

अंग्रेजी-हिंदी मशीनी अनुवाद तंत्रों की मूल्यांकन पद्धतियाँ

¹ सुमेध खुशालराव हाडके, ² गिरीश नाथ झा

¹ शोध छात्र, भारतीय भाषा केन्द्र, जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली, भारत।

² प्रोफेसर, विशिष्ट संस्कृत अध्ययन केंद्र, जवाहरलाल नेहरू विश्वविद्यालय, नई दिल्ली, भारत।

सारांश

मशीनी अनुवाद शब्द का प्रयोग 'कंप्यूटर अनुवाद' के लिए किया जाता है, जिसमें किसी पाठ या वाक्य का एक प्राकृतिक भाषा से दूसरी प्राकृतिक भाषा में स्वचालित अनुवाद किया जाता है। मशीनी अनुवाद के क्षेत्र में अनुसंधान की शुरुआत 1950 के बाद शुरू हुई थी। उसके बाद से विश्व स्तर पर भाषा वैज्ञानिक और कंप्यूटर वैज्ञानिकों ने विविध भाषाओं के लिए मशीनी अनुवाद के तंत्र विकसित किए गए। साथ में उनके मूल्यांकन के लिए अनेक पद्धतियों का भी विकास किया। यह मूल्यांकन पद्धतियाँ मशीनी अनुवाद तंत्रों के गुणवत्ता की जाँच करने के लिए तथा विशिष्ट तंत्र की अनुवाद प्रक्रिया को बेहतर बनाने के लिए उपयोगी साबित हुई है। इस शोध आलेख में मानव मूल्यांकन और स्वचालित मूल्यांकन पद्धतियों का विवेचन किया गया है जो अंग्रेजी से हिंदी मशीनी अनुवाद तंत्रों के लिए कार्य करती हैं। साथ ही यह भी विवेचन किया गया है कि भारतीय भाषाओं के लिए उपलब्ध मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए कौन सी पद्धति सर्वश्रेष्ठ है।

मूल शब्द: मशीनी अनुवाद, प्राकृतिक भाषा संसाधन, कृत्रिम बुद्धि, मानव मूल्यांकन, स्वचालित मूल्यांकन।

1. प्रस्तावना

सूचना प्रौद्योगिकी के क्षेत्र में पिछले कुछ दशकों में शीघ्रता से विकास हुआ है। इसमें सूचनाओं के संचार के लिए कंप्यूटर का प्रयोग कर मानव और मशीन के बीच संबंध स्थापित किया गया है। इससे कंप्यूटर का प्रयोग अधिक होने लगा है। कंप्यूटरीकरण के बढ़ते प्रयोग से भाषाओं का भी विकास हो रहा है जिसके लिए भाषाओं का स्थानीयकरण (Language Localization) संकल्पना का उदय हुआ है। इसमें स्थानीय भाषाओं का भाषा प्रौद्योगिकी के साथ विकास किया जा रहा है, जिसके लिए विभिन्न भाषाई सामग्री और उपकरणों का निर्माण किया जा रहा है। भारत एक बहुभाषिक देश होने के चलते इसमें भाषाई विकास के लिए देश तथा विदेश में विविध प्रयास किए जा रहे हैं। इसका प्रतिफल है कि आज इन भाषाओं में कई मशीनी अनुवाद तंत्रों का निर्माण किया गया ताकि मनुष्य किसी भी सूचना को तुरंत प्राप्त कर सकें तथा उसका अनुवाद किसी भी भाषा में पढ़ सकें। इसके लिए मशीनी अनुवाद आज बहुत कारगर साबित हुआ है। मुख्यतः मशीनी अनुवाद कृत्रिम बुद्धि और प्राकृतिक भाषा संसाधन का क्षेत्र है, जिसमें एक भाषा से दूसरी भाषा में अनुवाद के लिए कंप्यूटर का प्रयोग किया जाता है। आज अनुवाद के लिए कई मशीनी अनुवाद के तंत्र विकसित हुए हैं। लेकिन यह अनुवाद तंत्र अनुवाद करने के लिए पूर्णतया सफल नहीं हैं, इनसे अनुवाद में अक्सर गलतियाँ होती हैं। अनुवाद केवल एक भाषा से दूसरे भाषा में रूपांतरण ही नहीं है बल्कि उसमें मूल भाषा के कथ्य का संप्रेषण लक्ष्य भाषा में होना चाहिए साथ ही अनुवाद बोधगम्य और अच्छी गुणवत्ता का होना चाहिए। इसलिए अनुवाद की गुणवत्ता को जाँचने के लिए कुछ मूल्यांकन पद्धतियों का विकास किया गया है। मशीनी अनुवाद का मूल्यांकन करना कठिन कार्य माना जाता है क्योंकि इसमें अनुवाद तंत्रों द्वारा अनूदित वाक्यों या कार्पस का स्वचालित मूल्यांकन करना होता है और फिर उसका मानव मूल्यांकन के साथ तुलना करना होता है, जिसमें समय और श्रम दोनों की आवश्यकता होती है। पहले यह मूल्यांकन कार्य केवल मानव द्वारा ही किया जाता था। वर्तमान में स्वचालित मूल्यांकन पद्धतियों का प्रयोग अधिक किया जा रहा है।

