

---

## इकाई 16 मशीनी अनुवाद : अवधारणा और विकास

---

### इकाई की रूपरेखा

- 16.0 उद्देश्य
- 16.1 प्रस्तावना
- 16.2 मशीनी अनुवाद : अर्थ एवं आवश्यकता
  - 16.2.1 मशीनी अनुवाद का अर्थ
  - 16.2.2 मशीनी अनुवाद की आवश्यकता
- 16.3 मशीनी अनुवाद के प्रकार
  - 16.3.1 पूर्णतः स्वचालित मशीनी अनुवाद
  - 16.3.2 अर्ध-स्वचालित मशीनी अनुवाद
- 16.4 अर्ध-स्वचालित मशीनी अनुवाद (SAMT) के विविध प्रकार
  - 16.4.1 मानव-सहायक मशीनी अनुवाद
  - 16.4.2 मशीन-सहायक मानव अनुवाद
- 16.5 मशीनी अनुवाद का विकास : पश्चिम में
  - 16.5.1 प्रारंभिक तथा अत्युत्साह का युग
  - 16.5.2 शून्य काल
  - 16.5.3 पुनरुत्साह का युग
  - 16.5.4 कुछ मशीनी अनुवाद तंत्र
- 16.6 मशीनी अनुवाद का विकास : भारत में
  - 16.6.1 भारतीय भाषाओं में अनुवाद
  - 16.6.2 अंग्रेजी-हिंदी मशीनी अनुवाद
- 16.7 इंटरनेट और मशीनी अनुवाद
- 16.8 सारांश
- 16.9 अभ्यास के लिए प्रश्न
- 16.10 उपयोगी पुस्तकें

---

### 16.0 उद्देश्य

---

प्रस्तुत इकाई को पढ़ने के बाद आप :

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

---

## 16.1 प्रस्तावना

---

पिछले खंडों की विभिन्न इकाइयों में आप अनुवाद की प्रक्रिया, प्रविधियाँ, प्रकार-क्षेत्र और साधन-उपकरणों आदि के बारे में पढ़ चुके हैं। कंप्यूटर प्रौद्योगिकी के रूप में आधुनिक विकास के कारण अनुवाद के साधन-उपकरणों में इसके माध्यम से प्राप्त होने वाले ऑनलाइन और ऑफलाइन कोश भी विकसित हो चुके हैं और उपलब्ध हैं। कंप्यूटर रूपी मशीन केवल ये साधन उपलब्ध कराने का ही माध्यम नहीं हैं। इससे अनुवाद कार्य भी कराया जा रहा है। कंप्यूटर जगत में इस विकास ने 'मशीनी अनुवाद' की अवधारणा को साकार कर दिखाया है। इसलिए प्रस्तुत खंड को हम 'मशीनी अनुवाद' के रूप में आपके सामने ला रहे हैं।

हालाँकि आप जानते ही हैं कि अनुवाद एक भाषिक प्रक्रिया है जिसमें स्रोत भाषा में दिए गए पाठ या मूल पाठ की सूचना को लक्ष्य भाषा में रूपांतरित किया जाता है। ऐसा करने में अनुवाद की तीन प्रक्रियाओं का सहारा लिया जाता है। ये प्रक्रियाएँ हैं – विश्लेषण, अंतरण; और पुनर्गठन। मूल पाठ को समझने और उसमें कथ्य को जानने में विश्लेषण की प्रक्रिया हमारी सहायता करती है। यह कार्य स्रोत भाषा के व्याकरणिक एवं प्रयोग के नियमों के आधार पर किया जाता है। इस प्रक्रिया के दौरान मानव अनुवादक स्रोत भाषा के पाठ में निहित संदेश अथवा सूचना को समझता है और 'कथ्य' तक पहुँचने की चेष्टा करता है। मानवकृत अनुवाद में यह कार्य मानव कंप्यूटर अर्थात् मानव अनुवादक का मस्तिष्क कार्य करता है और मशीनी अनुवाद में यह कार्य मशीन करती है। लेकिन इसके लिए कंप्यूटर को तकनीकी से समृद्ध करना पड़ता है। इसके अलावा अन्य आयाम भी हैं, जिनके बारे में जानकारी अपेक्षित है। यहाँ हम मशीनी अनुवाद के बारे में सैद्धांतिक जानकारी दे रहे हैं ताकि अवधारणा के स्तर पर अनुवादकों को यह बोध हो सके कि मशीनी अनुवाद क्या है? उसका स्वरूप और आवश्यकता क्या है? यह कितने प्रकार का होता है? देश-विदेश में इसके विकास का संदर्भ क्या है और इसमें इंटरनेट की क्या भूमिका है? इस इकाई में हमने इन सभी प्रश्नों के संदर्भ में विचार किया है। आइए, अब हम यह जानें कि मशीनी अनुवाद क्या है?

---

## 16.2 मशीनी अनुवाद : अर्थ और आवश्यकता

---

### 16.2.1 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 समायोजन। 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 निर्माण (जेनरेशन) की प्रक्रिया से गुजरते हुए उन अनूदित तत्वों को लक्ष्य भाषा की व्याकरणिक संरचना के

अनुसार समायोजित कर लक्ष्य भाषा में अनूदित पाठ के रूप में प्रस्तुत किया जाएगा।

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। लक्ष्य भाषा के पाठ को 'अनूदित पाठ' कहा जाता है। स्रोत और लक्ष्य भाषा 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।

## 16.22 मशीनी अनुवाद की आवश्यकता

अब तक आप अनुवाद के महत्व और उसकी प्रासंगिकता के बारे में पढ़ चुके हैं। अनुवाद की परंपरा बहुत पुरानी है और उसका महत्व भी स्वयंसिद्ध है। सामाजिक, सांस्कृतिक, राजनीतिक, 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 'स्वांतः सुखाय' होता था। इसके अंतर्गत धार्मिक ग्रंथों का अनुवाद आता है। आधुनिक युग में अनुवाद एक संगठित व्यवसाय के रूप में उभरकर आया है। यह आधुनिक युग की सामाजिक, राजनीतिक और आर्थिक आवश्यकताओं के द्वारा प्रेरित है। आज अनुवाद की आवश्यकता इतनी अधिक बढ़ गई है कि केवल मानव अनुवादकों के द्वारा 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 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 राजनीतिक, सांस्कृतिक और 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 राष्ट्र संघ 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।

### 16.3 मशीनी अनुवाद के प्रकार

कंप्यूटर के विकास ने मशीनी अनुवाद को संभव बना दिया है। अब कंप्यूटर या मशीन द्वारा अनुवाद कार्य किया जा सकता है। कंप्यूटर ने अनुवाद की प्रक्रिया को सरल एवं त्वरित बना दिया है। कंप्यूटर अनुवाद प्रणाली की जटिलता मुख्य रूप से स्रोत और लक्ष्य भाषा 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 सिद्ध होती है। उदाहरण के लिए, “आर्यभाषा परिवार” की दो भाषाओं (बांग्ला-हिंदी, मराठी-हिंदी आदि) में परस्पर अनुवाद करना। 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।

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”।

#### 16.3.1 “पूर्णतः स्वचालित मशीनी अनुवाद”

“पूर्णतः स्वचालित मशीनी अनुवाद” (Fully Automatic Machine Translation - FAMT) 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। वैसे इसके प्रयोग की सफलता के बारे में वैज्ञानिकों में मतैक्य नहीं 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 एम.यू. (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।

