

Report of the work done

हिन्दी अध्ययन-अध्यापन के संदर्भों में गणनामूलक भाषाविज्ञान:

(Computational Linguistics:) एक अध्ययन।

सूचना प्रौद्योगिकी (Information Technology) के इस युग में संसार एक विश्वग्राम (Global Village) बन गया है। इसने भाषा-भौगोलिकता और सभ्यता-संस्कृति की सीमाओं को तोड़ अपने को विश्वव्यापी रूप प्रदान किया है। अतः शिक्षा के क्षेत्र में भी इन्हीं विषयों की माँग बढ़ गयी है। फलस्वरूप गणनामूलक भाषाविज्ञान (Computational Linguistics) जैसे विज्ञान की नयी शाखाओं का विकास अनिवार्य बन रहा है। विज्ञान की इस शाखा का यह अंतर्विषयक (interdisciplinary) स्वरूप भाषाविदों तथा कंप्यूटर वैज्ञानिकों को एक साथ मिलकर काम करने का मौका प्रदान करता है।

आज इस क्षेत्र का जबरदस्त विकास हो रहा है, इसका एक प्रमाण है गूगल द्वारा गूगल असिस्टेंट (Google Assistant) का निकालना। कृत्रिम बुद्धि (Artificial Intelligence), प्राकृतिक भाषा संसाधन (Natural Language Processing), मशीनी अनुवाद (Machine Translation), वाक् संसाधन (Speech Synthesis), दृश्य संसाधन (Vision Processing), वाक् अभिज्ञान (Speech Recognition) आदि विधाओं की सहायता से मानव भाषा संप्रेषण में गूगल ने बहुत महत्वपूर्ण सफलता प्राप्त की है। उनका समर्थन है कि कई विदेशी भाषाओं में आपस में मानव मनोभावों और विचारों को समझकर काम निभाने में गूगल असिस्टेंट 93.9% शुद्धता (accuracy) रखता है। यह एक गंभीर कदम है। यह सॉफ्टवेयर हिन्दी में भी काम करता है। पर अब हिन्दी में इसका

कार्यक्षेत्र अनुवाद तथा वाक् से पाठ (Text to Speech) तक सीमित है।

प्रस्तुत शोध का विषय है “हिन्दी अध्ययन-अध्यापन के संदर्भों में गणनामूलक भाषाविज्ञान:(Computational Linguistics:) एक अध्ययन।” पढ़ाई की सुविधा के लिए इसे चार अध्यायों में बाँटा है। वे यथाक्रम सूचना प्रौद्योगिकी युग में विदेशी भाषाओं के मुकाबले हिन्दी का स्थान, गणनामूलक भाषाविज्ञान की परंपरा एवं विकास-यात्रा, गणनामूलक भाषाविज्ञान की प्रयुक्तियाँ: हिन्दी अध्ययन के विशेष संदर्भ में, गणनामूलक भाषाविज्ञान की उपलब्धियाँ एवं संभावनाएँ आदि शीर्षक में हैं।

गणनामूलक भाषाविज्ञान की उपलब्धियों और संभावनाओं के बारे में चर्चा करते हुए ज़्यादातर हिन्दी के अध्येताओं तथा शोधार्थियों का ध्यान गणनामूलक भाषाविज्ञान की ओर आकृष्ट करना प्रस्तुत शोध का मुख लक्ष्य रहा है। गणनामूलक भाषाविज्ञान का उद्भव और विकास, दुनिया के मशहूर दार्शनिक तथा भाषावैज्ञानिक नोआम चॉमस्की (Noam Chomsky) तथा उनके वैश्विक व्याकरण जैसे सिद्धान्त, यूनिकोड जैसी सूचना-प्रौद्योगिकी की आधुनिक उपलब्धियाँ, कंप्यूटर साधित भाषा अधिगम (Computer-assisted Language Learning CALL), कंप्यूटर साधित भाषा शिक्षण (Computer-assisted Language Teaching : CALT), अंग्रेज़ी तथा हिन्दी के बीच होनेवाली कंप्यूटर-अनुवाद में उत्पन्न होनेवाली समस्याएँ तथा हल आदि प्रस्तुत किया है। नवीनतम आविष्कारों से परिचित रहना तथा जहाँ हो सके अपने सुझाव सहित प्रस्तुत करने का प्रयास हुआ है।

