

## भाषा प्रौद्योगिकी क्षेत्र-क्रांति की ओर....

✧ सुमेधा शुक्ला

आज का युग प्रौद्योगिकी का युग है। आज इस बात की प्रबल आवश्यकता है कि सूचनाओं का आदान-प्रदान कम से कम समय में विश्व के एक छोर से दूसरे छोर तक कैसे तीव्र गति से पहुँचाया जाए? प्रौद्योगिकी जिस तेजी से बदल रही है उन नवीनतम सूचनाओं पर अपनी पहुँच बनाए रखना बहुत जरूरी है एवं हर भाषा को तकनीकी से जोड़ने के लिए सूचना प्रौद्योगिकी की आवश्यकता पड़ती है। इंटरनेट इसके लिए एक आसान माध्यम है जिसके द्वारा हम सूचना एवं कम्प्यूटर से जुड़े हैं। आज समय आ गया है जब एक से अधिक भाषा में सम्पर्क करना जरूरी हो गया है एवं कम्प्यूटर विज्ञानी व भाषा विज्ञानी एक साथ मिलकर भाषा को तकनीकी से जोड़ने की कोशिश कर रहे हैं।

कम्प्यूटर नेटवर्क और दूरसंचार का महत्त्व बहुत तेजी से बढ़ रहा है एवं दूरसंचार का भी कम्प्यूटराइजेशन हो रहा है तथा मोबाइल फोन व पेजर की महत्ता दिनों-दिन बढ़ती जा रही है और उसी प्रकार कम्प्यूटर को LAN (Local Area Network) MAN (Metropolitan Area Network) एवं WAN (Wide Area Network) नेटवर्कों की सहायता से आपस में जोड़ा गया है।

आज कम्प्यूटर हिन्दी में ही नहीं विश्व की किसी भी भाषा में परिचालित किये जा सकते हैं। सरकारी क्षेत्र में कम्प्यूटरों पर हिन्दी के उपयोग की गति धीमी है परन्तु भारतीय रेल ने यात्री टिकट एवं आरक्षण सूचियाँ बड़े पैमाने पर जारी की है। बैंक भी इस दिशा में तेजी से अग्रसर है। इसके अतिरिक्त अधिकांश कम्पनियों के द्विभाषी एवं त्रिभाषी शब्द संसाधन के पैकेज बाजार में उपलब्ध हैं जिनमें-टाटा कन्सलटेंसी सर्विस का 'शब्दमाला', सोनाटा का 'मल्टीवर्ड', काल्स का 'सुलेख', सॉफ्टेक कम्पनी का 'अक्षर' व हिन्दीट्रोन का 'आलेख' प्रमुख हैं। इनके द्वारा किसी भी पर्सनल कम्प्यूटर पर हिन्दी व अन्य भारतीय भाषाओं में शब्द संसाधन का कार्य किया जा सकता है, किन्तु भाषा सम्बन्धी अध्ययन व विश्लेषण के लिए डाटा संसाधन व डेस्क प्रकाशन की सुविधा होना बहुत आवश्यक है। डाटा संसाधन के लिए लोटस I, II, III जैसे सॉफ्टवेयर पैकेजों का प्रयोग किया जाता है। आई.आई.टी. कानपुर ने भी 'जिस्ट' (Graphics and intelligence based Script Technology) प्रणाली के माध्यम से आज डाटा संसाधन से सम्बन्धित सभी कार्य हिन्दी एवं अन्य भारतीय भाषाओं में सरलता के साथ सम्पन्न किए हैं।

भारत सरकार के इलैक्ट्रॉनिकी विभाग ने सभी भारतीय भाषाओं के लिए एक समान कोड का महत्त्व स्वीकारा है। यह कोड अंग्रेजी के मानक कोड आस्की-7 (ASCII-7) का विस्तृत रूप है जिसमें कि 7 बिट कोड में अंग्रेजी के सभी अक्षर, संख्या व प्रतीक समाहित हो जाते हैं। भारतीय भाषाओं की वर्णमाला के लिए कम से कम 8 बिट के कोड की आवश्यकता होती है, इलैक्ट्रॉनिकी विभाग, भारत सरकार ने ब्राह्मी लिपि पर आधारित भारत की सभी लिपियों का एक समान कोड इस्की-8 (ISCII-8) नाम से तैयार किया है। इसमें रोमन लिपि के सभी अक्षरों के साथ-साथ भारतीय भाषाओं के भी सभी अक्षरों को समाहित किया है, इसका सबसे बड़ा लक्ष्य यह है कि इसके माध्यम से किसी भी भारतीय लिपि के जरिये सभी भारतीय भाषाओं में अध्ययन विश्लेषण का कार्य किया जा सकता है, इसके अतिरिक्त एक ही पृष्ठ पर हिन्दी अंग्रेजी एवं किसी भी भारतीय भाषा में एक साथ ही कुंजीपटल (Key-Board) द्वारा कार्य किया जा सकता है।

