# मशीन अनुवाद की गुणवत्ता में सुधार
# Quality Improvement of Machine Translation

**दीप्ति चोपड़ा[1], निशीथ जोशी[2], इति माथुर[3]**
**Deepti Chopra, Nisheeth Joshi, Iti Mathur**
कम्प्यूटर साइंस विभाग, वनस्थली विद्यापीठ, राजस्थान, भारत
deeptichopra11@yahoo-co-in
nisheeth-joshi@rediffmail-com
mathur_iti@rediffmail-com

**सारांश**

मशीन अनुवाद का मुख्य उद्देश्य भाषाओं द्वारा उत्पन हुए अवरोधों को दूर करना है। इस लेख में हमने चर्चा की है कि किस प्रकार सरलता के साथ मशीन अनुवाद की गुणवत्ता में सुधार किया जा सकता है। इसमें हमने अंग्रेजी से हिंदी में अनुवाद का प्रयोग किया है और इसका मूल्यांकन मानव एवं स्वचालित मेट्रिक द्वारा किया है।

**ABSTRACT**

Main aim of machine translation is the removal of barriers that arise due to languages. In this article we have discussed how with ease we can improve the quality of machine translation. For this we have used English to Hindi machine translation and evaluated it using human and automatic metrics.

**मुख्य शब्द :** मशीन अनुवाद, टेक्स्ट पुनर्लेखन, मूल्यांकन

**Key words :** Machine Translation, Text Rewriting, Evaluation

## परिचय

मशीन अनुवाद एक भाषा से दूसरी अन्य भाषा में अनुवाद के कार्य [15][16][18][21] को कहा जाता है। अनुवाद का कार्य व्यापारियों द्वारा नए स्थानों की यात्रा करने के लिए और उस क्षेत्र के लोगों के साथ बातचीत करने के लिए इस्तेमाल किया जाता था। हर कोई ज्ञान या जानकारी का भंडार है तथा किसी की भावनाओं, विचारों, राय, कुछ तथ्यों को आम जनता के साथ व्यक्त करने के लिए मशीन अनुवाद द्वारा सुलभ बना सकते हैं। भारतीय संदर्भ में, अनुवाद पाली, प्राकृत, देवनागरी और अन्य क्षेत्रीय भाषाओं में पवित्र ग्रंथों के अनुवाद के साथ शुरू हुआ। यह दुनिया भर में नैतिक मूल्यों, लोकाचार, परंपरा, मान्यताओं और संस्कृति के संचारण करने में मदद करता है। आज के आधुनिक युग में, सूचना प्रोद्योगिकी के विकास के साथ दुनिया सिमट गयी है। अत्याधुनिक प्रौद्योगिकी की वर्तमान स्थिति के साथ, दुनिया वैश्विक गांव में तब्दील हो गयी है और इस वर्तमान परिदृश्य में, अनुवाद की भूमिका अतीत की तुलना में और भी अधिक महत्वपूर्ण हो गयी है। आज हम दुनिया के किसी कोने में बैठे हमारे मित्र से बिना उसकी भाषा जाने संपर्क कर सकते हैं।

मशीन अनुवाद की गुणवत्ता में सुधार के कई तरीकों में से एक टेक्स्ट पुर्नलेखन भी हैं। इसके अलावा नेम एंटिटी की पहचना (एनईआर) भी एक विकल्प है। एनईआर में व्यक्तिवाचक संज्ञाओं या संस्थाओं के नाम को ढूँढने और उन्हें विभिन्न श्रेणियों में वर्गीकृत करने का काम [5][6][7][8] किया जाता है। अनुवाद के समय वाक्य सरल या जटिल हो सकता है। अनुवाद के समय मशीन को जटिल वाक्यों को अनुवाद करने में कठनाई होती है तथा सरल वाक्यों का अनुवाद आसानी

