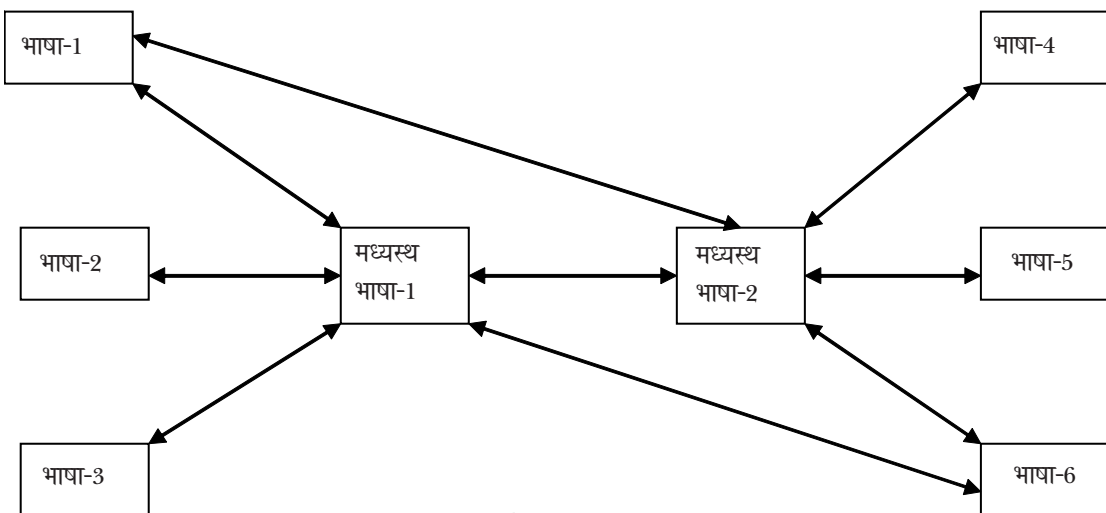
तकनीकी दृष्टि से और भाषा के संदर्भ में देखा जाए तो बहुभाषिक मशीनी अनुवाद करते समय प्रत्येक भाषा-युग्म के लिए अंतरण घटक निर्मित करने की आवश्यकता नहीं पड़ती। उदाहरण के लिए, अगर चार भाषाओं में अनुवाद-कार्य के लिए चार भाषा-युग्म हों तो हमें 13 शब्दकोशों की जरूरत होगी। अगर इन चारों भाषा-युग्मों का 'प्रत्यक्ष अनुवाद पद्धति' अथवा 'अंतरण अनुवाद पद्धति' आधारित मशीनी अनुवाद किया जाए तो निम्नलिखित आरेख के अनुसार 13 अनुवाद मंडल विकसित करने होंगे :



अगर इन्हीं चार भाषाओं में अनुवाद-कार्य के लिए चार भाषा-युग्मों का 'अंतरभाषिक अनुवाद पद्धति' के अंतर्गत मध्यस्थ भाषा आधारित मशीनी अनुवाद किया जाए तो निम्नलिखित आरेख के अनुसार केवल 8 अनुवाद मॉड्यूल विकसित करने की आवश्यकता होगी :



इस पद्धति से यह भी संभावना बनी है कि अनुवाद में कभी-कभी दो अंतरभाषिक मशीनी अनुवाद पद्धतियों का प्रयोग किया जाए। हो सकता है कि इन दोनों में से कोई एक अंतरभाषिक पद्धति स्रोतभाषा की अपेक्षाकृत अधिक विशेषताओं का प्रयोग करे और दूसरी अंतरभाषिक पद्धति में लक्ष्यभाषा की विशेषताओं की भरमार हो। उस स्थिति में एक भाषा के वाक्यों का दो चरणों में लक्ष्यभाषा के अपेक्षाकृत अधिक नजदीक वाक्यों में अनुवाद किया जाता है। पहली भाषा के वाक्यों को लक्ष्यभाषा के उन वाक्यों में परिवर्तित किया जाता है जो उस भाषा के बहुत निकट होते हैं। यह पद्धति इस तरह से भी व्यवस्थित की जा सकती है कि दूसरे अंतरभाषिक अनुवाद में लक्ष्यभाषा की उस विशिष्ट शब्दावली का ज्यादा प्रयोग किया जाए जो लक्ष्यभाषा के ज्यादा नजदीक अथवा समरूप हो। इस प्रकार, दो अंतरभाषिकों का इस्तेमाल करके मशीनी अनुवाद की गुणवत्ता में सुधार किया जा सकता है। दो अंतरभाषिकों के इस्तेमाल को निम्नलिखित आरेख के द्वारा स्पष्ट किया जा सकता है :



अपने आदर्श रूप में अंतरभाषिक मशीनी अनुवाद प्रणाली के आधार पर यदि विश्व-ज्ञान निरूपण की कोई सफल सार्वभौम भाषा विकसित हो जाती है तो एक ही युक्ति से एक से अधिक भाषाओं के लिए मशीन अनुवाद पद्धति विकसित करना संभव हो सकेगा। तब केवल दो भाषाओं के बीच ही नहीं बल्कि विभिन्न भाषाओं के बीच परस्पर अनुवाद की व्यवस्था संभव हो सकेगी। किंतु यह एक जटिल कार्य है। मध्यस्थ भाषा के स्वरूप को लेकर भी समस्या है। इसके अलावा, संबंधित भाषाओं की समानताओं और विषमताओं को सीधे-सीधे नियमों के साथ नहीं जोड़ पाने की भी सीमा इस पद्धति के प्रभावी अनुप्रयोग के मार्ग में बाधक है। इसलिए इस क्षेत्र में आर्थी विश्लेषण और कृत्रिम बुद्धि पर बहुत अधिक अनुसंधान करने की जरूरत है। वहीं इस पद्धति के प्रयोक्ताओं का यह भी मानना है कि यह मशीनी अनुवाद पद्धति संपूर्ण अनुवाद करने के स्थान पर स्रोतभाषा पाठ का लक्ष्यभाषा में केवल सार प्रस्तुत कर देता है।

अंतरण मशीनी अनुवाद पद्धति की तुलना में अंतरभाषिक मशीनी अनुवाद पद्धति को श्रेयस्कर माना जाता है क्योंकि इस पद्धति के अनुसार बहुभाषायी अनुवाद संभव है। इसके अलावा, इस पद्धति के अनुसार पहले स्रोतभाषा पाठ को समझा जाता है, तत्पश्चात् उसका भावानुवाद किया जाता है। इसलिए इस पद्धति का अनुवाद अपेक्षाकृत अधिक सरल और सहज है। मध्यस्थ भाषायी पद्धति की पूरी तकनीक विकसित न होने की स्थिति में अंतरण पद्धति ठीक है। किंतु अगर प्रत्येक भाषा को भली प्रकार से समझकर उसे माध्यम भाषा में प्रस्तुत कर दिया जाए तो अंतरभाषिक मशीनी अनुवाद पद्धति बेहतर मानी जाती है।

1970 के दशक में विकसित मेटल (METAL) अनुवाद परियोजना में अंतरभाषिक मशीनी अनुवाद पद्धति को व्यवहार में लाया गया। इस दौरान स्टेनफोर्ड में भी रोजर शैंक (Roger Schank) और यॉरिक विल्क्स (Yorik Wilks) ने आरंभिक अंतरभाषिक मशीनी अनुवाद पद्धतियाँ तैयार कीं। वहीं 1980 के दशक में इसे काफी महत्व दिया गया और इस क्षेत्र में काफी शोध किया गया। अब उत्त्रेख्त (Utrecht~) में 'वितरित भाषा अनुवाद' (Distributed Language Translation –DLT) में काफी कार्य हुआ है।

बहुभाषा-भाषी देशों अथवा बहुभाषी मशीनी अनुवाद प्रणाली को विकसित करने की दृष्टि से अंतरभाषिक मशीनी अनुवाद पद्धति बहुत उपयोगी है। अनुवाद की मानवीय प्रक्रिया के सबसे अधिक निकट होने के कारण भी यह पद्धति महत्वपूर्ण मानी जाती है। अंतरभाषिक विधि से चीन, जापान, इंडोनेशिया, मलेशिया एवं थाईलैंड की भाषाओं में परस्पर अनुवाद के लिए मशीनी अनुवाद प्रणाली विकसित हो रही है। कुछ अन्य मशीनी अनुवाद पद्धतियों की चर्चा अगली इकाई में की जाएगी।

13.3 मशीनी अनुवाद पद्धति विकास का आधुनिक चरण : कृत्रिम बुद्धि

'अंतरभाषिक मशीनी अनुवाद पद्धति' (Interlingua Machine Translation System) वस्तुतः वाक्य-विश्लेषण तकनीक पर आधारित थी। बाद में इस पद्धति में 'कृत्रिम बौद्धिकता' (Artificial Intelligence) नामक एक नया आयाम जोड़ा गया है। कृत्रिम बुद्धि का अर्थ है - मशीनों की बुद्धिमत्ता। कृत्रिम बुद्धि कंप्यूटर विज्ञान की एक शाखा है। कृत्रिम बुद्धि, मानव बुद्धि से जुड़े कुछ विशिष्ट प्रकार के कार्यों का अनुकरण करने वाली कंप्यूटर तकनीकों का अध्ययन है। यह मशीन का बुद्धिमान अमूर्त एजेंट है जो अपने परिवेश को समझता है और उसे सौंपे गए काम को ज्यादा से ज्यादा सफलतापूर्वक तरीके से संपन्न करता है। जॉन मैकार्थी (John McCarthy) ने इस शब्द को 1956 में गढ़ा था और इसे बुद्धिमान मशीनें बनाने का विज्ञान और प्रौद्योगिकी कहा था। मशीनों को बुद्धिमान बनाने संबंधी दृष्टिकोण मानव-मास्तिष्क के अध्ययन संबंधी तंत्रिका विज्ञान (Neurology) के क्षेत्र में हुई हाल ही की खोजों और सूचना का "साइबरनेटिक्स" नामक नए गणितीय सिद्धांत और गणित आधारित मशीन (अर्थात् अंकीय/डिजिटल कंप्यूटर) की खोज पर आधारित है। जहाँ डिजिटल कंप्यूटर गणितीय तर्क-शक्ति के अमूर्त सार पर आधारित है वहीं साइबरनेटिक्स विज्ञान का वह क्षेत्र है जिसमें मशीन और मानव के बीच स्वतः नियंत्रण और स्थायी समझ के साथ-साथ संचार का तुलनात्मक अध्ययन शामिल है। इन अध्ययनों में सूचना-संचार संबंधी मशीनों और प्राणियों एवं मानव-मास्तिष्क एवं तंत्रिका तंत्र के बीच अध्ययन शामिल है। ज्ञान प्रस्तुतिकरण और ज्ञान इंजीनियरी, कृत्रिम बुद्धि अनुसंधान का केंद्र हैं। इसके द्वारा प्राकृतिक भाषा को कंप्यूटर भाषा द्वारा समझने में बड़ी मदद मिलती है।

कंप्यूटर अनुवाद पद्धति को कृत्रिम बुद्धि से युक्त करने की संकल्पना इस तथ्य पर आधारित है कि इसे एक सीमा तक बुद्धियुक्त बनाया जा सकता है और उसमें भाषा संबंधी मूलभूत व्याकरणिक नियमों अथवा सूत्रों को संचित किया जा सकता है। 'कृत्रिम बुद्धि', मानव-बुद्धि की कार्य-प्रणाली और निर्णय-प्रक्रिया का बौद्धिक अनुकरण करती है, जिसे कंप्यूटर की भाषा में 'कंप्यूटर-अनुकरण' कहा जा सकता है। इसमें वाक्यपरक, अर्थपरक और संदर्भपरक तत्वों को समन्वित रूप से देखने का प्रयास किया जाता है। कृत्रिम बुद्धि के कारण कंप्यूटर भी मानव की भाँति निर्णय और आउटपुट दे सकता है। वास्तव में मानव-मस्तिष्क में सूक्ष्म और अव्यक्त रूप में विद्यमान ज्ञानानुभव रूपी तर्कों, नियमों एवं प्रतिबंधों को कंप्यूटर के स्मृति कोश में डालने पर कंप्यूटर अनुवाद पद्धति का सार्थक परिचालन संभव हो पाता है। इन्हें गणित के एल्गोरिद्म अथवा सूत्रों के जरिए कंप्यूटर के स्मृति कोश (मेमोरी) में डाल देने पर मानव की भाँति निर्णय और आउटपुट संभव होता है। गणितीय एल्गोरिद्म विधि वास्तव में प्रश्नों को हल करने की गणितीय और/अथवा तार्किक संक्रियाओं की चरण-दर-चरण क्रियाविधि है। इस तरह, कृत्रिम बुद्धि के अंतर्गत कंप्यूटर को मानव की भाँति तर्क करने, समस्याओं का हल खोजने और स्वयं निर्णय लेने के साथ-साथ "परीक्षण और अशुद्धि पद्धति" (Trial and Error Method) से स्वयं सीख लेने योग्य बनाया जाता है। आज कंप्यूटर अनुवाद पद्धति के कुछ मॉडलों में कृत्रिम बुद्धि संबंधी तकनीकों का उपयोग किया जा रहा है।

कंप्यूटर अनुवाद के क्षेत्र में कृत्रिम बुद्धि के सार्थक उपयोग के बावजूद यह स्वीकार किया जाता है कि व्याकरणिक लक्षणों और संदर्भ से जुड़ी व्यापक भाषिक अभिव्यक्तियों को पूरी तरह से नियमबद्ध कर पाना अत्यंत कठिन कार्य है। यही कारण है कि आजकल सीमित व्यवहार-क्षेत्रों या प्रयोजनों के लिए कंप्यूटर अनुवाद प्रणाली विकसित करने और उनसे संबंधित प्रयुक्तियों के अनुवाद की संभावना पर अधिक बल दिया जा रहा है। सीमित व्यवहार-क्षेत्रों में प्रयुक्त होने की वजह से उनमें द्वि-अर्थकता नहीं होती अथवा बहुत ही कम होती है। इससे जहाँ क्षेत्र-विशेष की समग्र शब्द-सूची को फीड करके कॉर्पस निर्माण किया जाता है, वहीं अनुवाद का एक विश्वसनीय व्याकरण विकसित करना भी सरल हो जाता है। परिणामस्वरूप कंप्यूटर द्वारा किए गए अनुवाद की शुद्धता एवं विश्वसनीयता की प्रतिशतता में भी बढ़ोत्तरी होती है।

13.4 मशीनी अनुवाद के प्रकार

अगर अनुवादक द्वारा ही अनुवाद-कार्य होता है तो वह 'विशुद्ध मानव अनुवाद' कहलाता है किंतु कंप्यूटर के विकास ने मशीनी अनुवाद को संभव बना दिया है। वास्तव में कंप्यूटर ने अनुवाद की प्रक्रिया को सरल एवं त्वरित बना दिया है। कंप्यूटर अनुवाद प्रणाली की जटिलता मुख्य रूप से स्रोत और लक्ष्यभाषा के व्याकरणिक स्तरों, उनके भाषिक पक्षों तथा शब्दावली पर निर्भर करती है। इस आधार पर कंप्यूटर अनुवाद को सिद्धांततः दो भागों में विभाजित किया जा सकता है। ये हैं - 'समान भाषा परिवार की दो भाषाओं के बीच अनुवाद' और 'दो अलग-अलग भाषा परिवारों की भाषाओं में परस्पर अनुवाद'। पहली विधि समान भाषा परिवार की दो अलग-अलग भाषाओं के बीच कंप्यूटर के जरिए अनुवाद करने से संबंधित है, जो अपेक्षाकृत सरल होती है। उदाहरण के लिए, 'आर्यभाषा परिवार' की दो भाषाओं (बांग्ला-हिंदी, मराठी-हिंदी आदि) में परस्पर अनुवाद करना इस प्रकार के अनुवाद में सहजता एवं स्वाभाविकता होने से प्रक्रिया अधिक सकारात्मक परिणाम वाली होती है। दूसरी विधि अलग-अलग परिवार की उन भाषाओं के बीच कंप्यूटर अनुवाद से संबंधित है जिनमें वाक्य-विन्यास और शब्दावली के स्तर पर काफी भिन्नता पाई जाती है। उदाहरण के लिए, अंग्रेजी से हिंदी में अनुवाद। किंतु जहाँ तक मशीन अनुवाद के प्रकारों का संबंध है, इसे मानव-मशीन सहभागिता के संदर्भ में देखा जाना चाहिए। अनुवाद में कंप्यूटर की सहभागिता के कारण अस्तित्व में आई 'कंप्यूटर अनुवाद' अथवा 'मशीनी अनुवाद' की संकल्पना दो प्रकार से मूर्त रूप प्राप्त करती है। ये हैं - 'पूर्णतः स्वचालित कंप्यूटर अनुवाद', और 'अर्ध-स्वचालित कंप्यूटर अनुवाद'।

13.4.1 पूर्णतः स्वचालित मशीनी अनुवाद (Fully Automatic Machine Translation - FAMT)

'पूर्णतः स्वचालित मशीनी अनुवाद' में आम तौर पर मानव का हस्तक्षेप आवश्यक नहीं समझा जाता है। यहाँ मानव का हस्तक्षेप का अर्थ यह है कि कंप्यूटर द्वारा प्रस्तुत किए गए अनुवाद में मानव-अनुवादक के स्तर पर

संशोधन-परिवर्धन नहीं किया जाता। इसमें अनुवाद की पूरी प्रक्रिया मशीन के द्वारा तय की जाती है और मनुष्य किसी भी तरह से और किसी भी रूप में कोई भूमिका नहीं निभाता। इस प्रकार के कंप्यूटर अनुवाद में व्यक्ति पूरी तरह से कंप्यूटर अनुवाद प्रणाली पर निर्भर रहता है। इसलिए इसे पूर्ण मशीनी अनुवाद भी कहा जाता है अर्थात् इसमें इनपुट डालकर प्राप्त आउटपुट एक अंतिम पाठ (अनूदित) स्वीकार कर लिया जाता है।

मौसम संबंधी सूचनाओं को अंग्रेजी से फ्रांसीसी में अनूदित करने के लिए मांट्रियल विश्वविद्यालय ने 1977 में जो टॉम मेटो (TAUM METEO) अनुवाद तंत्र विकसित किया था वह मूलतः पूरी तरह से स्वचालित मशीनी अनुवाद प्रणाली थी। इसी प्रकार, वैज्ञानिक एवं तकनीकी प्रलेखों के जापानी और अंग्रेजी के बीच अनुवाद के लिए 1985 में विकसित एमयू (MU) मशीनी अनुवाद प्रणाली भी पूरी तरह से स्वचालित मशीनी अनुवाद प्रणाली ही थी। क्षेत्र की दृष्टि से सीमित होने के कारण इस प्रकार की मशीनी अनुवाद प्रणालियों में सभी शब्द और वाक्यों के अर्थ सीमित एवं सुनिश्चित रहते हैं। यह पद्धति भी पूर्णतया आदर्श अनुवाद नहीं दे पाई है। वैसे इसके प्रयोग की सफलता के बारे में वैज्ञानिकों में मतैक्य नहीं है। कुछ का यह मानना है कि सफल पूर्ण मशीनी अनुवाद कभी भी संभव नहीं होगा जबकि कुछ का विचार इसके विपरीत है। वे यह मानते हैं कि इस प्रकार का पूर्ण मशीनी अनुवाद कभी न कभी अवश्य संभव होना है।

13.4.2 अर्ध-स्वचालित मशीनी अनुवाद (Semi-Automatic Machine Translation – SAMT)

अनुवाद प्रक्रिया मूलतः बौद्धिक प्रक्रिया है और मनुष्य ही एकमात्र प्राणी है जो इस क्षमता से संपन्न है, जबकि कंप्यूटर एक मशीन मात्रा है, व्यक्ति नहीं। विश्वभर में प्रचलित कोई भी मशीनी अनुवाद प्रणाली ऐसी नहीं है, जिसमें मानव का हस्तक्षेप बिलकुल न हो। इस तरह मशीनी अनुवाद प्रणालियाँ सिद्धांततः पूरी तरह से स्वचालित और सामान्य उद्देश्यीय नहीं हैं। इसलिए 'अर्ध-स्वचालित मशीनी अनुवाद' की अवधारणा बनी रहती है। अर्ध-स्वचालित मशीनी अनुवाद और उसके विविध प्रकारों के बारे में आगे विस्तार से चर्चा की जा रही है।

13.5 अर्ध-स्वचालित मशीनी अनुवाद (SAMT) के विविध प्रकार

जिस प्रकार अनुवादक, स्रोतभाषा पाठ की अभिव्यक्ति के अभिप्राय को ग्रहण करते हुए उसका लक्ष्यभाषा में पुनर्कथन करता है उसी प्रकार कंप्यूटर के जरिए स्रोतभाषा पाठ की अभिव्यक्ति के अभिप्राय को ग्रहण करते हुए उसका लक्ष्यभाषा में पुनर्कथन संभव है। किंतु कंप्यूटर यह पुनर्कथन अपने आप नहीं करता। इसमें मशीन के साथ मानव की सहभागिता अवश्यभावी है। वह सहभागिता मानव की सहायक के रूप में भी हो सकती है और कंप्यूटर की सहायक के रूप में इस्तेमाल करने संबंधी भी। मानव-अनुवादक और कंप्यूटर की इस परस्पर सहभागिता के परिप्रेक्ष्य में अर्ध-स्वचालित मशीनी अनुवाद को 'मशीन-साधित मानव अनुवाद' (Machine-Assisted Human Translation – MAHT) और 'मानव-साधित मशीनी अनुवाद' (Human-Assisted Machine Translation – HAMT) के रूप में वर्गीकृत किया जाता है। मशीनी अनुवाद के क्षेत्र में विश्व में जो अधिकतर प्रणालियाँ विकसित की गई हैं, उनमें इस प्रकार की सहायता की बराबर स्थिति बनी रहती है।

13.5.1 मानव-सहायक मशीनी अनुवाद

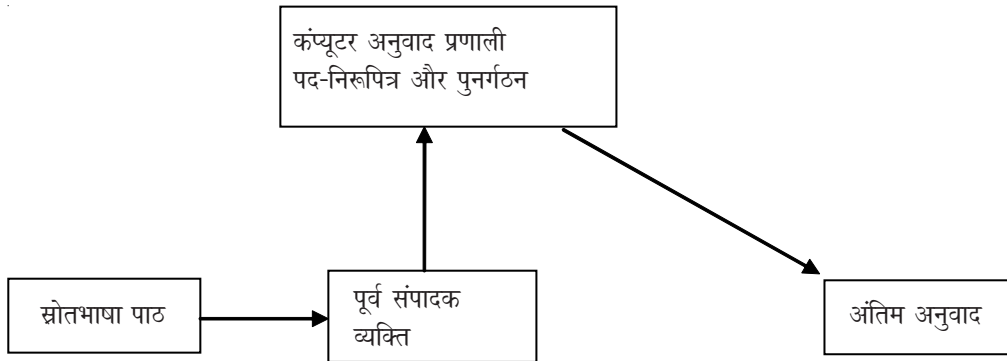
'मानव-साधित मशीनी अनुवाद' प्रणाली मानव-नियंत्रित अनुवाद है जिसमें कंप्यूटर अनुवाद प्रणाली मनुष्य के सहायक के रूप में काम करती है। इसके अंतर्गत अनुवाद कार्य तो कंप्यूटर द्वारा संपन्न होता है किंतु इस प्रक्रिया में मानव की सहायता ली जाती है। इसमें मनुष्य की सहभागिता-भूमिका कहीं न कहीं निरंतर बनी रहती है। यहाँ सहभागिता का अर्थ है - संदिग्धार्थक अंशों के अनुवाद का मानव द्वारा संपादन। इसमें संदिग्धार्थक वाक्यों के निहितार्थ को ध्यान में रखते हुए मानव की सहायता से ठीक करके मूल पाठ की श्लिष्ट संरचनाओं का अर्थ निर्धारण किया जाता है। मानव द्वारा अनुवाद का संपादन, कंप्यूटर द्वारा अनुवाद किए जाने से पूर्व भी हो सकता है, कंप्यूटर द्वारा अनुवाद किए जाने के दौरान भी और कंप्यूटर-अनुवाद के पश्चात भी। इसके अलावा, अनुवाद-पूर्व मानव-मशीन में अंतर्क्रिया करते हुए भी अनुवाद संभव है। इस दृष्टि से मानव-साधित कंप्यूटर अनुवाद संपादन के प्रमुख रूप इस प्रकार हैं :

- पूर्व संपादन (Pre-editing)
- पश्च संपादन (post-editing)
- अंतरा-संसाधन संपादन (intra processing editing) और
- अंतर्क्रियात्मक संपादन (interactive editing)

(i) **पूर्व संपादन प्रणाली**

पूर्व-संपादन प्रणाली वस्तुतः कंप्यूटर द्वारा अनुवाद किए जाने से पूर्व संपादन किए जाने से संबंधित है। इसमें मशीन द्वारा अनुवाद किए जाने से पूर्व ही स्रोतभाषा पाठ को व्यक्ति द्वारा संपादित किया जाता है। इस प्रणाली के अंतर्गत व्यक्ति स्रोतभाषा पाठ को पढ़ता है और उसे इस प्रकार संपादित करता है कि कंप्यूटर प्रणाली उसका स्वतः ही अनुवाद कर दे। अनुवाद-पूर्व संपादन कार्य करने की वजह से उस व्यक्ति-विशेष को 'मशीन अनुवाद-पूर्व संपादक' (Machine Translation Pre-Editor) भी कहा जाता है। पूर्व-संपादन के दौरान वह स्रोतभाषा सामग्री में व्याप्त कठिनाइयों, यथा शब्दों की वर्तनी को सही करना, योजकों का सही चुनाव तथा विभिन्न पदबंधों में सह-संबंधता अथवा एनाफ़र की कमी आदि दूर करता है। पूर्व-संपादन के दौरान ही जटिल व्याकरणिक संरचनाओं, अनेकार्थकताओं, संदिग्ध शब्दों एवं समस्यापरक वाक्यगत विशिष्टताओं का भी समाधान किया जाता है। किंतु इस बात का ध्यान रखा जाता है कि मूल पाठ को संपादित करने के दौरान और अशुद्धियों को दूर करने की प्रक्रिया में मूल पाठ की संरचना भंग न हो। पूर्व संपादन कार्य अधिकांशतः कागज पर छपी स्रोतभाषा सामग्री पर व्यक्ति ही करता है। इसके अलावा, स्रोतभाषा सामग्री का संपादन कंप्यूटर में विशेष प्रोग्राम की सहायता से भी हो सकता है।

चूँकि इसमें लेखन नियमों का ध्यान रखना पड़ता है और उनके आधार पर स्रोतभाषा सामग्री को संपादित करना पड़ता है इसलिए यह काफी जटिल प्रक्रिया होती है, जबकि मानव-भाषा प्रयोग नियम-शासित न होकर नियम-प्रभावित होती है। इस समस्या को ध्यान में रखते हुए यह युक्ति अपनाई जाती है कि कंप्यूटर में पाठ का विश्लेषण करने वाले प्रोग्राम के जरिए पूर्व-संपादन कार्य कर लिया जाए। इस विश्लेषण के मूल में यह दृष्टि रहती है कि स्रोतभाषा सामग्री के शब्द और व्याकरणिक संरचना अगर मानक भाषा के अनुसार नहीं है तो उनका पूर्व-संपादन कर लिया जाए। फ्रांस में टाइटस (TITUS) मशीनी अनुवाद प्रणाली में पूर्व-संपादन प्रणाली प्रयुक्त होती है। पूर्व संपादन प्रणाली को निम्नलिखित आरेख से समझा जा सकता है :



आरेख नं. 13.8

वैसे पाठ को मशीन के द्वारा स्वीकार्य रूप में परिवर्तित करने की प्रक्रिया में बहुत अधिक समय लग जाता है। इसके अलावा, इसमें स्रोतभाषा पाठ के अर्थ एवं आशय की सूक्ष्मता से 'मशीन अनुवाद पूर्व-संपादक' द्वारा परिवर्तन की संभावना भी बनी रहती है। इस प्रकार के कतिपय कारणों से पूर्व-संपादन प्रणाली को बेहतर नहीं माना जाता है।

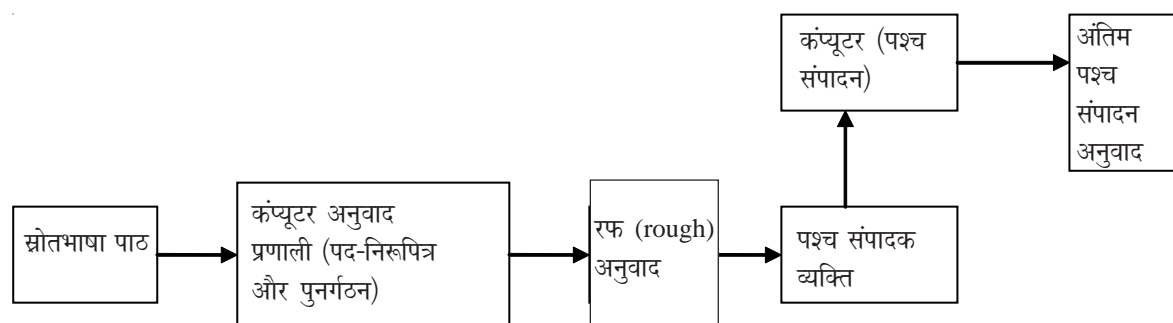
(ii) **पश्च संपादन प्रणाली**

पश्च-संपादन प्रणाली के अंतर्गत मशीनी अनुवाद में कंप्यूटर से अधिकाधिक काम लिया जाता है। इस प्रणाली में स्रोतभाषा पाठ का कंप्यूटर द्वारा अनुवाद किया जाता है और उसके बाद मानव-अनुवादक इस अनुवाद की जाँच