### 16.3.2 अर्ध-स्वचालित मशीनी अनुवाद

अनुवाद प्रक्रिया मूलतः बौद्धिक प्रक्रिया है और मनुष्य ही एकमात्र प्राणी है जो इस क्षमता से संपन्न है। जबकि कंप्यूटर एक मशीन मात्र है, व्यक्ति नहीं। विश्व-भर में प्रचलित कोई भी मशीनी अनुवाद प्रणाली ऐसी नहीं है जिसमें मानव का बिल्कुल भी हस्तक्षेप आवश्यक न हो। इस तरह मशीनी अनुवाद प्रणालियाँ सिद्धांततः पूरी तरह से स्वचालित और सामान्य उद्देश्यीय नहीं हैं। इसलिए “अर्ध-स्वचालित मशीनी अनुवाद” (Semi & Automatic Machine Translation – SAMT) की अवधारणा बनी रहती है। अर्ध-स्वचालित मशीनी अनुवाद और उसके विविध प्रकारों के बारे में इकाई के अगले भाग में विस्तार से चर्चा की जा रही है।

## 16.4 अर्ध-स्वचालित मशीनी अनुवाद (SAMT) के विविध प्रकार

जिस प्रकार अनुवादक, स्रोत भाषा पाठ की अभिव्यक्ति के अभिप्राय को ग्रहण करते हुए उसका लक्ष्य भाषा में पुनर्कथन करता है, उसी प्रकार कंप्यूटर के जरिए स्रोत भाषा पाठ की अभिव्यक्ति के अभिप्राय को ग्रहण करते हुए उसका लक्ष्य भाषा में पुनर्कथन संभव है। किंतु कंप्यूटर यह पुनर्कथन अपने आप नहीं करता है। इसमें मशीन के साथ मानव की सहभागिता अवश्यभावी है। यह सहभागिता मानव के 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-सहायक मानव अनुवाद”; और “मानव-सहायक मशीनी अनुवाद” के रूप में वर्गीकृत किया जाता है। मशीनी अनुवाद के क्षेत्र में विश्व में जो अधिकांश प्रणालियाँ विकसित की गई हैं, उनमें इस प्रकार की सहायता की बराबर स्थिति बनी रहती है।

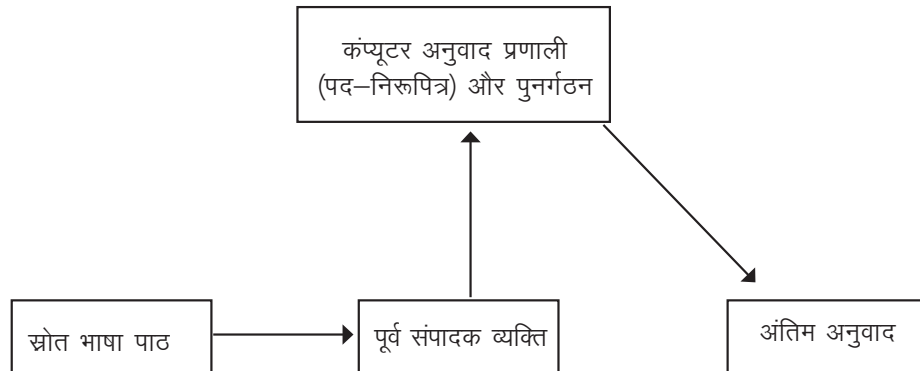
### 16.4.1 “मानव-सहायक मशीनी अनुवाद”

“मानव-सहायक मशीनी अनुवाद” (Machine-Assisted Human Translation - MAHT) 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 किया जाता 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 को श्लेष 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-सहायक कंप्यूटर अनुवाद संपादन के प्रमुख रूप इस प्रकार हैं :

- पूर्व संपादन (Pre-editing)
- पश्च संपादन (Post-editing)
- अंतरा-प्रोसेसिंग 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 स्रोत भाषा पाठ को व्यक्ति द्वारा संपादित किया जाता है। इस प्रणाली के अंतर्गत व्यक्ति स्रोत भाषा 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 स्रोत भाषा सामग्री में व्याप्त कठिनाइयों को दूर कर देता है। पूर्व-संपादन की स्थिति के दौरान ही जटिल व्याकरणिक संरचनाओं, अनेकार्थकताओं, संदिग्ध शब्दों एवं समस्यापरक वाक्यगत विशिष्टताओं का भी समाधान कर दिया 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 स्रोत भाषा सामग्री पर व्यक्ति द्वारा किया जाता है। इसके अलावा, स्रोत भाषा सामग्री का संपादन कंप्यूटर में विशेष प्रोग्राम की सहायता से भी किया जा सकता है।

चूँकि इसमें लेखन नियमों का ध्यान रखना पड़ता है और उनके आधार पर स्रोत भाषा सामग्री को संपादित करना पड़ता है, इसलिए यह काफी जटिल प्रक्रिया सिद्ध 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 स्रोत भाषा सामग्री के शब्द और व्याकरणिक संरचना अगर भाषा मानक के अनुसार नहीं है तो उनका पूर्व-संपादन कर लिया जाए। फ्रांस में TITUS मशीनी अनुवाद प्रणाली में पूर्व-संपादन प्रणाली को प्रयुक्त किया जाता है। पूर्व संपादन प्रणाली को निम्नलिखित आरेख के माध्यम से समझा जा सकता है :



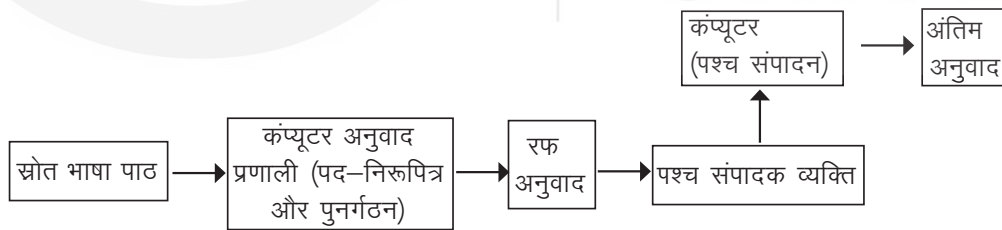
वैसे पाठ को मशीन के द्वारा स्वीकार्य रूप में परिवर्तित करने की प्रक्रिया में बहुत अधिक समय लग जाता है। इसके अलावा, इसमें स्रोत भाषा पाठ के अर्थ एवं आशय की सूक्ष्मता से “मशीन अनुवाद पूर्व-संपादक” द्वारा परिवर्तन की संभावना भी बनी रहती है। इस प्रकार के कतिपय कारणों से पूर्व-संपादन प्रणाली को बेहतर नहीं 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 स्रोत भाषा 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 कहा जाता है। 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 संपन्न किया जा सके। कंप्यूटर अनुवाद प्रणाली की पश्च संपादन तकनीक को निम्नलिखित प्रकार से संचालित किया जाता है :