से हो जाता है। अगर अनुवाद से पहले मशीन जटिल वाक्य को परख कर उसको सरल वाक्य में परिवर्तित कर दे तो मशीन अनुवाद की गुणवत्ता में काफी सुधार हो जायेगा। हालांकि इस पुर्नलेखन से वाक्य की संख्या बढ़ जाती है, लेकिन यह भी मशीन अनुवाद प्रणाली के प्रदर्शन को बेहतर बनाता है। इस लेख में, हमने इसी पर चर्चा की है कि कैसे टेक्स्ट पुर्नलेखन से मशीन अनुवाद की गुणवत्ता में सुधार लाया जा सकता है।

### मशीन अनुवाद के विभिन्न प्रकार

मशीन अनुवाद के लिए इस्तेमाल किये गए नीचे दिये दृष्टिकोण हैं:

1 नियम आधारित मशीन अनुवाद
  1.1 स्थानांतरण आधारित मशीन अनुवाद
  1.2 इन्टरलिंगुअल मशीन अनुवाद
  1.3 शब्दकोश आधारित मशीन अनुवाद
2 सांख्यिकीय मशीनी अनुवाद
3 हाइब्रिड मशीन अनुवाद

नियम आधारित मशीन अनुवाद में, स्रोत के बारे में भाषाई जानकारी और लक्ष्य टेक्स्ट अर्थ और वाक्यात्मक विश्लेषण का उपयोग कर अनुवाद किया जाता है। इस तकनीक में, हम लक्ष्य टेक्स्ट के साथ स्रोत टेक्स्ट की संरचना करते हैं। स्थानांतरण आधारित मशीन अनुवाद में प्रयुक्त भाषा जोड़ी का उपयोग किया जाता है, जिससे कि लक्ष्य भाषा में अनुवाद किया जा सकता है। इनटर्रलिंगुअल मशीन अनुवाद में, स्रोत टेक्स्ट सबसे पहले ईन्टरलिंगुआ लक्ष्य भाषा में बदल दिया जाता है और फिर वह लक्ष्य भाषा में परिवर्तित कर दिया जाता है। वर्तमान में शब्दकोश प्रविष्टियों के रूप में शब्दकोश आधारित मशीन अनुवाद भी किया जाता है।

सांख्यिकीय मशीन अनुवाद समानांतर कोष का उपयोग करता है, जिसमें सांख्यिकीय पद्धति का उपयोग करके अनुवाद शामिल है। उदाहरण आधारित मशीन अनुवाद में अनुवाद के लिए कुछ घटकों को चुना जाता है, और उसके बाद इन घटकों को लक्ष्य भाषा में अनुवाद किया जाता है और फिर सभी अनुवादित घटकों को पुनः व्यवस्थित कर दिया जाता है। हाइब्रिड मशीन अनुवाद मशीन में सभी प्रकार के अनुवाद के तरीकों को जोड़ कर अनुवाद किया जाता है, जो एक बेहतर अनुवाद प्रस्तुत करता है।

### पूर्व कार्य

Aluisio एवं अन्य [1] के द्वारा पठनीयता मूल्यांकन तकनीक की मदद से टेक्स्ट सरलीकरण प्रदर्शन किया गया है। उन्होंने सरलीकरण दो प्रकार (प्राकृतिक और मजबूत) से किया है। प्राकृतिक सरलीकरण में बुनियादी स्तर पर लोगों द्वारा सरलीकृत टेक्स्ट समझा जा सकता है। मजबूत सरलीकरण में मौलिक स्तर पर लोगों द्वारा सरलीकृत टेक्स्ट समझा जा सकता है। उन्होंने अंग्रेजी में टेक्स्ट बेहतर समझने के लिए, सात नए मैट्रिक्स विकसित किये हैं। उनका पियर्सन कोरिलेशन 0.8502 और मीन एब्सोलुट एरर 0.3478 थी।

Bott और Saggion (2009) [2] ने स्पेनिश के लिए सरल वाक्य युक्त समानांतर कोष का निर्माण किया है। उन्होंने प्रशिक्षण के लिए 2456 शब्दों व 110 वाक्यों का इस्तेमाल किया है और परीक्षण के लिए 1840 शब्दों व 145 वाक्यों का इस्तेमाल किया है। उनका रिकोल स्कोर करीब 80.9% था।