करता है। अर्थात् अनुवाद को संपादित करने का कार्य कंप्यूटर द्वारा अनुवाद करने के पश्चात् किया जाता है। इसलिए इस प्रणाली को 'पश्च संपादन प्रणाली' कहा जाता है। वैसे इसे 'अनुवादोत्तर संपादन' भी कहते हैं। पश्च संपादन प्रणाली में कंप्यूटर को मानव-नियंत्रित अनुवाद के एक अनिवार्य साधन के रूप में प्रयुक्त किया जाता है ताकि अपेक्षाकृत अधिक मात्रा में और जल्दी से अनुवाद कार्य सम्पन्न किया जा सके। कंप्यूटर अनुवाद प्रणाली की पश्च संपादन तकनीक निम्नलिखित प्रकार से संचालित होती है :

- स्रोतभाषा सामग्री को कंप्यूटर द्वारा पठनीय रूप में परिवर्तित किया जाता है। यानी स्रोतभाषा पाठ को कंप्यूटर में फीड किया जाता है।
- उसके बाद स्रोतभाषा सामग्री को अनूदित करने के लिए कंप्यूटर अनुवाद प्रणाली को आवश्यक आदेश दिए जाते हैं। बैच प्रोसेसिंग अनुवाद प्रणाली अनूद्य सामग्री को संसाधित करके कंप्यूटर स्क्रीन पर अपरिष्कृत (rough) अनुवाद प्रस्तुत कर देती है।
- मूल पाठ और अपरिष्कृत (rough) अनूदित सामग्री को मानव-अनुवादक को सौंपा जाता है। मानव-अनुवादक उस अपरिष्कृत (rough) अनुवाद का पुनरीक्षण-संपादन करके उसे अंतिम रूप प्रदान करता है। ऐसा करके वह 'मशीन अनुवाद पश्च-संपादक' (Machine Translation Post-Editor) का दायित्व निभाता है। उल्लेखनीय है कि यह संपादन कार्य सीधे कंप्यूटर पर भी किया जा सकता है और मुद्रित कागज पर भी।
- पश्च संपादन की प्रक्रिया में लक्ष्यभाषा की संरचना को ध्यान में रखा जाता है और अनूदित पाठ को सहज स्वीकार्य बनाने के लिए भाषिक अपेक्षाओं के अनुरूप उसमें कुछ जोड़ा अथवा छोड़ा भी जाता है।

पश्च संपादन प्रणाली को निम्नलिखित आरेख के माध्यम से समझा जा सकता है :



आरेख नं. 13.9

मशीनी अनुवाद के संदर्भ में पश्च-संपादन प्रणाली को बेहतर माना जाता है क्योंकि इसमें मानव-अनुवाद की तुलना में इसमें कम समय लगता है। इसके अलावा, इससे मानवीय कुशलता में अभिवृद्धि भी होती है और साथ ही बड़ी मात्रा में अनुवाद-कार्य शीघ्र संपन्न हो सकता है। किंतु यह भी ध्यान में रखना जरूरी है कि अनुवादोत्तर संपादन विशेष कौशल की माँग करता है। इसके लिए अनुवादक का सुयोग्य एवं जानकार होना आवश्यक है। उसे विषय-क्षेत्र का ज्ञान, लक्ष्यभाषा का विशेषज्ञ और दोनों भाषाओं का व्यतिरेकी ज्ञान-संपन्न होना चाहिए। यानी उसे मूल अनुवादक की भाँति इन सभी में पारंगत होना चाहिए। हालाँकि पश्च संपादन प्रणाली की सिद्धांततः महत्वपूर्ण विशेषता यह है कि 'मशीन अनुवाद पश्च संपादक' को स्रोतभाषा की जानकारी होना अपेक्षित नहीं है लेकिन, व्यवहार में यह देखा गया है कि अगर पश्च संपादक स्रोतभाषा का जानकार नहीं है तो इस प्रणाली द्वारा किए गए अनुवाद को संपादित करने में बड़ी मुश्किल होती है। 'सिस्ट्रान' (SYSTRAN) और 'लोगोस' (LOGO) आदि सभी प्रमुख कंप्यूटर अनुवाद प्रणालियों में पश्च-संपादन घटक समाविष्ट है।

(iii) अंतरा-संसाधन संपादन प्रणाली :

कंप्यूटर में संसाधन (processing) के समय अनूद्य सामग्री का संपादन संभव है। कंप्यूटर द्वारा अनुवाद की इस प्रक्रिया को 'अंतरा-संसाधन संपादन प्रणाली' कहा जाता है। यह कंप्यूटर अनुवाद प्रक्रिया के दौरान मानव-मशीन परस्पर सहयोग पर आधारित प्रणाली है। अंतरा-संसाधन कार्य करने वाला व्यक्ति 'अंतरा-संसाधन संपादक'

(Intra-processing Editor) कहलाता है। अंतरा-संसाधन संपादक संसाधन के दौरान अंतर्क्रिया करता है। इस दौरान वह उन द्विअर्थकताओं (ambiguities) का समाधान करता है जिन्हें प्रोसेसिंग करते समय मशीन स्वयं हल नहीं कर पाती।

अंतरा- संसाधन संपादन प्रणाली में द्विभाषिक प्रश्न पूछे जाते हैं। किंतु ध्यान देने की बात यह है इसके अंतर्गत परस्पर अंतर्क्रिया मानव और कंप्यूटर, दोनों स्तरों पर हो सकती है। कहने का तात्पर्य यह है कि प्रश्न कंप्यूटर द्वारा भी पूछे जा सकते हैं और मानव द्वारा भी। इस तरह, अंतरा-प्रोसेसिंग के हमें दो रूप नजर आते हैं - (1) मशीन का नियंत्रण में होना (यानी मशीन द्वारा प्रश्न पूछना और व्यक्ति द्वारा यह इंतजार करना कि मशीन प्रश्न पूछे); और (2) मानव नियंत्रित अंतरा-संसाधन (यानी व्यक्ति द्वारा प्रश्न पूछना और कंप्यूटर द्वारा यह इंतजार करना कि व्यक्ति प्रश्न पूछे)। कंप्यूटर अनुवाद के क्षेत्र में विकसित प्रणालियों में इन दोनों प्रकार के परस्पर सहयोग के उदाहरण मिलते हैं। जैसे, TransActive के नाम से जानी जाने वाली एल्प्स, सीटीएस (ALPS CTS) कंप्यूटर अनुवाद प्रणाली प्रथम प्रकार की अंतरा-प्रोसेसिंग का उदाहरण है तो कनाडा की टर्मिम (Termium) नामक कंप्यूटर अनुवाद प्रणाली दूसरे प्रकार की अंतरा-प्रोसेसिंग का।

अंतरा- संसाधन के स्तर पर मानव-मशीन सहयोग को बनाए रखने के लिए विशेष प्रकार का इंटरफेस निर्मित करने की आवश्यकता होती है। इस प्रणाली के अंतर्गत पूछे जाने वाले प्रश्न भिन्न प्रकार के होते हैं। उदाहरण के लिए, मानव से स्रोतभाषा की इकाई की द्विअर्थकता वाले शब्द-विशेष के साथ लक्ष्यभाषा के सह-संबंधित शब्द को भी पूछा जा सकता है। यह स्रोतभाषा पाठ का अर्थ उस भाषा में उपलब्ध कराने किए कह सकता है जो अनुवाद प्रणाली में अर्थ को व्यक्त करता हो। जैसे, 'normal' शब्द के संदर्भ में यह द्विभाषिक प्रश्न पूछा जा सकता है कि इसका क्या अनुवाद किया जाए - "प्रसामान्य", "प्राकृत", "मानक", "सामान्य अथवा साधारण", "अभिलंब" या फिर "नॉर्मल"। एक प्रकार से मशीन इनपुट के तुरंत बाद चयन हेतु विकल्प प्रस्तुत कर प्रश्न व्यक्ति के समक्ष करती है।

शब्द में अर्थ और भाव के स्तर पर द्विअर्थकता को अंतरा-प्रोसेसिंग के दौरान मानव-मशीन सहयोग के द्वारा दूर करना सरल है, किंतु स्रोतभाषा पाठ की संदर्भपरक द्विअर्थकता भी समस्या खड़ी करते हैं। उदाहरण के लिए, यह संदर्भपरक द्विअर्थकता अंग्रेजी के 'he', 'she', 'it' सर्वनामों (pronouns) के संदर्भ में विशेष तौर पर देखी जा सकती है। जैसे, निम्नलिखित वाक्य को देखा जा सकता है जिसमें NTR शब्द तो दो बार प्रयुक्त हुआ है, लेकिन यह पता लगाना जरा मुश्किल है कि ये दोनों किस-किस के लिए प्रयुक्त हुए हैं :

'My dog caught hold of a cat. **It** later developed friendship with **it**.'

ऐसी स्थिति में अंतरा-प्रोसेसिंग प्रणाली का उपयोग करने वाला अनुवादक यह निर्णय ले सकता है कि दोनों NTR किन-किन संदर्भों में प्रयुक्त हुए हैं। इसी प्रकार अंग्रेजी के 'he' और 'she' सर्वनामों के परिप्रेक्ष्य में संदर्भपरक द्विअर्थकता को दर्शाने वाले निम्नलिखित वाक्य देखे जा सकते हैं :

My brother has got a new boss. **He** appreciates **his** managerial skill.

'My sister is a follower of Mother Teresa. **She** practices **her** ideology in **her** life.'

उक्त स्थिति में सर्वनामों के लिए कंप्यूटर द्वारा सुझाए गए कई विकल्पों (व्यक्ति विशेष के संदर्भ में, व्यक्ति विशेष से संबंध को बताने के संदर्भ में अथवा अन्य संदर्भ में प्रयुक्त विकल्प) में से सही विकल्प चुनना होगा। किंतु कंप्यूटर प्रणाली के लिए पूरे स्रोतभाषा पाठ में से इस प्रकार के संदर्भित शब्द ढूँढना कठिन है। संसाधन के दौरान मानव-मशीन परस्पर सहयोग द्वारा वाक्यपरक द्विअर्थकता को दूर करना कठिन हो सकता है और इसके लिए जो कंप्यूटर घटक तैयार करने होंगे वे महँगे होंगे, क्योंकि प्रभावी कंप्यूटर-अनुवाद प्रक्रिया के लिए वाक्यपरक संरचना वृक्षों की तुलना करना प्रत्येक व्यक्ति के लिए सरल कार्य नहीं है। अर्थपरक विश्लेषण द्वारा द्विअर्थकता को भली प्रकार से दूर किया जा सकता है, इसलिए वाक्यपरक द्विअर्थकता की स्थिति में समुचित भावार्थ को चुनने के लिए अंतरा-संसाधन संपादन प्रणाली के अंतर्गत मानव का हस्तक्षेप होता ही है।

(iv) अंतर्क्रियात्मक संपादन प्रणाली

कंप्यूटर द्वारा अनुवाद कार्य करने में मानव-मशीन सहयोग को अंतर्क्रियात्मक स्तर पर भी देखा जा सकता है। इस प्रणाली के अंतर्गत जब व्यक्ति कंप्यूटर को अपनी (स्रोत) भाषा में अनूद्य सामग्री (इनपुट) स्वयं और सीधे देता है तो कंप्यूटर व्यक्ति के साथ अंतर्क्रिया शुरू कर देता है। इस अंतर्क्रिया के अंतर्गत कंप्यूटर आवश्यकतानुसार वाक्यवार अपनी भाषा में स्रोत पाठ के शब्दों के संबंध में व्यक्ति से स्पष्टीकरण माँगता है। मान लीजिए स्रोतभाषा में 'chest' शब्द आता है तो कंप्यूटर प्रणाली अंतर्क्रिया के दौरान यह पूछ सकती है कि :

The word 'chest' means :

- 1) 'a persons's body part'
- 2) 'a stronge box'

NUMBER >>

उक्त उदाहरण में कंप्यूटर प्रणाली शब्द-विशेष के अर्थ का अर्थ-संदर्भ जानने के लिए अंतर्क्रिया करती है। ऐसा भी हो सकता है कि अनूद्य सामग्री में संदर्भपरक द्विअर्थकता हो। इस द्विअर्थकता को दूर करने के लिए कंप्यूटर-मानव अंतर्क्रिया हो सकती है। मान लीजिए अंग्रेजी के पाठ में 'he' शब्द आता है तो कंप्यूटर यह पूछ सकता है कि :

The word 'he' refers to

- 1) Anil
- 2) my husband
- 3) the lion

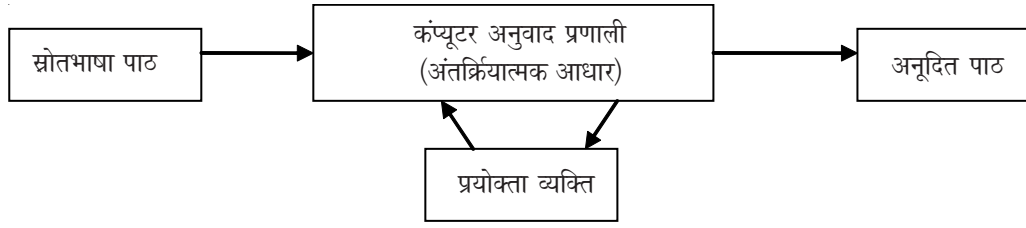
NUMBER >>

इस अंतर्क्रिया में जब व्यक्ति इस प्रकार के विकल्पों में से उपयुक्त विकल्प का चयन करके कंप्यूटर को उत्तर दे देता है तो कंप्यूटर अनूद्य वाक्य अथवा पाठ का विश्लेषण के पश्चात उसका लक्ष्यभाषा में अनुवाद प्रस्तुत कर देता है। मशीन के साथ अंतर्क्रिया करते हुए अनुवाद की दिशा में आगे बढ़ने वाले व्यक्ति को 'अंतर्क्रिया मशीन संपादक' (Interactive Machine Editor) कह सकते हैं।

मानव और मशीन की आपसी अंतर्क्रिया स्थूल रूप से कंप्यूटर अनुवाद-पूर्व संपादन अथवा अंतरा-प्रोसेसिंग संपादन मानव-मशीन सहयोग आधार वाली प्रतीत होती है। लेकिन ध्यान देने योग्य है कि इस प्रणाली में मानव-मशीन का सहयोग अनुवाद के स्तर पर नहीं होता है। यह अंतर्क्रिया अनुवाद-कार्य प्रारंभ होने से पूर्व होती है और उसके बाद स्रोतभाषा पाठ की लक्ष्यभाषा में प्रस्तुति होती है। उल्लेखनीय है कि इस दृष्टि से किए गए अनुवाद में अनुवादोत्तर संपादन या पश्च संपादक/अनुवादक की जरूरत नहीं रह जाती।

अंतर्क्रियात्मक अनुवाद प्रणाली में व्यक्ति को लक्ष्यभाषा का बोध होना आवश्यक नहीं होता है क्योंकि यह अंतर्क्रियात्मक पूर्व-संपादन परस्पर सहयोग स्रोतभाषा के स्तर पर होता है, न कि लक्ष्यभाषा के स्तर पर। इसके लिए व्यक्ति को भाषाविज्ञान, कंप्यूटर प्रौद्योगिकी अथवा अनुवाद कार्य का विशेष ज्ञान भी अपेक्षित नहीं होता है। इसलिए बिना अनुवादकीय सहयोग के ही अनूदित सामग्री प्राप्त की जा सकती है।

अंतर्क्रियात्मक दृष्टिकोण आधारित कंप्यूटर अनुवाद प्रणाली बहुत थोड़ी और/अथवा निजी प्रकृति की सामग्री और उसके तुरंत अनुवाद के लिए उपयुक्त है। ज्ञान-आधारित कंप्यूटर अनुवाद दृष्टिकोण के आधार पर अनेक भाषाओं में अनुवाद में भी यह उपयोगी सिद्ध होती है। जब स्रोतभाषा सामग्री इतनी अधिक तकनीकी प्रकृति की हो कि उसके व्यावसायिक अनुवादक को भी ज्ञान के क्षेत्र-विशेष के विशेषज्ञों की सहायता की जरूरत है तो भी इस प्रणाली का उपयोग किया जा सकता है। इस प्रणाली के अंतर्गत अनुवाद की प्रक्रिया का आरेखीय निरूपण निम्नलिखित प्रकार से किया जा सकता है :



आरेख नं. 10.9

13.5.2 मशीन-साधित मानव अनुवाद (Machine Assisted Human Translation - MAHT)

‘मशीन-साधित मानव अनुवाद’ (MAHT) प्रणाली को ‘कंप्यूटर साधित मानव अनुवाद’ (Computer Assisted Human Translation - CAHT) अथवा ‘कंप्यूटर साधित अनुवाद’ (Computer Assisted Translation - CAT) भी कहा जाता है। इस प्रकार के अनुवाद में अनुवादक की भूमिका केवल संपादन के स्तर पर ही न होकर, मशीनी अनुवाद की समस्त प्रक्रिया के दौरान बनी रहती है। इसमें अनुवाद संबंधी अधिकतम कार्य मुख्य रूप से मानव-अनुवादक द्वारा किया जाता है और आवश्यकता पड़ने पर अनुवादक कंप्यूटर की सहायता ले लेता है। इस प्रणाली में कंप्यूटर, अनुवाद कार्य में सहायता करने वाला कार्य-केंद्र (work-station) बन जाता है। इसमें प्रत्येक अनुवादक व्यक्तिगत कार्य-केंद्र बना सकता है और कार्य-विशेष की जरूरतों के अनुसार उसे रूपांतरित भी कर सकता है। इस प्रकार कंप्यूटर-सहायक मानव अनुवाद में लचीलापन है जो अनुवादक को स्वतंत्रता प्रदान करता है। इससे अनुवादक के समय की बचत होती है। वैसे, इस प्रकार की प्रौद्योगिकी के बारे में लोगों को ज्यादा जानकारी नहीं है। इसके अलावा, यह प्रौद्योगिकी अनुवादकों को अधिकांशतः सुलभ भी नहीं है।

मशीन-साधित मानव अनुवाद प्रणाली एक व्यापक अवधारणा है जिसमें सरल-जटिल, हर प्रकार के साधन-उपकरण शामिल हैं। इसमें बड़ी सावधानी से तैयार किए गए व्याकरण जाँच सॉफ्टवेयर, सही शब्दों वाले मानक शब्दकोश सॉफ्टवेयर, समनुक्रमणिका, अनुवाद स्मृति-कोश, इलेक्ट्रॉनिक शब्दकोश और शब्दावली डाटाबेस के साथ-साथ समरूप कार्यक्रम आदि विशिष्ट प्रकृति के कंप्यूटर कार्यक्रमों के रूप में मानव-अनुवादक के लिए अनुवाद सहायिकाएँ शामिल हैं। मानव द्वारा इनका उपयोग करके अनुवाद की परिशुद्धता में सुधार किया जा सकता है। ‘लोगोस’, ‘वेडेनर’ और ‘एल्पस’ आदि कंपनियों के विभिन्न उत्पाद मशीन-साधित मानव अनुवाद पर ही आधारित हैं। आज ‘ट्रांसलेटर मैनेजर’ (Translator Manager), ‘ट्रांसलेटर्स वर्कबेंच’ (Translators Workbench), ‘ऑप्टिमाइजर’ (Optimizer), ‘ट्रेडॉस’ (Trados) और ‘यूरोलैंग’ (Eurolang) आदि कंप्यूटर कार्यक्रमों में अनुवादक स्मृति-कोश प्रबंधक जैसे साधन-उपकरणों को व्यवहार में लाया जा रहा है। अनुवादक की प्रमुख सहायिकाएँ इस प्रकार है :

(1) शब्दकोश स्कैन सॉफ्टवेयर (Dictionary Scan Software)

शब्दकोश स्कैन सॉफ्टवेयर से अनुवादक को शब्दों की वर्तनी की जाँच करने में मदद मिलती है। इसकी सहायता से उसे यह पता चल जाता है कि सामग्री में कहीं शब्दों की वर्तनी संबंधी कोई दोष तो नहीं है। हालाँकि शब्दकोश जाँच सॉफ्टवेयर, वर्तनी जाँच सॉफ्टवेयर (Spell Check Software) का ही एक रूप कहा जा सकता है, किंतु यह उससे भिन्नता भी रखता है। वर्तनी जाँच सॉफ्टवेयर भाषा-विशेष के शब्दकोश (अर्थात् एकभाषिक शब्दकोश) से जुड़ा होता है और यह अधिकांशतः कंप्यूटर प्रणालियों में उपलब्ध ही होता है, जबकि शब्दकोश जाँच सॉफ्टवेयर द्विभाषिक शब्दकोश (Bilingual dictionary) से जुड़ा होता है। इसके अलावा, अविद्यमान अथवा अप्राप्य (missing) शब्दों की पहचान करना भी इस शब्दकोश स्कैन सॉफ्टवेयर की उल्लेखनीय विशेषता रहती है।

(2) इलेक्ट्रॉनिक शब्दकोश, शब्द-संग्रह और शब्दावली डाटाबेस (Electronic Dictionaries, Glossaries and Terminology Databases)

इलेक्ट्रॉनिक शब्दकोश, शब्द-संग्रह और शब्दावली डाटाबेस अनुवादकों को यह सुविधा इलेक्ट्रॉनिक रूप में उपलब्ध कराते हैं। कंप्यूटर पर उपलब्ध होने वाले इलेक्ट्रॉनिक अथवा डिजिटल शब्दकोश स्थूल रूप से प्रकाशित शब्दकोश

से भिन्न नजर नहीं आते, किंतु व्यावहारिक धरातल पर देखा जाए तो इनका रूप प्रकाशित शब्दकोशों की तुलना में भिन्न आवश्यक है। वास्तविकता यह है कि मुद्रित कोश में शब्द और उसके अर्थ आदि को खोजने में बड़ा समय लग जाता है जबकि इलेक्ट्रॉनिक शब्दकोश किसी शब्द को टाइप करने से भी कम समय लेते हैं। यह सॉफ्टवेयर शब्दों के अर्थ को अपने आप ही कंप्यूटर स्क्रीन पर प्रस्तुत कर देता है या फिर आवश्यकतानुसार अपेक्षित कुंजियों का प्रयोग करने पर भी शब्दों की प्रविष्टियों को कंप्यूटर स्क्रीन पर प्रस्तुत कर देता है। इलेक्ट्रॉनिक शब्दकोश से लिंक के जरिए संबंधित आँकड़ों को भी जल्दी देखा-पढ़ा जा सकता है। इनके अनुप्रयोग का महत्व इस दृष्टि से भी है कि विविध प्रकार की सामग्रियों के अनुवाद के दौरान कई प्रकार के शब्दकोशों को साथ-साथ देखना भी संभव हो जाता है। इस प्रकार के सॉफ्टवेयरों में से कुछ में यह सुविधा भी होती है कि अनुवाद कार्य के दौरान यदि अनुवादक कोई नया शब्दावली-युग्म बनाता है तो वह उसे शब्दावली डाटाबेस में शामिल भी कर सकता है, जबकि सामान्य शब्दकोश में यह कर्तई संभव नहीं। कुछ अपेक्षाकृत अधिक उन्नत प्रणालियों में अनुवादकों को यह जानकारी भी उपलब्ध हो जाती है कि उसने अनूद्य सामग्री में अनुवाद स्मृति कोश इकाइयों के भीतर और बाहर जो स्रोत अथवा लक्ष्यभाषा युग्म का इस्तेमाल किया है क्या वह सही है अथवा नहीं।

इलेक्ट्रॉनिक शब्दकोश सॉफ्टवेयर अथवा सी.डी.रोम के रूप में और इंटरनेट आदि के जरिए मिलते हैं। ये एकल अथवा द्विभाषी कोश के रूप में उपलब्ध होते हैं। सॉफ्टवेयर और सी.डी. रोम के रूप में उपलब्ध इलेक्ट्रॉनिक शब्दकोश को कंप्यूटर में डाला जा सकता है। इंटरनेट के जरिए इस्तेमाल की जाने वाली यह इलेक्ट्रॉनिक शब्द-सूची शब्दावली डाटाबेस के रूप में उपलब्ध होती हैं। इंटरनेट के जरिए इन इलेक्ट्रॉनिक शब्दकोशों तक पहुँच संभव हो पाती है। उदाहरण के लिए, सर्च इंजन 'गूगल' से कई भाषाओं के विविध प्रकार के एकल-भाषी और द्विभाषी शब्दकोशों तक पहुँच संभव हो पाती है। वैसे, कुछेक ऐसे शब्दावली डाटाबेस भी हैं जिनका इंटरनेट के जरिए इस्तेमाल करने के लिए ऑन-लाइन ग्राहक बनना (सब्सक्राइब करना) पड़ सकता है। ऑक्सफोर्ड इंग्लिश डिक्शनरी इसी प्रकार का 'ऑन-लाइन अभिदाता कोश' (On-line Subscriber Dictionary) है। ऑन-लाइन शब्दकोश अनुवादक के लिए अपने कॉर्पोरा में से सामग्री को व्यवस्थित करके प्रस्तुत करता है। इसमें अनुवादक किसी एक मुख्य शब्द से संबंधित सभी शब्दों अथवा भाषा विशेष के सभी शब्दों के लिए भी आदेश दे सकता है। इस तरह अनुवादक की सूचना तक इसकी तुरंत पहुँच संभव हो जाती है।

(3) समनुक्रमणिका (Concordance)

कंप्यूटर अनुवाद में 'समनुक्रमणिका' भी मानव-अनुवादक के लिए विशेष तौर पर सहायक उपकरण का काम करती है। समनुक्रमणिका से अनुवादक को शब्द-विशेष की विभिन्न अर्थ-छटाओं का बोध हो जाता है। वैसे समनुक्रमणिका, शब्दकोशों और शब्द-संग्रहों जैसे अनुवाद के साधन-उपकरणों की प्रतिस्थापक (substitute/replacement) नहीं है। समनुक्रमणिका मूल सामग्री के अनुवाद में भिन्न प्रकार की मदद करती है तथा अतिरिक्त जानकारी उपलब्ध कराती है।

समनुक्रमणिका शब्द-संसाधित कंप्यूटर कार्यक्रम है। शब्द/वर्ण-प्रयोगों की सूची उपलब्ध कराने वाला वह सुनिश्चित कॉर्पस है जिसमें शब्दों की विभिन्न अर्थ-छटाओं की व्याख्या दी हुई होती है। उदाहरण के लिए, ये उपसर्ग अथवा प्रत्यय के रूप में शब्द का अंग भी हो सकते हैं, पूर्ण शब्द भी अथवा शब्द-समूह भी। शब्दों अथवा पूर्वसर्गों के बारे में सांख्यिकीय आँकड़ें उपलब्ध कराना, शब्दों के प्रयोग अथवा वर्णक्रम आदि के रूप में शब्दों का वर्गीकरण करना, शब्द-विशेष के प्रयोग के सुनिश्चित संदर्भ को पहचानना और उसके आधार पर अर्थ-निर्धारण में मदद करना आदि समनुक्रमणिका के विशिष्टकार्य हैं। इसमें शब्द संबंधी सूचना संचित होती है और यह सूचना बिना क्रम से किसी भी समय परामर्श के लिए उपलब्ध रहती है। समनुक्रमणिका के डाटाबेस को भंडारित भी किया जा सकता है। 'Longman Lexicon of Contemporary English' कंप्यूटर पर तैयार की गई इलेक्ट्रॉनिक समनुक्रमणिका का एक उदाहरण है।

समनुक्रमणिका से अर्थ स्पष्ट रूप से परिभाषित होता है। इसलिए यह विशेष प्रकृति की सामग्री के अनुवाद में खास तौर से उपयोगी है। समनुक्रमणिका शाब्दिक संगतता को सुनिश्चित करती है। इससे अनुवादक का अनूद्य सामग्री पर ज्यादा नियंत्रण बना रह जाता है। किंतु भाषा की अनेकार्थकता और लक्षणा संबंधी समस्याओं से जूझ

रहे साहित्यानुवादकों के लिए समनुक्रमणिका अधिक सहायक सिद्ध नहीं होती। सभी प्रकार के अनुवादों में स्पष्ट रूप से संभावित भूमिका निभाने वाले कुछ साहित्यानुवादक समनुक्रमणिका का भी इस्तेमाल करते हैं।

(4) अनुवाद स्मृति-कोश प्रबंधक (Translation Memory Manager - TMM)

अनुवाद स्मृति-कोश, मानव-अनुवादक की सहायता करने वाला कंप्यूटर डाटाबेस सॉफ्टवेयर कार्यक्रम है। इस सॉफ्टवेयर के अंतर्गत कंप्यूटर डाटाबेस में स्रोतभाषा पाठ और उसकी एक अथवा उससे अधिक भाषाओं में पहले से ही अनूदित सामग्री संचित होती है। नई सामग्री का अनुवाद करते समय इस डाटाबेस में से आवश्यकता के अनुसार पूर्णतया अथवा अंशतः अनूदित अंश प्राप्त किया जा सकता है।

अनुवाद स्मृति-कोश कार्यक्रम स्रोतभाषा सामग्री को व्यवस्थित की जाने वाली इकाइयों में विभाजित करता है जिन्हें खंड कहा जाता है। ये खंड छोटी इकाइयों के रूप में स्रोतभाषा के वाक्य अथवा वाक्यांश भी हो सकते हैं या फिर बड़ी इकाइयों के रूप में अनुच्छेद अथवा पूरा पाठ भी। समानता को दृष्टि से यह सॉफ्टवेयर कार्य को तीन स्तरों पर देखता है। ये स्तर है-शब्दावली के स्तर पर समानता, वाक्यांश के स्तर पर समानता और छोटे-सरल वाक्यों के स्तर पर समानता। शब्दावली के स्तर पर सॉफ्टवेयर यह देखता है कि कहीं शब्दों के अर्थ भिन्न तो नहीं हैं। वाक्यांश के स्तर पर वह यह देखता है कि क्या भावों, विचारों अथवा कथ्य को समान शब्दों से व्यक्त किया गया है अथवा नहीं। छोटे और सरल वाक्यों के मिलान से अनुवादक पुनरावृत्ति से बचता है और उसे कम परिश्रम करना पड़ता है। इससे अनुवाद में द्विअर्थकता की आशंका भी नहीं रह जाती। जो शब्द अनुवाद स्मृति-कोश में नहीं होते हैं, यह सॉफ्टवेयर उनका अनुवाद अपनी शब्दावली के आधार पर प्रस्तुत करता है।