यूरोपीय भाषाओं में अनेक मशीनी अनुवाद तंत्र विकसित किए गए हैं और उनके मूल्यांकन के लिए कई मूल्यांकन पद्धतियाँ भी विकसित की गई हैं। यही मूल्यांकन पद्धतियों का प्रयोग भारतीय भाषाओं के लिए किया जाए तो मूल्यांकन के परिणाम सही प्राप्त नहीं होंगे। भारतीय भाषाएँ अत्यधिक विभक्ति-प्रत्ययों वाली भाषाएँ हैं, जिसमें अधिक रूप-रचना, अपेक्षाकृत शब्दक्रम मुक्त और कर्ता-कर्म-क्रिया के वाक्य संरचना पर आधारित है। इसके अलावा इसमें कई शैलीगत भेद भी हैं। अंग्रेजी भाषा रूप-रचना की दृष्टि से अधिक स्थिर भाषा है तथा इसकी वाक्य संरचना कर्ता-क्रिया-कर्म पर आधारित है और हिंदी भाषा की वाक्य संरचना कर्ता-कर्म-क्रिया पर आधारित है। इन दोनों भाषाओं की संरचना अलग होने के वजह से मशीनी अनुवाद तंत्र निर्माण में भी विभिन्न सामग्री तथा उपकरणों का समावेश किया जाता है इसलिए इनका मूल्यांकन करना अनुवाद के दृष्टि से आवश्यक हो जाता है।

2. मशीनी अनुवाद मूल्यांकन

आज मशीनी अनुवाद तंत्रों के अनुवाद की गुणवत्ता की जाँच करने तथा अनुवाद परीक्षण के लिए अनेक मूल्यांकन पद्धतियों का प्रयोग किया जाता है। जब मार्केट में कोई नया उत्पाद आता है तो कोई भी कंपनी अपने उत्पाद की जाँच कर उसका परीक्षण करती है। वैसे ही, मशीनी अनुवाद तंत्रों के मूल्यांकन की आवश्यकता होती है। मशीनी अनुवाद का मूल्यांकन एक वाक्य के स्तर पर और दूसरा कार्पस के स्तर पर किया जाता है। इसमें अधिकतर वाक्य स्तर पर मूल्यांकन किया जाता है, क्योंकि इससे मूल्यांकन के परिणाम सही प्राप्त होते हैं। मशीनी अनुवाद के मूल्यांकन को भी दो श्रेणियों में विभाजित किया गया है— मानव मूल्यांकन (Human evaluation) और स्वचालित मूल्यांकन (Automatic evaluation)।

2.1 मानव मूल्यांकन (Human Evaluation)

मशीन द्वारा अनूदित वाक्यों का मूल्यांकन जब मनुष्यों द्वारा किया जाता है तो इसे मानव मूल्यांकन कहते हैं। मानव मूल्यांकन के लिए मानव अनुवादक की आवश्यकता होती है जो दोनों भाषाओं का