आई.आई.टी. कानपुर ने टी.डी.आई.एल. (Technology development for Indian Language) नाम से अंग्रेजी व हिन्दी अनुवाद के लिए एक सहायक सिस्टम बनाया गया है, जिसमें आसानी से अंग्रेजी-हिन्दी अनुवाद हो सकता है।

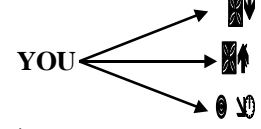
सी.डेक भी इस क्षेत्र में काफी कार्य कर रहा है, वह एस.टी.टी. (Speech to text) व टी.टी.एस. (Text to Speech) सिस्टम पर कार्य कर रहा है जिसे अन्तर्गत मानव द्वारा बोले गए शब्दों को कम्प्यूटर लिखित रूप में दिखाएगा एवं मानव द्वारा लिखित Word File को कम्प्यूटर द्वारा बोलकर प्रस्तुत किया जायेगा।

इसके अतिरिक्त विभिन्न संस्थाओं द्वारा एक ऐसा interface भी बनाया जा रहा है, जिसमें दो लोग एक-दूसरे की भाषा को न जानते हुए भी आपस में वार्तालाप कर सकते हैं, इसके लिए Unicode में एक डाटाबेस बनाया जा रहा है जो सभी भाषाओं पर समान रूप से कार्य करता है।

भाषा प्रौद्योगिकी भाषा विज्ञान के अध्ययन का एक नया आयाम है जिसके अन्तर्गत इलैक्ट्रॉनिक्स के माध्यम से भाषा का लेखन, पाठन, मुद्रण की नई तकनीकी, लिपि का प्रौद्योगिकी से तालमेल एवं भाषा शिक्षण के उपकरणों का विकास आदि समाहित हैं जिनके लिए संवेदनशील कम्प्यूटरीकृत प्रयोगशालाएँ प्रयोग में लायी जा रही हैं। भाषा प्रौद्योगिकी की संभावना को आज भाषा से जुड़े प्रत्येक क्षेत्र से जोड़ने की कोशिश की जा रही है। इसे द्वितीय भाषा शिक्षण से जोड़ा जाए तो यह अंतरभारतीय और अन्तर्राष्ट्रीय स्तर पर सभी भाषा-भाषियों को एक-दूसरे के निकट लाने में सहायक हो सकता है। इसके लिए सभी भाषाओं के छोटे-छोटे नमूने लिए जाते हैं और उनका आपस में व्यतिरेकी विश्लेषण किया जाता है और उसमें जो व्यतिरेक मिले उसे बाकी तत्त्वों में समानता को देखते हुए एक डेटाबेस बनाकर आज ऐसी तकनीक विकसित की जा रही है जिससे उन विरोधी

तत्त्व को देखते हुए आसानी से पहचाना जा सके, जो अन्य भाषा भाषी को समझने में सहायक होगा। इसके लिए मशीन को उन भाषाओं का वाक्यक्रम दिखाना होता है। उदाहरण-हिन्दी भाषा में वाक्यक्रम SOV (SUBJECT, OBJECT, VERB), अंग्रेजी भाषा में वाक्यक्रम SVO (SUBJECT, VERB, OBJECT)

इसके पश्चात् अंग्रेजी से हिन्दी में कार्य करते समय अंग्रेजी के 'YOU' शब्द के संदर्भ में हमें हिन्दी के तू, तुम व आप शब्दों में से किस उपयुक्त शब्द का चयन किया जाए? इसके लिए मशीन को 'YOU' के संदर्भ में तीनों शब्दों को बताना होता है।