- स्रोत भाषा सामग्री को कंप्यूटर द्वारा पठनीय रूप में परिवर्तित किया जाता है। यानी स्रोत भाषा पाठ को कंप्यूटर में फीड किया जाता है।
- उसके बाद स्रोत भाषा 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 रफ अनुवाद प्रस्तुत कर देती है।
- मूल पाठ और रफ अनूदित सामग्री को मानव-अनुवादक को सौंप दिया जाता है। मानव-अनुवादक उसका पुनरीक्षण-संपादन करके उसे अंतिम रूप प्रदान करता है। ऐसा करके वह “मशीन अनुवाद पश्च-संपादक (Machine Translation Post-Editor)” का दायित्व निभाता है। यह संपादन कार्य सीधे कंप्यूटर पर भी किया जा सकता है और मुद्रित कागज पर भी।
- पश्च संपादन की प्रक्रिया में लक्ष्य भाषा की संरचना को ध्यान में रखा जाता है और अनूदित पाठ को सहज 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:



मशीनी अनुवाद के संदर्भ में पश्च-संपादन प्रणाली को बेहतर माना जाता है क्योंकि 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, लक्ष्य भाषा का विशेषज्ञ और 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 स्रोत भाषा की जानकारी होना अपेक्षित नहीं है लेकिन, व्यवहार में यह देखा गया है कि अगर पश्च संपादक स्रोत भाषा का जानकार नहीं है तो इस प्रणाली द्वारा किए गए अनुवाद को संपादित करने में बड़ी मुश्किल होती है। "सिस्ट्रान" (SYSTRAN) और "लोगोस" (LOGOS) आदि सभी प्रमुख कंप्यूटर अनुवाद प्रणालियों में पश्च-संपादन घटक का समावेश मिलता है।

**(iii) अंतरा-प्रोसेसिंग संपादन प्रणाली :** कंप्यूटर में प्रोसेसिंग के समय अनूद्य सामग्री का संपादन संभव है। कंप्यूटर द्वारा अनुवाद की इस प्रक्रिया को "अंतरा-प्रोसेसिंग संपादन प्रणाली" कहा जाता है। यह कंप्यूटर अनुवाद प्रक्रिया के दौरान मानव-मशीन परस्पर सहयोग पर आधारित प्रणाली है। अंतरा-प्रोसेसिंग कार्य करने वाला व्यक्ति "अंतरा-प्रोसेसिंग संपादक" (Intra-processing Editor) कहलाता है। अंतरा-प्रोसेसिंग संपादक, प्रोसेसिंग के दौरान अंतर्क्रिया करता है। इस दौरान वह उन द्व्यर्थकताओं (ambiguities) का समाधान करता है जिन्हें प्रोसेसिंग करते समय मशीन स्वयं हल नहीं कर पाती।

अंतरा-प्रोसेसिंग संपादन प्रणाली में द्विभाषिक प्रश्न पूछ 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-प्रोसेसिंग (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 ALPS CTS कंप्यूटर अनुवाद प्रणाली प्रथम प्रकार की अंतरा-प्रोसेसिंग का उदाहरण है तो कनाडा की Termium नामक कंप्यूटर अनुवाद प्रणाली दूसरे प्रकार की अंतरा-प्रोसेसिंग का।

अंतरा-प्रोसेसिंग 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 स्रोत भाषा की इकाई की द्व्यर्थकता वाले शब्द-विशेष के साथ लक्ष्य भाषा के सह-संबंधित शब्द को भी पूछा जा सकता है। यह स्रोत भाषा पाठ का अर्थ उस भाषा में उपलब्ध कराने के लिए 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"।

शब्द में अर्थ और भाव के स्तर पर द्व्यर्थकता को अंतरा-प्रोसेसिंग के दौरान मानव-मशीन सहयोग के द्वारा दूर करना सरल है। किंतु स्रोत भाषा पाठ की संदर्भपरक द्व्यर्थकता भी समस्या खड़ी करते हैं। उदाहरण के लिए, यह संदर्भपरक द्व्यर्थकता 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 'it' शब्द तो दो बार प्रयुक्त हुआ है, लेकिन कंप्यूटर के द्वारा यह पता लगाना जरा मुश्किल है कि ये दोनों 'it' किस-किसके 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 'it' किन-किन संदर्भों में प्रयुक्त हुए हैं। इसी प्रकार अंग्रेजी के 'he' और 'she' सर्वनामों के परिप्रेक्ष्य में संदर्भपरक द्वयर्थकता 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 दर्शाने के संदर्भ में या फिर अन्य संदर्भ में प्रयुक्त विकल्प) में से सही विकल्प चुनना होगा। किंतु कंप्यूटर प्रणाली के लिए पूरे स्रोत भाषा पाठ में से इस प्रकार के संदर्भित शब्द ढूँढना कठिन है। प्रोसेसिंग के दौरान मानव-मशीन परस्पर सहयोग द्वारा वाक्यपरक द्वयर्थकता 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 (Trees) की तुलना करना प्रत्येक व्यक्ति के लिए सरल कार्य नहीं है। अर्थपरक विश्लेषण द्वारा द्वयर्थकता को भली प्रकार से दूर किया जा सकता है। इसलिए वाक्यपरक द्वयर्थकता की स्थिति में समुचित भावार्थ को चुनने के लिए अंतरा-प्रोसेसिंग 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 आवश्यकत अनुसार, वाक्यवार अपनी भाषा में स्रोत पाठ के शब्दों के संबंध में व्यक्ति से स्पष्टीकरण माँगता है। जैसे, मान लीजिए स्रोत भाषा 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 द्वयर्थकता हो। इस द्वयर्थकता 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

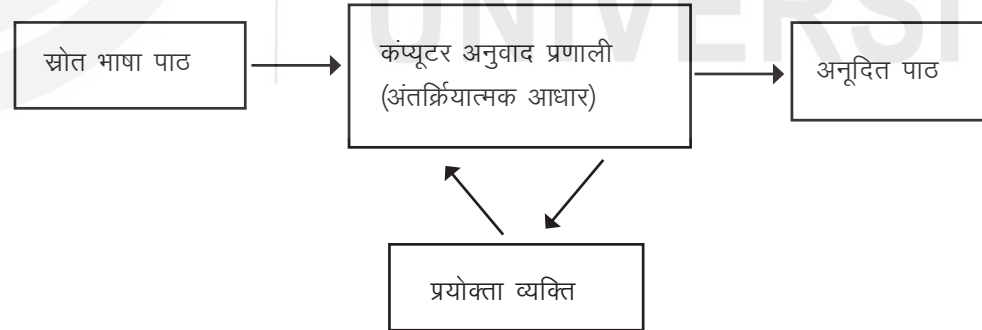
NUMBER>>

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 लक्ष्य भाषा में अनुवाद प्रस्तुत कर देता है। मशीन के साथ अंतर्क्रिया करते हुए अनुवाद की दिशा में आगे बढ़ने वाले व्यक्ति को “अंतर्क्रियात्मक मशीन संपादक” (Interactive Machine Editor) कहा जा सकता है।

मानव और मशीन की 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 स्रोत भाषा पाठ की लक्ष्य भाषा में प्रस्तुति तक होती है। उल्लेखनीय है कि इस दृष्टि के आधार पर किए गए अनुवाद में अनुवादोत्तर संपादन या पश्च संपादक/अनुवादक की जरूरत नहीं रह जाती।