जानकार होता है। मानव मूल्यांकनकर्ता तंत्रों का मूल्यांकन करते समय अनुवाद की बोधगम्यता, सहजता और अर्थ की उपयुक्तता को देखते हुए मूल्यांकन करता है। मानव मूल्यांकन सभी मूल्यांकन पद्धतियों में अधिक उपयुक्त पद्धति मानी जाती है। लेकिन यह समय और श्रम के दृष्टि से कठिन कार्य है जिसमें एक सप्ताह या एक माह भी लग सकता है। साथ ही मूल्यांकन सांख्यिकीय परिणाम की दृष्टि से निष्पक्ष होना चाहिए। इसके लिए मशीन से प्राप्त अनुवाद का मूल्यांकन एक से अधिक मूल्यांकनकर्ता द्वारा किया जाना चाहिए ताकि मूल्यांकन सही हो।

मानव मूल्यांकन के लिए अनेक मूल्यांकन पद्धतियों में श्रेणियों (Grading) का प्रयोग किया जाता है। उसमें 7 पॉइन्ट ग्रेडिंग स्केल, 5 पॉइन्ट ग्रेडिंग स्केल, अल्पैक द्वारा 0-9 ग्रेडिंग स्केल और एडवांस्ड रिसर्च प्रोजेक्ट्स एजेंसी (एआरपीए) द्वारा 0-5 ग्रेडिंग स्केल आदि शामिल हैं। इसमें प्रमुख रूप से मानव मूल्यांकन के लिए प्रारंभ में सन 1966 में अल्पैक समिति [Automatic Language Processing Advisory Committee (ALPAC)] के रिपोर्ट में मानव अनुवाद के साथ मशीनी अनुवाद का तुलनात्मक अध्ययन के लिए सुबोधता (Intelligibility) और परिशुद्धता (Fidelity) को मानदंड मानते हुए (0-9) ग्रेडिंग स्केल का प्रयोग किया गया था। इस पद्धति के अध्ययन से ज्ञात हुआ कि मानव और मशीनी अनुवाद की गुणवत्ता का मूल्यांकन के लिए अत्यधिक विश्वसनीय आकलन इस पद्धति से किया जा सकता है।

एडवांस्ड रिसर्च प्रोजेक्ट्स एजेंसी ने [Advanced Research Projects Agency (ARPA)] 1991 में मानव भाषा प्रौद्योगिकी कार्यक्रम के अंतर्गत मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए एक पद्धति विकसित की थी। जिसमें मूल्यांकन के लिए विभिन्न सांख्यिकीय, नियम आधारित और मानव-सहायता आधारित दृष्टिकोण का परीक्षण किया गया था। इसमें मूल्यांकन के लिए अनुवाद में सहजता (Fluency) और पर्याप्तता (Adequacy) को मानदंड मानते हुए (0-5) ग्रेडिंग स्केल का प्रयोग किया गया था। उसके बाद मशीनी अनुवाद से निर्गत अनुवाद का मानव मूल्यांकन के लिए RPA मूल्यांकन कार्यक्रम ने इसे मानक पद्धति के रूप में प्रयोग किया गया था इसलिए मानव मूल्यांकन के लिए यह पद्धति भी प्रचलित रही है।

सामान्यतः मानव मूल्यांकन की मानक श्रेणियों में (0-4) स्केलिंग का प्रयोग किया जाता है। जिसमें 4-उत्कृष्ट अनुवाद के लिए (All meaning or perfect), 3-अच्छे अनुवाद के लिए (most meaning or Good), 2-समझने योग्य अनुवाद के लिए (much meaning or Comprehensible but has quite a few errors), 1-समझने में कठिन (little meaning) और 0-अस्वीकार्य/अबोधगम्य (Incomprehensible) इत्यादि स्केलिंग स्कोर का प्रयोग होता है। मानव मूल्यांकनकर्ता लक्ष्य भाषा के अनुवाद को देखता है और उसे जाँचने के बाद तय करता है कि अनुवाद सही है या गलत। उस आधार पर प्रत्येक वाक्य को स्कोर देता है। मानव मूल्यांकन की सबसे महत्वपूर्ण विलक्षणता यह है जब दो मानव मूल्यांकनकर्ता एक ही पाठ को देखते हुए दो अलग-अलग मूल्यांकन करते हैं तो उसके मूल्यांकन के स्कोर अलग-अलग आते हैं। इसका अर्थ यह है कि मशीन से निर्गत अनुवाद का मूल्यांकन व्यक्तिपरक निर्भर करता है। फिर भी मानव द्वारा मूल्यांकन स्वचालित मूल्यांकन से अधिक विश्वसनीय माना जाता है। इसमें कुछ समस्याओं के चलते तथा इस मूल्यांकन का पुनर्प्रयोग न होने के वजह से वर्तमान में स्वचालित मूल्यांकन पद्धतियों का प्रयोग अधिक किया जाता है जो अधिक शीघ्रता से मशीनी अनुवाद का मूल्यांकन करती है।