जब अनुवादक अनुवाद करना शुरू करता है तो अनुवाद स्मृति-कोश सॉफ्टवेयर यह देखता है कि क्या स्रोत इकाई उसके डाटाबेस से मेल खाती है। यदि वह मेल खाती है तो सॉफ्टवेयर उस स्रोत इकाई को कंप्यूटर स्क्रीन पर दिखा देता है ताकि अनुवादक उसका इस्तेमाल कर सके। अनुवादक उस अनूदित अंश को यथावत भी ले सकता है या फिर स्रोतभाषा सामग्री से पूरी तरह से मेल खाने वाला अंश बनाने के लिए उसमें संशोधन भी कर सकता है। लेकिन यदि अनुवाद स्मृति-कोश में संकलित सामग्री मूल से मेल नहीं खाती या अनुवादक को वह सही अनुवाद नहीं लगता तो वह उसका नया अनुवाद कर देता है। नए अनुवाद अथवा संशोधित अंश को कंप्यूटर सॉफ्टवेयर में नई इकाइयों के रूप में डाटाबेस में शामिल करने की व्यवस्था भी इस सॉफ्टवेयर में है। यानी जब अनूदित इकाई का अनुवाद पूरा हो जाता है और सॉफ्टवेयर उस नए अनुवाद को संकलित-संग्रहीत कर लेता है तो वह अगली इकाई की दिशा में आगे बढ़ जाता है। कंप्यूटर डाटाबेस में भंडारित की गई इन नई इकाइयों का भविष्य में भी इस्तेमाल हो सकता है और अनूदित किए जा रहे पाठ के शेष अंश में भी। इस सॉफ्टवेयर में स्रोतभाषा इकाइयाँ और उनके अनुवाद के साथ-साथ अनूदित इकाई बनाने की तारीख, इकाई-विशेष को अंतिम बार इस्तेमाल किए जाने की तारीख, अनुवादक का नाम आदि सूचनाएँ भी शामिल होती हैं।

अनुवाद स्मृति-कोश प्रबंधक का इस्तेमाल दो प्रकार से हो सकता है - (क) परस्पर सहयोग विधि, और (ख) स्वचालित विधि। परस्पर सहयोग विधि में अनूदित सामग्री को कंप्यूटर में डाला जाता है तो उसकी स्क्रीन पर पूर्व-अनूदित सामग्री के अनूदित अंश नजर आ जाते हैं। अनुवादक उन अंशों को एक-एक करके चुन लेता है और अनुवाद करता चलता है। अनुवाद स्मृति-कोश प्रबंधक प्रत्येक चुने हुए अंश को अपने स्मृति-कोश में खोजता है और पूरी तरह से अथवा आंशिक रूप से समान अनुवादों को अलग विंडोज़ पर दिखा देता है। अनुवादक उन अनुवादों को यथावत ग्रहण कर सकता है, उसमें संशोधन कर सकता है या फिर उसे ग्रहण नहीं भी कर सकता। वहीं स्वचालित विधि के अंतर्गत कार्यक्रम समस्त दूरी स्रोतभाषा सामग्री को स्वचालित प्रक्रिया अपनाते हुए संपन्न करता है और अपने स्मृति-कोश में उपलब्ध सामग्री का लक्ष्यभाषा में अनुवाद प्रस्तुत कर देता है। यह विधि उस तरह की सामग्री के लिए ज्यादा उपयोगी है जिसमें मूल सामग्री की ज्यादा पुनरावृत्ति हो, क्योंकि बार-बार प्रयुक्त होने वाली इकाई को कंप्यूटर प्रणाली एक ही मानकर चलती है।

अनुवाद स्मृति-कोश कार्यक्रम आम तौर पर निम्नलिखित के सहयोग से संपन्न हो पाता है :

- अनुवादक/संपादक (जो लक्ष्यभाषा पाठ के फॉर्मेट को बनाए रखता है;
- पाठ इकाई स्थान निर्धारक (localizer);

- शब्दकोश प्रबंध व्यवस्था के लिए शब्दावली साधन-उपकरण;
- नए पाठों के विश्लेषण के लिए स्वचालित प्रणाली; और
- अनूदित और अनूद्य शब्दों की संख्या, भाषा आदि की जानकारी देने वाले सांख्यिकीय साधन-उपकरण।

तकनीकी मैनुअलों जैसी विशिष्टप्रकृति की सामग्री अथवा एक जैसी सामग्री का बार-बार अनुवाद किए जाने जैसे मामले में अनुवाद स्मृति-कोश सॉफ्टवेयर डाटाबेस विशेष तौर पर उपयोगी सिद्ध होता है। लेकिन कुछ ऐसे भी अनुवाद-स्मृति कोश कार्यक्रम हैं जिनमें डाटाबेस सृजन की व्यवस्था नहीं होती है। वहीं कुछ अनुवाद स्मृति-कोश प्रबंधक स्टैंडअलोन (एकल) परिवेश में भी काम करते हैं। अनुवाद स्मृति-कोश प्रणाली का अगर नियमित रूप से लगातार इस्तेमाल किया जाए तो यह भविष्य में ज्यादा उपयोगी सिद्ध होती है, क्योंकि इससे अनुवादक कम परिश्रम करके ज्यादा अनुवाद कर पाता है। भारत सरकार के सी-डेक, पुणे द्वारा अंग्रेजी-हिंदी अनुवाद के लिए तैयार किए गए 'मंत्र राजभाषा' नामक कंप्यूटर अनुवाद तंत्र में इंटरनेट और इंटरनेट के अलावा स्टैंडअलोन परिवेश में भी काम किया जा सकता है।

अनुवाद-स्मृति कोश, सरल-साधारण भाषापरक इकाइयों वाले स्रोतभाषा पाठ के अनुवाद के संदर्भ में तो उपयोगी सिद्ध होता है, किंतु यह पाठ को उसमें निहित विचारों, अवधारणाओं अथवा समग्र संदेश के संदर्भ में मानव-अनुवादक की भाँति समग्र रूप में समझने की क्षमता से संपन्न नहीं होता है। मानव-अनुवादक स्रोतभाषा सामग्री के कथ्य को पुनर्व्यवस्थित आदि कर सकता है, क्योंकि लक्ष्यभाषा की प्रकृति और उसके सामाजिक-सांस्कृतिक संदर्भ पाठ को सुसंगत बनाने अथवा व्यापक रूप से सरल-सुकर बनाने के लिए भिन्न विषयवस्तु की माँग करते हैं। अनुवाद स्मृति-कोश प्रबंधन प्रणाली का एक यह भी दोष है कि इसके डाटाबेस के प्रभावी और सही तरीके से इस्तेमाल करने के लिए प्रशिक्षण की आवश्यकता होती है। डाटाबेस का इस्तेमाल करने में समय भी लगता है। वास्तव में अनलंकृत और निश्चित व्याकरणिक संरचना वाली सामग्री के अनुवाद की प्रक्रिया को प्रभावी और गुणवत्तापूर्ण बनाने के लिए तैयार किए अनुवाद स्मृति-कोश कार्यक्रम मानव-अनुवादक के सहायक तो कहे जा सकते हैं लेकिन वे उनके प्रतिस्थापक नहीं हैं।

(5) भाषा सर्च इंजन सॉफ्टवेयर (Language Search Engine Software) :

अनुवाद स्मृति-कोश प्रबंधक का व्यापक विस्तार हमें 'भाषा सर्च इंजन सॉफ्टवेयर' के रूप में भी नजर आता है। यह इंटरनेट-आधारित प्रणाली है और यह इंटरनेट सर्च इंजन की भाँति काम करती है। नाम से भले ही यह सॉफ्टवेयर भाषा खोज (सर्च) से संबंधित प्रतीत होता है किंतु यह इंटरनेट पर भाषा खोजने का काम नहीं करता है। वस्तुतः, यह सर्च इंजन स्रोत इकाई के आधार पर पहले से अनूदित उन वाक्यों, वाक्यांशों, अनुच्छेदों आदि की अपने व्यापक अनुवाद स्मृति भंडार में खोज करता है जो स्रोत इकाई से मेल खाते हैं। भाषा सर्च इंजनों को आधुनिक सर्च प्रौद्योगिकी को बढ़ावा देने की दृष्टि से तैयार किया गया है। यह मुख्यतः व्यापक अनुवाद स्मृति भंडार पर निर्भर करता है। उल्लेखनीय है कि "अनुवाद स्मृति-कोश सॉफ्टवेयर" की तुलना में यह काफी बड़ा होता है।

(6) समरूप कॉर्पोरा सॉफ्टवेयर (Aligned Corpora Software) :

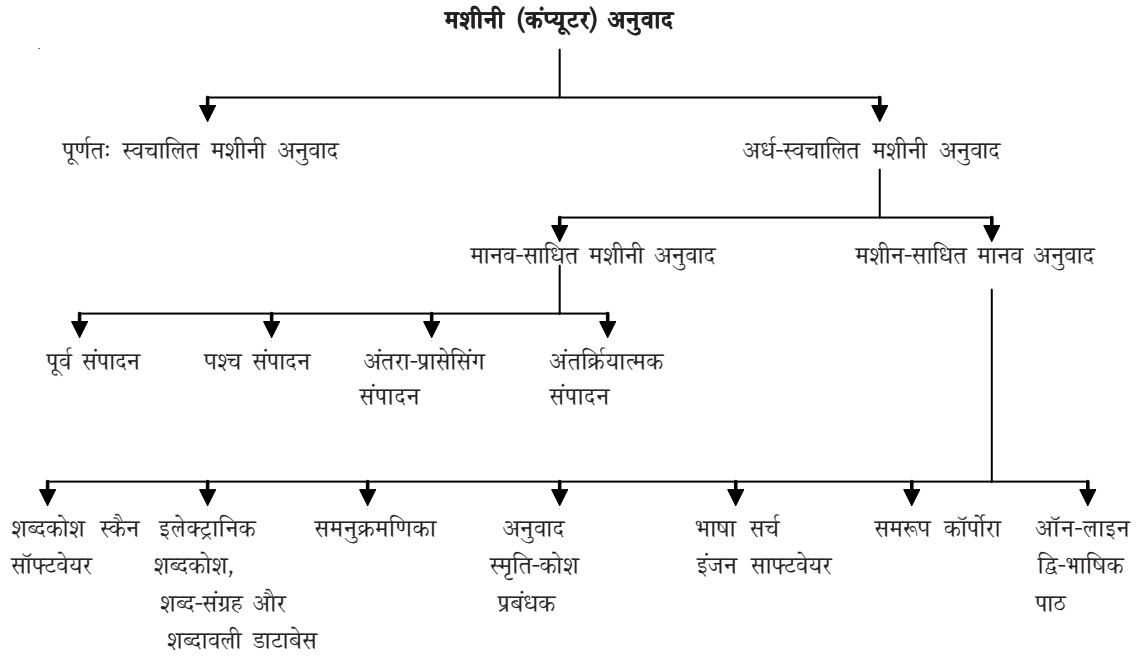
समरूप सॉफ्टवेयर पूरे अनुवाद का कॉर्पोरा होता है। स्रोतभाषा और लक्ष्यभाषा पाठ इकाइयों में विभाजित होते हैं। यह सॉफ्टवेयर यह निर्धारित करता है कि संबंधित विषयवस्तु वाले अनुवाद स्मृति डाटाबेस को तैयार करने के लिए कौन-सी इकाइयाँ एक-दूसरे से जुड़ी हुई हैं। कई समरूप सॉफ्टवेयरों में यह व्यवस्था है कि ठीक से मेल नहीं खाने वाली इकाइयों को अनुवादक स्वयं ही समरूप व्यवस्थित कर देता है। अनुवादक ने जो अनुवाद तैयार किया है, उसे अनुवाद स्मृति-कोश कार्यक्रम में डाल दिया जाता है ताकि भविष्य में अनुवाद कार्य करने के लिए मदद ली जा सके।

(7) ऑन-लाइन द्विभाषिक पाठ

द्विभाषिक कॉर्पोरा में आम तौर पर मानव-अनुवादक द्वारा पहले से ही किए अनुवाद और उनके मूल पाठ संचित होते हैं। इलेक्ट्रॉनिक रूप में भंडारित इस प्रकार की अनूदित सामग्री को द्वि-पाठ (bi-text) कहा जाता है। इससे

अनुवादक को बाद में कोई और सामग्री का अनुवाद करने में मदद मिलती है क्योंकि इसमें सुनिश्चित अभिव्यक्तियों का पहले से ही किया हुआ अनुवाद उपलब्ध होता है। अनुवाद की बढ़ती माँग के कारण विभिन्न भाषाओं में इस प्रकार के उपलब्ध अनुवादों के बेहतर ऑन-लाइन कॉर्पोरा से देश-विदेश के संगठनों और संस्थाओं को बड़ी मदद मिलती है।

कंप्यूटर अनुवाद के विभिन्न प्रकारों/उप-प्रकारों को आरेख द्वारा इस प्रकार चित्रित किया जा सकता है :



आरेख नं. 13.11

13.6 कंप्यूटर अनुवाद प्रौद्योगिकी के संदर्भ में अनुवादकों के प्रकार

जीवन के प्रत्येक क्षेत्र में अनुवाद की आवश्यकता ने इसे एक स्वैच्छिक क्रिया के अतिरिक्त रोजगार का माध्यम भी बना दिया है। प्रयोजनमूलक क्षेत्रों में अनुवादकों की बढ़ती माँग से रोजगार की संभावनाओं में निरंतर वृद्धि हो रही है। आज के समाज में अनुवाद एवं अनुवादक की आवश्यकता ने अनुवाद कार्य को संगठित रूप प्रदान कर दिया है। अनुवाद कार्य के इस संगठित रूप को मशीन अनुवाद के संदर्भ में भी देखा जा सकता है। मशीन अनुवाद प्रणाली हमें मुख्य रूप में “मानव-सहायक मशीन अनुवाद” (HAMT) और “मशीन-सहायक मानव अनुवाद” (MAHT) के रूप में नजर आती है। वैसे, यह प्रश्न उठना स्वाभाविक है कि इन अनुवाद प्रणालियों का प्रयोक्ता वर्ग कौन है? यानी किस प्रकार के अनुवादक इस प्रणाली का प्रयोग करते हैं।

मशीन की सहायता से अनुवाद करने वाले प्रयोक्ता वर्ग को मुख्यतः तीन वर्गों में विभाजित किया जा सकता है। ये हैं - (1) समूह में काम करने वाले पेशेवर अनुवादक (Professional Translators working in Teams), (2) स्वतंत्र पेशेवर अनुवादक (Independent Professional Translators); और (3) गैर-पेशेवर अनुवादक (Occasional Non-Professional Translators)।

13.6.1 समूह में काम करने वाले पेशेवर अनुवादक (Professional Translators working in Teams)

आज अनुवाद संगठित रूप धारण कर चुका है। इसलिए यह कार्य संगठित प्रयासों से भी संपन्न किया जाने लगा है। जहाँ बड़ी मात्रा में अनुवाद कार्य किया जाता है, वहाँ काफी अधिक अनुवादकों की सेवाएँ ली जाती हैं। वे समूह के रूप में अनुवाद कार्य करते हैं। ये वास्तव में सक्षम अनुवादक होते हैं। पश्चिमी विकसित देशों में समूह के रूप में काम करने वाले अनुवादकों के लिए कार्य-केंद्रों (Work stations) का चलन है। भारत में इस प्रकार

के कार्य-केंद्र व्यवहार में आने लगे हैं। कार्य-केंद्रों पर कई भाषाओं से और कई भाषाओं में अनुवाद करना संभव हो पाता है। उदाहरण के लिए, “यूरोलैंग्वेज ऑप्टिमाइजर” (Euro-language Optimizer) नामक अनुवादक कार्य-केंद्रों में स्रोतभाषा के रूप में छह भाषाओं (अंग्रेजी, फ्रांसीसी, जर्मन, इतालवी और स्पैनिश) तथा 17 लैटिन भाषाओं (लैटिन अक्षरों में लिखी जाने वाली लगभग सभी भाषाओं) में अनुवाद कार्य किया जाता है।

इसमें अनुवाद-कार्य मुख्य रूप से शब्द संसाधन, दूरसंचार और शब्दकोशों आदि अनुवाद सहायिकाओं पर आधारित होता है। समूह में काम करते हुए अनुवादक कार्य-केंद्र पर इंटरनेट के स्थानीय नेटवर्क से जुड़े होते हैं। ऐसा भी हो सकता है कि अनुवादक कार्य-केंद्र बाहरी स्थानीय नेटवर्क पर अथवा दूर-दराज स्थित कंप्यूटर में स्थापित किया गया हो जो दूरसंचार के माध्यम से संचालित हो। इसमें प्रत्येक अनुवादक को कार्य-स्थल पर ही द्विभाषिक शब्दावली और ‘अनुवाद स्मृति-कोश’ जैसे साधन-उपकरण उपलब्ध करा दिए जाते हैं। अनुवादकों के ये साधन-उपकरण ‘पाठ-संसाधक’ (text processor) से पूरी तरह से जुड़े होते हैं। अनुवादकों से यह अपेक्षा की जाती है कि वे सौंपे गए स्रोत पाठ का अनुवाद करके उसके अनूदित अंशों को मशीन अनुवाद सर्वर में डाल दें।

कार्य-केंद्र पर अनुवाद के दौरान स्रोतभाषा की अनूद्य सामग्री को सर्वर में विश्लेषण करते हुए संसाधित किया जाता है। इस विश्लेषण के दौरान सॉफ्टवेयर स्रोतभाषा पाठ को शब्दावली डाटाबेस और अनुवाद स्मृति-कोश में उपलब्ध वाक्यों से कुंजीपटल शॉर्टकटों से जोड़ देता है। इसके बाद उस सामग्री को संबंधित किट के साथ अनुवादक के कार्य-केंद्र पर भेजा जाता है। सर्वर सहायता उपकरण में आम “बहुभाषिक शब्दकोश डाटाबेस” (multi-lingual database) और “अनुवाद स्मृति-कोश” उपलब्ध रहता है जिसमें पहले से अनूदित अंश संचित रहते हैं। इस किट में शब्दकोश और अनुवाद स्मृति-कोश के उप-सेट के साथ-साथ विकल्प के रूप में बैच आधारित मशीन अनुवाद प्रणाली से प्राप्त अपरिष्कृत (rough) अनुवाद भी शामिल होता है। अनुवादकों तक यह सामग्री स्थानीय नेटवर्क के जरिए पहुँच जाती है। अनुवाद कार्य करने के दौरान वे अनुवादक शब्दकोश-किट को नए शब्दों अथवा शब्द-युग्मों से समृद्ध-संपन्न भी करते हैं। जब उनके द्वारा अनुवाद कार्य पूरा हो जाता है तो वे उसे अद्यतन किट के साथ मशीन अनुवाद सर्वर में भेज देते हैं। नए अनुवाद-युग्मों को सर्वर में अवस्थित ‘अनुवाद स्मृति-कोश’ में डाल दिया जाता है। अनुवादक द्वारा शब्दकोश में जोड़े गए शब्दों आदि को ‘बहुभाषिक शब्दावली डाटाबेस’ की व्यवस्था करने वाला व्यक्ति सर्वर में डालकर अनुवाद संबंधी डाटाबेस को अद्यतन कर लेता है। इससे अनुवादकों द्वारा किए जाने वाले अनुवादों की मात्रा (उत्पादकता) में पर्याप्त बढ़ोत्तरी होती है।

13.6.2 स्वतंत्र पेशेवर अनुवादक (Independent Professional Translators)

स्वतंत्र पेशेवर अनुवादक भी मशीन के सहयोग से अनुवाद कार्य करते हैं। वे मशीन पर भी अनुवाद कर सकते हैं और कंप्यूटर की सहायता लेते हुए भी अनुवाद कर सकते हैं। हालाँकि उनके काम के लिए मशीन की जो सुविधा उपलब्ध होती है वह अपेक्षाकृत कम सुविधाजनक, कम शक्तिशाली और कम लागत वाली होती है। इनमें अनूद्य सामग्री का रूपपरक विश्लेषण (morphological analysis) और शब्दों अथवा वाक्यांशों अथवा बार-बार आने वाले पूरे-पूरे वाक्यों की स्वतः जाँच (automatic look-up) शामिल है। यह सभी अथवा कई वाणिज्यिक पाठ और दस्तावेजों के लिए उपयुक्त होती है। इसमें भी अनुवाद-कार्य मुख्य रूप से शब्द संसाधन, दूरसंचार और शब्दकोशों आदि अनुवाद सहायिकाओं पर आधारित होता है। इनके अलावा, पेशेवर स्वतंत्र अनुवादक डिक्शनरी स्कैन, समनुक्रमणिका और पाठ-संबंधित शब्दावलियों जैसे विकल्पों का भी प्रयोग करते हैं। हालाँकि डिक्शनरी स्कैन, वर्तनी जाँच का ही एक रूप होता है किंतु यह द्विभाषिक शब्दकोश से जुड़ा होता है और अविद्यमान अथवा अप्राप्य (missing) शब्दों को पहचानता है। समनुक्रमणिका कार्यक्रम अनुवादक द्वारा स्रोतभाषा पाठ का अनुवाद किए जाने से पूर्व उसे भली प्रकार से समझने अथवा कठिन शब्दों अथवा वाक्यांशों की सभी अर्थ-छटाओं को समझने का अच्छा साधन-उपकरण सिद्ध होता है। पाठ-संबंधित शब्दावली एक प्रकार से केवल शब्दों अथवा वाक्यांशों की सूची होती है। लेकिन यह द्विभाषिक होती है जिसमें शब्द, वाक्यांशों और उनके अनुवाद होते हैं। इससे अनुवादक को स्रोतभाषा के शब्द-विशेष का लक्ष्यभाषा में समतुल्य अर्थ मिल जाता है। तकनीकी प्रकृति की सामग्री के अनुवाद के लिए ये शब्दावलियाँ विशेष तौर पर उपयोगी सिद्ध होती हैं। पेशेवर स्वतंत्र अनुवादकों को सामान्यतः अपने अनुवाद को उसी फॉर्मेट में प्रस्तुत करना होता है, जिस फॉर्मेट में स्रोत सामग्री होती है। स्वाभाविक है कि

वह फॉर्मेट अलग-अलग किस्म का होता है। इस प्रकार का अनुवाद करने वालों को हालाँकि केंद्रीकृत आम बहुभाषिक शब्दकोश डाटाबेस तो उपलब्ध नहीं होता लेकिन वे शब्दावली संबंधी फाइलों तक पहुँच सकते हैं, उन्हें आसानी से सृजित कर सकते हैं, संशोधित कर सकते हैं और उनमें शब्द आदि जोड़ भी सकते हैं। लिंगवाटेक का मर्करी/टर्मैक्स (Mercury/Termex) और “विनटूल” (WinTool) इसी प्रकार के अनुवाद-उपकरण कार्यक्रम हैं जो पेशेवर स्वतंत्र अनुवादकों के लिए लाभदायक हैं।

13.6.3 गैर-पेशेवर अनुवादक (Occasional Non-Professional Translators)

गैर-पेशेवर अनुवादक सामान्यतः दोनों भाषाओं-स्रोत और लक्ष्यभाषा-अथवा संभवतः केवल एक ही भाषा (अर्थात् स्रोतभाषा) में ही दक्ष होता है। गैर-पेशेवर अनुवादक वे हैं जिन्होंने अनुवाद को रोजगार का माध्यम नहीं बनाया हुआ है और न ही वे स्वातंत्र्य से सुखाय अनुभूति से अनुप्राणित होकर अनुवाद कर्म करते हैं। ये वे अनुवादक होते हैं जो जरूरत पड़ने पर कभी-कभी अनुवाद कर लेते हैं। इस प्रकार के अनुवादकों की सहायता के लिए बाजार में ‘समरूप पाठ’ और ‘ऑन-लाइन द्विभाषिक पाठ’ जैसे मशीन-सहायक साधन-उपकरण उपलब्ध हैं। उदाहरण के लिए, मैकिनतोष और पर्सनल कंप्यूटर पर चलने वाले लैंग्वेज इंजीनियरिंग के ‘एम्बेसेडर’ में अंग्रेजी-जापानी, अंग्रेजी-फ्रांसीसी, अंग्रेजी-स्पैनिश और फ्रांसीसी-जापानी के पत्रों के फॉर्मों के लगभग 200 और 450 वाक्यों अथवा अनुच्छेदों के टैपलेट उपलब्ध हैं। ‘समरूप सॉफ्टवेयर’ में पहले से अनूदित पाठ होते हैं। इनमें से कई सॉफ्टवेयरों में यह प्रावधान मिलता है कि ठीक से मेल नहीं खाने वाली इकाइयों को अनुवादक स्वयं समरूप व्यवस्थित कर लेता है। अनुवादक द्वारा इस तरह तैयार अनुवाद को भविष्य में अनुवाद कार्य में मदद के लिए अनुवाद स्मृति-कोश में डाल दिया जाता है। ‘ऑन-लाइन द्विभाषिक पाठ’ में आम तौर पर मानव-अनुवादक द्वारा पहले से ही किए अनुवाद और उनके मूल पाठ संचित होते हैं। इलेक्ट्रॉनिक रूप में भंडारित इस प्रकार की अनूदित सामग्री को द्वि-पाठ (bi-text) कहा जाता है।

मशीन-सहायक उपकरण के रूप में ऐसे भी द्विभाषी उपकरण उपलब्ध हैं जो उन्हें स्रोतभाषा के कुछेक उन शब्दों को जानने में मदद करते हैं जिन्हें अनुवादक नहीं जानते। ये साधन-उपकरण उन गैर-पेशेवर अनुवादकों के लिए विशेष तौर पर सहायक सिद्ध होते हैं जो द्विभाषी तो होते हैं लेकिन वे मूल भाषा में अनुवाद नहीं कर पाते हैं। इन शब्दकोशों में सामान्य शब्दों के साथ-साथ लक्ष्यभाषा के थिसॉरस, पर्याय कोश, संयुग्मक (conjugator), शैली-जाँचकर्ता (style checker) आदि साधन-उपकरण भी शामिल हैं। इस प्रकार के साधन-उपकरणों का प्रयोग करके भी गैर-पेशेवर अनुवादक अनुवाद कार्य कर सकते हैं, लेकिन उनके द्वारा किए गए अनुवाद के संपादन-संशोधन की आवश्यकता बनी रहती है।

ध्यान देने योग्य यह भी है कि वे मशीन अनुवाद से संबंधित सॉफ्टवेयर अथवा हार्डवेयर उपकरण हैं जिन्हें उस स्थिति में अनुवाद के लिए प्रयुक्त किया जाता है जब अनूद्य सामग्री को मशीन द्वारा पठनीय रूप (machine readable form) में शामिल नहीं किया जाता। अन्य शब्दों में गैर-पेशेवर अनुवादक इनका उपयोग करते हुए अनूद्य सामग्री का सीधे अनुवाद करता है। यह सीधे अनुवाद कागज पर लिखते हुए किया जा सकता है और कंप्यूटर पर टाइप करते हुए भी।

निष्कर्षतः इसमें कोई संदेह नहीं कि वैश्विक पटल पर विभिन्न भाषाओं के समृद्ध बहुभाषी सहायता उपकरणों की अनुपलब्धता, बहुभाषिक लेक्सिकल डाटाबेस निर्मित करने की आवश्यकता, स्रोतभाषा के मूल वर्ण (sacred characters), एक समय में केवल एक ही भाषा-युग्म में काम करने जैसी कतिपय सीमाएँ इस प्रौद्योगिकी की कुछ प्रमुख समस्याएँ हैं। अब तक विकसित साधन-उपकरणों में गति, श्रम-साध्य अनुवाद कार्य में दक्षता को बढ़ाने की तकनीकों का विकास करने और अनुवादकों की कार्य-प्रणालियों एवं कार्य-स्थितियों आदि में सुधार करने की आवश्यकता है ताकि वे ज्यादा प्रभावशाली तरीके से तथा सुविधाजनक ढंग से कार्य कर सकें। मशीन अनुवाद के संदर्भ में अनुवादकों की श्रम-दक्षता को बढ़ाने की तकनीक विकसित करने के साथ-साथ और उदाहरण-आधारित मशीन अनुवाद को परस्पर वार्ता-आधारित मशीन अनुवाद (Dialogue-based Machine Translation) से जोड़ने की आवश्यकता है ताकि मशीन की सहायता से अधिक और गुणवत्तापूर्ण अनुवाद संभव हो सके।

13.6 सारांश

इस इकाई के अध्ययन से यह आपको स्पष्ट हो गया होगा कि विभिन्न प्रकार की कंप्यूटर अनुवाद पद्धतियों की विविधताओं ने मानव-मशीन सहयोग के स्वरूप और सीमा को उभारा है। अपनी कतिपय सीमाओं के बावजूद ये कंप्यूटर अनुवाद पद्धतियाँ सार्थक हैं। इन पद्धतियों ने वस्तुतः मानव-मशीन सहयोग की अवधारणा को पुष्ट किया है। ये पद्धतियाँ मानव-अनुवादक का परिश्रम और समय बचाती हैं तथा अपेक्षाकृत तीव्र गति से अनुवाद संभव बनाती हैं। इनकी सहायता से मनुष्य की अनुवाद क्षमता में अभिवृद्धि होती है। इन विविध प्रकार की मशीनी अनुवाद पद्धतियों को विकसित करने का मुख्य उद्देश्य यह रहा है कि कंप्यूटर की सहायता से किया गया अनुवाद अर्थ के स्तर पर ज्यादा समतुल्य हो और समुचित अर्थ संप्रेषण की दृष्टि से सार्थक सिद्ध हो। ऐसे में स्रोतभाषा सामग्री का लक्ष्यभाषा में सहज-स्वाभाविक भाव-संप्रेषण ही अनुवाद की यथातथ्यता और गुणवत्ता का आधार बन जाता है। इस कसौटी पर किसी भी मशीन अनुवाद प्रणाली की गुणवत्ता इस आधार पर आँकी जा सकती है कि वह कितनी तीव्र गति से और परिशुद्धता के साथ अनुवाद रूपी आउटपुट देने के साथ-साथ कितनी प्रयोक्ता-सहज है।

कंप्यूटर अनुवाद प्रणाली में मानव-सहभागिता के चार स्तरों पर संभव है-पूर्व संपादन अथवा अंतर्क्रिया और मशीनी अनुवाद प्रक्रिया के दौरान अथवा उसके द्वारा अनुवाद कार्य संपन्न कर लेने के पश्चात। किंतु इसका यह अभिप्राय नहीं कि किसी भी कंप्यूटर अनुवाद प्रणाली-विशेष में मानव-सहभागिता की इनमें से किसी एक युक्ति को ही अपनाया जा सकता है। वास्तव में कंप्यूटर अनुवाद प्रणाली के सफल कार्य-संचालन के लिए इन सभी को अथवा इनमें से किन्हीं दो-तीन युक्तियों को भी अपनाया जा सकता है। मुख्य बात यह है कि जब मानव का हस्तक्षेप समाप्त हो जाता है और अनुवाद कार्य के लिए जो कुछ शेष रह जाता है तो मानव-सहभागिता के पश्चात के समस्त कार्यों को भली प्रकार से संपन्न करने की मशीन में क्षमता की अपेक्षा रहती है। वास्तव में विशुद्ध मशीनी अनुवाद आज भी मानव के लिए एक चुनौती बना हुआ है। मशीनी अनुवाद के संदर्भ में हमने अनुवादकों की कार्य-प्रवृत्ति पर भी चर्चा की और पाया कि कुछ अनुवादक मशीनी अनुवाद के संदर्भ में समूह बनाकर काम करते हैं और कुछ यह कार्य निजी रूप में करते हैं। उल्लेखनीय है कि आजकल अनुवाद एक व्यवस्था के रूप में तेजी से उभर रहा है; तथा इसमें पेशेवर और गैर-पेशेवर दोनों ही मशीनी अनुवाद को गति दे रहे हैं।

नई कंप्यूटर प्रौद्योगिकियों के विकास से आज अनुवाद के क्षेत्र में महत्वपूर्ण परिवर्तन आया है, जो उत्तरोत्तर और उन्नत तथा विकसित ही होगा। लेकिन यह एक नया व्यवसाय नहीं है। यह अनुवादकों की उत्पादकता और गुणवत्ता को बढ़ाने का महत्वपूर्ण साधन है। वहीं नई प्रौद्योगिकी को मानव-सहायक और उसका इष्टतम उपयोग करने की संभावना वाली उस प्रौद्योगिकी का रूप देने की आवश्यकता है जो अपेक्षाकृत अधिक यथार्थ, परिशुद्ध, संगत और उत्पाद के रूप में सहायक सिद्ध हो। अनुवादकों से यह अपेक्षा की जाती है कि वे नई प्रौद्योगिकियों को अपनाएँ और उसमें शिक्षित-प्रशिक्षित हों ताकि वे अपनी उत्पादकता और गुणवत्ता बढ़ाने के लिए इन प्रौद्योगिकियों की संभावनाओं और क्षमताओं का अधिकतम लाभ उठा सकें। वस्तुतः मशीनी अनुवाद प्रणाली में अनुवादकों की भी महती भूमिका है।

13.8 अभ्यास के लिए प्रश्न

1. मशीनी अनुवाद पद्धति विकास के दो चरण कौन-से हैं तथा प्रत्यक्ष मशीनी अनुवाद पद्धति किन कारकों पर आधारित होती है?
2. मशीनी अनुवाद की अंतरण पद्धति की अंतरभाषिक पद्धति से क्या भिन्नता है?
3. मशीनी अनुवाद पद्धति में कृत्रिम बुद्धि से आप क्या समझते हैं? क्या मशीन से आदर्श अनुवाद संभव है? स्पष्ट कीजिए।
4. मानव-साधित कंप्यूटर अनुवाद संपादन के प्रमुख रूप कौन से हैं?