अंतर्क्रियात्मक अनुवाद प्रणाली में व्यक्ति को लक्ष्य भाषा का बोध होना आवश्यक नहीं होता है क्योंकि यह अंतर्क्रियात्मक पूर्व-संपादन परस्पर सहयोग, स्रोत भाषा के स्तर पर होता है न कि लक्ष्य भाषा 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 स्रोत भाषा सामग्री इतनी अधिक तकनीकी प्रकृति की हो कि उसके व्यावसायिक अनुवादक को भी ज्ञान के क्षेत्र-विशेष के विशेषज्ञों की सहायता की जरूरत हो तब भी इस प्रणाली का उपयोग किया जाता है। इस प्रणाली के अंतर्गत अनुवाद की प्रक्रिया का आरेखीय निरूपण इस प्रकार से किया जा सकता है :



“मानव सहायक मशीनी अनुवाद” के बारे में जानने के बाद आइए अब हम “मशीन-सहायक मानव अनुवाद” (MAHT) प्रणाली के बारे में जानें।

#### 16.4.2 “मशीन-सहायक मानव अनुवाद”

“मशीन-सहायक मानव अनुवाद” प्रणाली को “कंप्यूटर सहायक मानव अनुवाद” (Computer Assisted Human Translation - CAHT) अथवा “कंप्यूटर सहायक 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। इससे अनुवादक के समय की बचत होती है।

इस प्रणाली में 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। इस प्रकार की कुछ प्रमुख सहायिकाएँ हैं :

- (1) **शब्दकोश स्कैन सॉफ्टवेयर (Dictionary Scan Software)** : इससे अनुवादक को शब्दों की वर्तनी की जाँच करने में मदद मिलती है।
- (2) **इलेक्ट्रॉनिक शब्दकोश, शब्द-संग्रह और शब्दावली डाटाबेस (Electronic Dictionaries, Glossaries and Terminology Databases)** : ये अनुवादकों को सॉफ्टवेयर या 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।
- (3) **समनुक्रमणिका (Concordance)** : यह शब्द/वर्ण-प्रयोगों की सूची उपलब्ध कराने वाला वह सुनिश्चित कॉर्पस है उसमें शब्दों की विभिन्न अर्थ-छटाओं की व्याख्या दी हुई होती है। वैसे यह शब्दकोशों आदि का प्रतिस्थापक नहीं है। शब्दों अथवा पूर्वसर्गों के बारे में सांख्यिकीय आँकड़े 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 इसके विशिष्ट कार्य हैं। इसमें शब्द संबंधी सूचना संचित/संकलित होती है और यह सूचना बिना क्रम से किसी भी समय परामर्श के लिए उपलब्ध रहती है। इसके डाटाबेस को भंडारित भी किया जा सकता है।
- (4) **अनुवाद स्मृति-कोश प्रबंधक (Translation Memory Manager - TMM)** : इसमें कंप्यूटर डाटाबेस में स्रोत भाषा पाठ और उसकी एक अथवा उससे अधिक भाषाओं में पहले से ही अनूदित सामग्री संचित होती है, जिन्हें आवश्यकता के अनुसार पूरा अथवा अंशतः अनूदित अंश के रूप में प्राप्त किया जा सकता है।
- (5) **भाषा सर्च इंजन सॉफ्टवेयर (Language Search Engine Software)** : यह सर्च इंजन स्रोत इकाई के आधार पर पहले से अनूदित उन वाक्यों, वाक्यांशों, अनुच्छेदों आदि की अपने व्यापक अनुवाद स्मृति भंडार में खोजने का काम करता है जो स्रोत इकाई से मेल खाते हैं।

- (6) **समरूप कॉर्पोरा सॉफ्टवेयर (Aligned Corpa Software)** : यह निर्धारित करता है कि संबंधित विषय-वस्तु वाला अनुवाद स्मृति डाटाबेस तैयार करने के लिए कौन-सी इकाइयाँ एक-दूसरे से संबंधित हैं। कई समरूप सॉफ्टवेयरों में यह व्यवस्था है कि ठीक से मेल नहीं खाने वाली इकाइयों को अनुवादक स्वयं ही समरूप व्यवस्थित कर दे। अनुवादक द्वारा इस तरह तैयार किए गए अनुवाद को 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।

---

## 16.5 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। सन 1933 में फ्रांसीसी वैज्ञानिक जार्ज आत्सूनी ने स्वचालित द्विभाषी कोश और 1939 में रूसी वैज्ञानिक पीटर त्रोयांस्की ने द्विभाषी कोश तथा एस्परांतो 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 डिजिटल 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 पूर्व-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 59 रूसी वाक्यों का अंग्रेजी में सावधानीपूर्वक 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 वाक्य-विन्यासीय 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।

### 16.5.1 प्रारंभिक तथा अत्युत्साह का 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 तत्कालीन 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। बीसवीं शताब्दी के 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।

पिछली शताब्दी के साठ के दशक में तत्कालीन 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 मूल रूप 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।

### 16.5.2 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 'एल्पैक समिति' (Automatic Language Processing Advisory Committee - ALPAC) 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 ही कम और चिंताजनक है।
- समकालीन स्थिति में मशीनी अनुवाद संभव नहीं है।
- मशीनी अनुवाद के लिए आधारभूत भाषा विश्लेषण सिद्धांतों का अभाव है। भाषा विश्लेषण मॉडलों के विकास पर बल देना आवश्यक है।
- समिति ने अनुवादकों की सहायता के लिए स्वचालित कोशों के विकास की सिफारिश की।

एल्पैक समिति की रिपोर्ट का प्रभाव अमेरिका में पूर्ण रूप से तथा तत्कालीन सोवियत संघ और इंग्लैंड में आंशिक रूप से हुआ। इसके बाद अमेरिका में मशीन 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। लेकिन 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 (Computational Linguistics) 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।

### 16.5.3 पुनरुत्साह का युग

एल्पैक समिति की रिपोर्ट आने के बाद अमेरिका में तो इस क्षेत्र में काम बिलकुल ठप्प हो गया किंतु यूरोप तथा अन्य देशों में छिटपुट 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 एवं 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 मशीन द्वारा अनुवाद किया गया। यह तंत्र अत्यधिक सफल रहा और इसने मशीनी अनुवाद के मृतप्रायः 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 मैन्युअलों 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 जिन विशेषताओं को स्वीकृति मिली; वे हैं :

- 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 छिटपुट 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। इन प्रयासों के माध्यम से कुछ विशिष्ट अनुवाद तंत्र विकसित हुए।

### 16.5.4 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। ग्रैनोब्ल विश्वविद्यालय में 'आरियान (ARIANE) टैक्सास विश्वविद्यालय में 'मेटल' (METAL), सारब्रुकें विश्वविद्यालय में 'सूसी' (SUSY), क्योतो (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 वित्त-पोषण 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 अनुवादक, के समान समकक्ष बनाने के प्रयासों को छोड़कर 'मानव अनुवाद के सहायक रूप' में ग्रहण किया गया है।
- इस युग में यूरोपीय समुदाय द्वारा वित्त-पोषित '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 प्रदर्शन किया गया। जापान में CICC द्वारा जापानी, चीनी, कोरियन, थाई भाषा, इंडोनेशियाई, भाषा और मलय में परस्पर अनुवाद की अंतरराष्ट्रीय परियोजना पर काम शुरू हो गया है।
- विभिन्न व्याकरण सिद्धांतों के विकास और 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।