2.2 स्वचालित मूल्यांकन (Automatic Evaluation)

जब अनुदित पाठ का मूल्यांकन मशीन की सहायता से गणितीय

आंकड़ों द्वारा किया जाता है तो उसे स्वचालित मूल्यांकन या मशीनी मूल्यांकन कहते हैं। स्वचालित मूल्यांकन वाक्य स्तर और कॉर्पस स्तर पर किया जाता है। वाक्य स्तर पर मूल्यांकन में अनुदित वाक्य के समूह का एल्गोरिथम द्वारा स्कोर निकालकर गणना की जाती है फिर उसका मानव मूल्यांकन के साथ तुलना की जाती है। उसी तरह कॉर्पस स्तर पर मूल्यांकन में वाक्यों से अधिक स्कोर के लिए मानव निर्णय (Human Judgment) और मेट्रिक निर्णय (Metric judgment) को एकत्रित किया जाता है और फिर कुल स्कोर से मूल्यांकन के परिणाम निकाले जाते हैं। इससे मशीनी अनुवाद तंत्र का मूल्यांकन होता है। मशीनी अनुवाद का स्वचालित मूल्यांकन प्रमुख रूप से अनुवाद की बोधगम्यता, सटीकता, सहजता और पर्याप्तता आदि घटकों पर निर्भर करता है। इन मानदंडों के द्वारा मूल्यांकन किया जाता है।

स्वचालित मूल्यांकन यह मानव मूल्यांकन की तुलना में अधिक शीघ्र और कम खर्च पर होता है जो किसी भी भाषाओं के मशीनी अनुवाद तंत्रों का मूल्यांकन करने के लिए आसान है। इसका प्रयोग वाक्य स्तर पर त्रुटियों के विश्लेषण, मशीनी अनुवाद तंत्र की तुलना करने और तंत्र को अनुकूल (System optimization) करने के लिए किया जाता है। कहा जाता है कि इसकी ग्रेडिंग मानव मूल्यांकन की तरह सही नहीं होती। लेकिन इसके मूल्यांकन के परिणाम मानव मूल्यांकन से अधिक निष्पक्ष रूप में मिलते हैं।

प्रथमतः मशीनी अनुवाद का मूल्यांकन करने के लिए 1956 में मिल्लर और बीबे-सेंटर ने (Miller G.A. & Beebe-Center J.G.) और 1965 में पफ्लिन ने (Pfallin S.M.) मानव मूल्यांकन को प्रस्तावित किया था। जिससे मशीनी अनुवाद का मूल्यांकन मानव द्वारा होने लगा लेकिन इसमें समय और श्रम की बचत के चलते बाद में स्वचालित मूल्यांकन पद्धतियों का प्रयोग अधिक होने लगा। वर्तमान में मशीनी अनुवाद के मूल्यांकन के लिए अनेक स्वचालित मूल्यांकन पद्धतियाँ हैं जिनका प्रयोग विभिन्न मशीनी अनुवाद तंत्र करते हैं। इसमें प्रमुख रूप से ब्लू (BLEU), निस्ट (NIST), मेटियोर (Meteor), एफ-परिमाण पद्धति (F-measure) और मेटियोर-हिंदी (Meteor-Hindi) आदि मूल्यांकन पद्धतियाँ शामिल हैं, जिनका नीचे विवेचन किया गया है।

2.2.1 ब्लू (Bleu)