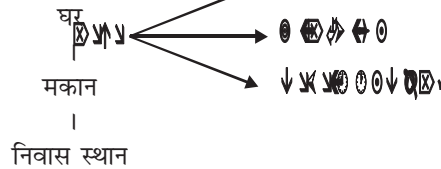
कम्प्यूटर द्वारा भाषा शिक्षण व मशीन अनुवाद इस क्षेत्र को और अधिक महत्त्व प्रदान करते हैं, भाषा शिक्षण में भारतीय भाषाओं को आपस में जोड़ने के लिए इनका प्रयोग किया जाता है जिसमें बहुत सी समस्याएँ उभर कर सामने आती हैं, क्योंकि हिन्दी के अतिरिक्त अन्य भारतीय भाषाओं में कारक चिह्न संज्ञा के साथ जुड़ जाते हैं, जैसे-

राम को बुखार है	हिन्दी	रामकु ज्वारा होयछि	उड़िया
रामेर ज्वार होयछे	बंगाली	रामला ताप आहे	मराठी

इस संदर्भ में मशीन को यह बताने की आवश्यकता होती है कि इन भाषाओं में परसर्ग संज्ञा के साथ ही जुड़ जाते हैं। इसके अतिरिक्त मराठी, गुजराती व दक्षिण भारतीय भाषाओं बहुत से शब्द हिन्दी वर्णमाला में नहीं हैं जिसके कारण सभी भाषाओं के साथ कार्य करने में कठिनाई का अनुभव होता है एवं इसके लिए संयुक्ताक्षरों का प्रयोग किया जाता है। इसकी सहायता से एक भाषा से दूसरी भाषा में आसानी से कार्य सम्पादित किया जा सकता है।

मशीन अनुवाद के लिए पैरलल कॉर्पोरा के प्रयोग से एक ऐसा प्रोग्राम बनाया जा सकता है जिसमें शब्दों की भ्रांतियों एवं बहुअर्थता के समाधान के साथ स्रोत भाषा की शब्दावली एवं व्याकरणिक संरचना का ज्ञान भी प्राप्त किया जा सकता है। इसके अतिरिक्त इलैक्ट्रॉनिक्स शब्दकोश के निर्माण का कार्य भी सफलतापूर्वक किया जा सकता है। मशीन अनुवाद करते समय भाषा के कुछ आयाम स्पष्ट रूप से सामने आते हैं। उदाहरण-

अनुवाद करते समय इन पक्षों को ध्यान में रखना अत्यन्त आवश्यक होता है तथा इस संदर्भ में कम्प्यूटर को भी इन बिन्दुओं से अवगत कराना होता है। इसके अतिरिक्त भाषा में बहुतायत से पाये जाने वाले एक शब्द का अर्थ की दृष्टि से समान, प्रचलित भिन्न रूपों का मशीन अनुवाद सम्बन्धी कार्य बहुत सावधानी से करना पड़ता है। उदाहरण-



कम्प्यूटर को घर शब्द के सम्बन्ध में मकान एवं निवास स्थान शब्द से भी परिचित कराना आवश्यक होता है।

अतः भाषा प्रौद्योगिकी नवीन संकल्पनाओं व अवधारणों के साथ विकास पथ पर अग्रसर हो रही हैं जिसमें तकनीक से जुड़ने के कारण भाषा नवीन प्रणालियों को स्वीकार कर रही है। जिसमें विभिन्न संस्थाएँ आगे बढ़कर सहयोग कर रही हैं एवं भाषा को संप्रेषणीय बनाने के लिए भाषा का सूचना प्रौद्योगिकी से जोड़ रही हैं।

#### संदर्भ ग्रन्थ :-

1. महरोत्रा, विजयकुमार-कम्प्यूटर के भाषिक अनुप्रयोग।
2. जगन्नाथ, वी.रा.-प्रयोग और प्रयोग दिल्ली ऑक्सफोर्ड यूनिवर्सिटी प्रेस।
3. ठाकुरदास, मशीनी अनुवाद : विधियाँ एवं प्रविधियाँ गवेषणा 63-64 केन्द्रीय हिन्दी संस्थान, आगरा।
4. Darbari Hemant - An efficient parsing technique for hindi language.