Candido जूनियर एवं अन्य [3] ने ब्राज़ीलियाई-पुर्तगाल के लिए नियम आधारित टेक्स्ट सरलीकरण का विकास किया है। उन्होंने पार्स संरचना को सरल बनाने के लिए सात नियमों का इस्तेमाल किया है। उन्होंने सिम्प्लिफिका नामित संलेखन उपकरण विकसित भी किया है। उनकी इस प्रक्रिया से एफ स्कोर 57.62 उत्पन हुआ है।

Belder एवं अन्य [4] ने टेक्स्ट सरलीकरण की एक विधि दी जिसमें उन्होंने शाब्दिक सरलीकरण और वाक्य की रचना के आधार पर सरलीकरण किया। इन दोनों में से जो ज्यादा बेहतर था उसको अंत में आगे की प्रक्रिया के लिए चुना गया। ऐसा करने के लिए उन्होंने इन्टिजर लीनियर प्रोग्रामिंग का उपयोग किया। इससे उनके वर्ड सेंस डिसअम्बिगुएशन में 11.7% का सुधार हुआ।

Glavas एवं अन्य [9] ने टेक्स्ट सरलीकरण प्रदर्शन के लिए अप्रासंगिक या महत्वहीन जानकारी की तथ्यात्मक घटनाओं और उन्मूलन निकालने की तकनीक का इस्तेमाल किया। उन्होंने एक घटना संकर्षण का विकास किया जिसमें घटना एंकर और घटना तर्क और दो तकनीकों, वाक्य सरलीकरण और इवेंट सरलीकरण, के द्वारा सरलीकृत टेक्स्ट को मूल टेक्स्ट से प्राप्त किया। उन्होंने ईएमएम समाचारपत्र से 100 समाचारों पर प्रयोग कर अपनी

प्रणाली का मूल्यांकन किया । मूल्यांकन मैन्युअल रूप से किया गया है। सर्वनाम एनाफोरा तकनीक में भी इसने बेहतर परिणाम दिए।

Jonnalagadda एवं अन्य [10] ने बायो-सिम्पलिफय नामक एक टेक्स्ट सरलीकरण प्रक्रिया का प्रस्ताव दिया है, उन्होंने गैर वाक्यात्मक परिवर्तन और वाक्यात्मक परिवर्तन का उपयोग कर टेक्स्ट सरलीकरण का प्रदर्शन किया। उन्होंने 1,100 वाक्यों पर परीक्षण किया और 1159 सरलीकृत वाक्य प्राप्त किये। उनके प्रयोग में 78. 86% का एफ स्कोर हासिल किया, और 2.9% निरपेक्ष त्रुटि कमी प्राप्त की। लिंक व्याकरण का उपयोग करके, उन्होंने 4.23% निरपेक्ष त्रुटि कमी प्राप्त की।

Klaper एवं अन्य [11] ने वेब से जर्मन या सरल जर्मन दस्तावेजों की मदद से समानांतर कोष विकसित किया। सांख्यिकीय मशीन अनुवाद प्रणाली में समानांतर कोष का इस्तेमाल करके जटिल वाक्य से सरलीकृत वाक्य उत्पन किये। उन्होंने जर्मन कोष का इस्तेमाल किया, जिसमें लगभग 7000 वाक्य (70000 टोकन) थे। उनके दृष्टिकोण से प्राप्त प्रेसिजन और रिकाल क्रमशः 27.7% और 5% हैं।

Klebanov एवं अन्य [12] ने प्राकृतिक भाषा टेक्स्ट को सरल बनाने की विधि को प्रस्तुत किया। इस विधि से शाब्दिक जानकारी प्राप्त की जा सकती थी। उनकी प्रस्तावित एल्गोरिथ्म के अनुसार, सबसे पहले नाम संस्थाओं (व्यक्ति के नाम) बीबीएन के पहचानकर्ता का उपयोग करते हुए पाया जाता है और फिर डिपेंडेसी संरचना मिनीपार का उपयोग कर प्राप्त की जाती है। उन्होंने इसका परीक्षण 128 वाक्यों पर किया गया और 55% प्रेसिजन प्राप्त किया।