5. मानव-साधित कंप्यूटर अनुवाद के पूर्व और पश्च संपादन में से कौन-सी पद्धति बेहतर है और क्यों?
6. “मशीन-साधित मानव अनुवाद” (MAHT) पद्धति में कौन से साधन-उपकरण शामिल किए जाते हैं? उल्लेख कीजिए।
7. “अनुवाद स्मृति-कोश प्रबंधक” क्या है? पाँच पंक्तियों में स्पष्ट कीजिए।
8. क्या “भाषा सर्च इंजन सॉफ्टवेयर” (Language Search Engine Software) इंटरनेट सर्च इंजन की भाँति काम करता है? स्पष्ट कीजिए।
9. क्या कंप्यूटर अनुवाद प्रौद्योगिकी के संदर्भ में अनुवादक भी अलग-अलग प्रकार के हो गए हैं? स्पष्ट कीजिए।
10. मशीन की सहायता से अनुवाद करने वाले प्रयोक्ता वर्ग को किन वर्गों में विभाजित किया जा सकता है?
11. मशीन अनुवाद पद्धति अनुवादकों के लिए किस प्रकार सहायक सिद्ध होती है?

13.9 शब्दावली

अंतरभाषिक, साफ्टवेयर, कृत्रिमबुद्धि, पद-व्याख्या, रूपपरक, इनपुट, आउटपुट, इंटरलिंगुआ, अनुमिति, साइबरनेटिक्स, लेक्सीकॉन, एल्गोरिद्म, अंतर्क्रिया, इंटरफेस,

13.10 कुछ उपयोगी पुस्तकें

- कृष्णकुमार गोस्वामी, अनुवाद विज्ञान की भूमिका, नई दिल्ली, राजकमल प्रकाशन।
- पूरनचंद टंडन, (संपा.), कंप्यूटर अनुवाद, नई दिल्ली, भारतीय अनुवाद परिषद।
- हरीश कुमार सेठी, ई-अनुवाद और हिंदी, नई दिल्ली, किताब घर।
- Gargesh, Ravinder and Krishna Kumar Goswami (Eds.), 2007, Translation and Interpreting, New Delhi, Orient Longman.
- Lawson, V., 1982, Practical Experience of Machine Translation, North Holland, Amsterdam
- Nirenburg, Sergei, 1987, Machine Translation: Theoretical and Methodological Issues, Cambridge, Cambridge University Press.
- Slocum, J., 1988, Machine Translation Systems, Cambridge, Cambridge University Press.

इकाई 14 मशीनी अनुवाद : प्रक्रिया एवं चुनौतियाँ

इकाई की रूपरेखा

- 14.0 उद्देश्य
- 14.1 प्रस्तावना
- 14.2 मशीनी अनुवाद की अवधारणा
- 14.3 मशीनी अनुवाद की प्रक्रिया
- 14.4 मशीनी अनुवाद और पूर्ण अनुवाद
- 14.5 मशीनी अनुवाद : एक चुनौती
- 14.6 मशीनी अनुवाद की कुछ और पद्धतियाँ
 - 14.6.1 नियम आधारित पद्धति
 - 14.6.2 कॉर्पस आधारित या प्रयोगाश्रित पद्धति
 - 14.6.3 हाइब्रिड मशीनी अनुवाद पद्धति
- 14.7 मशीनी अनुवाद के अन्य मुद्दे
 - 14.7.1 भाषा संसाधन
 - 14.7.2 प्रणाली पद्धति
 - 14.7.3 अनुप्रयोग
- 14.8 सारांश
- 14.8 अभ्यास के लिए प्रश्न
- 14.9 शब्दावली
- 14.10 कुछ उपयोगी पुस्तकें

14.0 उद्देश्य

मशीनी अनुवाद का क्षेत्र एक उभरता हुआ अनुवाद क्षेत्र है; अतः इसमें अनेक संकल्पनाएँ सामने आ रही हैं। नए-नए प्रयोग किए जा रहे हैं। मशीनी अनुवाद के क्षेत्र में अब तक न तो कोई संकल्पना निर्णायक अथवा अंतिम सिद्धांत के रूप में सफल सिद्ध हो पाई है और न ही कोई निर्धारित पद्धति स्पष्टतः सफल कही जा सकती है। मशीनी अनुवाद के ऐसे अनेक पहलुओं पर इस इकाई में अलग-अलग मत देखने को मिलेंगे। मशीनी अनुवाद की अवधारणा, प्रक्रिया तथा मशीनी अनुवाद के समक्ष मुद्दों पर अध्ययन के उपरांत आप मशीनी अनुवाद के विकास एवं उसकी प्रक्रियात्मक जटिलताओं के विषय में भी हम चर्चा करेंगे; साथ ही मशीनी अनुवाद से जुड़ी चुनौतियों तथा अनुवाद संबंधी विभिन्न मुद्दों पर भी विस्तार से विमर्श करेंगे। प्रस्तुत इकाई के अध्ययन के उपरांत आप :

- मशीनी अनुवाद की मूल 'अवधारणाओं' और पारिभाषिक शब्दावली को समझ सकेंगे;
- मशीनी अनुवाद की प्रक्रिया एक संश्लिष्ट प्रक्रिया है, यह जान पाएंगे;
- मशीनी अनुवाद प्रणाली बनाने में आने वाली मूलभूत भाषिक और तकनीकी समस्याओं का परिचय प्राप्त कर सकेंगे;
- मशीनी अनुवाद प्रणाली बनाने की विभिन्न पद्धतियों और उनकी संरचना को समझ सकेंगे;
- मशीनी अनुवाद की जटिलता एवं चुनौतियों को जान सकेंगे; तथा आप

- यह भी जान सकेंगे कि मशीनी अनुवाद का विकास आधुनिक बहुभाषी संसार के निरंतर बढ़ते बहु-आयामी संप्रेषण को कैसे एक प्रभावकारी दिशा और गति दे सकता है।

14.1 प्रस्तावना

प्रस्तुत खंड की पिछली दो इकाइयों में आपने मशीनी अनुवाद के स्वरूप, क्षेत्र तथा मशीनी अनुवाद की कुछ पद्धतियों एवं प्रकारों पर अध्ययन किया। आपने यह भी पढ़ा कि मशीनी अनुवाद की शुरुआत लगभग 20वीं शताब्दी के मध्य से होकर अब तक अनेक चरणों से गुजरकर एक विकसित प्रणाली के रूप में स्थापित हो रही है। मशीनी अनुवाद का क्षेत्र भारतीय भाषाविदों एवं कंप्यूटर-विज्ञानियों के लिए रुचि और चुनौती का क्षेत्र रहा है। अनेक मशीनी अनुवाद तंत्र तथा प्रविधियाँ विकसित हो चुकी हैं। हालाँकि इन सबके अभी तक मिश्रित परिणाम ही प्राप्त हुए हैं। पिछली इकाइयों में आपने यह भी पढ़ा कि मशीनी अनुवाद के विभिन्न प्रकार उसकी अनुप्रयोज्यता और प्रयुक्त प्रौद्योगिकी एवं प्रविधि के आधार पर ही सामने आते हैं। कुछ पद्धतियों की चर्चा पिछली इकाई में की गई है। मशीनी अनुवाद के विभिन्न प्रकारों पर भी प्रकाश डाला जा चुका है। इस इकाई में हम मशीनी अनुवाद की प्रक्रिया, उसकी व्यावहारिकता और मशीनी अनुवाद के समक्ष मुद्दों और चुनौतियों तथा कुछ और पद्धतियों पर विस्तार से चर्चा करेंगे। मशीनी अनुवाद पद्धतियों के अंतर्गत हम इस इकाई में नियम आधारित पद्धति, कॉर्पस पद्धति तथा हाइब्रिड पद्धति पर विशेष रूप से चर्चा करेंगे। कॉर्पस आधारित पद्धति के अंतर्गत हम उदाहरण आधारित मशीनी अनुवाद तथा सांख्यिकीय आधारित मशीनी अनुवाद पर विशेष रूप से ध्यान केंद्रित करेंगे। प्रत्यक्ष, अंतरण तथा अंतरभाषिक पद्धतियों पर आप पहले ही पढ़ चुके हैं। मशीनी अनुवाद की वर्तमान स्थिति विभिन्न प्रकारों, पद्धतियों और मानव तथा मशीन के अंतरसंबंध एवं मशीनी अनुवाद के भविष्य को लेकर उठाए जाने वाले प्रश्नों पर भी आप गहराई से मनन कर सकेंगे। इस चर्चा में हम यह भी जानने की कोशिश करेंगे कि किस प्रकार इनके संभावित हल ढूँढकर हम मशीनी अनुवाद की दिशा में अधिक स्वावलंबी होकर अनुवाद क्षेत्र की लगातार बढ़ती माँग को पूरा कर सकेंगे। वास्तव में मशीनी अनुवाद से जुड़े मुद्दे और उसके क्रियान्वयन में आने वाली चुनौतियाँ ही इस क्षेत्र को अधिक रुचिकर तथा शोध के उपयोगी विषय के रूप में प्रस्तुत करती हैं। मुद्दों पर गौर करने और चुनौतियों से पार पाने का मतलब ही है कि हम कुछ जटिल समस्या का हल खोज पाने में समर्थ हो रहे हैं। इससे मशीनी अनुवाद को भविष्य में अधिक व्यावहारिक बनाने में महत्वपूर्ण उपलब्धियाँ हासिल होंगी।

14.2 मशीनी अनुवाद की अवधारणा

पिछली इकाइयों में मशीनी अनुवाद की संकल्पना पर हुई चर्चा के अनुक्रम में यह स्पष्टतः ज्ञात हो चुका है कि मशीनी अनुवाद (MT) मशीन द्वारा किया जाने वाला अनुवाद कार्य है और यह मशीन आधुनिक संसार के सबसे लोकप्रिय यंत्रों में से एक 'कंप्यूटर' है। अतः इसे हम कंप्यूटर अनुवाद भी कह सकते हैं। सामान्य तौर पर देखें तो मशीनी अनुवाद (MT) संगणक प्रौद्योगिकी का वह अनुप्रयोग है जिसमें स्रोतभाषा (SL) से दूसरी लक्ष्यभाषा (TL) में अनुवाद प्राप्त किया जाता है। ये भाषा युग्म कोई-सी भी दो भाषाएँ हो सकती हैं। वास्तव में, इसके विकास के शुरुआती चरण में, एक भाषा से दूसरी भाषा में अनुवाद उपलब्ध कराने वाले कंप्यूटर प्रोग्राम के लिए 'मैकेनाइज्ड अनुवाद' का प्रयोग होता था। इस बारे में भी मतभेद रहा है कि मशीनी अनुवाद में सम्मिलित प्रक्रिया की प्रकृति को कौन-सा शब्द अच्छी तरह से वर्णित कर सकता है। किंतु मशीनी अनुवाद अब इस उद्देश्य के लिए प्रयोग किया जाने वाला मानक शब्द बन चुका है (केथ, 2005)। सारांशतः मशीनी अनुवाद कंप्यूटर सॉफ्टवेयर द्वारा किया जाने वाला अनुवाद है और मशीनी अनुवाद प्रणाली एक कंप्यूटर प्रोग्राम है जो 'इनपुट' के तौर पर स्रोतभाषा का पाठ लेता है और उसका लक्ष्यभाषा के पाठ में अनूदित परिणाम (आउटपुट) देता है। यह स्पष्ट है कि मशीनी अनुवाद के मुख्य मुद्दे एक मशीन अर्थात् कंप्यूटर प्रोग्राम को प्राकृतिक भाषा समझने में सक्षम बनाने से संबंधित हैं। यह वस्तुतः एक कृत्रिम साधन कंप्यूटर को मानव भाषा का ज्ञान एवं समझ प्रदान करने के बराबर है। इस रूप में स्पष्टतः मशीनी अनुवाद विषय का एक महत्वपूर्ण मुद्दा, भाषा क्या है, कहाँ अवस्थित है, इसका स्वरूप क्या है, इसके प्रयोग की प्रक्रिया क्या है, और कंप्यूटर में मानव भाषा के इस ज्ञान या समझ (इसके प्रयोग की

जानकारी सहित) को किस प्रकार अंतरित किया जाए, इत्यादि संबंधित हैं। मानव भाषा की वास्तविक प्रकृति भाषा विचारकों, भाषाविदों और वैज्ञानिकों की पूरी समझ से अभी भी परे है। भाषा एक जटिल संरचना है और आंतर एवं अंतर-भाषिक संप्रेषण में बौद्धिकता एक आधारभूत तत्व है। इसमें व्यावहारिक एवं अभिव्यक्तिगत, साभिप्राय तथा सांस्कृतिक भिन्नताएँ इत्यादि, जटिलता के मुख्य स्रोत हो सकते हैं। भाषा की समझ (और प्रयोग) स्वयं भाषा संरचना की समझ तक सीमित नहीं है अपितु सांसारिक अनुभवों और सांसारिक ज्ञान पर आधारित होती है। मशीनी अनुवाद (MT) का दूसरा चरण संगणक विज्ञान और तकनीक या प्रौद्योगिकी से संबंधित है; जिसमें प्रणाली निर्माण और विकास, प्रोग्रामिंग एवं कृत्रिम बौद्धिकता आदि के मुद्दे सम्मिलित हैं। मशीनी अनुवाद (MT) के मुद्दे पूर्णतः अंतरविषयक और बहुविषयक हैं। भाषा और भाषाविज्ञान से लेकर, अनुवाद अध्ययन, दर्शनशास्त्र, तर्कशास्त्र, मनोभाषाविज्ञान, समाजभाषाविज्ञान, कंप्यूटर विज्ञान और अभियांत्रिकी, सूचनाविज्ञान और प्रौद्योगिकी, गणितशास्त्र, सांख्यिकी, संज्ञानात्मक विज्ञान जैसे विविध विषयों की विशेषज्ञता मशीनी अनुवाद के शोध और विकास के लिए अत्यधिक महत्वपूर्ण है। इन विविध विषयों की विशेषज्ञताओं की उपलब्धि एवं इसमें सामंजस्य मशीनी अनुवाद की एक बड़ी चुनौती है। इस अर्थ में मशीनी अनुवाद विषयक अध्ययन एवं शोध, न केवल मशीनी अनुवाद प्रणालियाँ और उपकरण उपलब्ध कराकर मानव जीवन की संप्रेषणात्मक कठिनाइयों को सहज कर सकती है अपितु एक अंतर्दृष्टि भी प्रदान कर सकती है जो मानव भाषा एवं मस्तिष्क के रहस्यों को स्पष्ट करने में मदद कर सकती है।

अनुवाद का अस्तित्व मानव समाज की बहुभाषिकता और विभिन्न भाषाओं में व्यापक संपर्क और अंतरक्रिया पर निर्भर है। बहुभाषिकता खासकर आधुनिक मानव समाज एवं सभ्यता की पहचान है। आधुनिक वैश्विक समाज में विभिन्न स्तरों पर भाषायी अनुवाद की आवश्यकता में चमत्कारिक रूप से बढ़ोत्तरी हुई है, जिसने मशीनी अनुवाद (MT) को पारंपरिक मानव अनुवाद के एक अत्यंत संभावनापूर्ण विकल्प के रूप में प्रस्तुत किया है। प्राकृतिक भाषा संसाधन या संगणक भाषाविज्ञान के क्षेत्र में गत शताब्दी में हुए विकास से मशीनी अनुवाद एक महत्वपूर्ण आधुनिक प्रौद्योगिकी के रूप में उभरा है। इसने मानव को श्रम-साधित एवं नीरस अनुवाद प्रक्रिया (खासकर गैर-साहित्यिक अनुवाद) से मुक्त होने की संभावनाएँ प्रदान की है। मशीनी अनुवाद ने अपनी यात्रा की शुरुआत पूर्णतः या अंशतः स्वतः अनुवाद पद्धति जैसे अनुवाद सहायक उपकरण प्रणालियाँ प्रदान की है। आरंभ में, मानव अनुवादकों में यह भय/संदेह भी रहा है कि क्या मशीनी अनुवाद कभी सफल हो पाएगा या यह मानव रोजगार को खत्म कर देगा। मशीनी अनुवाद के क्षेत्र में इन विषयों पर बहुत कुछ कहा गया है और वर्तमान में सामान्य अवधारणा यह है कि अनुवादकों को मशीनी अनुवाद प्रणालियों और अनुवाद सहायक उपकरणों से पूर्णरूप से अवगत होना आवश्यक है ताकि वे अपने भय और संदेह की स्थिति से उबर सकें और उनका अपने लाभ के लिए प्रयोग कर सकें (हचिन्स, 1997)। मशीनी अनुवाद के भविष्य और इसकी भूमिका के बारे में चाहे जो भी सोचा और कहा जाए, आधुनिक बहुभाषिक समाज में भाषिक सीमाओं को तोड़ने में एवं एक अंतरविषयक और बहुविषयक शैक्षणिक विषय के रूप में मशीनी अनुवाद एक महत्वपूर्ण स्थान ग्रहण कर चुका है।

मशीनी अनुवाद प्रणाली का विकास मुख्यरूप से प्रशासनिक पर केंद्रित किया गया। आर्थिक, वैज्ञानिक एवं तकनीकी और इसी प्रकार के दूसरे क्षेत्रों से संबंध रखने वाले उन तकनीकी दस्तावेजों जिनमें भाषा जटिलता और रचनात्मकता के स्तर पर कम-से-कम कठिनाइयाँ होती हैं, मानव भाषा और संप्रेषण की जटिलता अभी भी उन वैज्ञानिकों के लिए चुनौती थी जो मशीनी अनुवाद की ज्ञान आधारित पद्धति के विकल्पों के विकास में प्रयासरत थे। इस बीच सूचना और तकनीकी विज्ञान के क्षेत्र में उन्नति से कंप्यूटर की संग्रहण और संसाधन क्षमता में वृद्धि हो रही थी। इसने मशीनी अनुवाद के वैकल्पिक प्रणाली के अनुसंधानकर्ताओं और विकासकर्ताओं के प्रयास को सकारात्मक गति प्रदान की। मशीनी अनुवाद प्रणाली विकास के क्षेत्र के कुछ विद्वानों और विशेषज्ञों ने ज्ञान-आधारित एवं पद्धतियों में सामंजस्य की संभावनाओं पर भी विचार करना शुरू किया और मशीनी अनुवाद पद्धति में मिश्रण/संकरण (Hyberdity) की अवधारणा आई हालाँकि, इन वर्षों में मशीन अनुवाद की सांख्यिकीय पद्धतियों के विकास की बड़ी संभावनाएँ सामने आई और दूसरी-तीसरी पीढ़ी की मशीन अनुवाद प्रणालियों के विकास के दौरान यह एक प्रभावी और व्यावहारिक विकल्प के रूप में उभर कर आया। इस खंड की पिछली इकाइयों में इन सब पर विस्तारपूर्वक चर्चा हो चुकी है।

मशीनी अनुवाद आधुनिक युग का एक बड़ा यथार्थ है और इस समय की वैज्ञानिक एवं तकनीक विकास के लिए एक चुनौतीपूर्ण महत्वाकांक्षा भी है। एक तरफ तो आज विभिन्न भाषाओं के लिए अनेकानेक विकास मशीनी अनुवाद उपकरण उपलब्ध हैं किंतु पूर्ण-स्वचालित सामान्य उद्देश्य की मशीनी अनुवाद उपकरण का विकास अभी बाकी है। आज जो मशीन अनुवाद उपकरण उपलब्ध हैं वे मुख्यरूप से कंप्यूटर-साधित/आश्रित अनुवाद (CAT) या मशीन-साधित/आश्रित अनुवाद (MAT) प्रणाली हैं। इन उपकरणों में मशीन-साधित मानव अनुवाद (MAHT) और मानव-साधित मशीनी अनुवाद (HAMT) उपकरण आते हैं (हचिन्स, 2003)। हाल के दशकों में पूरे विश्व में विभिन्न भाषा-युग्मों को लेकर मशीनी अनुवाद पर अनुसंधान और विकास की गतिविधियों में गुणात्मक वृद्धि हुई है। फिर भी अब तक मशीनी अनुवाद पर अधिकांश अनुसंधान कार्य इसमें निहित जटिलताओं के प्रभावी निदान की अपेक्षा इन जटिलताओं के विविध स्रोतों को ही उजागर करते हैं। अनुवाद एक जटिल, उबाऊ और साधन-आश्रित कार्य है और इसमें मशीन-साधित पद्धति से हस्तक्षेप करने का यही एक बड़ा कारण है। अनुवाद के लिए समृद्ध और विविध ज्ञान स्रोतों की आवश्यकता होती है जिसमें न केवल भाषिक अपितु भाषा प्रयोग के सामाजिक-सांस्कृतिक, इत्यादि आयामों की गहन जानकारी आवश्यक है। स्रोतभाषा पाठ के अर्थ का लक्ष्यभाषा पाठ में अंतरण संबंधित भाषा के व्याकरण और शब्दकोश के ज्ञान के अतिरिक्त उन भाषाओं के प्रयोग (व्यावहारिक और वैश्विक ज्ञान) के ज्ञान से निर्धारित होता है। अर्थ निर्धारण एवं निरूपण में भाषा प्रयोग के संदर्भ ज्ञान की बड़ी भूमिका है, परंतु साथ ही इन्हें पूर्णतः समझना और विधिवत मशीन में प्रयुक्त करना एक बड़ी समस्या है। सामान्य ज्ञान को समझना एवं इन्हें नियमबद्ध करना एक मुश्किल कार्य है और मशीन अनुवाद प्रणाली विकास के लिए एक बड़ी चुनौती। भाषिक जानकारी (ध्वनि, शब्द, पदबंध, वाक्य और अर्थ की व्यवस्था) नियमबद्ध रूप में बहुत हद तक प्राप्त भी की जा चुकी है और इस दिशा में काफी प्रयास हो रहे हैं किंतु सामान्य/विश्व-ज्ञान (empirical knowledge) को नियमबद्ध करने की दिशा में कोई खास सफलता नहीं मिली है। मशीनी अनुवाद के शोध एवं विकास के क्षेत्र में यह आज भी एक बड़ा मुद्दा और कठिन चुनौती है।

14.3 मशीनी अनुवाद की प्रक्रिया

मशीनी अनुवाद (एम.टी) मूलतः कंप्यूटेशनल भाषाविज्ञान का एक अनुप्रयोग है जिसमें कंप्यूटर के प्रयोग द्वारा पाठ को एक प्राकृतिक भाषा से दूसरी प्राकृतिक भाषा में अनुवादित किया जाता है। इसमें स्रोत और लक्ष्यभाषा की प्रकृति समान या असमान हो सकती है। अनुवाद प्रक्रिया की जटिलता भाषायुग्म की भिन्नताओं पर निर्भर करती है। यह जटिलता विभिन्न कारणों से खड़ी हो सकती है। इनमें वाक्य संरचना के रूपों, व्याकरण रूपों, संदर्भिक द्वि-अर्थकता या वाक्य में पाई जाने वाली शाब्दिक द्वि-अर्थकता, लेखन शैली के रूपों, मुहावरों या लोकोक्तियों का प्रयोग, सांस्कृतिक भिन्नता आदि होते हैं। ये ऐसे प्रयोग हैं जिनसे मूल स्रोतपाठ एक विशिष्टता लिए रहता है और इसका दूसरी भाषा में अंतरण करते समय जटिलताएँ उत्पन्न होती हैं।

अनुवाद संबंधी प्रौद्योगिकी पत्रिकाओं में मशीनी अनुवाद को दस सबसे चुनौतीपूर्ण प्रौद्योगिकियों की सूची में रखा गया है। मशीनी अनुवाद के विस्तार में जाने से पूर्व हम देखेंगे कि अनुवाद स्रोतभाषा में व्यक्त संदेश के लिए लक्ष्यभाषा में निकटतम सहज समतुल्य संदेश को प्रस्तुत करना है, यह समतुल्यता पहले अर्थ के स्तर पर होती है फिर शैली के स्तर पर।

सामान्य अनुवाद के उद्देश्य, प्रविधि तथा उससे संबंधित चुनौतियों पर विचार करने से हम देखते हैं कि अनुवाद में हम उस समतुल्य को पाना चाहते हैं जो स्रोतभाषा पाठ में निहित होता है तथा अनुवाद प्रक्रिया द्वारा संप्रेषित किया जाता है। यहाँ समतुल्यता अर्थ, रूप, शैली और प्रभाव के संबंध में है। समतुल्यता प्रकृति में औपचारिक रूप से गतिशील हो सकती है। नाइडा के अनुसार तात्विक समतुल्यता रूप और तत्व दोनों में केवल संदेश पर केंद्रित रहती है, इसका लक्ष्य होता है पाठकों तक अधिक से अधिक लक्ष्यभाषा के संदर्भ को पहुँचाना। तथापि, परिवर्तनात्मक समतुल्यता समतुल्य प्रभाव के सिद्धांत पर आधारित होती है अर्थात् प्राप्तकर्ता और संदेशकता के बीच का संबंध वैसा ही होता है जैसाकि मूल प्राप्तकर्ताओं और स्रोतभाषा संदेशकता के बीच होता है।

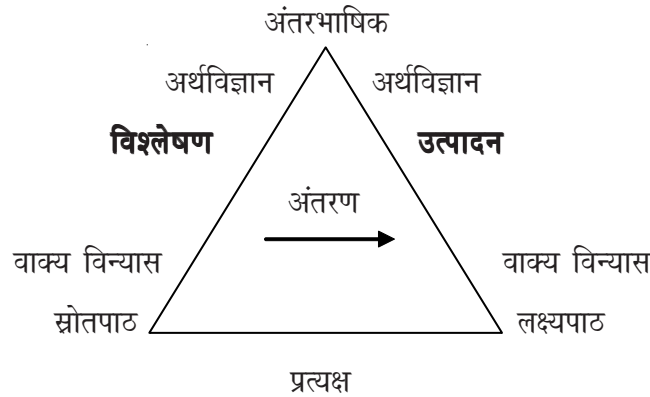
यदि हम मानवीय आधार पर अनुवाद प्रक्रिया का विश्लेषण करें तो हमें यहाँ दिए गए उपकार्यों की प्रक्रिया की तरह विश्लेषण करना होगा;