---

## 16.6 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 तंजावूर 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 प्राकृतिक भाषा संसाधन (NLP) के अंतर्गत मुख्य रूप से मशीनी अनुवाद के क्षेत्र में कार्य पिछली शताब्दी के अस्सी के दशक में शुरू किया गया था।

### 16.6.1 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 (आई.आई.टी.) कानपुर में संपन्न एशियाई भाषाओं में कंप्यूटर संसाधन (Computer Processing in Asian Languages) 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 पाँच अनुसारकों पर कार्य हुआ। ये थे - तेलुगु-हिंदी, कन्नड़-हिंदी, पंजाबी-हिंदी, बांग्ला-हिंदी और मराठी-हिंदी अनुसारक। ये अब संचार और सूचना प्रौद्योगिकी विभाग (Technological Development in Indian Language-TDIL) की वेबसाइट पर उपलब्ध हैं। 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.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।

### 16.6.2 अंग्रेजी-हिंदी मशीनी अनुवाद

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

### 1. 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 मैनुअलों 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 एजेंसी 'प्रगत संगणन 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। 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।

### 2. मात्रा (Matra), सी.डैक. मुंबई

यह अनुवाद 'तंत्र' प्रगत संगणन 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 इंटरैक्टिव संरचना एडिटर 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। मात्रा-प्रो के लिए प्रयोगकर्ताओं को लाइसेंस लेना अपेक्षित रहा 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।

### 3. 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 वित्त-पोषण 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 करके राजभाषा विभाग की वेबसाइट पर उपलब्ध करा दिया। 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।

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

आंग्ल भारती का विकास भारतीय प्रौद्योगिकी संस्थान, (आई.आई.टी) कानपुर और प्रगत संगणन 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 से उपयोग किया जाता है। यह अंग्रेजी से हिंदी 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 परिवर्धन हुआ है। आंग्ल भारती के मॉडल के आधार पर अंग्रेजी-पंजाबी, अंग्रेजी-उर्दू, अंग्रेजी-बांग्ला, अंग्रेजी-मलयालम, अंग्रेजी-तेलुगु, अंग्रेजी-नेपाली आदि पर कार्य हो रहा है।

#### 5. अनुवादक (सुपर इंफोसॉफ्ट, नई दिल्ली)

मैट, मात्रा तथा मंत्रा – ये तीनों अनुवाद तंत्र सरकारी वित्तपोषण के आधार पर विकसित किए गए हैं। इसके विपरीत, 'अनुवादक' अनुवाद तंत्र एक निजी कंपनी (सुपर इंफोसॉफ्ट, 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 क्षेत्र के कार्यालयों, बैंकों आदि के अंग्रेजी सूचनात्मक साहित्य का अनुवाद करता है। इससे वार्षिक रिपोर्टों, समाचारिकाओं (बुलेटिन), परिपत्रों, अधिसूचनाओं आदि का अनुवाद किया जा सकता है।

'अनुवादक' एक ऐसा अनुवाद तंत्र है जो कृत्रिम बुद्धिमत्ता वाले अनुवाद इंजन 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 जाँच 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।



## 6. यूनिवर्सल नेटवर्किंग लैंग्वेज (UNL) – भारतीय प्रौद्योगिकी संस्थान, 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 यूनिवर्सल 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 आधिकारिक 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 यूनिवर्सल कोश तैयार हो चुका है। हिंदी से यू.एन.एल. में अनुवाद करने के लिए एनकन्वर्टर (Enconvertor) और यू.एन.एल. से हिंदी में अनुवाद करने के लिए डिकन्वर्टर (Deconvertor) 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।

### 6. 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 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

मशीनी अनुवाद  
: अवधारणा और  
विकास

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।

## 16.7 इंटरनेट और अनुवाद

इंटरनेट का विकास कंप्यूटर के क्षेत्र में एक महत्वपूर्ण परिघटना है। इंटरनेट वेब-आधारित प्रौद्योगिकी है जो आधुनिक समाज और सामाजिक व्यवस्था का अभिन्न अंग बन रही है। इंटरनेट के माध्यम से सूचना का व्यापक प्रसार हो जाता है, जिससे समाज के आम आदमी के सामाजिक-आर्थिक जीवन में परिवर्तन संभव हुआ है। जीवन-व्यवहार के विविध क्षेत्रों में इंटरनेट प्रौद्योगिकी के अनुप्रयोग से लोग लाभान्वित हो रहे हैं। इन्हीं में से एक प्रमुख क्षेत्र भाषा और अनुवाद भी है। इंटरनेट तकनीक ने जहाँ आम व्यक्ति को उसकी भाषा में आवश्यक जानकारी उपलब्ध कराना संभव कर दिखाया है वहीं कंप्यूटर अनुवाद प्रणाली को 'वेब अनुवाद' के रूप में विकसित करने में भी महत्वपूर्ण योगदान दिया है। इस क्षेत्र में हुआ विकास इंटरनेट-प्रयोक्ता को ऑनलाइन अनुवाद तक की सुविधा उपलब्ध करा रहा है।

मानव द्वारा किए जाने वाले अनुवाद और कंप्यूटर अनुवाद के परिप्रेक्ष्य में इंटरनेट को मुख्य तौर पर एक भाषा की सामग्री का कंप्यूटर के माध्यम से दूसरी भाषा में अनुवाद करने और तत्संबंधी जानकारी उपलब्ध कराने का प्रमुख स्रोत माना जाता है। कंप्यूटर अनुवाद के संदर्भ में इंटरनेट इस सूचना का भंडार है कि वे कौन-कौन से मशीन अनुवाद सॉफ्टवेयर और पोर्टेबल अनुवाद प्रणालियाँ हैं जो कंप्यूटर पर एक भाषा से दूसरी भाषा में अपने आप ही अनुवाद कर देते हैं। इंटरनेट पर जहाँ इस प्रकार के सॉफ्टवेयरों की जानकारी मिल जाती है वहीं यह भी पता चल जाता है कि वे सॉफ्टवेयर किन-किन कंपनियों द्वारा निर्मित किए गए हैं अथवा किनके द्वारा उनका विपणन (मार्केटिंग) किया जा रहा है। इंटरनेट पर सर्फ़िंग के दौरान इन सॉफ्टवेयरों और कंपनियों आदि के बारे में पता चल जाता है। उससे यह भी जानकारी मिल जाती है कि कौन-सा सॉफ्टवेयर किस भाषा से और किस भाषा में अनुवाद कर सकता है। साथ ही यह भी पता चलता है कि वे पोर्टेबल अनुवाद प्रणालियाँ कौन-सी हैं जो केवल शब्दशः (वर्ड फॉर वर्ड) अनुवाद करती हैं या फिर केवल मूल वाक्यांशों का ही अनुवाद करती हैं। ये सॉफ्टवेयर पूरे-पूरे वाक्यों का अनुवाद करने की क्षमता रखते हैं।

इंटरनेट के जरिए ऑनलाइन संपर्क करने पर ऑनलाइन अनुवाद की सुविधा भी संभव हो पाती है। आज पश्चिम में विकसित विभिन्न अनुवाद प्रणालियों के 'वेब' संस्करण भी तैयार किए जा चुके हैं। सबसे पहले विकसित कंप्यूटर अनुवाद प्रणाली 'सिस्ट्रॉन' (SYSTRAN) पर आधारित 'आल्टा विस्ता' (www.altavista.com) के जरिए कई प्राकृतिक भाषाओं के लिए ऑनलाइन अनुवाद करने की सुविधा विकसित की गई है। 'आल्टा विस्ता' से