Klerke एवं अन्य [13] ने डेनिश समाचार से सरलीकृत वाक्य उत्पन्न कर समानांतर कोष का निर्माण किया। प्रेसिजन और रिकाल टीफ (टर्म फ्रिक्वेंसी) आईडीएफ और कोसाईन समानता स्कोर के अनुसार उनको स्कोर रहा 90.8% और 84.7%।

Mirkin एवं अन्य [14] ने मशीनी अनुवाद की गुणवत्ता में सुधार लाने के लिए स्रोत पुनर्लेखन प्रणाली प्रस्तुत की। प्रारंभ में, स्रोत वाक्य को लक्ष्य वाक्य में मशीन अनुवाद प्रणाली द्वारा अनुवाद कर दिया गया और फिर इस स्रोत और लक्ष्य वाक्य की गुणवत्ता का आकलन किया गया। ऐसा करने के लिए उन्होंने एक मशीन लर्निंग प्रणाली का विकास किया जो सही ढ़ंग से लक्ष्य वाक्य को स्रोत वाक्य में अनुवाद करने का एक कांफिडेंस आंकलन देती है। उन्होंने 2008 के समाचारों के डेटा का उपयोग कर अंग्रेजी स्पेनिश अनुवाद के लिए प्रदर्शन किया। उन्होंने 440 स्रोत वाक्यों का उपयोग किया और तीन स्पेनिश विशेषज्ञों के समक्ष अनुवाद प्रस्तुत किया। उन्होंने यह पाया, कि स्रोत वाक्य 20.6% और पुनः लिखा गया वाक्य 30.4% बार अनुवाद के लिए पसंद किया गया है और 49% वाक्यों के लिए सीधे अनुवाद किया जा सकता था।

Paetzold एवं अन्य [17] ने शाब्दिक और वाक्यात्मक सरलीकरण तकनीक की मदद से टेक्स्ट सरलीकरण प्रदर्शन किया। प्रशिक्षण के लिए, स्रोत और सरलीकृत स्रोत वाक्य के 133 हजार जोड़े सरल विकिपीडिया कॉर्पस से लिए गए। एसआरआईएलएम टूलकिट सरलीकरण के लिए, 3 ग्राम भाषा मॉडल का निर्माण करने के लिए इस्तेमाल किया गया। परीक्षण के लिए, 130 वाक्य सरल विकिपीडिया कॉर्पस से चयन किया गए। मूल्यांकन पांच इंसानों की मदद और ब्लीयू स्कोर से किया गया। वाक्यात्मक और शाब्दिक सरलीकरण के लिए, 0.342 ब्लीयू स्कोर प्राप्त हुआ।

Siddharthan [19] ने टेक्स्ट सरलीकरण के लिए रिजेंट नाम की एक प्रणाली विकसित की। स्टैनफोर्ड पार्सर का स्रोत वाक्य की टाइप निर्भरता प्रतिनिधित्व प्राप्त करने के लिए प्रयोग किया गया। उन्होंने एक कनवर्टर डिजाइन किया जो डीसिन्टेस अंकन में अनुवाद कर सकता है। इससे यह पाया गया कि सरलीकरण आपरेशन की 9% अंक से वृद्धि हुई और 5% संशोधित वाक्यों की संख्या अंक की वृद्धि हुई।

सिद्धार्थन [20] ने टेक्स्ट सरलीकरण प्रणाली के लिए एक रुपरेखा प्रस्तुत की। इसमें टेक्स्ट सरलीकरण दो प्रकार से किया गया - वाक्यात्मक सरलीकरण और शाब्दिक सरलीकरण। वाक्यात्मक सरलीकरण में तीन चरण शामिल हैं -विश्लेषण, परिवर्तन और पुर्नजनन। शाब्दिक सरलीकरण में तीन चरण शामिल हैं -विश्लेषण,

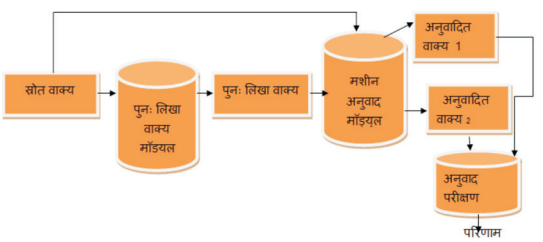
पेराफ्रेजिन्ग और पर्यायवाची प्रतिस्थापन। उन्होंने वाक्यात्मक सरलीकरण के लिए विभिन्न तकनीकों का सुझाव दिया।