ब्लू मूल्यांकन पद्धति को आई.बी.एम ने विकसित किया है। यह पहली मेट्रिक पद्धति है जो मशीनी अनुवाद के मूल्यांकन का मानव मूल्यांकन की गुणवत्ता के साथ तुलना करने के लिए प्रयोग की जाती है। वर्तमान में इसका स्वचालित मूल्यांकन के लिए सबसे अधिक प्रयोग किया जाता है। इसमें एन-ग्राम के आधार पर अलग-अलग वाक्यों के स्कोर की गणना करने के लिए ज्यामितीय साधन का प्रयोग करती है और स्कोर की गणना करने के बाद फिर अंतिम स्कोर के लिए पूरे कॉर्पस से गणितीय आंकड़ों द्वारा औसत स्कोर निकाला जाता है, जिससे मूल्यांकन के परिणाम निकलते हैं। इस मेट्रिक पद्धति से गणना निम्नलिखित सूत्र द्वारा की जाती है।

$$BLEU = \min \left(1, \frac{MT \text{ output-length}}{\text{reference-length}} \right) \times \exp \left(\sum_{n=1}^N w_n \log(p_n) \right)$$

उपरोक्त सूत्र में दिया है कि मशीनी अनुवाद से निर्गत अनुवाद और संदर्भ अनुवाद के माध्यम से ब्लू के परिणाम निकलते हैं। इसका प्रयोग अंग्रेजी से हिंदी मशीनी अनुवाद तंत्र मूल्यांकन के लिए अधिक करते हैं जिससे मूल्यांकन के सटीक परिणाम प्राप्त होते हैं।

2.2.2 निस्ट (NIST)

निस्ट यह ब्लू मेट्रिक का संशोधित रूप है। इसका विकास राष्ट्रीय मानक एवं प्रौद्योगिकी संस्थान, अमेरिका ने किया है इसी संस्थान

का ही नाम इस मूल्यांकन पद्धति को दिया गया है। यह मूल्यांकन पद्धति BLEU मेट्रिक पर आधारित है, लेकिन इसमें कुछ परिवर्तन किया है। इसमें एन-ग्राम (N-gram) की गणना करते समय प्रत्येक एन-ग्राम को समान महत्व देने के बजाय उनके अनुवाद की बोधगम्यता के अनुसार महत्व दिया जाता है। ब्रेविटी पेनल्टी (Brevity penalty) का प्रयोग कर निर्गत अनुवाद से प्रति शब्द के लिए एक एन-ग्राम का प्रयोग करते हुए फिर उसे संदर्भ अनुवाद के एन-ग्राम के साथ मिलान (match) किया जाता है और अंत में औसत स्कोर निकाला जाता है। इसमें स्कोर के लिए 1 से 5 संख्या का प्रयोग किया जाता है। इसमें गणना निम्नलिखित सूत्र द्वारा की जाती है।

$$NIST = BP * \sum_{n=1}^N \left(\frac{\sum_{n\text{-gram} \in \text{matching}} \text{match}(n\text{-gram})}{\sum_{n\text{-gram}' \in \text{translation}} \text{match}(n\text{-gram}')} \right)$$

2.2.3 मेटियोर (Meteor)

मेटियोर मूल्यांकन पद्धति को कार्नेगी मेलॉन विश्वविद्यालय ने विकसित किया है। यह यूनियोग्राम सुव्यतता (Precision) और यूनियोग्राम स्मरण क्षमता (recall) पर आधारित ग्रेडिंग के लिए (0-4) स्कोर का प्रयोग किया जाता है। इसे ब्लू की कमियों को दूर करने के लिए बनाया गया है। इस मेट्रिक पद्धति में मशीनी अनुवाद एवं मानवीय संदर्भ अनुवाद में कई तरह से मिलान (Matching) किया जाता है। संदर्भ पाठ या अनुवाद पाठ के प्रत्येक स्तर पर मूल शब्दों के साथ मिलान किया जाता है जो मशीनी अनुवाद तथा संदर्भ अनुवाद में समान हो। संदर्भ अनुवाद से पर्यायवाची शब्दों के प्रयोग की गणना करने के लिए वर्डनेट (Wordnet) का प्रयोग किया जाता है। इस मेट्रिक पद्धति का कार्यान्वयन (Implementation) मॉड्यूलर के रूप में होता है क्योंकि एल्गोरिदम में जो शब्द मेल खाते हैं उस शब्द को मॉड्यूलर में परिवर्तित किया जाता है और मिलान न होने पर उसमें किसी शब्द को जोड़ने की सुविधा भी होती है। अंत में इन सभी ग्रेडिंग का औसत निकाल कर अंतिम स्कोर प्राप्त किया जाता है।

