

## मैथिली रूपिमिक विश्लेषक : एक व्युत्पादन आधारित अभिगम

सरोज कुमार झा<sup>1</sup>, विजय कुमार कौल<sup>2</sup>, पीयूष प्रताप सिंह<sup>3</sup>, श्वेतांगी कुमारी<sup>4</sup>

सूचना एवं भाषा अभियांत्रिकी केंद्र, भाषा विद्यापीठ, महात्मा गाँधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा, महाराष्ट्र, भारत।

### सारांश

यह शोधपत्र मैथिली से हिंदी मशीनी अनुवाद को लक्ष्य करते हुए उसके एक आंशिक भाग रूपिमिक विश्लेषक के निर्माण हेतु प्रस्तुत है जो संपूर्ण विश्लेषक के आंशिक भाग के निर्माण के उपरांत उनसे संबन्धित चुनौतियों पर केन्द्रित है। इस रूपिमिक विश्लेषक का निर्माण मैथिली से अन्य भाषा में मशीनी अनुवाद के दृष्टिगत निर्मित है। विश्लेषक के निर्माण की प्रक्रिया मिश्रित अभिगम पर आधारित है जो अधिकांशतः शब्दों में प्रयुक्त उपसर्ग एवं प्रत्यय की पहचान कर के उसके रूप को विश्लेषित कर परिणाम प्रस्तुत कर सके। इस शोध पत्र में मैथिली भाषा हेतु रूपिमिक विश्लेषक के निर्माण कार्य को संस्कृत के अनुरूप निर्मित करने की कोशिश की गई है जिसमें व्युत्पादन स्वरूप को कृदंत, तद्धित, स्त्री एवं समास के अनुरूप वर्णित किया गया है। साथ ही संबंधित परिचर्चा में मैथिली भाषा के चिह्नित समस्त स्वरूप एवं code switching एवं code mixing के दृष्टिकोण से भी समाधान प्रस्तुत है। विश्लेषण के स्वरूप को सर्वप्रथम संज्ञा, विशेषण, क्रिया एवं क्रिया-विशेषण शब्द वर्ग के अनुसार प्रयोग किया गया है। एवं विश्लेषण निर्माण कार्य में कार्पस आधारित अभिगम की भी मदद ली गई है। विश्लेषक निर्माण को चरण दर चरण प्रस्तुत किया गया है जिसमें उपसर्ग एवं प्रत्यय को सर्वप्रथम रेखांकित कर रूपावली के रूप में निर्माण किया गया है तत्पश्चात् सभी धातु शब्दों की सूची निर्माण कर, उपसर्ग एवं प्रत्यय के संग धातु/नाम के संयोजन के आधार पर चिह्नित किया गया है। यह विश्लेषण यौगिक शब्दों के आधार पर है जबकि रूढ शब्दों को यथावत रखा गया है। इसमें समास में प्रयुक्त संज्ञा-समास, क्रिया-समास, Collocations, मुहावरों/कहावतों एवं आवृतिक शब्दों को भी नियमानुसार प्रस्तुत किया गया है। समस्त निर्माण कार्य जावा आधारित प्लेटफार्म पर निर्मित है।

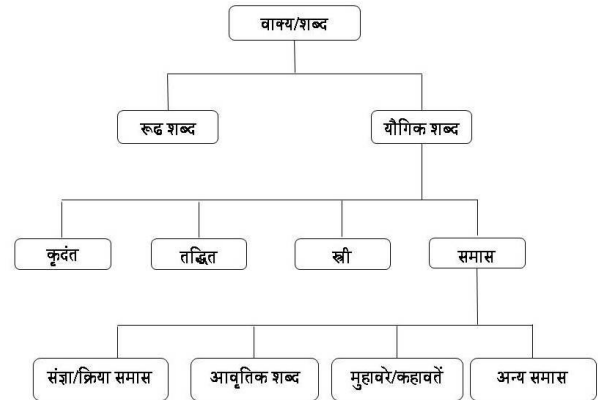
**मूल शब्द :** रूपिमिक विश्लेषक, मशीनी अनुवाद, कृदंत, तद्धित, उपसर्ग एवं प्रत्यय।