## प्रस्तावित कार्य

सबसे पहले, हमने 15000 अंग्रेजी वाक्यों को वेब से लेकर उनका हिन्दी में अनुवाद किया। यह वाक्य पर्यटन क्षेत्र के थे। इसके साथ हमने अंग्रेजी के विकिपीडिया से बनाया गया सरलीकरण कोर्पस लिया। इससे हमने मोसीस मशीन अनुवादक टूलकिट द्वारा हमारी अंग्रेजी की वाक्य सरलीकरण प्रणाली का विकास किया।

इसके बाद हमने हमारे पर्यटन के 15000 अंग्रेजी वाक्यों का इस अनुवादक प्रणाली द्वारा सरलीकरण किया। इसके बाद जो सरलीकृत अंग्रेजी वाक्य हमें मिले हमने उनका भी हिंदी अनुवाद किया। अब हमारे पास दो अलग-अलग तरह के कोर्पस थे। पहला 15000 अंग्रेजी-हिंदी के वाक्य और दूसरा इन 15000 अंग्रेजी वाक्यों का सरलीकरण और उनका हिंदी अनुवाद।

इन दोनों कोर्पस के साथ हमने मोसीस मशीन अनुवादक टूलकिट की मदद से अनुवाद प्रणाली विकसित की। चित्र 1 इस प्रक्रिया को दर्शाता है तथा तालिका 1 अनुवादक के लिए उपयोग में लाये गए कोर्पस के आंकड़ों को दर्शाती है।




**चित्र 1: अंग्रेजी हिन्दी अनुवाद का प्रदर्शन**

**तालिका 1: अनुवाद प्रणाली के लिए उपयोग में लाये गए कोर्पस का विवरण**

| कोर्पस | अंग्रेजी-हिंदी सामानांतर कोर्पस | |
|---|---|---|
| वाक्य | 15000 | |
| | अंग्रेजी | हिंदी |
| शब्द | 6863569 | 693421 |
| यूनिक शब्द | 1900847 | 1291354 |

**तालिका 2 अनुवादित अंग्रेजी मूल स्रोत टेक्स्ट और अनुवादित सरलीकृत टेक्स्ट का विवरण**

| | अनुवादित मूल स्रोत टेक्स्ट | सरलीकृत टेक्स्ट |
|---|---|---|
| वाक्यों की संख्या | 500 | 500 |
| टोकन की संख्या | 6543 | 7654 |

## मूल्यांकन एवं परिणाम

यह समझने के लिये कि हमारे द्वारा विकसित प्रणाली किस प्रकार काम कर रही है, हमने 500 वाक्य और एकत्रित किये। इनको पहले वाक्य सरलीकरण प्रणाली को दिया और फिर अंग्रेजी और सरलीकृत अंग्रेजी वाक्य को अनुवाद प्रणाली को अनुवाद हेतु दिया। तालिका 2 अनुवादित अंग्रेजी मूल स्रोत टेक्स्ट और अनुवादित सरलीकृत टेक्स्ट को दर्शाती है। अनुवादित वाक्यों के मूल्यांकन के लिए हमने उन्हें एक मानव अनुवादक को दिया। मानवीय मूल्यांकन के लिए हमने जोशी एवं अन्य द्वारा विकसित मैट्रिक का उपयोग किया जो दस बिन्दुओं पर मूल्यांकन करती है। यह दस बिंदु हैं :-