2.2.4 एफ-परिमाण पद्धति (F & Measure Metric)

इसे न्यूयॉर्क विश्वविद्यालय ने विकसित किया है। मानव द्वारा मूल्यांकित वाक्यों और संदर्भ वाक्यों के मिलान के लिए एन-ग्राम (N-gram) स्कोर की गणना की जाती है जिसमें एन-ग्राम के अधिकतम मिलान (Maximum matching) के लिए ग्राफ थ्योरी (Graph theory) का प्रयोग किया जाता है। इसमें मूल अनुवाद के साथ मानव द्वारा अनुवाद कितना सही है यह जाँचने के लिए एन-ग्राम परिणाम के दोनों के स्कोर के साथ तुलना की जाती है। साथ ही दोनों के बीच समान घटकों का मूल्यांकन कर विश्लेषण किया जाता है। शुद्धता (Precision) और स्मरण क्षमता (Recall) का महत्वपूर्ण योगदान इस पद्धति में किया गया है। इस मूल्यांकन पद्धति का प्रयोग अंग्रेजी से हिंदी मशीनी अनुवाद तंत्रों के मूल्यांकन के लिए भी किया जाता है।

2.2.5 मेटियोर-हिंदी परिमाण पद्धति (Meteor-Hindi Metric)

यह मेटियोर का एक संशोधित संस्करण है, जिसमें मूल्यांकन के दृष्टि से हिंदी के लिए विशेष गुणों का समावेश किया है। अंग्रेजी से हिंदी मशीनी अनुवाद तंत्र के मूल्यांकन के लिए इसकी सहायता ली जाती है। इसके मूल्यांकन के स्कोर की मानव मूल्यांकन स्कोर के साथ तुलना की जाती है। मेटोयोर-हिंदी का स्कोर (0.703) ब्लू (0.271) से मिलान करने से ज्ञात होता है की मेटोयोर-हिंदी के परिणाम काफी अधिक है। अंग्रेजी से हिंदी अनुवाद के मूल्यांकन के

लिए इसे अधिक गुणवत्ता आधारित बनाने के लिए मेटोयोर-हिंदी में निम्न मॉड्यूल का समावेश किया गया है। 1) स्थानीय शब्द समूह (LWG): जिसमें मूल शब्द और उसके संबंधित प्रकार्य शब्द शामिल है। 2) उपवाक्य मिलान (Clause Match): उपवाक्य को एक वाक्यांश के रूप में परिभाषित किया जाता है, जिसमें कम से कम एक क्रिया और एक कर्ता शामिल है। 3) शब्दभेद मिलान (POS Matcher) से प्रत्येक शब्द की गणना की जाती है साथ ही पीओएस टैग से भी गणना की जाती है। इस संशोधित मेटोयोर-हिंदी मूल्यांकन पद्धति का प्रयोग हिंदी में उचित मूल्यांकन के लिए किया जाता है।

यह सभी स्वचालित मूल्यांकन पद्धतियों का प्रयोग मशीनी अनुवाद मूल्यांकन के लिए किया जाता है। जिसमें मशीनी अनुवाद तंत्र के गुणवत्ता की जाँच करने के लिए संख्यात्मक मेट्रिक पद्धति का प्रयोग किया जाता है इससे मानव अनुवाद के साथ तुलना कर कुछ हद तक मूल्यांकन के परिणाम प्राप्त किए जा सकते हैं। इसके लिए अधिक से अधिक गणितीय सूत्रों या एल्गोरिथम का प्रयोग किया जाता है। ये सभी पद्धतियों का प्रयोग अंग्रेजी से हिंदी में उपलब्ध मशीनी अनुवाद तंत्रों का मूल्यांकन करने के लिए किया जाता है। इनसे पाए जानेवाले परिणाम निष्पक्षता से प्राप्त होते हैं इसलिए इनका प्रयोग मूल्यांकन के लिए अधिक किया जाता है।