### प्रस्तावना

मशीनी अनुवाद का निर्माण आधुनिक समय की आवश्यकता बन चुकी है। अनुवाद की प्रक्रिया से आज कोई भी क्षेत्र अछूता नहीं है। यह प्रक्रिया मानवीय दृष्टिकोण से सामान्य एवं सरल प्रतीत होती है परंतु मशीनी परिप्रेक्ष्य में यह उतनी ही जटिल प्रतीत होती है। सर्वप्रथम संगणक वैज्ञानिकों के समक्ष यह एक चुनौती है कि जिस प्रकार हम अनुवाद करने के लिए जिस दृष्टिकोण का उपयोग करते हैं उसी प्रकार क्या हम मशीन को यह किस प्रकार से समझा सकते हैं कि कैसे सर्वप्रथम शब्द, पद एवं उसके उपरांत पदबंध एवं वाक्य के सही क्रम को अनुवाद के दृष्टिकोण से भी देखने का प्रयास करें। परंतु अंग्रेजी से हिंदी एवं अन्य भारतीय भाषाओं में शब्द क्रम एवं शब्द के रूपिमिक स्वरूप के कारण अनुवाद की प्रक्रिया थोड़ा जटिल स्वरूप धरण कर लेती है क्योंकि भारोपीय भाषाएँ अपने स्वरूप को हमेशा रूपसाधित स्तर पर अर्थिय भिन्नताओं के अनुरूप प्रस्तुत करती रहती हैं। यही कारण है कि मैथिली से अंग्रेजी एवं अन्य गैर-भारोपीय भाषाओं में सरल अनुवादिक दृष्टिकोण अपनाने हेतु रूपिमिक विश्लेषक का निर्माण आवश्यक है।

उक्त संकल्पना के अंतर्गत मैथिली भाषा में अनुवाद की परंपरा को विकसित करने के उद्देश्य से मैथिली रूपिमिक विश्लेषक का निर्माण प्रक्रिया अग्रसर है। इस रूपिमिक विश्लेषक का निर्माण व्युत्पादन के स्वरूप में संस्कृत एवं हिंदी का मैथिली भाषा में आगत शब्द एवं मैथिली के मूल रूप के अनुरूप विकसित करने की चेष्टा की गई है। जैसा कि हम जानते हैं कि मैथिली भी संस्कृत से ही व्युत्पादित भाषा है जो हिंदी के अति निकट मानी जाती है। इसी आधार पर यदि व्युत्पादन के स्वरूप को देखें तो यह संस्कृत के कृदंत, तद्धित, स्त्री एवं समास के अनुरूप ही विकसित हुई है। इस शोध पत्र में कृदंत एवं तद्धित के आधार पर उपसर्ग एवं प्रत्यय को चिह्नित कर व्युत्पादित करने की प्रक्रिया को सरल रूप में दर्शाने का प्रयास किया गया है। साथ ही सामासिक शब्दों के भिन्न भेदों के रूप में भी इसे विकसित किया गया है। उक्त

रूपिमिक विश्लेषक निम्न स्वरूपों में विभक्त होकर व्युत्पादन की व्यवस्था को रेखांकित करता है जो निम्न रूप में प्रस्तुत है-



चित्र सं 1

उपरोक्त चित्र सं 1 के अंतर्गत निम्न वर्गों के आधार पर व्युत्पादन की प्रक्रिया का अनुसरण किया गया है। जिसके अंतर्गत यौगिक शब्दों को व्युत्पादन हेतु भिन्न रूपों में चयनित किया गया है। उपरोक्त वर्णित स्वरूपों के अनुरूप व्युत्पादन की प्रक्रिया इन सभी स्तरों के आधार पर प्रस्तुत होती है। मैथिली रूपिमिक विश्लेषक के निर्माण में इन सभी के अतिरिक्त कुछ अन्य स्वरूप जैसे- तत्सम, तद्धित, देशज एवं विदेशज को भी संलग्न किया गया है।

मातृभाषा होने के कारण तथा शिक्षित वर्ग के प्रभाव के कारण संस्कृत से बहुत से शब्द मैथिली में अविकृत रूप में निरंतर आते रहे हैं। ऐसे शब्द सब तत्सम कहलाते

हैं। संस्कृत में इसे कुछ इस प्रकार कहा भी गया है “तत् संस्कृतम् तेन समम् तत्सम्” अर्थात् संस्कृत से किसी भी भाषा में आए हुए शब्द तत्सम कहलाते हैं। उदाहरण – जल, देह, दिन, मास, काल, एवं मेघ। अर्थात् जिस शब्द के स्वनात्मक स्वरूप जैसे संस्कृत में वैसे ही मैथिली में भी रहे, तत्सम् कहलाते हैं। उसी प्रकार मूल एवं वाद्य भाषा से आए हुए शब्द जब मैथिलीभाषी के जिह्वा के अनुरूप वाक प्रक्रिया में ढलते-ढलते स्वान संरचना को विशुद्ध मैथिली में परिवर्तित हो जाता है तब वह तद्भव कहलाता है। जो उच्च एवं निम्न के प्रयोग में विभक्त है। उच्च- हस्त, मुख, ग्राम, पुत्र आदि एवं निम्न- हाथ, मुह, ग्राम, पूत आदि। मैथिली भाषा के समस्त शब्दों के स्वरूप को चयनित एवं विश्लेषित करने हेतु हमें इन सभी वर्गों के शब्दों को आधारभूत स्थिति में चयनित करना होगा जिससे शब्दों को वर्ग के अनुरूप विश्लेषित किया जा सके।