यह सुविधा विभिन्न यूरोपीय भाषाओं में परस्पर अनुवाद के लिए है। इसके अंतर्गत 900 शब्दों तक के सीमित पाठ का किसी भी यूरोपीय भाषा में परस्पर अनुवाद किया जा सकता है। इसी प्रकार आज विश्व की महत्वपूर्ण भाषाओं में बेव पर अनुवाद करने की सुविधा प्राप्त है। याहू, हॉटमेल, गूगल आदि विभिन्न सर्विस प्रोवाइडर मूलतः सर्च इंजन का कार्य तो करते ही हैं, साथ ही, अनुवाद की सुविधा के लिंक भी उपलब्ध कराते हैं। वेब पर अनुवाद करने के लिए कई साइट विकसित हो चुकी हैं। इन साइटों के आधार पर 'वेब' पृष्ठों अथवा निवेशित स्रोत भाषा सामग्री का लक्ष्य भाषा में इंटरनेट के द्वारा अनुवाद संभव हो पाता है। आज अनेक वेबसाइटों के द्वारा निःशुल्क (मुफ्त) ऑनलाइन अनुवाद सेवाएँ उपलब्ध कराई जा रही हैं जिसके मूल में व्यावसायिक दृष्टि काम करती है। अपने व्यापार में बढ़ोतरी करना इनका मूलभूत उद्देश्य है। वेबसाइट के द्वारा किए जाने वाले अनुवाद कार्य की गुणवत्ता की जाँच करके कोई भी इच्छुक व्यक्ति-संस्था आदि उन सॉफ्टवेयरों अथवा शब्दकोशों आदि खरीदने का निर्णय ले सकते हैं।

इसके अलावा, अनुवाद विषयक कई अन्य पक्ष भी हैं जो इंटरनेट से प्रत्यक्ष अथवा परोक्ष रूप से जुड़े हुए हैं। इंटरनेट, अनुवाद और भाषा उपकरणों की जानकारी उपलब्ध कराने का साधन भी है। कई अनुवादक चाहते हैं कि उन्हें सॉफ्टवेयर उपलब्ध कराने वाली कंपनियों के बारे में पता चल जाए ताकि उन्हें अनुवाद कार्य में मदद मिल सके। इंटरनेट से प्राप्त जानकारी के आधार पर पता चलता है कि वैश्विक पटल पर आज 'लोगोस' (LOGOS) (<http://www.logos.it/>), 'वर्ड मैजिक सॉफ्टवेयर' (Word Magic Software) (<http://www.wordmagicsoft.com/>), आदि द्वारा अनुवादकों की सहायता के लिए साधन-उपकरण उपलब्ध हैं। इंटरनेट पर सर्फिंग के दौरान इन कंपनियों के उत्पादों, वेबसाइट पते आदि की जानकारी भी प्राप्त की जा सकती है।

इंटरनेट विभिन्न भाषाओं के शब्दकोश उपलब्ध कराने का संपर्क स्थल भी सिद्ध होता है। इंटरनेट के माध्यम से जो कोश उपलब्ध होते हैं उन्हें 'ऑनलाइन शब्दकोश' कहा जाता है। इन इंटरनेट कोशों से संपर्क स्थापित करके जहाँ कोई भी आम व्यक्ति भाषा-विशेष को बेहतर तरीके से समझ सकता है, वहीं अनुवादक को अनुवाद-कार्य के दौरान इनका उपयोग करने की सुविधा प्राप्त हो जाती है। वेबसाइट होस्ट जहाँ तक संभव हो सकता है ज्यादा से ज्यादा ऑनलाइन शब्दकोश सुविधा उपलब्ध कराने का प्रयत्न करते हैं। [www.word2word.com](http://www.word2word.com) आदि वेबसाइट पर भाषावार सूची उपलब्ध है जिसमें से अपेक्षित भाषा पर क्लिक करने पर ऑनलाइन कोश से संपर्क स्थापित हो जाता है और अपेक्षित जानकारी प्राप्त की जा सकती है।

इंटरनेट भाषाओं से संबंधित निःशुल्क सॉफ्टवेयरों (Free Language Softwares) का संपर्क स्रोत भी है। सूचना प्रौद्योगिकी से जुड़ी विश्व की कई कंपनियों ने भिन्न-भिन्न भाषाओं से संबंधित सॉफ्टवेयर विकसित किए हैं। इनकी सहायता से भाषा-विशेष को सीखने में मदद मिलती है। इसके अलावा, ये सॉफ्टवेयर अनुवाद-कार्य में भी सहायक रहते हैं। इन सॉफ्टवेयरों का विकास करने वालों ने निःशुल्क भाषा सॉफ्टवेयरों के संस्करण भी तैयार किए हैं। इन निःशुल्क सॉफ्टवेयरों को इंटरनेट से डाउनलोड करने की सुविधाएँ उपलब्ध हैं।

कंप्यूटर के क्षेत्र में भाषाओं को पहचानने से संबंधित सॉफ्टवेयर (Language Recognition Systems) भी विकसित किए जा चुके हैं। इस प्रकार के कंप्यूटर कार्यक्रमों की सहायता

से यह पता चल जाता है कि शब्द विशेष किस भाषा से संबंधित है और वह किस भाषा से आया है। इस प्रकार की जानकारी से अनुवादकों को शब्द विशेष के निहित अर्थ का स्पष्ट रूप से बोध हो जाता है। इंटरनेट भाषा पहचान प्रणालियों की जानकारी का साधन भी है। इंटरनेट पर इनकी सूची उपलब्ध है जिनकी सहायता से भाषा को पहचानने में मदद मिल सकती है।