1. स्रोतपाठ के अर्थ को समझना; और
2. लक्ष्यभाषा पाठ में इस अर्थ को उचित रूप से गठित करना।

यद्यपि यह पूरी प्रक्रिया सामान्य-सी प्रतीत होती है, तथापि इसमें जटिल ज्ञानात्मक प्रक्रिया सम्मिलित रहती है। स्रोतपाठ के अर्थ को समझने अथवा विकोडीकरण करने में अनुवादक को दिए गए संदर्भ में पाठ में अंतर्निहित अर्थ की व्याख्या और विश्लेषण करना होता है जो स्रोतभाषा के व्याकरण, अर्थग्रहण वाक्य-विन्यास, मुहावरों आदि के साथ-साथ उसके बोलने वालों के सांस्कृतिक रूपों के गहन अध्ययन में निहित होता है। लक्ष्यभाषा में समान रूप से पुनर्गठित और कोडीकृत करने के लिए, अनुवादक को लक्ष्यभाषा के व्याकरण और लक्ष्यभाषा पाठकों के सांस्कृतिक रूपों की विस्तारपूर्वक जानकारी होनी चाहिए।

मशीनी अनुवाद के संदर्भ में जब यह बात आती है तब यह चुनौतीपूर्ण होता है कि कैसे एक कंप्यूटर पर प्रोग्राम किया जाए कि वह पाठ को मानव अनुवादक की तरह 'समझे' और लक्ष्यभाषा में नए पाठ की 'रचना' करे जो ऐसा प्रतीत हो कि लक्ष्यभाषा के किसी व्यक्ति द्वारा लिखा गया है। अर्थात् लक्ष्य पाठक उस रचित पाठ को स्वाभाविक रूप से ग्रहण करे।

एम.टी की कार्यप्रणाली में आने वाली समस्याओं को कई प्रकार से दूर किया गया है। साधनों और आवश्यकताओं के बढ़ने के साथ इसका क्षेत्र भी बढ़ता जा रहा है। मशीनी अनुवाद की पद्धतियों को वर्गीकृत करने के लिए विभिन्न मानक बनाए गए हैं। तथापि सबसे प्रचलित वर्गीकरण भाषायी स्तर (और उत्पाद) को ध्यान में रखकर 'सिस्टम' की जरूरत के अनुसार अनुवाद प्रस्तुत करने के लिए किया गया है। इस पूरी प्रक्रिया को निम्नलिखित एम.टी. पिरामिड द्वारा समझा जा सकता है जिसे 'वोकोइस त्रिकोण' कहते हैं।



आरेख 14.1 वोकोइस त्रिकोण

सामान्य तौर पर कहा जाए तो पिरामिड का निचला भाग उन माध्यमों को प्रस्तुत करता है जो लक्ष्यवाक्य प्रस्तुत करने के क्रम में स्रोत वाक्य का कोई भाषिक विश्लेषण नहीं करते हैं। हम त्रिकोण में जैसे-जैसे ऊपर की ओर बढ़ते हैं एम.टी का क्षेत्र बढ़ता जाता है अर्थात् जैसे ही हम ऊपर की तरफ बढ़ते हैं सिस्टम ज्यादा से ज्यादा विश्लेषण करता है। अंत में पिरामिड के शीर्ष पर स्रोतवाक्य की प्रासंगिक एवं अर्थगत विशेषता कार्यान्वित होती है जो स्रोतवाक्य में विन्यासगत व्याख्या के पश्चात लक्ष्य वाक्य उत्पादन में सहायक होती है।

14.4 मशीनी अनुवाद और पूर्ण अनुवाद

मशीनी अनुवाद के संदर्भ में अक्सर यह प्रश्न आम सुनने को मिलता है कि क्या कभी कोई मशीन पूर्ण अनुवाद कर पाएगी। इस प्रश्न का उत्तर 'सकारात्मक' या 'नकारात्मक' रूप में देना कठिन है। कुछ वर्ष पहले तक इसका उत्तर न में दिया जाता था परंतु समय के साथ-साथ प्रयास ने इसे 'सकारात्मक' में तबदील कर दिया है। लेकिन अभी भी एक लंबा रास्ता तय करना बाकी है। यह बहुत से अन्य कारणों पर भी निर्भर करता है; जैसे कि प्रयुक्ति क्षेत्र, अपेक्षा, उद्देश्य, प्रयोगकर्ता की संलग्नता और प्रक्रिया। कार्यविस्तार में एम.टी. के संकलन को स्वीकृत करने की नम्यता (लोच) मौजूद होती है। कुछ ऐसे भी क्षेत्र हैं जो शब्द-भंडार के नियम एवं भाषा और व्याकरणिक

संरचनाओं की दृष्टि से सीमित हैं; जैसे-मौसम पूर्वानुमान, ज्योतिषशास्त्र पर आधारित दैनिक प्रकथन। ऐसा कार्यक्षेत्र (domain) जो नियंत्रित स्रोतभाषा के प्रयोग की अनुमति दे अथवा लक्ष्य पक्ष पर स्वीकृत प्रतिस्थापनों के क्षेत्र को सीमित करे तो एम.टी की बेहतर गुणवत्ता की अपेक्षा की जा सकती है। इन विषय-क्षेत्रों में एम.टी उपयोगी रहा है और कुछ भाषा-युग्मों के लिए बहुत से व्यावसायिक उत्पाद उपलब्ध हैं। दूसरी महत्वपूर्ण चिंता प्रयोगकर्ता के अपेक्षा स्तर की है - जहाँ यह विषम भाषी सूचना की पुनःप्राप्ति, संदेश, प्रसार, यात्रा के दौरान संप्रेषण, ब्राउजिंग और कुछ अन्य कार्यों के लिए प्रयुक्त होता है जहाँ सूचना हस्तांतरण औपचारिक संदेश पहुँचाने की अपेक्षा अधिक महत्वपूर्ण होता है। यहाँ अपरिष्कृत एम.टी भी संभवतः पर्याप्त है। तीसरा दृष्टिकोण जो अंतिम नहीं है तथापि जहाँ तक यह उद्देश्य की पूर्ति करे पर्याप्त है। एक उत्साही यात्री के लिए अगर एक सुलभ अनुवाद यंत्र छोटी से छोटी जरूरत के लिए संपर्क करने और रास्ता तलाशने में मदद कर सकता है तो अपरिष्कृत क्षमता का एम.टी सुलभ एम.टी यंत्र के तौर पर उपयोगी है। वहीं चिकित्सा क्षेत्र के लिए छोटी-सी गलतफहमी या विसामान्य अनुवाद-घातक हो सकता है। मानव हस्तक्षेप भी अच्छी गुणवत्ता के एम.टी अनुवाद को पाने में मदद करता है। प्रयोगकर्ता का हस्तक्षेप पूर्व, मध्य और उत्तर अनुवाद जैसे एम.टी के विभिन्न स्तरों पर हो सकता है। पूर्व या पहले अनुवाद में प्रयोगकर्ता का हस्तक्षेप नियंत्रित भाषा, सही निवेश (उच्चारण और व्याकरण के स्तर पर) ऐसे शब्दों की पहचान जिन्हें अनुवाद किए जाने की आवश्यकता नहीं है (उदाहरणार्थ-नामित तत्व) और मुहावरों की सीमा निर्धारण करना आदि और यह विभिन्न प्रणालियों पर निर्भर करता है कि इस स्तर पर प्रयोगकर्ता के कितने हस्तक्षेप की अपेक्षा होती है। अनुवाद प्रक्रिया के दौरान कुछ विद्वान पदबंध के खास पद की द्विअर्थकता के समाधान के लिए सही दिशा में प्रयोगकर्ता के हस्तक्षेप की माँग कर सकते हैं। उत्तर मशीन अनुवाद में प्रयोगकर्ता से अपेक्षा की जा सकती है कि वह अनूदित पाठ-विन्यास के आधार पर एम.टी. सिस्टम द्वारा दिए गए विकल्पों में से सबसे सही या उपयुक्त का चयन करे। प्रयोगकर्ता की ग्राह्यता और मशीन अनुवाद निर्गम के सुधार के लिए तत्परता जैसे प्रयास आने वाले वर्षों में एम.टी को और बेहतर बनाने में मदद कर सकते हैं।

14.5 मशीनी अनुवाद : एक चुनौती

मशीनी अनुवाद अपने आप में एक अत्यंत संश्लिष्ट प्रक्रिया है तथा अभी तक बहुत सारे ऐसे मुद्दे हैं जिन पर गहन विचार-विमर्श हो रहा है। यह तो स्पष्ट ही है कि अनुवाद व्यावहारिक दृष्टि से भी एक मानवीय गतिविधि है और मशीन मनुष्य की इस श्रमसाध्य गतिविधि में सहायक हो सकती है। किस सीमा तक मशीन पर मनुष्य आश्रित हो सकता है अथवा मशीन से प्राप्त होने वाले पाठ को किस सीमा तक आदर्श अनुवाद माना जा सकता है यह अभी एक बड़ी बहस का मुद्दा है। इसका मूल कारण है कि मशीन के सामने वे तमाम पक्ष बाधा बनकर खड़े हो जाते हैं जिन्हें मानव मस्तिष्क अपनी सहजात वृत्ति से हल कर लेता है। मशीनी अनुवाद की चुनौतियाँ शब्दावली, व्याकरण और संकल्पनात्मक संरचना के रूप में भाषाओं की भिन्नता पर निर्भर करती है। इन चुनौतियों को निम्न प्रकार से वर्गीकृत किया जा सकता है :

- (क) **स्रोत वाक्य में निहित द्विअर्थकता;** जैसे-“He saw a girl with blue eyes” यहाँ द्विअर्थकता इस रूप में है कि नीली आँखें किसकी है; उस लड़के की जिसने लड़की को देखा या उस लड़की की जिसे लड़के ने देखा।
- (ख) **दोनों संबंधित भाषाओं के बीच की संरचनात्मकता विविधता :** हिंदी और जापानी भाषाओं में कर्ता-कर्म-क्रिया संरचना होती है वहीं अंग्रेजी में SVO और आइरि I भाषा में VSO है।
- (ग) **विशिष्टता में भिन्नता :** अंग्रेजी भाषा में ‘अंकल’ चाचा, ताऊ, मामा और फूफा कई सामाजिक रिश्तों/ परिस्थितियों के लिए प्रयोग होता है। इसी तरह हिंदी का कल अंग्रेजी के ‘tomorrow’ या ‘yesterday’ दोनों के लिए प्रयोग होता है। जिन भाषाओं में निर्धारक (Article) नहीं होता उनसे निर्धारक रखने वाली भाषा में अनुवाद मुश्किल होता है। उदाहरणार्थ, इस वाक्य का अनुवाद करने में “ताज एक सुंदर स्मारक है” और खासतौर से अनूदित वाक्य में “The Taj is a beautiful monument” में “The” निर्धारक रखने की आवश्यकता है।
- (घ) मुहावरे या लोकोक्तियों को एक भाषा में पहचानना और दूसरी भाषा में इसका पुनः प्रतिस्थापन करना; जैसे : Box office, दूर के ढोल सुहाने।

- (इ) शब्द एक से ज्यादा अर्थ रख सकते हैं। अर्थात् शाब्दिक प्रतिस्थापन में विषय का अर्थ या प्रसंग समझे बिना किसी शब्द के अर्थ की व्याख्या करना कठिन कार्य होता है।
- (च) कभी-कभी दो भाषाओं के बीच कथ्य के भागों का एक के लिए 'एक' शब्द-भेद नहीं मिलता है। उदारहरण के लिए, "He entered the room" वाक्य में क्रिया विषयक भाग 'entered' को हिंदी में 'प्रवेश किया' अनुवाद किया जाएगा जो संज्ञा+क्रिया (प्रवेश+क्रिया) का एक पद होगा।
- (छ) भाषाओं की सांस्कृतिक भिन्नता। "Prime minister is coming", इस वाक्य को "प्रधानमंत्री आ रहे हैं" के रूप में अनूदित किया जाएगा। सांस्कृतिक भिन्नता के कारण सम्मान देने के लिए यहाँ क्रियामूलक भाग बहुवचन में बदल जाएगा।

अतः प्रोग्रामर को जटिल भाषा और उन्हें प्रोग्रामिंग भाषाओं के माध्यम से किस प्रकार एनकोड (कोडित) किया जाए इसकी समझ होनी आवश्यक है। मानव भाषा व्यवहार के समकालिक रूप को समझना और उस तक पहुँचना मुश्किल ही नहीं, नामुमकिन है, क्योंकि निरंतर परिवर्तनशील भाषा एक अन-आच्छादित और सक्रिय इकाई है परंतु अभी तक इसे पूरी तरह समझा जाना बाकी है।

चुनौतियों और आर्थिक प्रयास के बावजूद वैश्विक और राष्ट्रीय स्तर पर ऐसे अनेक क्षेत्र हैं जहाँ अनुवाद की तत्काल आवश्यकता होती है परंतु उसके पर्याप्त मानवीय और दूसरे सहायक संसाधनों का भारी अभाव है। आवश्यकताएँ इतनी तत्कालिक एवं महत्व की हैं कि उन्हें नज़र-अंदाज करना व्यक्ति और समाज के हित में नहीं है। यही कारण है कि मशीनी अनुवाद समय और पूँजी के अनावश्यक व्यय को बचाता है। अतः मशीनी अनुवाद की आवश्यकता नितांत न्यायसंगत हो जाती है। अनुवाद कार्य की गति में तेजी, भाषाओं के विभाजन को पाटना, मानवीय प्रयासों को कम करना तथा सूचना प्रसारण आदि में भागीदारी जैसे कुछ और भी कारण शामिल हैं जो मशीनी अनुवाद को वर्तमान विश्व में और प्रासंगिक बनाते हैं। मशीनी अनुवाद संबंधी कुछ और मुद्दों तथा चुनौतियों पर हम इस इकाई में आगे विशेष रूप से चर्चा करेंगे।

पिछली इकाई में अनुवाद की कुछ पद्धतियों पर चर्चा हुई है। इसी क्रम में कुछ अन्य पद्धतियों पर हम यहाँ विचार करेंगे ताकि अनुवाद संबंधी चुनौतियों को सार्थक रूप से हल किया जा सके। इसी से मशीनी अनुवाद की व्यावहारिकता से जुड़े प्रश्नों पर भी हम और नजदीकी से गौर कर सकेंगे।

14.6 मशीनी अनुवाद की कुछ और प्रविधियाँ

एम.टी की चुनौतियों को कई प्रकार से सुलझाया गया है। तकनीकों के आधार पर इन्हें मुख्यतः चार वर्गों में बाँटा जा सकता है :

1. प्रत्यक्ष, 2. नियम आधारित, 3. कॉर्पस आधारित और 4. हाइब्रिड तकनीक।

प्रत्यक्ष अनुवाद को नियम आधारित पद्धति का आरंभिक चरण भी कहा जा सकता है। इस पर पिछली इकाई में विस्तार से चर्चा की जा चुकी है और यह स्पष्ट किया जा चुका है कि यह मशीनी अनुवाद की प्रथम पीढ़ी थी। इसमें अनुवाद प्रक्रिया को कुछ चरणों में तोड़ा जाता है और प्रत्येक चरण एक निर्दिष्ट कार्य करता है। इस पद्धति में वाक्य स्तर की 'पार्सिंग' नहीं होती है और शब्द से शब्द स्तर का अनुवाद द्विभाषिक शब्दकोश के प्रयास से होता है। यह बहुत ही सतही रूपात्मक-वाक्यात्मक विश्लेषण करता है। अगर हम इसमें सम्मिलित परंपरागत स्तर या मूल एलगॉरिद्म/प्रमेय को देखें तो वे निम्न प्रकार होंगे :

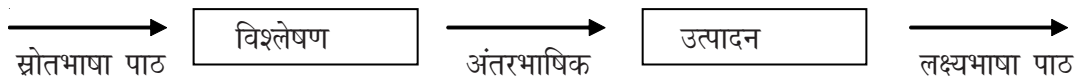
- वाक्य की सीमारेखा पहचानना और निर्धारित करना।
- वाक्य प्रतीक बनाना और वाक्य निर्मित करने वाले शब्दों को रखना।
- शब्दकोशों में शब्दों को ढूँढ़ने से पहले बड़े अक्षरों को छोटे अक्षरों (विभक्त) के रूप में परिवर्तित करना।
- द्विभाषी शब्दकोश में शब्द प्रति शब्द ढूँढ़ना।

- द्वि-अर्थकता की चुनौती से बचने के लिए एवं उपयुक्त शब्दार्थ लाने के लिए स्वतःशोध प्रणाली का प्रयोग करना।
- अज्ञात शब्दों, प्रतीकों या वाक्य के दूसरे असंचालित भाग को अनूदित निर्गम में उसी रूप में ही अंतरित (यथावत) करना; तथा।
- उत्तर-संपादन के लिए इंटरफेस।

यह मात्रा उन भाषा-युग्मों के लिए उपयुक्त होता है जो व्याकरण और वाक्य संरचना के आधार पर समान हों। यह तकनीक शब्द विभक्ति संचालन द्वि-अर्थकता के समाधान और शब्दक्रम जैसी चुनौतियों का सामना करती है, क्योंकि इसमें गंभीर वाक्यात्मक या सांकेतिक विश्लेषण नहीं होता। यह प्रारंभिक पद्धति आज प्रयोग नहीं की जाती है। वर्तमान में यह कई भाषा-युग्मों के लिए कार्य करती है और <http://babelfish.altavista.com> वेबसाइट पर इसकी सुविधा उपलब्ध है। यह प्रणाली द्वि-अर्थकता के समाधान, शब्दक्रम और अनुरूपताओं की चुनौतियों से ग्रस्त है। समय के साथ सिस्ट्रॉन अंतरण आधारित पद्धति की ओर अग्रसर हो गया है और उपरोक्त समस्याओं को सुलझाने के लिए यंत्रावली समाविष्ट कर ली है। भारत में विकासाधीन अनुसारक प्रणाली भारतीय भाषाओं के मध्य अनुवाद के लिए है और भारतीय भाषाओं के बीच की समानताओं की सुविधा के आधार पर आगे बढ़ रही है।

अंतरभाषिक मशीन अनुवाद पद्धति पर भी हमने पिछली इकाई में विचार किया था। इस पद्धति से अनूदित किया जाने वाला स्रोतभाषा पाठ एक अंतर्भाषा में परिवर्तित हो जाता है जो भाषायी स्वतंत्रता की पुनः प्रस्तुति का सार है। यह एक कृत्रिम भाषा है जिसे कुछ लोग छद्म (Psuedo) भाषा की संज्ञा देते हैं। इसे तटस्थ रखने के लिए यथासंभव प्रयास किए गए हैं। इस प्रकार हम कह सकते हैं कि यह स्रोत के साथ-साथ लक्ष्य स्वतंत्र भाषा भी है। यह यथासंभव स्रोतभाषा के अर्थ को पुनः प्रस्तुत करने का प्रयास करती है तब लक्ष्यभाषा अंतर्भाषिक से उत्पन्न होती है। नियम आधारित मशीन अनुवाद के प्रतिमान में भी अंतरभाषिक पद्धति प्रत्यक्ष अनुवाद पद्धति और अंतरण अनुवाद पद्धति के विकल्प के रूप में कार्य करती है। यह एक ऐसी क्रियाविधि के निर्माण की दिशा में एक प्रयास है जो भाषाओं के एक से अधिक युग्मों के बीच अनुवाद को सुसाध्य बना सके। अंतरभाषिक अनुवाद अंतरण आधारित प्रविधि का एक विशिष्ट उदाहरण है जहाँ अंतरण घटक गौण रहते हैं। अंतर्निहित अर्थों का कोडीकरण और प्रस्तुतीकरण एक कठिन कार्य है और यहाँ तक कि गहन प्रस्तुतिकरण भी मूल से भिन्न होता है। इस प्रयास की पहुँच और प्रस्तुति केवल भाषिक स्तर तक हो पाई है किंतु अर्थगत प्रस्तुतिकरण तक पहुँचने की दिशा में निरंतर प्रयास जारी है।

अंतरभाषिक शब्दकोश भाषा के प्रत्येक शब्द की प्रकृति और आचरण के विषय में अतिरिक्त जानकारी रखता है। इसमें घटनाएँ और क्रियाएँ सम्मिलित होती हैं जो अंतरभाषिक प्रस्तुतीकरण को यथासंभव स्रोत और लक्ष्यभाषा से स्वतंत्र निर्माण में मदद करता है।



आरेख 14.2 अंतरभाषिक मशीन अनुवाद

उपरोक्त चित्र अंतरभाषिक एम.टी पद्धति का प्रतीकात्मक रेखाचित्र प्रस्तुत करता है। इसमें मुख्यतः दो घटक हैं - विश्लेषक और संश्लेषक। विश्लेषक भाग स्रोत वाक्य के शाब्दिक, वाक्यगत और अर्थगत संरचना का विश्लेषण करता है। यह दिए गए पाठ का विहित अंतरभाषिक रूप में अर्थ व्यक्त करता है। संश्लेषक अर्थ के अंतरभाषिक निरूपण एक या एक से अधिक वाक्य का निर्माण करता है। के.ए.एन.टी, सी-स्टार और यू.एन.एल. (यूनिवर्सल नेटवर्किंग लैंग्वेज) पर आधारित पद्धतियाँ अंतरभाषिक मशीन अनुवाद पद्धति के उदाहरण हैं। यहाँ हम सी-स्टार 1999 से अंतरभाषिक आधारित पद्धति का उदाहरण देखेंगे :

स्रोत : I will be arriving on July 8th.

अंतरभाषिक : देना-सूचना-समयवाचक (टेम्पोरेल) पहुँचना

[Who =kk I, time =kk (July, md 8)]

लक्ष्य : मैं 8 जुलाई को आऊँगा।

अंग्रेजी से दूसरी भारतीय भाषाओं में अनुवाद के लिए आई.आई.टी कानपुर की आंग्ल भारती पद्धति आभासी (स्यूडो) अंतरभाषिक पद्धति का प्रयोग करती है। भारतीय भाषाएँ समान कर्ता-क्रिया-कर्म (SVO) क्रम का अनुकरण करती हैं और कभी-कभी पूर्ववर्ती संज्ञा से जुड़ी हुई परसर्गों का प्रयोग करती हैं यह सूडो-अंतरभाषिक पुनः प्रस्तुतिकरण भारतीय भाषाओं के युग्मों के अनुरूप हो जाता है। अंतरभाषिक अनुवाद के संदर्भ में नीचे दिए गए बिंदु विचारणीय हैं :

- प्रत्येक भाषा-युग्म के लिए कोई अंतरण घटक नहीं बनाया जाता।
- अंतरभाषिक की अभिकल्पना बहुत जटिल है। इसका कारण है कि पूर्ण अंतरभाषिक निरूपण बनाने की कोई स्पष्ट प्रविधि नहीं है।
- वृहत्तर क्षेत्र के लिए भी अंतरभाषिक निरूपण विकसित किया जाना लगभग असंभव है।

अंतरण पद्धति पर भी पिछली इकाई में विस्तार से चर्चा की जा चुकी है। यहाँ केवल इसके प्रक्रियात्मक पक्ष पर ही हम अपना ध्यान केंद्रित करेंगे। यह स्पष्ट हो चुका है कि एक निरपेक्ष अंतरभाषिक निरूपण के रूपांकन में आने वाली जटिलता को कम करने के लिए अंतरण पद्धति एक उपयोगी साधन प्रस्तुत करती है। यह एक अतिरिक्त 'अंतरण' घटक को सम्मिलित करता है, जो इसे त्रिस्तरीय प्रक्रिया बनाता है- विश्लेषण, अंतरण और संश्लेषण या उत्पादन। यहाँ 'विश्लेषण' की भूमिका स्रोतभाषा से संबंधित वाक्यगत संरचना के पुनः प्रस्तुतिकरण की समझ और प्रस्तुति तक सीमित रह जाता है, जबकि 'संश्लेषण' घटक की भूमिका लक्ष्य केंद्रित उत्पाद बनाने के आसपास संगठित होती है। 'अंतरण' घटक स्रोत के प्रतिरूपण से लक्ष्यभाषा के प्रतिरूपणों में नियमावली बनाता है। इसे दो भिन्न भाषा बोलनेवालों के बीच कार्य करने वाले मानवी व्याख्याता के कार्य के समान समझा जा सकता है। यहाँ एक अतिरिक्त स्तर के 'परिचय' की आवश्यकता पड़ती है, तथापि यह विश्लेषण और संश्लेषण के स्तरों पर कार्य साधन की जटिलता को नियंत्रित करता है। अर्थात् इन्हें अंतरभाषिक पद्धति के लिए अपेक्षित भाषा स्वतंत्रता की आवश्यकता न होकर एक विशेष भाषा से संबंधित होना होता है। इसमें तीन प्रकार के शब्दकोशों और व्याकरणों की आवश्यकता होती है : स्रोतभाषा, लक्ष्यभाषा और स्रोत-लक्ष्य अंतरण के लिए। इस संदर्भ में निम्नलिखित बिंदु हमारे सामने आते हैं :

- कॉर्पस आधारित पद्धति में बहुभाषिक कॉर्पस के बृहत् खंडों की आवश्यकता होती है जो आसानी से उपलब्ध नहीं होते हैं।
- फिर भी व्याकरण आधारित पद्धतियों को सावधानीपूर्वक व्याकरण की रूपरेखा तैयार करने वाले सक्षम भाषाविज्ञानी की आवश्यकता होती है।
- संबंधित भाषा-युग्मों की जरूरत के अनुसार भाषिक नियम निर्धारित किए जाने की आवश्यकता होती है।
- परस्पर संबंधित भाषाओं के मध्य अनुवाद के लिए सतही-अंतरण मशीनी अनुवाद जैसी एक तकनीक काफी हो सकती है।

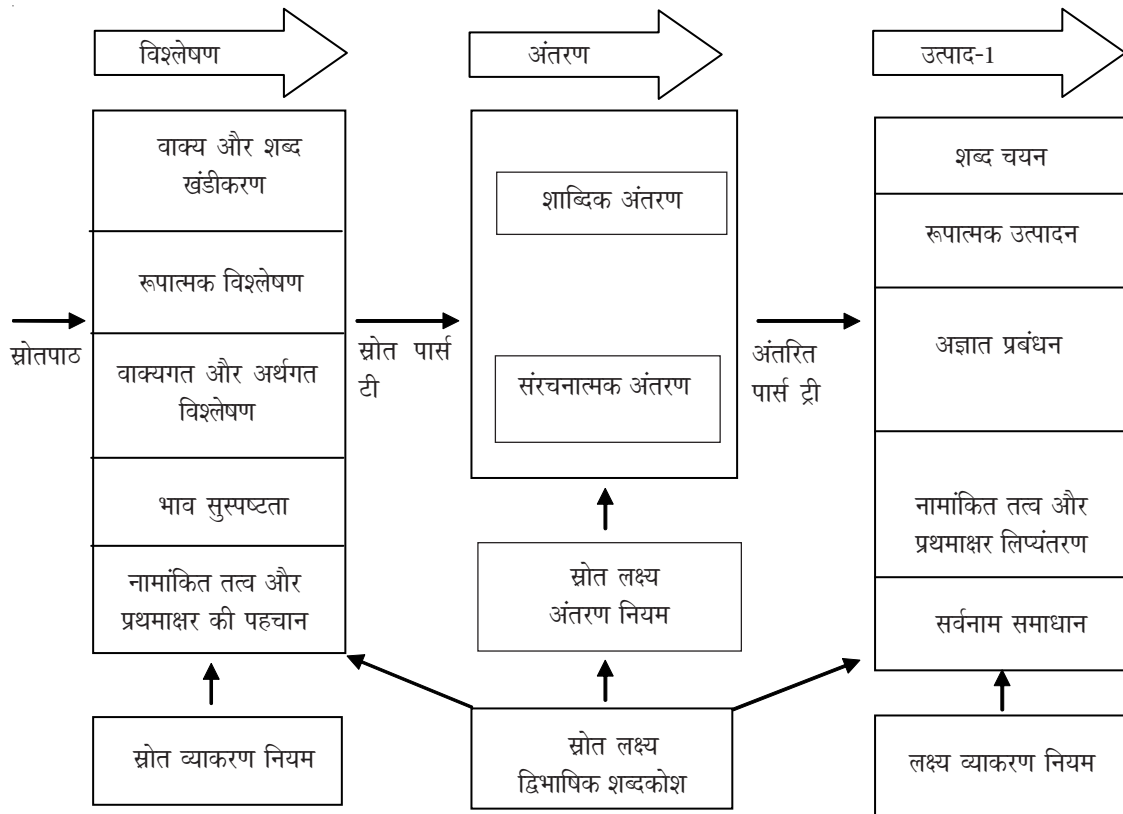
अब तक हमने मशीनी अनुवाद की पद्धतियों के अंतर्गत प्रत्यक्ष अनुवाद पद्धति, नियम आधारित पद्धति, अंतरभाषिक मशीनी अनुवाद पद्धति तथा अंतरण पद्धति पर चर्चा की। मशीनी अनुवाद की पद्धतियों में नियम आधारित अनुवाद पद्धति, कॉर्पस आधारित अथवा प्रयोगाश्रित पद्धति तथा हाइब्रिड पद्धति का भी अपना विशिष्ट स्थान है।