देवनागरी लिपि एक वैज्ञानिक लिपि है। भारतीय भाषाएँ विश्व की अनेक भाषाओं की तुलना में वाक्य-विज्ञान, ध्वनि विज्ञान और रैखिक दृष्टि से अधिक सुनियोजित है। अमरीकि कंप्यूटर वैज्ञानिक श्री रिक ब्रिगज़ की यह धारणा है कि संस्कृत भाषा कंप्यूटर प्रोग्राम की दृष्टि से आदर्श भाषा है। देवनागरी लिपि में कंप्यूटर पर कार्य करना कठिन नहीं है, किंतु आवश्यक सॉफ्टवेयर के अभाव में कुछ साल पूर्व तक इसकी कल्पना भी नहीं की जा सकती थी।

आज कम्प्यूटर, लैपटॉप, स्मार्टफोन तथा टैबलेट आदि डिजिटल उपकरण हर-एक भारतीय के भी दैनिक जीवन का हिस्सा बन चुके हैं। आजकल लगभग इन सभी उपकरणों में हिन्दी में काम करना सम्भव है। यूनिकोड सिस्टम ने हिन्दी को सभी कम्प्यूटिंग डिवाइस तक पहुँचा दिया है। यूनिकोड सिस्टम के कारण कम्प्यूटर पर हिन्दी एवं अन्य भारतीय भाषाओं में काम करना अंग्रेजी जैसा ही सरल हो गया है। इसी कारण अब इंटरनेट पर हिन्दी समाचार पत्रों तथा वेबसाइटों की भरमार है।

यूनिकोड आधारित वेबसाइटों को देखने के लिये पाठक के पास सम्बन्धित फॉण्ट होने की अनिवार्यता भी नहीं है। अगर कोई वेबसाइट यूनिकोड में है तो उसे किसी भी यूनिकोड सक्षम कम्प्यूटर पर देखा जा सकता है। यूनिकोड की लोकप्रियता संसार भर में बढ़ती आ रही है तथा इसके साथ ही हिन्दी तथा अन्य भारतीय भाषाओं में भी वेबसाइट, ब्लॉग, ऑनलाइन वेब आधारित औजारों तथा सुविधाओं का प्रयोग बढ़ता जा रहा है। ईमेल में सीधे हिन्दी में सम्प्रेषण किया जा रहा है। मोबाइल फोन पर भी हिन्दी तथा अन्य भारतीय भाषाओं में संक्षिप्त सन्देश

(SMS) तथा इंटरनेट संचार किया जाने लगा है। एण्ड्रॉइड जैसे मोबाइल ऑपरेटिंग सिस्टम में भी इण्डिक यूनिकोड का समर्थन आ गया है। किन्तु इन सब के बावजूद भी प्राकृतिक भाषा संसाधन (Natural Language Processing) के क्षेत्र में अब भी हमें एक लंबा सफर तय करना है। मशीन अनुवाद के कुछ लाभ और इसकी समस्याएँ तथा अपने कुछ सुझाव नीचे दिये जा रहे हैं-

मशीनी अनुवाद के लाभ

- अत्यन्त अल्प समय में अनुवाद - इससे तुरन्त पाठ का आशय सामने आ जाता है। अधिकांश स्थितियों में इतना ही काफी होता है।
- हर समय (चौबीसों घंटा) उपलब्ध है। हर जगह उपलब्ध है।
- कम या शून्य खर्च ।
- गोपनीयता एवं निजत्व (प्राइवेसी) की रक्षा।
- एक ही प्रोग्राम अनेकों भाषाओं से अनेकों भाषाओं में अनुवाद कर देता है जबकि अनुवाद करने वाला किसी एक भाषा से किसी दूसरी भाषा में ही अनुवाद कर सकता है।
- किसी पाठ को मशीन द्वारा अनुवाद करके उसको किसी व्यक्ति द्वारा सुधार लेना एक सस्ता एवं व्यावहारिक उपाय है।
- विश्व में करोड़ों लोगों के बीच सीधे संवाद बनेगा। इसके राजनीतिक, सामाजिक और वाणिज्यिक लाभ कल्पनातीत होगा।
- मशीनी अनुवाद के उपरोक्त गुणों के कारण विदेशी भाषा सीखने की आवश्यकता ही समाप्त होने वाली है।