इंटरनेट अपभाषा और बोलचाल की भाषा (Slang and Colloquialisms) सॉफ्टवेयरों की जानकारी का साधन भी है। अपभाषा अथवा बोलचाल की भाषा, वास्तव में आम बातचीत की भाषा के वे प्रयोग हैं जो मानक भाषा से बहुत भिन्न होते हैं। इन्हें बोलचाल में आम तौर पर व्यवहार में लाया जाता है, किंतु जीवन-व्यवहार के औपचारिक अवसरों पर या फिर भाषा के लिखित रूप में ये प्रयोग उपयुक्त नहीं माने जाते। उदाहरण के लिए, हिंदी में 'टुल्ला', 'फट्टे मारना' आदि शब्दों का बोलचाल के स्तर पर प्रयोग होते नजर आ जाता है किंतु मानक भाषा में इस प्रकार के शब्दों/वाक्यांशों का प्रयोग नहीं किया जाता है। वहीं ऐसा भी देखने में आता है कि मानक भाषा में प्रयुक्त कुछेक सहज-सामान्य शब्द भी अपभाषा और चलताऊ भाषा में भिन्न अर्थ देते हैं। जैसे, 'फेंकना' शब्द चलताऊ अर्थ में 'गप्प मारना' या 'दूर की हाँकना' के अर्थ को व्यक्त करता है। इसी प्रकार अंग्रेजी के 'chunk of wood' जैसे चलताऊ शब्दों को लिया जा सकता है, जिसका निहितार्थ 'लकड़ी का टुकड़ा' न होकर 'किसी काम का न होना' है। पहले अपभाषा शब्दों को शब्दकोश में शामिल नहीं किया जाता था और अगर वे शामिल भी हों तो उन्हें हटा दिया जाता था। किंतु अब इस प्रकार के शब्दकोश उपलब्ध हैं। ये स्वयं में विशिष्ट शब्दकोश हैं। इनमें केवल अपभाषा शब्द, उनकी परिभाषाएँ, उद्धरण और उन शब्दों के उद्भव संबंधी विस्तृत जानकारी दर्ज होती है। एरिक पैट्रीज (Eric Patridge) का 'A Dictionary of Slang and Unconventional English: Colloquialisms and Catch Phrases, Solecisms and Catechism, Nicknames, Vulgarisms and such Americanisms as have Naturalised' इसी प्रकार का विशिष्ट शब्दकोश है जिसमें अंग्रेजी के अपशब्द और अपरंपरागत शब्द शामिल हैं। भाषा-विशेष में अपभाषा प्रयोगों के स्पष्ट और सही अर्थ को समझने तथा उन प्रयोगों के प्रभाव को लक्ष्य भाषा पाठ में बनाए रखने की दृष्टि से ये कोश काफी महत्वपूर्ण सिद्ध होते हैं। इस संदर्भ में तैयार किए गए अपभाषा और चलताऊ भाषा सॉफ्टवेयर की इंटरनेट पर जानकारी भी उपलब्ध है। वैसे, इन स्रोतों की जानकारी के साथ इंटरनेट वेबसाइट होस्ट यह भी सूचित करना नहीं भूलते कि इनमें से कुछ साइटों पर 'प्रतिकूल शब्द' (cuss words) आदि भी दिए हुए हैं। इसलिए इस वेब-पृष्ठ को सुरक्षित सर्फ पंजीकरण से संबद्ध किया गया है ताकि ब्लॉकिंग सॉफ्टवेयर का इस्तेमाल करने वाले माता-पिता अपने बच्चों को इस वेब-पृष्ठ से दूर रख सकें। इसका कारण यह है कि माँ-बाप को यह आशंका होती है कि कहीं बच्चे अपभाषा शब्दों को ग्रहण करके उन्हें व्यवहार में न लाना शुरू कर दें। इसलिए वेब-पृष्ठ को सुरक्षित सर्फ पंजीकरण से संबद्ध करके वेबसाइट होस्ट बच्चों को अपभाषा सिखाने से बचाना चाहते हैं।

अगर अनुवादक को शब्द की व्युत्पत्ति का पता लगाना हो तो वह इस जानकारी को इंटरनेट के जरिए भी प्राप्त कर सकता है। इंटरनेट पर 'शब्द व्युत्पत्तिमूलक स्रोत' (word etymology source) की जानकारी उपलब्ध कराने के लिए शब्दों की व्युत्पत्ति संबंधी पृष्ठ भी मिलते हैं। ये पृष्ठ शब्दों के उद्भव और उनके इतिहास के बारे में जानकारी

का जरिया बनते हैं। व्युत्पत्ति कोश इस जानकारी का स्रोत रहते हैं कि भाषा विशेष में शब्द विशेष कहाँ से अथवा किस भाषा से आया है? वह कब और कैसे गढ़ा गया? समय के साथ-साथ उसके रूप और अर्थ में क्या परिवर्तन हुए अथवा आए? भाषाविदों, व्याकरणविदों और अनुवादकों आदि के लिए इस प्रकार के कोशों और उनमें शामिल शब्दों की ऐतिहासिक सूचनाएँ विशेष रूप से उपयोगी सिद्ध होती हैं।

इंटरनेट 'वाक् पहचान और उच्चारण संश्लेषण' (Voice Recognition and Speech Synthesis) संबंधी सॉफ्टवेयरों के बारे में सूचना देने का भी आधार बनता है। 'वाक् पहचान और उच्चारण संश्लेषण' में दो पक्ष समाहित हैं – वाक् की पहचान (Voice recognition); और उच्चारण संश्लेषण (Speech synthesis)। इनमें से पहला आयाम कंप्यूटर द्वारा एक भाषा के उच्चरित रूप (अर्थात् वाचिक प्रतीकों) को ग्रहण करने और उसे कंप्यूटर प्रणाली द्वारा लिप्यंतरित करने से संबंधित है। इस प्रकार का कार्य करने वाली प्रणाली को कंप्यूटर की शब्दावली में इसे 'वाक् से पाठ' (Speech to Text) भी कहते हैं। दूसरा आयाम कंप्यूटर द्वारा एक भाषा के लिखित रूप (अर्थात् भाषिक प्रतीकों) को ग्रहण करने और उसका कंप्यूटर प्रणाली द्वारा वाचन करने से संबंधित है। उच्चारण संश्लेषण अर्थात् पाठ को वाक् में बदलने का कार्य करने वाली प्रणाली को कंप्यूटर की शब्दावली में इसे 'पाठ से वाक्' (Text to Speech) भी कहते हैं। आज 'वाक् से पाठ' और 'पाठ से वाक्' जैसे क्षेत्रों में कंप्यूटर के अनुप्रयोग संबंधी सॉफ्टवेयर विकसित किए जा चुके हैं और किए जा रहे हैं।

इसके अलावा, लिप्यंतरण का क्षेत्र भी महत्वपूर्ण है जो इंटरनेट के जरिए और उससे प्राप्त जानकारी के आधार पर संपन्न किया जा सकता है। इंटरनेट पर लिप्यंतरण संबंधी वेबसाइटों की जानकारी भी उपलब्ध है। इन वेबसाइटों को एक लिपि से दूसरी लिपि में अंतरण करने संबंधी समस्या के समाधान के रूप में देखा जा सकता है। यह लिप्यंतरण इस तथ्य पर निर्भर करता है कि वेबसाइट को किस लिपि से किस लिपि में अंतरित करने के लिए विकसित किया गया है। लिप्यंतरण से भाषा-विशेष के शब्दों के उच्चारण का बोध हो जाता है। इसके अलावा, लिप्यंतरण संबंधी ये वेबसाइट उन लोगों के लिए विशेष तौर पर उपयोगी हैं जो स्रोत भाषा-विशेष की लिपि को लक्ष्य-भाषा विशेष की लिपि में लिप्यंतरित करना चाहते हैं ताकि वे उस भाषा-विशेष को भली प्रकार से बोल सकें।

इंटरनेट के जरिए हमें उच्चारण-लहजों और हास्यास्पद अनुवाद विषयक जानकारी देने वाले सॉफ्टवेयर/वेबसाइटों के बारे में भी पता चल सकता है। कुछ ऐसी ऑनलाइन स्वचालित अनुवाद प्रणालियाँ हैं जिनमें बोलचाल के क्षेत्रीय लहजे को व्यक्त करने का भी प्रयास किया जाता है। साथ ही हास्यास्पद अनुवाद संबंधी प्रणालियों की सूची भी इंटरनेट पर देखी जा सकती है। ये वे ऑनलाइन स्वचालित अनुवाद प्रणालियाँ हैं जिन्हें हास्य-विनोद और मनोरंजन के लिए डिजाइन किया गया है। वैसे इंटरनेट/वेबसाइट-होस्ट यह स्वयं स्वीकार करते हैं कि हालाँकि भाषा में बोलियों का अपना महत्त्व है किंतु इस सूची को केवल मनोरंजन को ध्यान में रखते हुए तैयार किया है।