14.6.1 नियम आधारित पद्धति

नियम आधारित पद्धति भाषा व्याकरण के नियमों का प्रयोग कर स्रोतपाठ का विश्लेषण करने और लक्ष्यभाषा के व्याकरणिक नियमों के आधार पर लक्ष्य भाग को निर्मित करने का समर्थन करती है। प्रत्यक्ष अनुवाद या शब्दकोश आधारित अनुवाद प्रक्रिया को नियम आधारित अनुवाद पद्धति के आरंभिक चरण के रूप में देखा जाता है। मशीनी अनुवाद पद्धतियों की द्वितीय पीढ़ी भाषाविज्ञान के नियमों और व्याकरण निरूपक तकनीकों के समावेश से गहन

भाषावैज्ञानिक विश्लेषण की ओर केंद्रित थी। यह धारणा इस तथ्य के आसपास घूमती है कि मूल जानकारी के विश्लेषण यदि अर्थगत नहीं हैं तो वाक्यगत और सार के बिना लक्ष्यभाषा में अनुवाद संभव नहीं है। इसे पिरामिड (14.1) में वाक्यगत और अर्थगत विश्लेषण के द्वारा दिखाया गया है। अधिक से अधिक जानकारी कूटबद्ध करने से हम इंटरलिंगुआ (एक मध्यस्थ भाषा) और अंतर घटक की समाप्ति के निकट पहुँचते हैं। आदर्शतः मूल इंटरलिंगुआ निरूपण के लिए यहाँ कोई अंतर घटक नहीं होना चाहिए।

सामान्यतः नियम आधारित प्रणाली शाब्दिक, वाक्यगत, अर्थगत और व्यावहारिक रीतियों में स्रोतभाषा का विश्लेषण करती है। इसमें रूपात्मक विश्लेषण, पाठ की पदव्याख्या, अर्थ स्पष्टता, 'नामक' तत्व की पहचान, बहु-संदेश अभिव्यक्तियों की पहचान और विश्व ज्ञान एवं व्यावहारिकता आदि सम्मिलित हैं। ठीक उसी प्रकार लक्ष्य पक्ष उत्पादन अर्थात् विश्लेषण का विपर्याय किया जाता है। उत्पादक पक्ष में परंपरागत रूप से पदव्याख्या वृक्ष-अंतरण (Tree-transfer), रूपात्मक संश्लेषण, नामक तत्वों, प्रथमाक्षरों से और बहु-संदेश अभिव्यक्तियों का अनुवाद या लिप्यंतरण, शाब्दिक अंतरण आदि के माध्यम से समतुल्य उत्पादन सम्मिलित है। इन दो महत्वपूर्ण चरणों (विश्लेषण और उत्पादन) के बीच एक मध्यस्थ प्रतीकात्मक निरूपण उत्पादित होता है। मध्यस्थ निरूपण की प्रकृति के आधार पर इस पद्धति को अंतरभाषिक मशीन अनुवाद या अंतरण आधारित मशीन अनुवाद कहा जाता है। नियम आधारित पद्धति रूपात्मक, वाक्यगत और अर्थगत जानकारी के साथ विस्तृत शब्दकोशों का प्रयोग करती है और नियमों के विस्तृत ढाँचे इसमें सहायता करते हैं। निम्नलिखित आरेख इस प्रक्रिया को स्पष्ट करता है :



आरेख 14.3 नियम आधारित मशीन अनुवाद पद्धति का अभिलक्षक आरेखीय ढाँचा

अंतरण-आधारित मशीन अनुवाद और अंतरभाषिक मशीन अनुवाद को हम नियम आधारित मशीन अनुवाद के उदाहरणों में विस्तार से देखेंगे।

14.6.2 कॉर्पस आधारित या प्रयोगाश्रित पद्धति

कॉर्पस आधारित पद्धति अनुवाद प्रस्तुत करने के लिए पूर्व अनूदित सामग्री के प्रयोग या सादृश्य के प्रयोग का समर्थन करता है। एकल और द्विभाषी कॉर्पस के बड़े खंडों की उपलब्धता के साथ एम.टी में कॉर्पस आधारित

पद्धतियों की ओर झुकाव की प्रवृत्ति है। उदाहरण आधारित एम.टी, मैमोरी आधारित एम.टी और सांख्यिकीय मशीन अनुवाद, ऐसी तकनीकें हैं जो प्रयोगाश्रित पद्धति का प्रयोग करती हैं। मूलतः ये सभी तकनीकें कॉर्पस या डेटाबेस के अनूदित उदाहरणों का प्रयोग करती हैं। यहाँ तक कि एम.टी को प्राप्त करने के लिए सांख्यिकीय मशीनी अनुवाद सांख्यिकीय नमूनों के लाभ को अंतस्थापित करने का प्रयास करता है। यह पिछले अवलोकनों द्वारा स्वयं ही शाब्दिक और संरचनात्मक प्राथमिकताओं को सीखने का प्रयास करता है।

(क) उदाहरण आधारित मशीन अनुवाद (ई.बी.एम.टी)

उदाहरण आधारित मशीन अनुवाद सादृश्य तंत्र की प्रणाली पर कार्य करता है। इसकी अवधारणा मानव अनुवादकों की अनुवाद प्रक्रिया से ली गई है। जब अनुवादक एक वाक्य का अनुवाद करता है तब वह उस वाक्य को वाक्यांशों या पदबंधों में खंडित कर देता है। फिर उन खंडित वाक्यों एवं पदबंधों का अनुवाद किया जाता है और अंत में पूर्ण वाक्य बनाने के लिए उन खंडित वाक्यांशों और पदबंधों को जोड़ा जाता है। अनुवादक इन खंडों को जोड़ते समय वाक्य को उत्कृष्ट बनाने का प्रयास भी करता है। इस संपूर्ण प्रक्रिया में कोई भी गहन भाषिक विश्लेषण मौजूद नहीं है। वाक्यांशों या पदबंधों का अनुवाद पूर्व अनुवादों के सादृश्यों से किया जाता है।

नए अनुवाद के आधार के रूप में पूर्व अनुवादों के उदाहरणों का पुनः प्रयोग करना ही उदाहरण आधारित मशीन अनुवाद की मूल अवधारणा है। यह अपनी मूल जानकारी के आधार के रूप में द्विभाषी कॉर्पस का प्रयोग करता है। इसे मशीन द्वारा सीखने के उदाहरण आधारित तार्किक दृष्टिकोण के कार्यान्वयन के रूप में देखा जा सकता है। यह मिलान साधारण शब्दकोशीय स्तर का मिलान या नमूना आधारित मिलान भी हो सकता है। नमूना आधारित पद्धति को कम प्रशिक्षित कॉर्पस की आवश्यकता होती है।

अब हम एक उदाहरण की मदद से इस तंत्र को देख सकते हैं। निम्न तालिका द्विभाषी अंग्रेजी-हिंदी कॉर्पस के एक उदाहरण को प्रस्तुत करती है। इन वाक्य-युग्मों में अंग्रेजी भाषा के साथ उनके हिंदी अनुवाद के वाक्य भी दिए गए हैं :

द्विभाषी कॉर्पस के उदाहरण

अंग्रेजी

Red Fort was built by Mughals

Taj Mahal was built by Shahjahan

हिंदी

लाल किला मुगलों द्वारा बनवाया गया था

ताजमहल शाहजहाँ द्वारा बनवाया गया था

उपवाक्यात्मक इकाई के अनुवादों को इन समांतर वाक्यों द्वारा सीखा जाता है। उदाहरण के लिए, उपरोक्त कॉर्पस से उदाहरण आधारित मशीन अनुवाद प्रणाली निम्नलिखित उपवाक्यात्मक इकाइयों को सीखती है :

“Red Fort” का समांतर

“was built by” का समांतर

“Taj Mahal” का समांतर

“लाल किला”

“द्वारा बनवाया गया था”

“ताजमहल”

हम कॉर्पस के इन वाक्यों में देखते हैं कि बोल्ड अक्षरों की उप-वाक्यात्मक इकाइयाँ अपरिवर्तनीय हैं; जबकि अन्य उप-वाक्यात्मक इकाइयाँ परिवर्तनीय हैं। इन परिवर्तनीय वाक्यांशों को समतुल्य वर्गों के रूप में देखा जाता है जिनका प्रयोग एक-दूसरे के स्थान पर किया जा सकता है। यह इस प्रकार अनुवाद करता है :

अंग्रेजी वाक्य <X> was built by <Y>

हिंदी वाक्य <X> <Y> द्वारा बनवाया गया था।

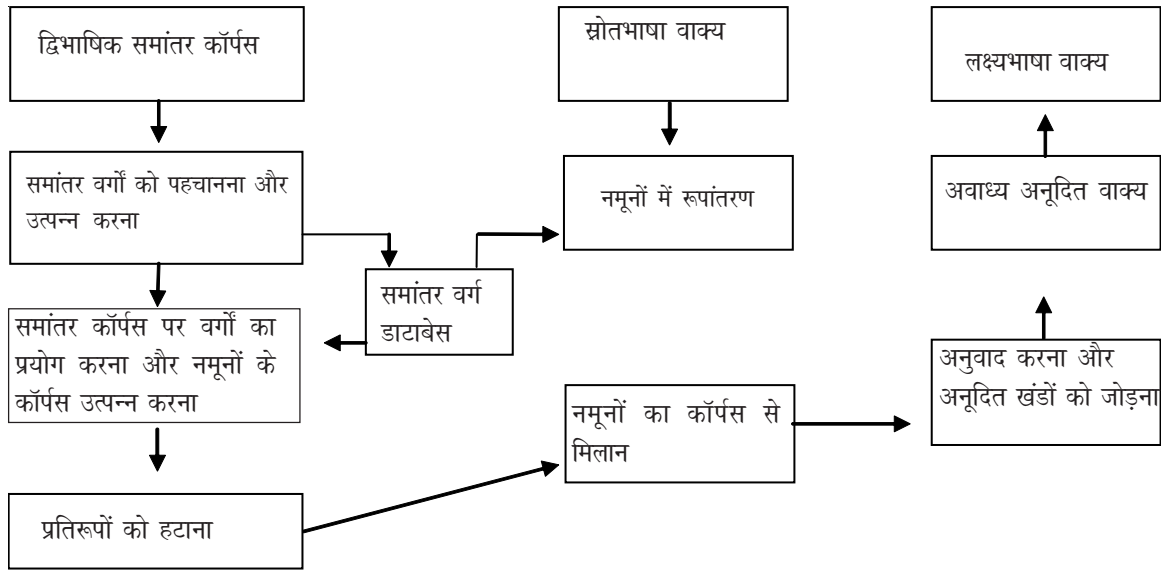
समतुल्य वर्गों में वर्गीकरण का आधार अर्थगत अथवा वाक्यगत हो सकता है। अर्थात् अर्थगत समतुल्य वर्ग कही गई जानकारी को इस प्रकार परिष्कृत करेंगे :

अंग्रेजी वाक्य <Building> was built by <Human or community name>

हिंदी वाक्य <इमारत> <व्यक्ति या समुदाय का नाम> द्वारा बनवाया गया था

इस प्रकार दिए गए समानता के वर्गीकरण के ढाँचे के साथ संपूर्ण समानांतर कॉर्पस नमूनों में बदल जाता है। अनूदित किया गया वाक्य नमूने में बदल जाता है और नमूनों के अनुक्रमणित अभ्यास को कॉर्पस के साथ मिलाया जाता है तथा मिलने वाली उप-वाक्यात्मक इकाइयाँ अनूदित की जाती हैं, और हम एक नए वाक्य “Qutab minar was built by Iltutmish” के समान नमूने “**कुतुब मीनार** इल्तुतमिश द्वारा बनवाया गया था” के वाक्य-रूप में अनुवाद प्राप्त करते हैं। यहाँ हम देख सकते हैं कि इस वाक्य को और परिष्कृत करने/बनाने की आवश्यकता है जैसा कि क्रिया भाग “बनवाया गया था” के स्थान पर इसे ‘स्त्रीवाचक’ (थी) बनवाई गई थी - में परिवर्तित करने की आवश्यकता है।

उदाहरण आधारित मशीन अनुवाद की संपूर्ण प्रक्रिया को नीचे दिए आरेख से चित्रित रूप में भी समझा जा सकता है :



आरेख 14.4 उदाहरण आधारित मशीन अनुवाद

उदाहरण आधारित मशीन अनुवाद का सुझाव सबसे पहले 1984 में माकोटो नगाओ ने दिया था। ई.बी.एम.टी. पदबंधीय क्रियाओं, जैसे उप-भाषिक प्रतिमानों के लिए सबसे उपयुक्त है।

पदबंधीय क्रियाएँ उच्च संदर्भ-आधारित अर्थ रखने के कारण एम.टी. के लिए चुनौतियाँ प्रस्तुत करती हैं। उनके विशेष संदर्भ से संबंधित अर्थों को संघटकों के अर्थ द्वारा व्युत्पन्न नहीं किया जा सकता है। पदबंधीय क्रियाओं का शब्द-प्रति शब्द अनुवाद अभिव्यक्ति या वाक्य के निहितार्थ से प्रायः दूर रहता है। इसका कोई सरल समाधान संभवतः ई.बी.एम.टी. ही उपलब्ध करा सकता है। उदाहरण के लिए ‘Puts’ और ‘On’ को अलग-अलग करने के बजाय ‘Puts on’ के रूप में ‘पहन’ शब्द की जानकारी ‘He puts on the clothes’ के वांछित अनुवाद को संभवतः प्राप्त करा सकती है।

शुरुआती ई.बी.एम.टी. (नगाओ) गाइजिन (वील - वे), सी.एम.यू. के सामान्यीकृत ई.बी.एम.टी. या जी-ई.बी.एम.टी., डी-3 (सुमिता) और एडगर (माइकल कार्ल) ई.बी.एम.टी प्रणालियों के उदाहरण हैं। उदाहरण आधारित मशीनी अनुवाद प्रणाली के संदर्भ में निम्नलिखित बिंदु द्रष्टव्य हैं :

- भाषिक ज्ञान की प्रचुरता या व्याकरण नियमों के परिचालन क्रम को बनाने की आवश्यकता नहीं है।
- खंड वाक्य अनुवादों को जोड़ने से बने परिणामी अनूदित वाक्य पूर्णतः एक दूसरे से नहीं जुड़ सकते हैं। वाक्यांशों में न मिलने वाले लिंग, उदाहरण, संख्या आदि हो सकते हैं। इसे सीमा मतभेद की समस्या के रूप में जाना जाता है।
- उच्च गुणवत्ता के सामान्य नमूनों को उत्पन्न करना आसान नहीं है।

(ख) सांख्यिकीय आधारित मशीन अनुवाद (एस.बी.एम.टी.)

सांख्यिकीय मशीन अनुवाद की एक और कॉर्पस आधारित पद्धति है। यह कुछ-कुछ ई.बी.एम.टी. के समान है जैसे कि यह भी पूर्ववर्ती अनुवाद कॉर्पस से शब्दों के पूर्व निर्धारित अर्थ) सीखती है। ई.बी.एम.टी. से भिन्न यह पद्धति वास्तविक द्विभाषिक कॉर्पस नहीं रखती और द्विभाषिक पाठ कॉर्पस आधारित सांख्यिकीय पद्धति का प्रयोग कर अनुवाद उत्पन्न करने का प्रयास भी करती है। द्विभाषिक कॉर्पस के कुछ प्रसिद्ध माध्यमों में कनाडा संसद के अंग्रेजी, फ्रांसीसी अभिलेख, यूरोपार्ल, यूरोपियन संसद के अभिलेख और भारतीय भाषाओं का एक समानांतर कॉर्पस ज्ञाननिधि शामिल है। एस.एम.टी. नियमों के रूप में क्रमवार रूप से कूटबद्ध करने के बजाय वास्तविक पाठ से स्वयं ही भाषा व्यवहार अध्ययन समझने की धारणा पर कार्य करता है। यह केवल सांख्यिकीय नमूनों और कॉर्पोरा से व्युत्पन्न मानकों के समूह का प्रयोग करता है। हम यह समझने का प्रयास करते हैं कि किस प्रकार दिए गए शब्द 'House' का अर्थ हिंदी में निर्धारित किया जाता है। नीचे की तालिका में एक द्विभाषिक समानांतर कॉर्पस में 'House' शब्द के लिए प्रतीक अर्थों की सूची दी गई है। मान लीजिए 'House' शब्द 1000 बार आया है और 700 बार 'घर', 120 बार 'मकान', 80 बार 'भवन', 60 बार 'सदन' और 40 बार 'हाऊस' के अर्थ में अनूदित किया गया है। नीचे दिया गया स्तंभ चित्रा दिखाता है कि संभाव्य शब्द किस प्रकार एक खास तरीके से अनूदित किया जा सकता है। यह जानकारी बिना किसी दूसरी जानकारी को ध्यान में रखकर उत्पन्न की जा रही है।

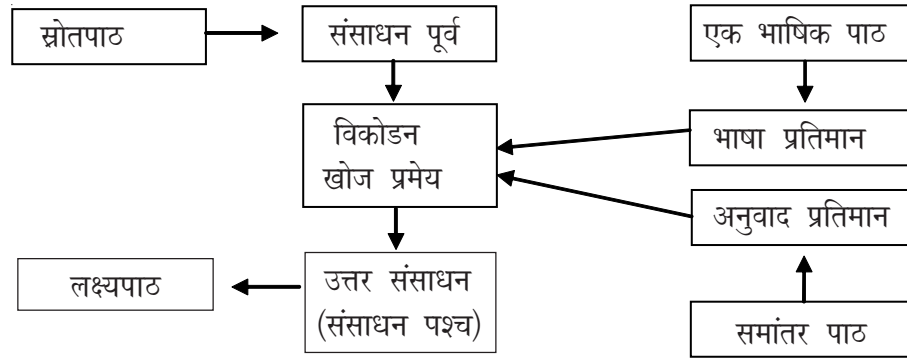
शब्दों का अनुवाद	उदाहरण वाक्य	संख्या	संभावित
घर	This is my house	700	70%
मकान		120	12%
भवन	The parliament house is in Delhi	80	8%
सदन	The lower house got adjourned in two hours only.	60	6%
हाऊस	The white house welcomed Indian president	40	4%

यह समान प्रकार के पाठों को अनुवाद करने एवं प्रभावशाली एम.टी. परिणाम पाने के लिए शीघ्रगामी और प्रभावी समाधान है। परन्तु इस प्रकार का कॉर्पोरा सहज ही उपलब्ध नहीं हैं और इसे उत्पन्न करना खर्चीला कार्य है। आई.बी.एम. शोधकर्ताओं द्वारा प्रस्तुत कैंडिड एस.एम.टी. प्रणाली ने मशीन अनुवाद की सांख्यिकीय पद्धति के लिए मार्ग प्रशस्त किया गया। गूगल कई भाषा युग्मों के लिए एम.टी. प्रणाली उपलब्ध कराता है, जो करोड़ों अनूदित वाक्यों के कॉर्पोरा पर आधारित है।

सांख्यिकीय मशीन अनुवाद में हमें लक्ष्यभाषा वाक्य T (टी) में अनूदित किए जाने के लिए स्रोत भाषा वाक्य S (एस) दिया जाता है। सांख्यिकीय मशीन अनुवाद बहुविकल्प माध्यम प्रतिमान पर आधारित है। यह प्रणाली बहुविकल्प अनुवाद वाक्य उत्पन्न कर सकती है और ऐसे वाक्य को पहचानना, जो स्रोतवाक्य S (एस) के अनुवाद में सबसे उपयुक्त होने पर अनुवाद की चुनौती होती है। यह निम्न सांख्यिकीय मापदंड के प्रयोग से संपन्न होता है :

$$t^* = \text{arg max } p(s/t) * p(t)$$

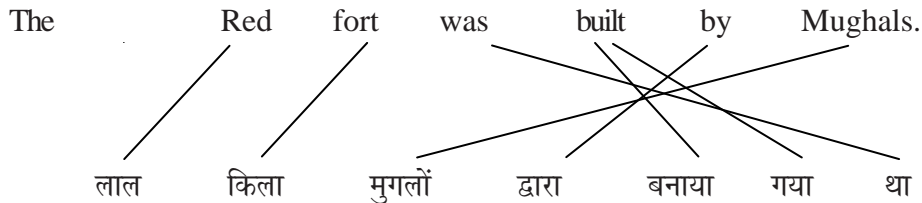
समीकरण (1) में, $p(s/t)$ अनुवाद प्रतिमान को और $p(t)$ भाषा प्रतिमान को प्रस्तुत करता है। अनुवाद प्रतिमान (टी.एम.) अनुवाद निष्ठा को सुनिश्चित करने की भूमिका निभाता है और भाषा प्रतिमान (एल.एम.) अनूदित निर्गम में धाराप्रवाह और व्याकरण की यथातथ्यता को सुनिश्चित करता है। विकोडक या खोजकर्ता कलनविधि (algorithm) समीकरण (1) में बताए गए के अनुसार इष्टतमीकरण (optimization) का कार्य निष्पादित करता है। इसकी भूमिका है संभाव्यता प्रारूप (TM & LM) द्वारा उत्पादित प्राप्तांक पर आधारित अनुवाद के निगमों में से सबसे अच्छे अनुवाद विकल्प का चयन करना। निम्नलिखित आरेख एस.एम.टी. में सम्मिलित विभिन्न घटकों के विशिष्टक्रम को दिखाया है।



आरेख 14.5 सांख्यिकीय मशीनी अनुवाद

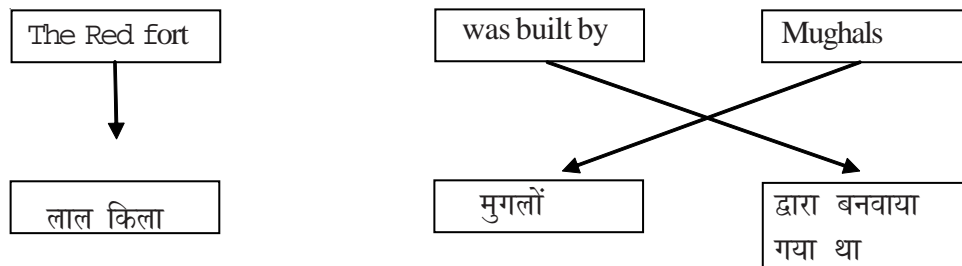
सांख्यिकीय प्रतिमान द्विअर्थकता की समस्या का अच्छा समाधान सुझाता है। त्रुटियों और नए आँकड़ों की मौजूदगी के बावजूद वे संतुलित और अच्छा कार्य करते हैं। हालांकि यह पद्धति शुरुआत में शब्द-प्रति-शब्द स्तर पर कार्य करती थी, किंतु बाद में इसने अपना दायरा बढ़ाया है और यह वाक्यांश से वाक्यांश के अनुवाद का प्रयोग करने में समर्थ हुई।

नीचे का उदाहरण (1) शब्द प्रति शब्द आधारित एस.एम.टी. (SMT) में सम्मिलित मापन और पुनःक्रमण को दिखाता है और उदाहरण (2) दिखाता है कि किस प्रकार शब्द से वाक्यांश आधारित एस.एम.टी. (SMT) में जाने पर मापन और पुनःक्रमण कम हो जाता है।



उदाहरण 1 : शब्द प्रति शब्द आधारित आदर्श

यहाँ वाक्यांश बहुशब्दिक इकाई है और यह जरूरी नहीं है कि वह भाषायी प्रक्रिया में वाक्यांश हो। उदाहरण के लिए “was built by” यहाँ वाक्यांश इकाई के रूप में माना गया है।



उदाहरण 2 वाक्यांश से वाक्यांश आधारित आदर्श

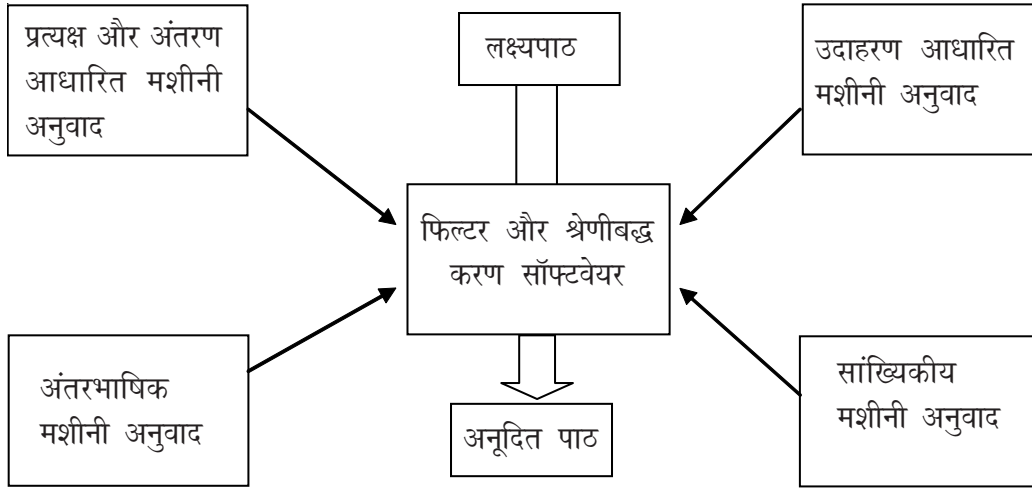
आजकल भाषिक सूचनाएँ निर्धारित स्तर पर सम्मिलित की जाती हैं और इस तरह वे एम.टी. पिरामिड के उच्च स्तर तक पहुँच रही हैं।

एस.एम.टी. सफलता कॉर्पस की शब्द भंडार क्षमता और मात्रा पर निर्भर करती है। अतः कॉर्पस निरूपक और प्रभाव-क्षेत्र में प्रयुक्त भाषिक संरचनाओं का क्षेत्र विस्तृत होना चाहिए।

14.6.3 हाइब्रिड मशीनी अनुवाद पद्धति

हाइब्रिड पद्धति मध्यम रास्ता अपनाती है और नियम आधारित और कॉर्पस आधारित दोनों पद्धतियों का लाभ लेने का प्रयास करती है। समय और शोधों के साथ यह सिद्ध हो चुका है कि कोई भी एक पद्धति पूर्ण मशीन

अनुवाद के उद्देश्य तक नहीं पहुँच सकी है। हाइब्रिड मशीन अनुवाद (एच.एम.टी.) सांख्यिकीय और नियम आधारित अनुवाद की कार्यविधि की शक्तियों का प्रयोग करती है। विभिन्न एम.टी कंपनियाँ अपने प्रयास को इस ओर ले जा रही हैं। उदाहरण के लिए, एशिया ऑनलाइन और सिस्ट्रॉन नियम और सांख्यिकीय दोनों का प्रयोग करने वाली हाइब्रिड पद्धति विकसित करने का दावा करते हैं। हाइब्रिड मशीनी अनुवाद पद्धति के लिए आगे दिया गया आरेख द्रष्टव्य है।



आरेख 14.6 हाइब्रिड मशीनी अनुवाद

कॉर्पस आधारित और नियम आधारित पद्धतियों का एकीकरण उनके क्रम या प्रयोग के अनुसार भिन्न होता है। कुछ निष्पादन की शुरुआत में नियम आधारित अनुवाद और तत्पश्चात् निर्गम को सही या रूपांतरित करने के लिए सांख्यिकी का प्रयोग करते हैं, जबकि कुछ सांख्यिकीय इंजन को निर्देशित या नियंत्रित करने के लिए पाठ की पूर्व-प्रक्रिया के नियमों का प्रयोग करते हैं। अनुवाद निर्गम के सुधारीकरण के लिए उत्तर प्रक्रिया भी नियमों की मदद से की जाती है। एक मध्यस्थ सॉफ्टवेयर विभिन्न एम.टी. प्रणालियों की उपयुक्तता के क्रम में उनके अनूदित विकल्पों की फिल्टरिंग और क्रमबद्धता को जोड़ने की भूमिका निभाता है। यह पद्धति अनुवाद के समय प्रचुर मात्रा में शक्ति, लचीलापन और नियंत्रण प्रदान करती है, क्योंकि एक अनुवाद प्रणाली के असफल होने पर दूसरा विकल्प कार्य में आ जाता है।

विभिन्न मशीनी अनुवाद पद्धतियों के नजदीकी परीक्षण से प्रत्येक पद्धति में आने वाली विभिन्न पेंचीदगियों का खुलासा हुआ तथापि कुछ और व्यावहारिक मुद्दे मशीनी अनुवाद को अत्यधिक चुनौतीपूर्ण बना देते हैं।

14.7 मशीनी अनुवाद के अन्य मुद्दे

इस प्रकार हमने देखा कि मशीनी अनुवाद प्राकृतिक भाषा संसाधन (NLP) का एक पूर्ण एवं अत्यंत चुनौतीपूर्ण अनुप्रयोगिक क्षेत्र है। अतः प्राकृतिक भाषा संसाधन से जुड़े सभी मुद्दे मशीनी अनुवाद के भी मुद्दे हैं। इसके अतिरिक्त भाषा, भाषाविज्ञान एवं अनुवाद अध्ययन के मुद्दे भी इसमें शामिल हैं। प्राकृतिक भाषा संसाधन का उद्देश्य कंप्यूटर को मानव भाषा में अच्छी तरह निपुण करना है ताकि मानव उसका उपयोग अपनी-अपनी भाषाओं में महत्वपूर्ण कार्यों के निष्पादन के लिए कर सके। मशीनी अनुवाद इन महत्वपूर्ण कार्यों में से कंप्यूटर के लिए सबसे जटिल कार्य है, क्योंकि इसमें भाषा संसाधन से जुड़े विभिन्न पक्ष जैसे भाषा-विश्लेषण, भाषा-सृजन, भाषा-समझ, भाषा-प्रयोग, इत्यादि सभी पहलू सम्मिलित हैं। कंप्यूटर को प्रारंभ से मूलतः संख्यात्मक आँकड़ों के विशाल संग्रहण एवं सटीक संसाधन के लिए जाना जाता है। अतः मानव भाषा संसाधन कंप्यूटर के लिए एक पूरी तरह से गैर-संख्यात्मक कार्य है। इस तरह हम देखते हैं कि कंप्यूटर द्वारा मानव भाषा का संसाधन कंप्यूटर के अन्य कार्यों से एक बहुत ही भिन्न और जटिल प्रक्रिया है। भाषा दार्शनिकों, भाषाविज्ञानियों, वैयाकरणों और विज्ञानियों द्वारा मानव भाषा को संपूर्ण रूप से समझने और नियमबद्ध करने की कोशिश अभी जारी है और भाषा को ठीक से समझ पाना और इसके सभी तथ्यों को मशीन के लिए सुव्यवस्थित और नियमबद्ध करना एक मुश्किल कार्य है। मानव