1. संज्ञाओं के लिंग व वचन का अनुवाद में प्रयोग।
2. मूल वाक्य में प्रयुक्त काल का अनुवाद में प्रयोग।
3. मूल वाक्य में प्रयुक्त वाच्या का अनुवाद में प्रयोग।
4. व्यक्तिवाचक संज्ञा की पहचान।
5. विशेषण व क्रिया विशेषण का मूल वाक्य में संज्ञा व क्रिया के अनुकूल प्रयोग।
6. अनुवाद में सही शब्दों/पर्याय का चयन।
7. अनुवाद में संज्ञा, क्रिया एवं सहायक क्रिया का क्रम।
8. अनुवाद में विराम चिन्हों का प्रयोग।
9. अनुवाद में मूल वाक्य् में प्रयुक्त महत्व पूर्ण भाग पर बल।
10. अनूदित वाक्य में मूल वाक्य में निहित अर्थ का सही समागम।

मानव अनुवादक को एक स्रोत वाक्य तथा उसका मशीन अनुवाद दिया गया और उनसे इस दोनों को पड़ने के बाद इन दस बिंदुओं की 0-4 में रेटिंग करने को कहा गया। अंत में इन सभी रटिंगों का औसत निकाल कर अंतिम स्कोर प्राप्त किया गया।

तालिका 3: अंग्रेजी हिन्दी अनुवाद का मूल्यांकन

| | अंग्रेजी-हिन्दी मशीन अनुवाद | सरलीकृत अंग्रेजी-हिन्दी मशीन अनुवाद |
|---|---|---|
| मनुष्य | 0.200832 | 0.610494 |
| ब्लीयू | 0.255135 | 0.341418 |
| मिटिओर | 0.198011 | 0.28586 |
| एफमेजिर | 0.121345 | 0.451014 |

तालिका 4: अंग्रेजी हिन्दी अनुवाद के मूल्यांकन के औसत का विवरण

| | अंग्रेजी-हिन्दी मशीन अनुवाद | सरलीकृत अंग्रेजी-हिन्दी मशीन अनुवाद |
|---|---|---|
| मनुष्य | 15.48608 | 55.76562 |
| ब्लीयू | 18.61117 | 32.9905 |
| मिटिओर | 16.14152 | 33.37778 |
| एफमेजिर | 33.19652 | 41.3059 |

तालिका 3 में अंग्रेजी-हिंदी मशीन अनुवाद तथा सरलीकृत अंग्रजी-हिंदी मशीन अनुवाद के मूल्यांकन का विवरण दिया गया है। तालिका 4 में इस मूल्यांकन के औसत का विवरण दिया गया है।

## निष्कर्ष

इस लेख में हमने अंग्रजी-हिंदी मशीन अनुवाद की गुणवत्ता में सुधार का एक उपाय सुझाया है। यहाँ हमने दर्शाया है कि अगर अंग्रेजी वाक्यों का सरलीकरण हो जाये तो उनका अनुवाद करना आसान है। ऐसा करने के लिए हमने तीन मशीन अनुवाद प्रणालियों को विकसित किया। पहली प्रणाली अंग्रेजी वाक्य को सरल अंग्रेजी में तब्दील करती है, दूसरी सरल अंग्रेजी वाक्य का हिंदी अनुवाद देती है तथा तीसरी प्रणाली अंग्रजी वाक्य का हिंदी अनुवाद देती है। दूसरी व तीसरी प्रणाली के मूल्यांकन से यह प्रदर्शित होता है कि अगर हम सरल अंग्रेजी वाक्य मशीन अनुवादक को दें तो वह अच्छे परिणाम देगा।

## Conclusion

In this article, we have shown a technique in improving English-Hindi machine translation. Here we have shown that if an English sentence can be simplified then its translation becomes easier. For this we have trained three machine translation systems. First system translated English sentence into simplified English sentence. Second system translated Simplified English sentence into Hindi and the third translated English sentence into Hindi. With evaluation of second and third system, it appears that if we provide a simplified English sentence to the machine translator then it would give better results.

## संदर्भ ( References)


[1] Aluisio, Sandra, et al. (2010), "Readability assessment for text simplification." Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications. Association for Computational Linguistics.

[2] Bott, Stefan, and Horacio Saggion, (2011), "An unsupervised alignment algorithm for text simplification corpus construction." Proceedings of the Workshop on Monolingual Text-To-Text Generation. Association for Computational Linguistics.

[3] Candido Jr, Arnaldo, et al. (2009), "Supporting the adaptation of texts for poor literacy readers: a text simplification editor for Brazilian Portuguese." Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications. Association for Computational Linguistics.