### साहित्यिक पुनरवलोकन

“डेवलपिंग ए फ़ाइनाइट-स्टेट मोर्फोलोजिकल एनलाइजर फॉर उर्दू एंड हिंदी”<sup>[1]</sup> में टीना बोगल, मरियम बट्ट एवं उनके साहयोगियों के अनुसार इस शोधपत्र में उर्दू के अंतर्गत भाषा से संबंधित एक कास्केड (cascade) के माध्यम से उर्दू और हिंदी दोनों के अंतर्निहित सामान्य निवारण हेतु अनुमति प्रदान करता है जो एक अलग ASCII प्रतिलेखन प्रणाली से पूर्णतः अलग स्क्रिप्ट का अनुवाद भी करता है। चूंकि यह लिप्यंतरण प्रणाली एक्सएफएसटी (XFST) उपकरण पर आधारित है जो उर्दू के शब्द की पुनरावृत्ति हेतु समाधान के लिए एक्सएफएसटी उपकरण (XFST) को प्रस्ताव के रूप में बताता है कि उर्दू/हिंदी में कितने भिन्न पुनरावृत्तिक पैटर्न की समस्या को बी एंड के (B&K) द्वारा प्रस्तावित फाइनाइट स्टेट विधियों के साथ सहज तरीके से निवारण किया जा सकता है। यह परिचर्चा मुख्य रूप से ParGram परियोजना के हिस्से के रूप में फाइनाइट स्टेट रूपिमिक विश्लेषक एवं एलएफजी (LFG) व्याकरण के बीच परिभाषित रूपिमिक-वाक्यविन्यास इंटरफेस के संबंध में जानकारी कहाँ और कैसे निरूपित किया जाना चाहिए, से संबंधित है।<sup>[2]</sup> “हिंदी मोर्फोलोजिकल एनलाइजर एंड जनरेटर” में विशाल गोयल एवं गुरप्रीत सिंह लेहल के अनुसार भाषाविज्ञान में रूपीमविज्ञान का क्षेत्र शब्दों की आंतरिक संरचनाओं का अध्ययन व्यक्त करता है। रूपिमिक विश्लेषण एवं सर्जन प्रकृतिक भाषा संसाधन में महत्वपूर्ण अंग है। इसमें विश्लेषण का अर्थ शब्दों को इनपुट के रूप में प्राप्त कर स्टेम (stem) एवं प्रत्यय को ढूँढना है। प्राप्त विश्लेषण वाक्य के शब्दों का अर्थ एवं वाक्यात्मक भूमिका की जानकारी प्राप्त करता है। हिंदी रूपिमिक विश्लेषक एवं सर्जक अपने डेटाबेस में सभी हिंदी के मूल शब्दों के लिए सभी संभावित सामान्य रूप से उपयोग किए जाने वाले शब्द के रूपों को संग्रह एवं उसपर परिचर्चा करता है। इन दिनों मेमोरी स्पेस न तो लागत के मामले में और न ही भंडारण आवश्यकताओं के मामले में समस्या का विषय नहीं है, यह दृष्टिकोण अन्य दृष्टिकोणों से बेहतर प्रदर्शन करेगा क्योंकि इसमें केवल रूट शब्द और रूपावली डेटाबेस में संग्रहीत किए जाते हैं। उन प्रविधियों में खोजने का समय सामान्य कारणों से अधिक है क्योंकि यह कम संग्रह स्थान घेरता है। यह प्रविधि मेमोरी के अनुरूप लगे समय के आधार पर परिशुद्धता को प्राथमिकता प्रदान करती है।<sup>[3]</sup> डेवलपिंग मोर्फोलोजिकल एनलाइजर फॉर साउथ-एशियन लैंग्वेज : एक्सपेरिमेंटिंग विथ हिंदी एंड गुजराती भाषाएँ में नीरज अस्वनी एवं राबर्ट गाइजाक्षुस के अनुसार स्टेमर्स और रूपिमिक विश्लेषकों के विकास में काफी सीमित मात्रा में काम किया गया है। इनमें से अधिकतर प्रविधि हस्तनिर्मित प्रत्यय-प्रतिस्थापन नियमों पर आधारित है लेकिन कुछ अलग से अपवादों को खोजने की कोशिश करते हैं। हालांकि ज्यादातर प्रविधियाँ प्रत्यय-प्रतिस्थापन या हटाने पर आधारित होती हैं। यह सब व्युत्पन्न स्टेमर्स के उदाहरण हैं जो उपसर्गों पर भी आधारित हैं। प्राप्त कार्पस एवं कोश में प्रस्तुत विधि का उपयोग एक घुमावदार शब्द से प्रत्यय-प्रतिस्थापन नियमों का एक सेट मूल रूप को प्राप्त करने के लिए किया जा सकता है। हिंदी भाषा के लिए एक प्रविधि विकसित की गई