किस प्रकार भाषा का प्रयोग करने में समर्थ है, यह अभी भी पूर्ण स्पष्ट नहीं है। हाल के दशकों में भाषिक और व्याकरणिक सिद्धांतों के क्षेत्र में महत्वपूर्ण कार्य हुए हैं जिससे इस दिशा में उन्नति हुई है और अनेक आवश्यक शोधकार्य चल रहे हैं। आज भी किसी भाषा-युग्म के लिए विवरणात्मक रूप से पूर्ण व्याकरण का अभाव है जो किसी भाषा विशेष के सभी नियमों को स्पष्ट रूप से प्रस्तुत करने का दावा कर सके। इस अर्थ में, और खासकर जो मशीन में प्रयोग के लिए उपयुक्त हो ऐसे व्याकरण सभी भाषाओं के लिए उपलब्ध होना जरूरी है। मानव भाषाओं में निहित जटिलता के अनेकानेक स्रोत हैं; मानव भाषा सदैव परिवर्तनशील है, समय और स्थान के परिवर्तन के साथ भाषा परिवर्तित होती जाती है, साथ ही अनेकानेक सामाजिक-सांस्कृतिक एवं अन्य प्रायोगिक विविधताओं के कारण भाषा के संरचनात्मक एवं अर्थ स्वरूपों में भिन्नताएँ देखने को मिलती हैं। इन सब कारणों से किसी भी मानव भाषा को एकरूप देना और उसका स्पष्ट और सटीक व्याकरण (खासकर संरचना एवं अर्थ/तात्पर्य निर्धारण के संदर्भ में) प्राप्त करना अत्यंत कठिन होता है।

वस्तुतः मशीनी अनुवाद एक अन्य चुनौती भी प्रस्तुत करता है। इसका संबंध संगणक तकनीक से हैं—जैसे इसकी प्रणाली के प्रारूप (सिस्टम डिजाइन) से जुड़े मुद्दे, यंत्र विकास, उनकी जटिलताएँ और परिणाम और अन्य संबंधित मुद्दे। आरंभ से मशीनी अनुवाद प्रणाली पद्धति की संरचना एवं उसके विकास की दिशा, कृत्रिम बुद्धि, सूचना तकनीक, ज्ञान निरूपण और अन्य संबंधित तकनीकों के क्षेत्र में अनुसंधान और विकास से प्रभावित रही हैं। मशीनी अनुवाद प्रणाली एवं कार्य को बेहतर और अधिक प्रभावकारी बनाने के लिए विभिन्न पद्धतियाँ एवं तकनीकें अपनाई गईं मशीनी अनुवाद प्रणाली विकास की अंतरभाषिक और अंतरण पद्धतियाँ पारंपरिक प्रत्यक्ष पद्धति से अधिक उपयुक्त एवं प्रभावकारी पाई गई हैं और इनका अनेक भाषा-युग्मों के मशीन अनुवाद प्रणाली विकास के लिए प्रयोग किया गया (जैसे भारतीय भाषा-युग्मों के संदर्भ में आंग्लभारती एवं अनुसारक मशीनी अनुवाद प्रणालियाँ)। दूसरे स्तर पर, संगणक तकनीक में विकास का एक महत्वपूर्ण और स्पष्ट परिणाम मशीन अनुवाद के ज्ञान-आधारित/भाषिक-ज्ञान आधारित पद्धतियों के एक प्रभावकारी विकल्प के रूप में सांख्यिकीय या आँकड़ा-आधारित प्रणालियों के विकास में दृष्टिगोचर होता है। परिणामस्वरूप, हाल के दशकों में सांख्यिकीय या आँकड़ा-आधारित मानव भाषा संसाधन पद्धतियाँ मशीनी अनुवाद के विकास को एक नई दिशा दे रही हैं जिसमें कई तरह के प्रयोग किए जा रहे हैं; जैसे-मशीनी अनुवाद की मिश्रित-पद्धति (हाइब्रिड एप्रोच) पारंपरिक ज्ञान-आधारित एवं सांख्यिकीय-आधारित पद्धतियों में सामंजस्य लाने का प्रयास है। आगे हम इन मुद्दों पर कुछ विस्तार से चर्चा करेंगे।

मशीनी अनुवाद से जुड़े मुद्दे जैसे समाज-सांस्कृतिक, आर्थिक, राजनीतिक, दार्शनिक और नैतिक प्रश्नों से संबंधित हैं। अलग-अलग क्षेत्रों से संबंधित इस तरह के सभी वैज्ञानिक अनुसंधानों, खोजों एवं परिवर्तनों से जुड़े होते हैं जो मानव के अंतरंग एवं जटिल समझी जाने वाली मानवीय गतिविधियों को 'मानव से स्वतंत्र' करने का प्रयास करते हैं जिससे मानव का अस्तित्व खतरे में प्रतीत होता है या उसकी सर्वोच्चता डगमगाती प्रतीत होती है। मशीनीकरण का रोजगार के लिए खतरा, मानव अनुवादकों को मशीन के हाथों में 'उपकरण' (अधीनस्थ) बनाना, रचनात्मकता, प्रवीणता और उत्पादकता के मानकों को पुनःपारिभाषित एवं पुनःरेखांकित करना आदि, कुछ ऐसे संवेदनशील मुद्दे हैं जिन्होंने विश्वभर में किसी न किसी रूप में मशीनी अनुवाद अनुसंधान, विकास और कार्यान्वयन को कुछ सीमा तक बाधित किया है। हालांकि, ये मुद्दे मशीनी अनुवाद के शैक्षिक पहलुओं से प्रत्यक्ष रूप से जुड़े नहीं लगते, तथापि इन्होंने मानव समाज में विचारों के रुख को जरूर प्रभावित किया है। फलस्वरूप, मशीन अनुवाद के विकास के लिए यथोचित शैक्षणिक एवं आर्थिक समर्थन का परिवेश बनाने की गति को प्रभावित किया है। ये मुद्दे इस प्रकार हैं :

14.7.1 भाषा संसाधन

मशीनी अनुवाद में भाषा मूल पहलू है जिसके मशीन परिवेश में सफल संसाधन पर मशीनी अनुवाद की सफलता प्रत्यक्ष रूप से निर्भर करती है। मानव में भाषा संरचना एवं समझ का ज्ञान जन्मजात गुण के रूप में निहित है और अपनी भाषा प्रयोग में हम इसका सहारा लेते हैं। मानव का अनुभव-संसार भी विविध और अथाह है जिससे भाषा प्रयोग में हमें सचेतन या अवचेतन ज्ञान-आधार; भाषिक और संदर्भात्मक (वैश्विक-ज्ञान) रूपों में सहयोग मिलता है किंतु इन ज्ञान आधारों की मूल प्रकृति को पूर्णतः सुलझाने एवं नियमबद्ध करने में भाषाविज्ञानियों और

संप्रेषण-सैद्धांतिकों को अंश मात्र ही सफलता मिली है। खासकर, भाषा-प्रयोग के संदर्भ में अर्थ की व्याख्याओं में निर्णायक भूमिका निभाने वाले वैश्विक ज्ञान (वर्ल्ड नॉलेज) की प्रकृति को समझना एवं नियमबद्ध करना मानव भाषा संसाधन एवं मशीनी अनुवाद अनुसंधानकर्ताओं और विकासकर्ताओं के लिए अभी भी एक बड़ी चुनौती है। सरल शब्दों में कहें तो भाषा ध्वनि-अर्थ युग्मन है। भाषिक ध्वनि की प्रकृति, अर्थ और महत्वपूर्ण रूप से उनके युग्मन की प्रकृति को लेकर भाषाविज्ञान एवं भाषा-दर्शन के इतिहास में शोध के रूप में रहा है। प्रजनक-सृजनात्मक (जेनेरेटिव) भाषाविज्ञान, मनोभाषाविज्ञान, तंत्रिका भाषाविज्ञान और संज्ञानात्मक भाषाविज्ञान के क्षेत्रों में आधुनिक शोध परिणामों से प्राप्त मानव भाषा की सूक्ष्म समझ ने भाषा संसाधन और मशीनी अनुवाद पद्धतियों के विकास में महत्वपूर्ण योगदान किया है।

भाषा क्या है, भाषा-ज्ञान का स्वरूप कैसा है, इसकी बनावट कैसी है, मानव इसका प्रयोग अर्थ/तात्पर्य के कोडीकरण और विकोडीकरण के लिए कैसे करता है, इत्यादि रोचक प्रश्न और इनके उत्तर ढूँढने के सतत प्रयास प्राकृतिक भाषा संसाधन और मशीनी अनुवाद प्रणालियों के विकास के केंद्र-बिंदु हैं। कोई व्यक्ति उस ज्ञान से कम ही परिचित होता है जिसे अपनी भाषा का एक निपुण प्रयोगकर्ता बनाता है। इस ज्ञान का अध्ययन भाषाविज्ञान में होता है जहाँ मानव भाषा के सभी पहलुओं, जैसे, भाषा-ज्ञान का स्वरूप, इसकी संरचना, प्रयोग-विधि इत्यादि का अध्ययन करता है। भाषा-ज्ञान का स्वरूप एक सुव्यवस्थित तंत्र (जैसा) है। यह ज्ञान ध्वनि, शब्द, वाक्य और अर्थ के स्तर पर उनके सुव्यवस्थित स्वरूप से संबंधित है जहाँ उनके संरचनात्मक व्यवस्था, अर्थ एवं उनके प्रयोगों के वैज्ञानिक अध्ययन के आधार पर उनके सिद्धांत एवं नियम निर्धारित होते हैं। सरल शब्दों में, भाषा-प्रयोग का ज्ञान तीन स्तरों पर संगठित है-शब्दकोश (लेक्सिकोन), वाक्यविज्ञान (सिंटैक्स) और वैश्विक-ज्ञान (एम्पिरिकल-नॉलेज)। पिछले कुछ दशकों में, शब्दकोश के स्वरूप एवं विभिन्न पहलुओं पर सैद्धांतिक, अनुप्रयुक्त और संगणक भाषाविज्ञान की दृष्टि से काफी शोध हुए हैं। आज शब्दों के अर्थ एवं पर्याय, उनके संचयन और प्रयोग-पद्धति, ई-शब्दकोश, बहुप्रकार्य शब्द, एवं बहु-शब्द अभिव्यक्ति, इत्यादि महत्वपूर्ण मुद्दों पर भाषा के मशीन संसाधन के दृष्टिकोण से अनेक उपयोग सिद्धांत एवं उपकरण हैं। वर्डनेट, फ्रेमनेट जैसी उपयोगी अवधारणाओं से, खासकर मशीन अनुप्रयोग के लिए, हमारी शब्द एवं उनकी व्यवस्था की समझ समृद्ध हुई है और उन्हें नियमबद्ध करने में काफी सफलता मिली है (फैलबाउम 1998)।

अर्थ निर्धारण और वैश्विक ज्ञान को नियमबद्ध करने की दिशा में उतनी उन्नति नहीं हो पाई है जितनी भाषा संप्रेषण के गूढ़ प्रश्नों को समझने एवं मशीनी अनुवाद प्रणाली की इनसे संबंधित जटिलताओं को दूर करने के लिए आवश्यक है। इस कारण, अर्थ-संदिग्धता एवं उसका निराकरण आज तक मशीनी अनुवाद की एक सबसे बड़ी चुनौती बनी हुई है। अर्थ-संदिग्धता मानव भाषा की एक स्वाभाविक विशेषता है जो सृजनशील लेखकों एवं अन्य भाषा प्रयोगकर्ताओं के लिए एक रूप में वरदान है तो मानव भाषा संसाधन एवं मशीनी अनुवाद शोधकर्ताओं के लिए दूसरे रूप में एक अभिशाप। मानव भाषा हरेक स्तर पर बहुअर्थक होती है और इस बहुअर्थकता की जटिलता विभिन्न रूपों में निहित होती है। मुश्किल से कोई ऐसा शब्द हो जो बहुअर्थक नहीं है। एक शब्द के कई अर्थ, प्रयोग, व्याख्याएँ, आशय, तात्पर्य या वर्गीय/व्याकरणिक कोटियाँ हो सकती हैं। मशीन संसाधन के संदर्भ में, भाषा का अर्थगत विश्लेषण, विशेषकर अर्थ-आषय संबंध कठिन चुनौतियों में से एक रहा है (केय और बोस्टेड 1996)। एक सामान्य अंग्रेजी वाक्य 'Is it alright' का अपना एक सरल अर्थ है किंतु इसकी व्याख्या विभिन्न संदर्भों में विभिन्न प्रकार से ही जा सकती है; जैसे 'is it alright?' (जो मैंने कहा क्या वह तुम्हारी समझ में आया?), is it alright? (क्या तुम इस प्रस्ताव से सहमत हो?), is it alright? (क्या अब हम समाप्त करें?) इत्यादि। इन विभिन्न व्याख्याओं की प्रकृति को निर्धारित करने के लिए भाषा-कारकों सहित अनेक कारक कार्य करते हैं। शोधार्थियों के लिए अभी भी यह एक चुनौती है कि कैसे इस अर्थ-आशय ज्ञान को मशीन संसाधन के लिए उपलब्ध कराया जाए। प्रायः उद्धृत एक सरल-सा दिखने वाला वाक्य 'time flies like an arrow' के व्याकरण नियमों के आधार पर दर्जनों आशय दिए जा सकते हैं। इसी प्रकार, एक सरल-सा दिखने वाला हिंदी वाक्य 'वह आम खाता है' का अनुवाद मात्र हिंदी व्याकरण नियमों से नहीं किया जा सकता है। एक मशीनी अनुवाद प्रणाली इस वाक्य के व्याकरणिक रूप से अनेक उपयुक्त अनुवाद दे सकता है। आश्चर्यजनक रूप से, इस हिंदी वाक्य के व्याकरणिक रूप से उपयुक्त अंग्रेजी अनुवादों में से कोई-एक बेतुका लग सकता है, जैसे : 'that is general account' (वह -

> that, आम ->general, खाता ->account, है -> is) (सिन्हा 2004)। यहाँ भाषा नियम एवं भाषा प्रयोग के बीच अस्पष्ट संबंध दिखाई दे रहा है; जिसमें कुछ भाषा प्रयोग व्याकरण नियमों से नियमबद्ध तो होते हैं किंतु व्यावहारिक दृष्टिकोण से अग्राह्य होते हैं, भाषाविदों के लिए एक चुनौती रहा है। यद्यपि, भाषा-सिद्धांतों से मानव भाषाओं की समानताओं को समझना और उन्हें नियमबद्ध करना सरल हुआ है किंतु हर-एक भाषा की अपनी विशिष्टता होती है जिसका अध्ययन उस भाषा एवं उसके अनुवाद-युग्म भाषा के संदर्भ में करने की आवश्यकता है। स्पष्ट रूप से, भिन्न-भिन्न भाषाएँ, अपनी-अपनी सांस्कृतिक, सामाजिक, इत्यादि परिवेश के कारण शाब्दिक स्तर पर भिन्न होती हैं। भाषाएँ अपनी संरचना में भी बहुत-सी भिन्नताएँ रखती हैं। मशीनी अनुवाद प्रणाली विकास के लिए इन भाषा-विशेष मुद्दों को समझना एवं इनका निराकरण करना आवश्यक है (सिन्हा एवं ठाकुर 2008)। इस तरह हम स्पष्ट देखते हैं कि मशीन अनुवाद की भाषा-संबंधित जटिलता इसके विकास के लिए एक बड़ा मुद्दा है, जिसका निराकरण किए बिना इस दिशा में अपेक्षित उन्नति नहीं की जा सकती है। इस दिशा में एक प्रभावकारी विकल्प मानव भाषा संसाधन की सांख्यिकीय पद्धति है जिसकी अब हम चर्चा करेंगे।

भाषा सुव्यवस्थित एवं नियमबद्ध है पर उतनी ही जटिल भी है। भाषा की जटिलता को समझने एवं इसके नियमबद्ध निराकरण की दिशा में भाषावैज्ञानिक शोध की धीमी रफ्तार के कारण प्राकृतिक भाषा संसाधन और मशीनी अनुवाद प्रणाली विकास से जुड़े विज्ञानियों को ज्ञान आधारित पद्धति के विकल्पों की खोज के लिए प्रेरित किया। खासकर, पिछली शताब्दी के बाद के दशकों से अब तक इस दिशा में काफी प्रगति हुई है जिससे मशीन अनुवाद में नई दिशाएँ एवं पद्धतियाँ आई हैं। मशीनी अनुवाद प्रणाली विकास से जुड़े विज्ञानियों को जब यह महसूस हुआ कि इसके लिए सुस्पष्ट व्याकरणिक नियम सुगमता और शीघ्रता से निकट भविष्य में आने वाली नहीं है तो उनकी पूरी क्षमता सांख्यिकीय मशीनी अनुवाद पद्धति के विकास में लग गई इसमें उस समय हो रहे कंप्यूटर तकनीकी में विकास का भी बड़ा योगदान था जिससे कंप्यूटर की संग्रहण एवं संसाधन शक्ति में पर्याप्त वृद्धि हो चुकी थी। इससे मशीनी अनुवाद अनुसंधान और तकनीकी विकास में प्रभावकारी परिवर्तन आया। इस पद्धति का मूल उद्देश्य भाषा के वास्तविक प्रयोग के उदाहरणों के बड़े आँकड़ों से सांख्यिकीय पद्धतियों के जरिए भाषा ज्ञान प्राप्त करना एवं उसका उपयोग भाषा संसाधन के जरिए भाषा-ज्ञान प्राप्त करना एवं उसका उपयोग भाषा संसाधन में करना है। इस पद्धति से कुछ शुरुआती सफलताएँ मिली और इस पर आधारित कुछ मशीनी अनुवाद प्रणाली (जैसे आई.बी.एम. की) विकसित और (व्यापारिक उद्देश्य के लिए भी) प्रयोग की जाने लगी (सोमर्स 2003)। यह जागरूकता आने पर कि पूर्णतः स्वचालित मशीनी अनुवाद प्रणाली आने में लंबा समय लग सकता है, HAMT (मानव साधित मशीन अनुवाद) और MAHT (मशीन साधित मानव अनुवाद) या CAT (कंप्यूटर साधित अनुवाद) के रूप में अंतर्क्रियात्मक (इंटरैक्टिव) मशीनी अनुवाद प्रणाली के विकास की दिशा में प्रयास किए जाने लगे। इस रूप में अनुवादकों के लिए अनुवाद सहायक उपकरण, अनुवाद स्मृति, ई-साधन और क्षेत्र-विशेष से संबंधित (domain specific) मशीनी अनुवाद प्रणालियाँ विकसित की जा रही हैं। बाद में, कुछ वैज्ञानिक मशीनी अनुवाद की अंशतः नियम-आधारित और अंशतः आँकड़ा-आधारित एक मिश्रित (हाइब्रिड) तकनीक की अवधारणा से भी प्रेरित हुए और इसके आधार पर मशीनी अनुवाद प्रणाली विकसित की जाने लगी है (सिन्हा 2004)।

14.7.2 प्रणाली युक्ति

प्रणाली युक्ति (सिस्टम-डिजाइन) मशीनी अनुवाद विकास में एक अन्य बड़ा मुद्दा है। मशीनी अनुवाद विकास के प्रणाली युक्ति (सिस्टम-डिजाइन) के मुद्दे मुख्यरूप से अभियांत्रिकी एवं प्रौद्योगिकी से जुड़े मुद्दे हैं। इनमें, सीमित ज्ञान एवं साधन से अधिकतम एवं सर्वोत्तम परिणाम प्राप्त करने की विधि, पूर्णरूप से स्वचालित प्रभावकारी मशीनी अनुवाद प्रणाली के अभाव में एक सामान्य रूप से उपयोगी मशीनी अनुवाद प्रणाली विकसित करने की विधि, इत्यादि जैसे कुछ महत्वपूर्ण मुद्दे आते हैं। एक व्यावहारिक मशीनी अनुवाद प्रणाली विकसित करने में विभिन्न तकनीकें (तरकीब) लगानी पड़ती है, जिनमें कई तरह के व्यवहार कौशल, जोड़तोड़ (मेनीपुलेशन) और समझौते शामिल होते हैं, जो इस पूरी प्रक्रिया को अभियांत्रिकी के विषय एवं मुद्दों से जोड़ते हैं (सिन्हा 2004)। स्पष्टतः, मशीनी अनुवाद प्रणाली का स्वरूप एवं डिजाइन इसके अनुप्रायोगिक उद्देश्य एवं परिवेश पर निर्भर करता है। मशीनी अनुवाद के अनुप्रायोगिक परिवेश पूर्णतः स्वचालित प्रणाली से लेकर, अर्द्ध-स्वचालित या अंशतः-स्वचालित प्रणाली तक हो सकते हैं। यद्यपि, मशीनी अनुवाद हमारे मन में एक ऐसी अनुवाद करने वाली मशीन को चित्रित करता

है जो बिना किसी मानवीय हस्तक्षेप/सहायता के एक भाषा से दूसरी भाषा में अपने आप अनुवाद कर दे। किंतु वास्तव में, वर्तमान में, इस तरह की कोई भी मशीन या उपकरण संसार के किसी भी कोने में उपलब्ध नहीं हैं जो पूर्णतः स्वचालित हो और बढ़िया अनुवाद देती हो। आज ऐसे उपकरण तो उपलब्ध हैं जो पूर्णतः स्वचालित अनुवाद करने का प्रयास करते हैं किंतु उससे प्राप्त अनुवाद अधिक अच्छा और संतोषजनक हो, ऐसा निश्चित रूप से नहीं कहा जा सकता है।

मशीन-साधित मानव अनुवाद (एम.ए.एच.टी.) में मानव की बड़ी भूमिका है जबकि मशीन, अनुवाद स्मृति, ई-स्रोतों/साधनों, आदि के रूप में सहायता उपलब्ध कराने वाले सहायक उपकरण की भूमिका निभाता है। इन अनुवाद उपकरणों (ट्रांसलेशन टूल्स) का उपयोग (मानव) अनुवादक अपने अनुवाद-कार्य शीघ्र और अधिक प्रभावी रूप में निष्पादित करने के लिए खासकर उन स्थितियों में करता है जहाँ अनुवाद कार्य मुख्यरूप से गैर-साहित्यिक उद्देश्य और बड़े पैमाने पर किया जाता है। इसके बिल्कुल विपरीत संकल्पना मानव-साधित मशीनी अनुवाद (एच.ए.एम.टी.) है जिसमें मशीन या कंप्यूटर सॉफ्टवेयर को बड़ी भूमिका निभाना अथवा अनुवाद का अधिकतम कार्य करना होता है। मानव हस्तक्षेप/सहयोग की आवश्यकता केवल पश्च-संसाधन (पोस्ट-प्रोसेसिंग) अथवा संपादन (एडिटिंग) में होती है। (मानव) अनुवादक या अनुवाद उपभोक्ता/प्रयोगकर्ता मशीनी अनुवाद प्रणाली द्वारा अनुवादित पाठ की त्रुटियों को सुधारकर उसे अधिक उपयुक्त बनाता है। 'अंतरक्रियापरक' मशीनी अनुवाद प्रणाली एक अन्य अवधारणा है जिसमें अनुवाद उपकरण स्वतः अनुवाद तो करता है, किंतु इसे विभिन्न चरणों पर अलग-अलग रूप में मानव हस्तक्षेप की आवश्यकता होती है। मानव हस्तक्षेप स्रोत पाठों के पश्च-संपादन, सविस्तार व्याख्यान करने, या फिर पश्च-संसाधन (प्रोस्ट-प्रोसेसिंग) कर अनुवादित पाठ को सुधारने, आदि के रूप में होते हैं जिससे अनुवाद उपकरण भाषा की जटिलताओं के समाधान की कठिनाइयों से बच जाता है। हम देखते हैं कि मशीन अनुवाद प्रणालियों के डिजाइन की पद्धति उसके प्रयोग उद्देश्य से निर्धारित होती है। उदाहरण के लिए, यदि अनुवाद का मुख्य उद्देश्य बहुत सारे गैर-साहित्यिक पाठों का कम लागत में शीघ्र और कामचलाऊ अनुवाद प्राप्त करना है तो अनुवाद्य पाठों की गुणवत्ता बड़ी अहम् मुद्दा नहीं होगी। ऐसे अनुवाद प्रायः स्रोतपाठों के सतही सार समझने के लिए होते हैं। अनुवाद उपभोक्ता/प्रयोगकर्ता की रुचि संपूर्ण स्रोतपाठ के उत्तम अनुवाद में हो भी सकती है और नहीं भी हो सकती है अथवा उन्हें स्रोतपाठ का कुछ भाग अपने तात्कालिक उद्देश्य के लिए रुचिकर या उपयोगी लग सकता है। ऐसे बहुत से मशीनी अनुवाद उपकरण आज उपलब्ध हैं जो हमें इस प्रकार की सुविधा प्रदान करते हैं (उदाहरणतः, ऑनलाइन अनुवाद उपकरण, गूगल ट्रांसलेशन, आंग्लभारती, अनुसारक, आदि)। फिर, अनुवाद का एक महत्वपूर्ण उद्देश्य क्षेत्र-विशेष से संबंधित पाठों को अनूदित करना होता है। क्षेत्र-विशेष पाठों के अनुवाद के लिए मशीनी अनुवाद प्रणालियों का डिजाइन सामान्य उद्देश्य वाली मशीनी अनुवाद प्रणालियों से बिल्कुल भिन्न होता है। और स्पष्टतः, चूँकि भाषा संबंधी जटिलताएँ क्षेत्र-विशेष के स्रोतपाठों में न्यूनतम होती हैं। इस कारण क्षेत्र-विशेष के लिए उपलब्ध मशीनी अनुवाद प्रणालियाँ अधिक सफल एवं प्रभावकारी पाई जाती हैं।

एक सामान्य व्यक्ति इस बात को लेकर उत्सुक होता है कि अनुवाद कार्य जो मानव के लिए भी एक कठिन प्रक्रिया है, एक मशीन (कंप्यूटर) कैसे कर लेती है। मशीन (कंप्यूटर सॉफ्टवेयर) एक कृत्रिम साधन है जिसमें स्वाभावतः न तो बुद्धि-विवेक और न ही भाषा निहित है। भाषा-अनुवाद परंपरागत रूप से एक मानव प्रक्रिया है जिसे प्रशिक्षित विशेषज्ञ ही कर पाते हैं। एक अच्छे अनुवादक के लिए स्रोत एवं लक्ष्य भाषाओं में न केवल दक्षता की आवश्यकता होती है अपितु इन दोनों भाषाओं के सामाजिक-सांस्कृतिक, इत्यादि पहलुओं से भी भलीभाँति अवगत होने की आवश्यकता होती है। अतः तार्किक रूप से, एक अच्छी मशीनी अनुवाद प्रणाली को भी अनुवाद भाषा-युग्मों (Translation Language Pair) के भाषायी एवं संदर्भिक ज्ञान से परिपूर्ण होना चाहिए। अनुवाद के लिए स्रोतभाषा पाठों का विश्लेषण एवं समझ और लक्ष्यभाषा में उनके समतुल्य पाठों की प्राप्ति जैसी ज्ञान-साधित प्रक्रियाएँ आवश्यक रूप से सम्मिलित हैं। यह ज्ञान-आधार (knowledge base) मूलतः (स्रोतभाषा व लक्ष्यभाषा दोनों के) शब्दार्थ (कोशीय-ज्ञान) और व्याकरण की ही नहीं अपितु अर्थ-संदर्भ ज्ञान और वैश्विक-ज्ञान की समग्रता है। अतः भाषा एवं अन्य प्रासंगिक ज्ञान-आधार उसमें कृत्रिम रूप से विकसित किया जाता है। दूसरे शब्दों में, कंप्यूटर को मानव भाषा के ज्ञान एवं प्रयोग की शिक्षा दी जाती है। भाषा-ज्ञान बुद्धि-विवेक से जुड़ा मुद्दा है। इस प्रकार मानव भाषा संसाधन एवं इसके मशीनी अनुवाद जैसे अनुप्रयोग के विषय आवश्यक रूप से अंतर्विषयक एवं बहु-विषयक

(Inter-disciplinary & multi-disciplinary) शोध का विषय है। एक कृत्रिम उपकरण कंप्यूटर में मानव भाषा का ज्ञान स्थापित करना, कृत्रिम-बुद्धि का एक अनोखा एवं जटिल अनुप्रयोग है। हाल के दशकों में इस पद्धति में विभिन्न सुधार हुए हैं और मशीन शिक्षण (machine learning) जैसे तकनीक के अनुप्रयोग से मशीनी अनुवाद प्रणाली विकास में एक नई दिशा का प्रारंभ हुआ है।