मशीनी अनुवाद की समस्याएँ

- अक्सर मशीनी अनुवाद प्रणालियों द्वारा प्रजनित संयुक्त एवं मिश्र वाक्य हिन्दी की प्रकृति के अनुसार नहीं आ पाते हैं। बहुत से प्रजनित वाक्यों में अन्विति का अभाव रहता है।
- अनुवाद प्रणाली द्वारा अनुवाद होने पर अक्सर शब्दों के प्रसंगानुकूल अर्थ न आकर सर्वाधिक प्रचलित अर्थ आ जाता है। कुछ शब्द ऐसे होते हैं जिनको कोई अर्थ उपलब्ध न होने पर उसे अंग्रेजी में ज्यो का त्यों मशीन ले लेती है।
- मिश्र वाक्यों के अनुवाद में सर्वाधिक गलतियाँ होती हैं। प्रणाली यह नहीं समझ पाती कि कौन-सा उपवाक्य प्रधान है और कौन-सा आश्रित है।
- सबसे बड़ी समस्या अनुवाद के क्षेत्र में मुहावरों और लोकोक्तियों को पहचानने की है।
- बहुअर्थीय वाक्यों और शब्दों के अनुवाद में भी समस्या रहती है।
- सांस्कृतिक तत्वों से भरपूर टेक्स्ट (जैसे कविता) का अनुवाद नीरस होगा।

मशीनी अनुवाद को उन्नत बनाने के सुझाव:-

- मशीनी अनुवाद के शब्दकोश को और अधिक विस्तृत बनाया जाना चाहिए। इसमें मुहावरों और लोकोक्तियों के लिए अलग से कोश बनाया जाना चाहिए।
- यदि नियम आधारित पद्धति से मशीनी अनुवाद प्रणाली का निर्माण किया गया है तो इसमें इसके व्याकरण संबंधी नियमों को और अधिक स्पष्ट

बनाना चाहिए।

- भाषा में हो रहे प्रयोगों को अधिक से अधिक नियम के रूप में बनाकर व्याकरण में देना चाहिए। यदि सांख्यिकी आधारित अनुवाद प्रणाली हो तो इसे अध्ययन के लिए अधिक से अधिक सभी क्षेत्रों का प्रतिनिधित्व देना चाहिए।
- बहुअर्थता से निपटने के लिए संदर्भ को समझने योग्य प्रणाली विकसित करने का प्रयत्न करना चाहिए।
- मशीनी अनुवाद प्रणाली को विभिन्न भाषायी संसाधनों से यथेष्ट मदद लेकर प्रणाली को समुन्नत बनाना चाहिए।
- भाषिक नियमों से संपूर्ण तथा कंप्यूटर के कृत्रिम बुद्धि (Artificial Intelligence) के अनुकूल सर्वोत्तम मशीन अनुवाद का software तैयार करें।
- एक अंतर्विषयक (interdisciplinary) स्वरूप के तौर पर भाषाविदों तथा कंप्यूटर वैज्ञानिकों को एक साथ मिलकर काम करने का मौका प्रदान करें।

उपर्युक्त बातों पर ध्यान लगाने से हम इस निष्कर्ष पर पहुँच सकते हैं कि सूचना प्रौद्योगिकी के इस नये युग में ज़्यादा से ज़्यादा भारतीय कंप्यूटर वैज्ञानिकों तथा भाषाविदों को गणनामूक भाषाविज्ञान की संभावनाओं की ओर और जागरूक रहना चाहिए। उनका लक्ष्य यह रहे कि भाषिक नियमों से संपूर्ण तथा कंप्यूटर के कृत्रिम बुद्धि (Artificial Intelligence) के अनुकूल सर्वोत्तम सॉफ्टवेयर (software) तैयार करें। इसके लिए सरकारी प्रयास भी और तेज करने होंगे। तभी हम हिन्दी को डिजिटल दुनिया की प्रमुख भाषा के रूप में स्थापित करने में सफल होंगे। वास्तव में अन्य विदेशी भाषाओं के साथ होड लगाने के लिए हिन्दी और देवनागरी लिपि पर्याप्त और सक्षम हैं।