भाषा और अनुवाद के संदर्भ में इंटरनेट के उपयोग के पक्ष यहीं तक सीमित नहीं हैं। इंटरनेट देश-विदेश के विभिन्न अनुवाद चिंतकों के अनुवाद संबंधी चिंतन की जानकारी उपलब्ध कराने के साथ-साथ अनुवाद एवं मशीन अनुवाद सैद्धांतिकी संबंधी ज्ञान का भंडार भी है। वहीं वह अनुवादकों/आशु अनुवादकों के लिए प्रभावी मंच भी है। यह

अनुवादकों और आशु-अनुवादकों को मंच भी उपलब्ध कराता है। कभी-कभी अनुवादक और आशु-अनुवादक इस व्यवसाय (अनुवाद कार्य) से जुड़े अन्य अनुवादकों से सहायता, नए संसाधनों का पता लगाने या फिर उनका परिचय प्राप्त करने आदि कई उद्देश्यों से संपर्क स्थापित करना चाहते हैं। इस संपर्क स्थापना में कंप्यूटर और इंटरनेट रूपी प्रमुख साधन अनुवादकों और आशु अनुवादकों के लिए मददगार सिद्ध होता है। इंटरनेट अनुवादकों के लिए मंचों की सूची और उन मंचों से इंटरनेट के जरिए संपर्क साधने की सुविधा भी उपलब्ध कराता है। उन मंचों में से कुछ तो ऑनलाइन लिस्ट सर्वर्स और फोरम हैं जिन्हें संचालित करते हुए अपनी जिज्ञासा निवारण के लिए प्रश्न पूछने पर उनसे उत्तर भी प्राप्त होते हैं।

वस्तुतः भाषा, अनुवाद और विशेष तौर पर कंप्यूटर अनुवाद के संदर्भ में इंटरनेट भाषा और अनुवाद संबंधी विभिन्न प्रकार की जरूरतों को पूरा करने के लिए अंतरराष्ट्रीय स्तर पर कई किस्म के नए उत्पादों की जानकारी देने का स्रोत-आधार है। वैसे इनके बारे में और अधिक जानकारी प्राप्त करने पर इंटरनेट संबंधी ये सामान्य निष्कर्ष सामने आते हैं कि इस पर वही जानकारी दी हुई होती है, जो वेबसाइट विकासकर्ताओं अथवा प्रयोक्ताओं द्वारा उपलब्ध कराई हुई होती है। इस संबंध में और जानकारी मिलने पर इंटरनेट-होस्ट इसे समय-समय पर अद्यतन भी करते रहते हैं। वहीं वे इस तथ्य को भी स्पष्ट रूप से स्वीकार करते हैं कि वे उन उत्पादों अथवा उनकी गुणवत्ता की कोई गारंटी नहीं देते हैं जिनके बारे में उन्होंने जानकारियाँ/सूचियाँ उपलब्ध कराई हैं। जहाँ निःशुल्क सॉफ्टवेयरों आदि की जानकारी दी हुई होती है वहीं वेबसाइट होस्ट यह भी बताने से नहीं चूकते कि अगर सूची में शामिल कंपनी सॉफ्टवेयर के इस्तेमाल के लिए कोई भुगतान लेती है तो उसकी जानकारी उन्हें तुरंत दी जाए। वैसे वेबसाइट होस्ट व्यवसायियों को वेबसाइट प्रायोजित करने का आमंत्रण भी देते हैं। वेबसाइट पर यह भी आह्वान भी मिलता है कि यदि लोगों को किसी नए उपयोगी संपर्कों के बारे में पता चलता है तो वे वेबसाइट होस्ट के साथ वेब के जरिए संपर्क कर सकते हैं। भाषा, अनुवाद और कंप्यूटर अनुवाद के संदर्भ में इंटरनेट के बारे में निष्कर्ष के तौर पर यही कहा जा सकता है कि यह कई दृष्टियों से संभावनाएँ प्रस्तुत करने का बेहतर साधन सिद्ध हुआ है।

## 16.8 सारांश

इस इकाई का अध्ययन करने के बाद आपने यह जाना कि मशीनी अनुवाद तंत्रों के विकास के संबंध में पिछले छह-सात दशकों में काफी कार्य हुआ है, साथ ही मशीन की प्रोसेसिंग तथा स्मृति क्षमता में भी अपार वृद्धि हुई है। इस दौरान भाषा विश्लेषण मॉडलों का भी विकास किया गया है।

प्रारंभिक युग में जहाँ 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.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।

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

---

## 16.9 अभ्यास के लिए प्रश्न

---

- 1) मशीनी अनुवाद के अर्थ और आवश्यकता को स्पष्ट कीजिए।
- 2) मानव और मशीन के संदर्भ में अनुवाद के प्रकार बताइए।
- 3) अर्ध-स्वचालित मशीनी अनुवाद (SAMT) के विविध प्रकार की चर्चा कीजिए।
- 4) मशीनी अनुवाद के प्रारंभिक युग को अत्युत्साह के युग की संज्ञा क्यों दी जाती है?
- 5) एल्यैक समिति की मुख्य सिफारिशें क्या-क्या थीं? एल्यैक समिति की रिपोर्ट के बाद मशीन अनुवाद का विकास किस प्रकार हुआ?
- 6) मशीन साधित अनुवाद (MAT) अनुवाद तंत्र पर प्रकाश डालिए।
- 7) यूनिवर्सल नेटवर्किंग लैंग्वेज से क्या तात्पर्य है? उससे मशीनी अनुवाद तंत्र का कैसे विकास होता है?
- 8) अनुवाद के संदर्भ में इंटरनेट की उपयोगिता के विभिन्न आयामों को उद्घाटित कीजिए।

---

## 16.10 उपयोगी पुस्तकें

---

- कृष्ण कुमार गोस्वामी, 2012, अनुवाद विज्ञान की भूमिका, नई दिल्ली : राजकमल प्रकाशन।
- ठाकुर दास, मशीनी अनुवाद : विधियाँ एवं प्रविधियाँ 'अनुवाद' अंक 169 भारतीय अनुवाद परिषद, नई दिल्ली।
- पूरनचंद टंडन, (संपा.), कंप्यूटर अनुवाद, नई दिल्ली, भारतीय अनुवाद परिषद।
- हरीश कुमार सेठी, ई-अनुवाद और हिंदी, नई दिल्ली, किताबघर।



- भाषा' (सूचना प्रौद्योगिकी और भारतीय भाषाएँ विशेषांक), 2002 में प्रकाशित मशीनी अनुवाद संबंधी लेख, केंद्रीय हिंदी निदेशालय, रामकृष्ण पुरम्, नई दिल्ली।
- 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. Prentice Hall.
- Sergei Nirenburg, Harold L. Somers and Yorick A. Wilks, Readings in Machine Translation, The Brain Science Connection, 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](http://www.freetranslation.com), [www.altavista.com](http://www.altavista.com), विदेशी भाषाओं तथा अंग्रेजी में परस्पर अनुवाद के लिए।
- [www.tdil@mit.gov.in](mailto:www.tdil@mit.gov.in) अनुसारकों तथा अंग्रेजी-हिंदी अनुवाद के लिए।

मशीनी अनुवाद  
: अवधारणा और  
विकास