[4] De Belder, Jan, and Marie-Francine Moens, (2010), "Text simplification for children." Proceedings of the SIGIR workshop on accessible search systems.

[5] Deepti Chopra and Sudha Morwal, (2013), "Named Entity Recognition in English using Hidden Markov Model", *International Journal on Computational Sciences & Applications* (IJCSA) Vo3, No.1, DOI : 10.5121/ijcsa.2013.3101,Pg 1-6 Available at: http://airccse.org/journal/ijcsa/papers/3113ijcsa01.pdf

[6] Deepti Chopra and Sudha Morwal, (2013), "Detection and Categorization of Named

Entities in Indian languages using Hidden Markov Model" International Journal of Computational Science and Information Technology (IJCSITY) Vol.1, No.1, Pg 25-32. Available at: http://airccse.org/journal/ijcsity/ papers/1113ijcsity04.pdf

[7] Deepti Chopra and Sudha Morwal, (2012), "Named Entity Recognition in Punjabi Using Hidden Markov Model", *International Journal of Computer Science & Engineering Technology* (IJCSET) ISSN : 2229-3345 Vol. 3 No. 12, Pg 616-620 Available at: http://www.ijcset.com/docs/IJCSET12-03-12-025.pdf

[8] Deepti Chopra, Nusrat Jahan and Sudha Morwal, (2012), "Hindi Named Entity Recognition By Using Rule Based Heuristics And Hidden Markov Model", *International Journal of Information Sciences and Techniques* (IJIST) Vol.2, No.6, DOI: 10.5121/ijist.2012.2604 Available at: http://airccse.org/journal/IS/papers/2612ijist04.pdf

[9] Glavaš, Goran, and Sanja Štajner, (2013), "Event-Centered Simplification of News Stories." Proceedings of the Student Workshop held in conjunction with RANLP.

[10] Jonnalagadda, Siddhartha, et al. (2009), "Towards effective sentence simplification for automatic processing of biomedical text." Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers. Association for Computational Linguistics.

[11] Klaper, David, Sarah Ebling, and Martin Volk, (2013), "Building a german/simple german parallel corpus for automatic text simplification." Proc. of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations.

[12] Klebanov, Beata Beigman, Kevin Knight, and Daniel Marcu, (2004), "Text simplification for information-seeking applications." On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE. Springer Berlin Heidelberg, 735-747.

[13] Klerke, Sigrid, and Anders Søgaard, (2012), "DSim, a Danish Parallel Corpus for Text Simplification." LREC.

[14] Mirkin, Shachar, et al. (2013), "SORT: An Interactive Source-Rewriting Tool for Improved Translation." ACL (Conference System Demonstrations). 2013.

[15] Joshi, Nisheeth, Hemant Darbari, and Iti Mathur, (2012), "Human and Automatic Evaluation of English to Hindi Machine Translation Systems." Advances in Computer Science, Engineering & Applications. Springer Berlin Heidelberg, 423-432..

[16] Joshi N, et al. (2014), "Implications of linguistic feature based evaluation in improving machine translation quality a case of english to hindi machine translation." *Indian ETD Repository@ INFLIBNET.*

[17] Paetzold, Gustavo H., and Lucia Specia. "Text Simplification as Tree Transduction.

[18] R Patel et al. (2013),"Reordering rules for English-Hindi SMT",ACL, August, 34-41

[19] Siddharthan, Advaith, (2011), "Text simplification using typed dependencies: A comparison of the robustness of different generation strategies." Proceedings of the 13th European Workshop on Natural Language Generation. Association for Computational Linguistics.

[20] Siddharthan, Advaith, (2002), "An architecture for a text simplification system." Language Engineering Conference, 2002. Proceedings. IEEE.

[21] Tyagi Shruti, Chopra Deepti, Mathur Iti , Joshi Nisheeth, (2015), "Classifier Based Approach For Improved Machine Translation" International Conference on Advances in Computer Engineering and Application 2015, organized by IMS Engineering College, 19th-20th March 2015, Ghaziabad. ❑