3. निष्कर्ष

निष्कर्ष रूप में कहा जा सकता है कि मानव मूल्यांकन और स्वचालित मूल्यांकन पद्धतियों का प्रयोग विभिन्न मशीनी अनुवाद तंत्रों के अनुवाद की गुणवत्ता की जाँच करने के लिए करते हैं। वास्तव में मानव द्वारा मूल्यांकन में अधिक समय और धन खर्च होता है तथा इसके निष्पक्षता पर संदेह होता है इसलिए इसका प्रयोग किसी विशेष मूल्यांकन के लिए किया जाता है। भारतीय भाषाओं के लिए उपलब्ध मशीनी अनुवाद तंत्रों का मूल्यांकन के लिए मानव मूल्यांकन पद्धति ही अधिक विश्वसनीय है। भारतीय भाषाओं में अक्सर अनेकार्थता की समस्या आती है जिसे केवल मानव ही मूल्यांकित कर सकता है। फिर भी भारतीय भाषाओं के लिए उचित स्वचालित मूल्यांकन पद्धति को विकसित करने की आवश्यकता है। स्वचालित मूल्यांकन पद्धतियाँ मूल्यांकन के दृष्टि से सही हैं, जिससे समय की बचत होती है और मूल्यांकन के डेटा का पुनर्प्रयोग किया जा सकता है। यह पद्धतियाँ यूरोपीय भाषाओं में उपलब्ध तंत्रों के मूल्यांकन के लिए सबसे उत्तम है जिससे अनुवाद की गुणवत्ता और बोधगम्यता की जानकारी मिलती है। इस जानकारी का प्रयोग तंत्र निर्माण तथा उसे अद्यतन (updatation) करने के लिए किया जाता है। अंतः में कह सकते हैं कि स्वचालित मूल्यांकन पद्धतियाँ अनुवाद की गुणवत्ता को परखने का सही तरीका है।

4. संदर्भ ग्रंथ

1. Malik Pooja, Gupta Abhilasha, Baghel Anurag, Key Issues in Machine Translation Evaluation of English-Indian Languages, International Journal of Engineering Research & Technology (IJERT). 2013; 2(10):3244-3245.
2. Malik Pooja, Baghel Anurag Singh. An Improvement in BLEU Metric for English-Hindi Machine Translation Evaluation, International Conference on Computing, Communication and Automation (ICCCA2016), 331-332.
3. https://en.wikipedia.org/wiki/Evaluation_of_machine_translation#Automatic_Language_Processing_Advisory_Committee (Retrieved on 16.06.2017)
4. https://en.wikipedia.org/wiki/Evaluation_of_machine_translation#Advanced_Research_Projects_Agency_.28ARP_A.29 (Retrieved on 16.06.2017)

5. Kalyani Aditi, Kumud Hemant, Singh Shashi Pal, Kumar Ajai, Darbari Hemant, Evaluation and Ranking of Machine Translated Output in Hindi Language using Precision and Recall Oriented Metrics, International Journal of Advanced Computer Research. 2014; 4(1):14, 54-55.
6. Avinesh PVS. Transfer Grammar Engine and Automatic Learning of Reorder Rules in Machine Translation, M.Sc. Computer Science Thesis. 2010, 14-15.
7. Miller G.A, Beebe-Center J.G. Some Psychological Methods for Evaluating the Quality of Translation, Mechanical Translations. 1956; 3:73-80.
8. Pfafflin S.M. Evaluation of Machine Translations by Reading Comprehension Tests and Subjective Judgments, Mechanical Translation and Computational Linguistics. 1956; 8:2-8.
9. Papineni K, Roukos S, Ward T, Zhu WJ. Bleu: a method for automatic evaluation of machine translation, RC22176 Technical Report, IBM T.J. Watson Research Center, 2001.
10. Doddington G. Automatic evaluation of machine translation quality using n-gram co occurrence statistics. Proceedings of the Human Language Technology Conference (HLT), San Diego. 2002, CA p. 128-132.
11. Denkowski M, Lavie A. Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems, Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation. 2011, 85-90.
12. Tomer N, Sinha D, Rai PK. Evaluating Machine Translation Evaluation's F-Measure Metric for English to Hindi Language Machine Translation. International Journal of Academy Research Computer Engineering and Technology. 2012; 1(7):151-156.
13. Ankush G, Sriram V, Rajeev S. METEOR-Hindi: Automatic MT Evaluation Metric for Hindi as a Target Language, Proceedings of ICON-2010: 8th International conference on Natural language processing, Macmillan Publishers, India.