शोध सारांश

वर्तमान युग सूचना का युग है और इस सूचना के युग में जीवन के लगभग सभी कार्य कंप्यूटर के द्वारा हो रहा है। आज के इस डिजिटल युग एवं भागमभाग भरे दुनिया में हर कोई अपना काम जल्दी एवं आसान ढंग से करना चाहता है।

पहले अनुवाद का कार्य मानव अनुवादक द्वारा किया जाता था जो बहुत महंगा और देरी से अनुवाद होता था। कोई मानव अनुवादक अधिक से अधिक दो या तीन भाषाओं का अनुवाद कर पाने में सक्षम हो पाता था। अनुवादक की उपलब्धता भी आसानी से नहीं होती थी। इसी समस्या को ध्यान में रखते हुए मशीनी अनुवाद का निर्माण हुआ।

इस शोध कार्य में हिंदी-अंगिका मशीन अनुवाद का निर्माण किया जाता है। जो हिंदी के पाठों का अंगिका के पाठों का स्वतः अनुवाद करता है। इस शोध कार्य में जिस मशीनी अनुवाद का निर्माण किया गया है वह शुद्धता के साथ हिंदी भाषा के वाक्यों का अनुवाद अंगिका में बिना भाव बदले करता है।

इस शोध के लिए शोध प्रविधि के रूप में तीन विधियों का प्रयोग किया गया है – भाषावैज्ञानिक शोध प्रविधि, गुणात्मक शोध प्रविधि एवं अभियांत्रिकी शोध प्रविधि।

भाषा वैज्ञानिक शोध प्रविधि के अंतर्गत हिंदी एवं अंगिका का व्यक्तिरेकी अध्ययन किया गया है। जिसमें हिंदी एवं अंगिका का ध्वनि, रूपिम, पद, पदबंध एवं वाक्य स्तर पर अध्ययन किया गया है।

स्वनिमिक स्तर पर हिंदी और अंगिका के कुछ असमानताएं है। जिस प्रकार उपसर्ग और प्रत्यय लगाकर हिंदी में शब्द निर्माण होता है उसी प्रकार अंगिका में भी उपसर्ग और प्रत्यय लगकर शब्द निर्माण होता है। उनके रूप हिंदी और अंगिका में भिन्न-भिन्न होते है जो स्वाभाविक है। अंगिका में 'श','ष'और' स' की जगह 'स' होता है। इसके उदाहरण – रमेश – रमेस, विष-विस। 'ष' की जगह कभी-कभी 'ख' भी होता है। 'र'की जगह कभी 'ड़'और 'ड़' की जगह 'र'हो जाता है। इसी प्रकार 'न' की जगह 'ल' और 'ल' की जगह 'न' होता है।

भाषाविज्ञान के सभी स्तर पर यह देखा गया है कि दोनों भाषाओं में क्या अंतर है और क्या सामान है। हिंदी और अंगिका में शब्द-क्रम समान होते हैं। अंगिका गुणात्मक शोध प्रविधि, शोध की वह विधि है जिसमें शोध में चरों (variables) का उनके गुणों के आधार पर

विश्लेषण किया जाता है। इस शोध में गुणात्मक शोध प्रविधि का प्रयोग भाषा वैज्ञानिक शोध के दौरान सहायक शोध प्रविधि के रूप में किया गया है। इसके लिए गुणात्मक विधि से डेटा का संग्रहण किया गया है। वैज्ञानिक और अभियांत्रिकी शोध प्रविधि के अंतर्गत हम सॉफ्टवेयर अभियांत्रिकी शोध के सभी चरणों के अनुसार शोध कर रहे है। समस्या ढूँढना, योजना बनाना, प्रविधि चुनना ये तो आरंभिक चरण हैं जो अत्यंत आवश्यक है और सभी प्रकार के शोध की शुरूआत इन्हीं से होती है। जब मशीन अनुवाद का विकास की बात आती है तो इसके लिए प्रत्यक्ष विधि का प्रयोग किया गया है। इसके लिए शब्द स्तर का डेटाबेस का निर्माण किया गया है। डेटा बेस में मुख्यतः उन शब्दों का संग्रह किया गया है जिसका रूप हिंदी एवं अंगिका में अलग-अलग होता है। इसके अतिरिक्त जो इस प्रविधि की विशेषता है। प्रोटोटाइप निर्माण, प्रोटोटाइप टेस्टिंग, डिज़ाइन, मूल्यांकन, रखरखाव एवं अपडेट के चरण का मूल्यांकन किया गया है। इस शोध के डेटा संग्रहण में प्राथमिक डेटा एवं द्वितीयक डेटा दोनों चाहिए था। प्राथमिक डेटा संग्रहण के तहत हमने अंगिका भाषी क्षेत्र में जाकर डेटा का संग्रहण किया। द्वितीयक डेटा संग्रहण के तहत अंगिका भाषा में लिखी हुई पुस्तक से शब्दों को लिया गया है और उन शब्दों के समकक्ष हिंदी शब्द लिखा गया है।

इस शोध में जिस मशीनी अनुवाद का निर्माण हुआ है वह शुद्धता के साथ हिंदी के वाक्यों का अनुवाद अंगिका में करता है। यह सरल वाक्य, संयुक्त वाक्य एवं मिश्रित वाक्यों का अनुवाद कर पाने में सक्षम है। यह लगभग 70 प्रतिशत शुद्धता के साथ अनुवाद करता है।