अब हम अनुवाद यंत्र की संकल्पना पर विचार कर सकते हैं। एक मशीनी अनुवाद प्रणाली वास्तव में कैसे कार्य करती है, यह समझने के लिए मशीनी अनुवाद प्रणाली के भीतर क्या और कैसे होता है अर्थात् उसकी आंतरिक संरचना कैसी होती है, यह किस ज्ञान-पृष्ठभूमि में कार्य करती है और यह ज्ञान एक मशीन परिवेश में कैसे निरूपित होता है, इत्यादि तथ्यों की समझ की आवश्यकता है। अनुवाद यंत्र संरचना से जुड़े इन मुद्दों को समझने के लिए मशीन अभियांत्रिकी एवं भाषाविज्ञान के कुछ आधारभूत तथ्यों एवं शब्दावलियों को समझना आवश्यक है। मशीनी अनुवाद प्रक्रिया के लिए आवश्यक ज्ञान-आधार (नॉलेज-बेस) के मशीन (कंप्यूटर) में निरूपण की युक्ति-संरचना (Design Architecture) अनुवाद यंत्र (Translation Engine) के नाम से जाना जाता है और मशीनी अनुवाद प्रौद्योगिकी का एक अहम् पहलू है। मशीनी अनुवाद प्रणाली की मुख्यतः तीन यंत्र-युक्तियाँ हैं जिन्हें तीन पद्धतियों के नाम से जाना जाता है : प्रत्यक्ष पद्धति, अंतरभाषिक पद्धति और अंतरण पद्धति। प्रत्यक्ष पद्धति सबसे पुरानी है और इसे मशीन के लिए सबसे सरल संरचना-यंत्र माना जाता है। यह पद्धति आज भी व्यवसायिक रूप से उपलब्ध कुछ मशीनी अनुवाद प्रणालियों में यह प्रणाली विशेष रूप से एक खास भाषा-युग्म के लिए उपयुक्त मानी जाती है। इसमें प्रणाली विकास के लिए अनुवाद-यंत्र को केवल उस सीमा तक स्रोत एवं लक्ष्यभाषा का ज्ञान-साधन (कोश-ज्ञान एवं व्याकरण-ज्ञान) दिया जाता है जितना अर्थ-संदिग्धता के निराकरण और सही समतुल्य पाठ प्राप्त करने के लिए आवश्यक है। अतः, स्रोतभाषा व लक्ष्यभाषा के बीच जितनी निकटता होती है, अनुवाद-प्रणाली की डिज़ाइन संरचना उतना ही सरल होती है।

अंतरभाषिक मशीनी अनुवाद-यंत्र द्विचरण पद्धति (Binary System) पर आधारित होता है। प्रथम चरण में, स्रोतभाषा पाठ को एक कृत्रिम अंतरभाषा में निरूपित किया जाता है और द्वितीय चरण में अंतरभाषिक निरूपण को लक्ष्यभाषा पाठ में प्राप्त किया जाता है। यह यंत्र-पद्धति एक स्रोतभाषा से अनेक लक्ष्य भाषाओं में एक साथ अनुवाद के लिए उपयुक्त मानी जाती है, खासकर जब लक्ष्य भाषाओं में एकरूपता एवं समानताएँ पाई जाती है। उदाहरण के लिए, अंग्रेजी से भारतीय भाषाओं में अनुवाद के लिए मशीनी अनुवाद प्रणाली के विकास में इस यंत्र-पद्धति को प्राथमिकता दी जा सकती है। अंतरभाषा कोई मानव भाषा नहीं होती बल्कि एक प्रकार की मशीन-निरूपित भाषा-संरचना होती है जो स्रोतभाषा और लक्ष्यभाषा दोनों की अर्थगत और संरचनात्मक समानताओं पर आधारित होती है। इस पद्धति में स्रोतभाषा से लक्ष्यभाषा तक एक चरण में पहुँचने के बजाय एक से अधिक चरणों में पहुँचना अधिक सरल है।

मशीनी अनुवाद-यंत्र विकास की तीसरी पद्धति अंतरण पद्धति है जो समृद्ध भाषावैज्ञानिक ज्ञान साधनों पर आधारित है। इस यंत्र-संरचना में अनुवाद प्रक्रिया के तीन चरण होते हैं। प्रथम चरण में स्रोतभाषा का विश्लेषण एवं प्रतिनिधित्व संरचना का निर्धारण किया जाता है। द्वितीय चरण अंतरण (ट्रांसफर) चरण है जो प्रथम चरण से प्राप्त स्रोतभाषा की प्रतिनिधि संरचना एवं लक्ष्यभाषा की संरचना की जानकारी के आधार पर लक्ष्यभाषा की प्रतिनिधि संरचना का निर्धारण करता है, और इस प्रकार तृतीय चरण में, द्वितीय चरण से प्राप्त परिणाम (आउटपुट) को लक्ष्यभाषा में अनूदित किया जाता है। अंतरण प्रणाली यंत्र सबसे पहले जापान में मशीनी अनुवाद प्रणाली विकास के लिए प्रयुक्त हुए जहाँ इसके अच्छे परिणाम भी प्राप्त हुए (केय और बोयटेट 1996)। इन प्रणालियों पर हम पहले भी इकाई 11 और 12 में चर्चा कर चुके हैं।

14.7.3 अनुप्रयोग

अनुवाद की बड़ी पुरानी एवं समृद्ध परंपरा है। यह अपने किसी न किसी रूप में मानव समाज की बहुभाषिकता से जुड़ी है और इसकी जड़े उतनी ही गहरी हैं जितनी मानव सभ्यता में बहुभाषिकता की। अनुवाद की उपयोगिता हर उस जगह है जहाँ भाषा की उपयोगिता है और जैसे-जैसे बहुभाषिकता तकनीकी विकास एवं अन्य आधुनिक विकास कारणों से मानव समाज में फैलती जा रही है अनुवाद की आवश्यकता भी बढ़ती जा रही है। सूचना

तकनीकी और ज्ञान-क्रांति के इस आधुनिक युग में, अनुवाद भाषा-अवरोध (Language Barrier) से मुक्त करने का एक अत्यंत सशक्त साधन है। परंतु साथ ही, अनुवाद कार्य के इस विशाल और बढ़ते कार्य-बोझ को बिना तकनीकी हस्तक्षेप के पूरा नहीं किया जा सकता और इसी में, मशीनी अनुवाद का महत्व एवं इसकी उपयोगिता निहित है।

अनुवाद का कार्यक्षेत्र व्यापक है, जैसे, मानव की ज्ञान-जिज्ञासा, ज्ञान सृजन और विस्तार, शिक्षा शोध, विकास एवं विस्तार एवं लोकतंत्रीकरण, सामाजिक-सांस्कृतिक मेल, व्यवसाय तथा प्रशासन (ई-कॉमर्स, ई-गवर्नेंस), के लिए भाषिक बाधाओं को तोड़कर मानव को एक दूसरे से जोड़ने से लेकर भूमंडलीकरण एवं स्थानीकरण के मुद्दों को समझने, ज्ञान-उत्पादों प्राकृतिक और बौद्धिक संसाधनों के सुव्यवस्थीकरण करने, मानव-विकास और गतिशीलता में सहयोग करने, बहुभाषिक और बहुसांस्कृतिक समाजों में सामाजिक राजनैतिक, धार्मिक सामंजस्य बनाने, बेहतर स्वास्थ्य सेवाओं, यात्रा एवं पर्यटन क्षेत्रों में सहायक होने, सामाजिक एवं विश्व शांति और संपन्नता के लिए भाषा-अनुवाद एक अत्यंत महत्वपूर्ण साधन है। आधुनिक समय में तो अनुवाद उद्देश्यों की संख्या और उनकी आत्यंतिकता बढ़ती ही जा रही है और अनुवाद के लिए उपलब्ध पाठ-सामग्री का बड़ा भाग साहित्यिक भाषा-पाठ का नहीं है जिन्हें विशिष्ट रूप से प्रशिक्षित सृजनात्मक लेखकों अथवा अनुवादकों से ही अनूदित करने की आवश्यकता हो। पाठ्यपुस्तकें, सरकारी और व्यावसायिक रिपोर्ट, सरकारी पत्रा-व्यवहार सामग्री, तकनीकी दस्तावेज इत्यादि ऐसी भाषा-सामग्री है जिनमें भाषा प्रयोग की कम से कम अथवा नहीं के बराबर लाक्षणिक सृजनात्मकता होती है और शब्दावली एवं संरचना के स्तर पर ये एकरूप और सरल होते हैं। इन्हें अलग तरह से अनूदित करने की आवश्यकता होती है और वे मशीनी अनुवाद के लिए सरल भी होते हैं। इसी कारण, सामान्य उद्देश्य मशीनी अनुवाद प्रणाली की अपेक्षा क्षेत्र-विशेष मशीनी अनुवाद प्रणालियों के विकास के प्रयत्नों को बेहतर परिणाम मिला है। उदाहरण के लिए, क्षेत्र-विशेष की एक मशीनी अनुवाद प्रणाली टॉम-मैटो (TAUM-METEO) को कनाडा के क्यूबेक प्रांत में मौसम समाचार (अंग्रेजी-फ्रेंच) अनुवाद करने के लिए सफलतापूर्वक प्रयोग किया जा रहा है। इस सफल मशीनी अनुवाद प्रणाली के विकास के पीछे अवश्य ही नागरिकों की भाषिक आकांक्षाओं की पूर्ति, शासन-तंत्र को लोकतांत्रिक बनाने और सामाजिक विविधताओं के बीच शांति एवं मैत्री स्थापित करने की आवश्यकताएँ जैसी प्रेरक शक्तियाँ रही होंगी। विश्व की सभी भाषाओं के लिए मशीनी अनुवाद प्रणाली के विकास के लिए इस प्रकार की या और अधिक प्रभावकारी प्रेरणाओं का होना अत्यावश्यक है। भारतवर्ष में, जहाँ विभिन्न राज्यों के बीच, राज्यों और केंद्रीय शासन के बीच एवं एक ही राज्य के विभिन्न भागों के बीच विभिन्न स्तरों पर संप्रेषण के लिए कई भाषाएँ प्रयुक्त होती हैं, भारतीय भाषाओं के बीच अनुवाद कार्य की प्रचुर संभावनाएँ हैं और जिसे भारतीय भाषाओं के लिए मशीन अनुवाद प्रणाली विकास से बड़े पैमाने पर पूरा किया जा सकता है। राज्यों, मंत्रालयों और विभागों के बीच कार्यालयी पत्र-व्यवहार कई भाषाओं में किया जाता है। भारतीय संसद में भी अनेक भाषाएँ बहस एवं कार्यवाहियों और प्रलेखन में (केवल अनूसूचित भाषाएँ) प्रयुक्त होती हैं। इतने बड़े पैमाने पर अनुवाद कार्य को संपन्न करने के लिए बड़ी और मंहगी प्रशिक्षित मानव शक्ति (अनुवादकों) की आवश्यकता है। इस दिशा में मशीनी अनुवाद प्रणाली अपने विभिन्न रूपों जैसे, अनुवाद सहायक उपकरण (Translation Support Tools), मशीन अनुवाद साधनों (Resources) आदि के रूप में तकनीकी सहायता देकर अनुवाद के जटिल एवं बोझिल कार्य को बहुत सरल बना सकती है और बना रही है (जैसे, सी-डैक-पुणे की मंत्रा मशीन अनुवाद प्रणाली, कार्यालयी-पत्र-व्यवहार के संदर्भ में)। विशेष रूप से भारत जैसे बहुभाषिक देशों के विभिन्न क्षेत्रों में, जैसे कि स्वास्थ्य सेवाओं, यात्रा और पर्यटन, बहुराष्ट्रीय संगठनों, आदि में सेवा की गुणवत्ता बढ़ाने में मशीनी अनुवाद विकास की अच्छी संभावना है। कुछ विकसित राष्ट्रों में जहाँ बहुभाषिकता की समस्याएँ मौजूद हैं, वहाँ यह कार्य पहले ही शुरू हो चुका है। संयुक्त राष्ट्र कार्यवाहियाँ, विश्व मीडिया रिपोर्टिंग इत्यादि इसके उदाहरण हैं।

दूसरी ओर, आज विशेषकर उन इंटरनेट खोजकों के लिए जिन्हें वेब पेजों और अन्य ऑनलाइन सामग्रियों के अचिलंब अनुवाद की आवश्यकता होती है, उनकी अनुवाद जरूरतों को पूरा करने के लिए बहुभाषिक ज्ञान-खोज प्रणालियों की तुरंत आवश्यकता है। कई ऑनलाइन सर्च इंजिन मशीनी अनुवाद सहायता उपलब्ध कराकर अपने प्रयोगकर्ताओं (यूजर्स) की इन आवश्यकताओं को कुछ हद तक पूरा कर रहे हैं (जैसे गूगल ट्रांसलेट, इत्यादि)। अंग्रेजी एवं कुछ मुख्य यूरोपियन भाषाओं के अतिरिक्त सभी भाषाओं में, सभी देशों में इंटरनेट की उपलब्धता एवं

इसकी सेवा की गुणवत्ता बढ़ाने के लिए (जो निश्चित रूप से आवश्यक है) संसार की सभी भाषाओं में बहुभाषिक सर्च इंजन, सूचना निष्कर्षण (Information Extaction) एवं प्राप्ति (Retrieval) प्रणालियाँ विकसित करनी होंगी। इससे विभिन्न भाषाओं के बीच मशीनी अनुवाद प्रणाली विकास की दिशा में शोध की उपयोगिता को और महत्व एवं बल मिलेगा। आज बहुत से लोग इंटरनेट पर मुफ्त उपलब्ध मशीनी अनुवाद सेवाओं से परिचित हैं। इस ऑनलाइन अनुवाद सेवा में अधिक से अधिक भाषा युग्मों को शामिल करने के प्रयास चल रहे हैं। आज यह कार्य मुख्यरूप से बड़े पैमाने पर भूमंडलीकरण और स्थानीकृत उद्योगों (localized industry) की आवश्यकताओं से प्रेरित है। ऐसा संभव है कि जिन भाषाओं के लिए कंप्यूटर संसाधन तकनीकों जैसी कोई उपयोगी साधन नहीं विकसित किए गए हैं, वे वैश्विक ज्ञान समाज में अपना स्थान खो सकती हैं अथवा ऐसी भाषाओं का अस्तित्व समाप्त भी हो सकता है जिससे मानव जीवन की सांस्कृतिक विविधता एवं संपन्नता पर नकारात्मक प्रभाव पड़ सकता है। मानव समाज सदैव ही बहुभाषिक रहा है। भूमंडलीकरण के कारण सिमटते आधुनिक संसार की चकाचौध के बावजूद, स्थानीय पहचान की जन-आकांक्षाएँ मूल एवं स्थानीय संस्कृतियों और भाषाओं के विकास के लिए एक बड़ी शक्ति रही है। भाषा, संप्रेषण और सूचना एक दूसरे से जुड़े हैं। भाषा ज्ञान सृजन, संरक्षण एवं विस्तार का मूल साधन है। आज के बहुभाषिक एवं सूचना तकनीक-आधारित अंतरराष्ट्रीय आदान-प्रदान के संदर्भ में, ज्ञान सृजन एवं डिजिटलाइजेशन, ऑनलाइन संपदा निर्माण (Resource creation), अर्थव्यवस्था का भूमंडलीकरण व्यावसायिक प्रतियोगिता, राजनयिक संबंध, इत्यादि महत्वपूर्ण कार्य भाषा एवं संप्रेषण विधि के प्रभावकारी प्रयोग पर निर्भर हैं। यह अंतरराष्ट्रीय संदर्भ एक बहुसांस्कृतिक संदर्भ है। दुनिया के किसी भी कोने में, किसी भी भाषा में, किसी भी जानकारी अथवा ज्ञान की सरलता से उपलब्धता आधुनिक विष्व नागरिक अधिकारों में से एक है। इस कारण अनुवाद एक अत्यावश्यकता का कार्य बनता जा रहा है और विश्व की सभी भाषाओं से सभी अन्य भाषाओं में अनुवाद की आवश्यकता मानव अनुवादकों पर असहनीय बोझ डालती जा रही है। आज स्थिति यह है कि सरकारी कार्यालयों, प्रकाशन-गृहों, व्यावसायिक एवं शिक्षण संस्थानों और वस्तुतः सभी जगह अनुवाद के लिए सामग्रियों का ढेर बढ़ता ही जा रहा है।

मशीनी अनुवाद के विकास का एक मुख्य उद्देश्य अनुवाद कार्य की बोझिल एवं असहनीय स्थिति से अनुवादकों को मुक्त करना है। मशीनी अनुवाद के कई सरल और उपयोगी उपकरण आज उपलब्ध हैं - जैसे, क्षेत्र-विशेष (मौसम, स्वास्थ्य, पर्यटन, कार्यालयी पत्राचार, इत्यादि) मशीनी अनुवाद प्रणाली। इस दिशा में धीमी किंतु प्रगति अवश्य हुई है और अधिक से अधिक परिष्कृत प्रणालियाँ बनाई जा रही हैं जो पेशेवर अनुवादकों से लेकर सभी प्रकार के उपयोगकर्ताओं, सामान्य नागरिकों, व्यावसायिकों, पर्यटकों और शिक्षा एवं शोध से जुड़े लोगों को लाभ पहुँचा रही हैं। उन्नत तकनीक एवं गुणवत्ता के साथ नई पीढ़ी की प्रणालियाँ विकसित की जा रही हैं। इन सब प्रयोग-क्षेत्रों के विकास के साथ मशीनी अनुवाद की समस्याएँ एवं चुनौतियाँ भी बढ़ रही हैं।

आज माना जा रहा है कि मशीनी अनुवाद की मुख्य चुनौती भाषायी संसाधन की अनुपलब्धता है जिसमें विभिन्न भाषाओं के लिए संगणनात्मक व्याकरण एवं भाषा-प्रयोग आँकड़ों (खासकर बहुभाषिक व समांतर) का अभाव शामिल है। बहुत बड़े कॉर्पस (सभी प्रकार के लिखित एवं मौखिक, एक-भाषिक, सामान्य उद्देश्य और क्षेत्र संबंधित, बहुभाषिक एवं समांतर, इत्यादि) के रूप में सभी भाषाओं के लिए भाषिक साधन विकसित करना एक चुनौतीपूर्ण कार्य है। कुछेक मुख्य यूरोपीय भाषाओं को छोड़कर अन्य सभी भाषाओं में मशीन-साध्य भाषिक संसाधनों (रिसोर्स) का अभाव है जिस कारण इन भाषाओं में मानव भाषा संसाधन एवं मशीनी अनुवाद प्रणाली विकास को आवश्यक गति नहीं मिल पा रही है। मशीनी अनुवाद की विभिन्न वैकल्पिक पद्धतियों (जैसे, सांख्यिकीय मॉडल) के लिए भाषिक संसाधन अत्यंत आवश्यक हैं। सरकारी विभागों, औद्योगिक तथा शैक्षणिक संस्थानों के साथ-साथ, सामान्य जनता भी मशीनी अनुवाद प्रणाली के विकास के प्रति सजग एवं आशान्वित है। विकसित एवं विकासशील देशों के चुनिंदा शैक्षणिक संस्थानों में प्रकृत भाषा संसाधन (NLP), भाषा प्रौद्योगिकी (Language Technology), मशीनी अनुवाद और अन्य संबंधित विषयों में पाठ्यक्रम भी चलाए जा रहे हैं जिससे इस क्षेत्र में प्रशिक्षित मानव कार्य-शक्ति की कमी की भरपाई की जा रही है। यह भी विचारणीय है कि अब तक मशीनी अनुवाद क्षेत्र में हुए शोध एवं अकादमिक प्रयास केवल सैद्धांतिक विमर्श तक ही सीमित है, जिसके कारण इस क्षेत्र में व्यावहारिक स्तर पर सफलताएँ नहीं मिल सकी है। यह एक बड़ा मुद्दा है।

मशीनी अनुवाद से जुड़े अन्य मुद्दों में मूल्यांकन का मुद्दा भी है जिसका यहाँ उल्लेख नहीं किया गया है। मशीनी अनुवाद का मूल्यांकन बहुत से कारणों पर आधारित हो सकता है, जैसे तकनीकी, उन्नति, बाजार की शक्ति और राजनैतिक प्रभाव इत्यादि। मशीनी अनुवाद के व्यापक स्तर पर प्रयोग में आ जाने के बाद इसके मूल्यांकन के लिए आवश्यक टूल (साधन) भी विकसित करना आवश्यक हो जाएगा।

14.8 सारांश

इस इकाई में हमने पाया कि मशीनी अनुवाद मूलतः मशीन द्वारा संसाधित अनुवाद है। इसमें मनुष्य की भी भूमिका रहती है। मशीनी अनुवाद भी सामान्य अनुवाद की तरह एक प्रक्रिया साधित उत्पाद के रूप में हमें प्राप्त होता है जिसके विभिन्न चरण होते हैं तथा इसकी सफलता अथवा मानकता के अनेक पैमाने हैं। मशीनी अनुवाद की अपनी सीमाएँ हैं और यही सीमाएँ उसके लिए चुनौतियाँ भी बन जाती है। इन सीमाओं से पार पाने के लिए विभिन्न पद्धतियों का विकास विश्व स्तर पर अलग-अलग समय और दृष्टिकोणों से किया गया है तथा यह कार्य लगातार जारी है। महत्वपूर्ण है कि मशीनी अनुवाद के लिए विभिन्न तंत्रों के साथ-साथ कोशीय संसाधनों की भी आवश्यकता होती है। इस इकाई में हमने नियम आधारित मशीन पद्धति, कॉर्पस आधारित पद्धति के साथ-साथ हाइब्रिड अथवा मिश्रित मशीन पद्धति पर चर्चा की है। अलग-अलग पद्धतियों की अनुवादनीय पाठ प्रकृति के परिवेश के कारण अलग-अलग उपादेयता है। कुछ पद्धतियाँ प्रचुरता से लचीली और नियंत्रण योग्य हैं जबकि कुछ पद्धतियों में कॉर्पस की पर्याप्तता से एक अच्छा अनुवाद हासिल करने की क्षमता होती है तथापि अभी तक कोई ऐसी सफल प्रणाली विकसित नहीं हुई है जो अंतिम रूप से स्वीकृत की जा सके।

मशीन अनुवाद के अनेक मुद्दों में से अभियांत्रिकी से जुड़ी जटिल समस्याएँ और मानव भाषा के स्वरूपों को समझने की चुनौती संभवतः मानविकी, विज्ञान, अभियांत्रिकी और तकनीकी विषयों की सबसे महत्वाकांक्षी एवं कठिन अंतर्विषयक शोध प्रश्नों में से एक है। दूसरा बड़ा मुद्दा मशीन अनुवाद का हमारे द्वारा नए अनुप्रयोगों में प्रभावी रूप से उपयोग है जिससे हमें अपनी भाषिक बाधाओं को समाप्त करने और अपना ज्ञान क्षेत्र बढ़ाने में सहायता मिल सकती है। भविष्य में नागरिकों के प्रभावी सशक्तिकरण के लिए उपकरणों के रूप में मशीनी अनुवाद से और भी अधिक लाभ होने की संभावना है।

यहाँ हमने कॉर्पस आधारित मशीनी अनुवाद के अंतर्गत उदाहरण आधारित मशीनी अनुवाद तथा सांख्यिकीय आधारित मशीनी अनुवाद की भी चर्चा की जो द्विअर्थकता जैसे मुद्दों को सुलझाने में सहायक हो सकते हैं। व्याकरणिक रूप से अक्षम वाक्यों में भी ये पद्धतियाँ किसी सीमा तक संप्रेषणपरक अनुवाद उत्पाद प्रदान करने में सहायक हो सकती हैं। आपने यह अनुभव किया होगा कि मशीनी अनुवाद की प्रक्रिया और प्रविधियाँ सतत शोध और विकास के लिए एक महत्वपूर्ण क्षेत्र है। यह एक ऐसा उद्यम है जिसमें भाषाविज्ञानी, कंप्यूटर-विशेषज्ञ और अनुवाद करने वाले सहभागिता से काम करें तो भविष्य में मशीनी अनुवाद की चुनौतियों को प्रभावी रूप से हल किया जा सकता है। लगातार विकास से आने वाले समय में मशीनी अनुवाद की उपयोगिता तो बढ़ेगी; इसके उत्पाद भी अधिक गुणात्मक होंगे।

मानव भाषा, वैज्ञानिक तकनीक एवं मशीनी अनुवाद से जुड़े कई महत्वपूर्ण मुद्दे एवं प्रश्न आज भी अनुत्तरित हैं और इन विषयों पर शैक्षणिक अनुसंधान चल रहा है। मशीनी अनुवाद के लिए अनुवाद प्रक्रिया में साधन स्वरूप मानव प्रयासों को कम या समाप्त करने का रास्ता अभी काफी लंबा है।

14.9 अभ्यास के लिए प्रश्न

1. मशीनी अनुवाद की संकल्पना स्पष्ट करते हुए सामान्य अनुवाद से उसकी तुलना कीजिए।
2. मशीनी अनुवाद के संदर्भ में कॉर्पस की भूमिका एवं महत्व पर प्रकाश डालिए।
3. मशीनी अनुवाद में आने वाली चुनौतियों तथा उनके समाधान पर एक लेख लिखिए।
4. सामान्य व्यवहार में प्रयोग की जाने वाली भाषा-शब्दावली के मशीनी अनुवाद में कौन सी प्रविधि अधिक

उपयोगी रहेगी, उदाहरणों सहित समझाइए।

5. मशीनी अनुवाद के संदर्भ में भाषा-संसाधन से क्या तात्पर्य है, तथा मशीनी अनुवाद में इससे क्या-क्या चुनौतियाँ उत्पन्न होती है?
6. मशीनी अनुवाद के तीन अनुप्रयोग कौन-कौन हैं; विस्तार से बताइए।
7. मशीनी अनुवाद प्रणाली विकास की पद्धतियों पर चर्चा कीजिए।
8. मशीनी अनुवाद की भाषायी जटिलता की व्याख्या कीजिए।
9. मशीनी अनुवाद की मुख्य चुनौतियों की पहचान करते हुए उनके संभावित समाधान क्या हैं? इस पर विचार कीजिए।
10. मशीनी अनुवाद प्रणाली और मानव-अनुवादकों के परस्पर सहयोग एवं अनुवाद के भविष्य पर विचार कीजिए।

14.10 शब्दावली

कॉर्पस, कार्पोरा, डोमेन, हाइब्रिड, पार्सिंग, प्रमेय, प्रतीकात्मक, विश्लेषक, संश्लेषक, अंतरभाषिक, प्रयोगाश्रित, सदृश्य-तंत्र, कलनविधि, इष्टतमीकरण, भाषा संसाधन, ई-रिसोर्स, मशीन अनुवाद इंजन/यंत्र, मशीन अनुवाद प्रणाली, प्रकृत भाषा संसाधन (एन.एल.पी.), वोकोइस त्रिकोण, सिस्टम।

14.11 कुछ उपयोगी पुस्तकें

- विनीत चैतन्य एवं राजीव संगल, अक्षर भारती, 1994, *नेचुरल लैंग्वेज प्रोसेसिंग : अ पाणिनियन प्रासपेक्टिव*, प्रेंटिस-हॉल ऑफ इंडिया, नई दिल्ली।
- सिन्हा, आर.एम.के., 2004 *आंग्लभारती एम.टी. सिस्टम मैनुअल्स*, आई.आई.टी. कानपुर।
- सिन्हा, आर.एम.के., 2004, “एन इंजीनियरिंग प्रॉस्पेक्टिव ऑफ मशीन ट्रांसलेशन”, इन द प्रोसीडिंग्स ऑफ इंटरनेशनल सिमपोजियम ऑन मशीन ट्रांसलेशन, एन.एल.पी. एण्ड ट्रांसलेशन सपोर्ट सिस्टम में, इस्ट्रांस-2004, टाटा मैकग्रॉहिल, नई दिल्ली,
- सिन्हा, आर.एम.के. और अनिल ठाकुर, 2008, हाऊ टु गेट बेस्ट रिजल्ट्स आउट ऑफ ए मशीन ट्रांसलेशन सिस्टम: ए केस स्टडी ऑफ इंग्लिश टु हिंदी ट्रांसलेशन, *कंप्यूटर सोसायटी ऑफ इंडिया (सी.एस.आई.) जर्नल खंड-38, भाग 4*.
- सिन्हा, आर.एम.के. और अनिल ठाकुर, 2008, ए स्टडी ऑफ द ट्रांसलेशन डाइवर्जेंस इन इंग्लिश एंड हिंदी एम.टी, *कंप्यूटर सोसायटी ऑफ इंडिया (सी.एस.आई.) जर्नल, खंड 38, भाग 3*.
- Arnold, D., L. Balkan, R. Lee Humphreys, S. Meijer, L. Sadler, 1994, *Machine Translation: An Introductory Guide* Manchester/Oxford: NCC Blackwell. viii, 240pp.
- Carl, Michael and Andy Way (eds.), 2003, *Recent Advances in Example-based Machine Translation* Dordrecht: Kluwer Academic Publishers. xxxi, 482pp.
- Hutchins, W.J., 1986, *Machine Translation: Past, Present, Future* Chichester: Ellis Horwood Ltd. New York: Halsted Press. 382pp.
- Hutchins, W. John & Harold L. Somers, 1992, *An Introduction to Machine Translation* &: London: Academic Press. xxi, 362pp.
- Koehn, Philipp, 2009, *Statistical Machine Translation*, Cambridge University Press. xii, 433pp.
- Somers, Harold (ed.), 2003, *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins Publishing Company. [xv], 349pp.
- Trujillo, Arturo, 1999, *Translation Engines: Techniques For Machine Translation*, Berlin: Springer

Verlag. xii, 303pp.

- Nagao, Makoto, 1989, *Machine translation: How Far Can It Go?* Oxford University Press. xii, 150pp.
- Wilks, Yorick, 2009, *Machine Translation: Its Scope and Limits* New York: Springer. x, 252pp.
- Arnold, D., Balkan, L., Meijer S., Humphreys, R. Lee & Sadler, L. (1994). *Machine Translation: An Introductory Guide*. NCC Blackwell, London.
- Jurafsky, Daniel and Martin, James (2004) (Second Edition). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson, Prentice Hall.
- Somers, Harold (2003). 'Machine Translation: Latest Developments'. In Ruslan Mitkov (edited) (2003). *The Oxford Handbook of Computational Linguistics*. OUP. pp: 512-527.
- Whitelock, Peter & Kilby, Kieran (1995). *Linguistic and Computational Techniques in Machine Translation System Design*. Second Edition. UCL Press. London.