है, जो यह दर्शाती है कि यह गुजराती भाषा को अनुकूलित करने के साथ कम से कम संबंधित भाषाओं के लिए यह विश्लेषक पोर्टेबल है।<sup>[4]</sup> “हिंदी डेरिवेशनल मोर्फोलोजिकल एनलाइजर” में निखिल कानहुप्रती, अभिलाषा इनुमेला एवं दीप्ति मिश्रा शर्मा के अनुसार हिंदी रूपीमविज्ञान के दृष्टिकोण से एक भारतीय भाषा है। जिसमें रूपिमिक विश्लेषक का निर्माण अब तक नगण्य रहा है। हिंदी भाषा को लेकर निर्मित यह एल्गोरिथम पूर्व के रूपसाधक विश्लेषक को व्युत्पादक विश्लेषक में परिवर्तित कर दो लक्ष्यों की प्राप्ति करता है। सर्वप्रथम यह सफलतापूर्वक रूपसाधक विश्लेषक में व्युत्पादक विश्लेषण को समाहित करता है एवं द्वितीय यह कि रूपसाधक विश्लेषण पूर्ववत रूपसाधक विश्लेषक के परिणाम को पहले से सक्षम रूप में रूपांतरित करने का प्रयास करता है। यह एल्गोरिदम कार्य को सम्पन्न करने के लिए पोर्टर्स (Porters) स्टेमर और क्रोवेट्ज (Krovetz) स्टेमर दोनों के मुख्य सिद्धांतों का उपयोग करता है। इस एल्गोरिदम द्वारा सामान्य प्रिंसीपल एवं रिकाल के साथ-साथ रूपसाधक विश्लेषण एवं परिणाम को भी दर्शाया जा सकता है।<sup>[5]</sup> हिंदी मोर्फोलोजिकल एनलाइसर एवं इंफ्लेक्सनल जनरेटर फॉर इंग्लिश टू हिंदी ट्रांसलेशन में पवन दीप सिंह, अर्चना कोरे, रेखा सुगंधी, गौरव आर्या एवं स्नेहा जादव के अनुसार अंग्रेजी से हिंदी सांख्यिकीय मशीन अनुवाद के लिए वाक्य रचनात्मक और मोर्फोलॉजिकल जानकारी को शामिल करने के लिए सरल और कम्प्यूटेशनल रूप से कई संपूर्ण अवधारणाएँ हैं जैसे (i) अंग्रेजी पद विच्छेदन हेतु सामान्य रूपांतरण का नियम (ii) एक साधारण प्रत्यय पृथक रूप से रूपसाधक जोड़े में वितरित रूपिमिक विश्लेषण का अंग है। हिंदी के लिए केवल द्विभाषी प्रशिक्षण डेटा और सीमित उपकरण की एक छोटी मात्रा के साथ एक उचित प्रदर्शन और पर्याप्त सुधार प्राप्त किए जा सकते हैं। यह शोध-पत्र प्राथमिक रूप से वितरित रूपीमविज्ञान ढांचे के अंतर्गत हिंदी के नाममात्र परिवर्तन का विश्लेषण प्रस्तुत करता है। इसमें वाक्य विन्यास के रूपसाधक वर्ग, रूपसाधक स्वरूप एवं रूपीम विज्ञान संसाधन की प्रक्रिया के विषय में विस्तृत चर्चा अपेक्षित है।<sup>[6]</sup> डेवलपमेंटऑफ़ मार्फ़ोलाजिकल एनलाइजर हिंदी में मयूरी रस्तोगी एवं पूजा खन्ना के अनुसार उक्त शोध-पत्र के परिपेक्ष्य में यह दृष्टिगत है कि यह रूपिमिक विश्लेषक हिंदी के मशीनी अनुवाद के निर्माण हेतु प्रस्तुत है। इस रूपिमिक-विश्लेषक में हिंदी के शब्दों को प्रविष्ट कर उन शब्द की सम्पूर्ण सूचना को प्राप्त किया जा सकता है। यह शब्द रूट शब्द के साथ उन शब्दों के सभी गुण जैसे पुरुष, वचन एवं संख्या आदि के वर्णन प्राप्त होते हैं। इसके अंतर्गत दो प्रकार की प्रविधि का प्रयोग किया गया है। जिसमें प्रथम नियम आधारित प्रक्रिया एवं द्वितीय कार्पस आधारित प्रक्रिया है। अभी तक किसी भी संस्थान द्वारा ऐसा कोई रूपिमिक विश्लेषक नहीं है जो रूपिमिक विश्लेषक के दोनों स्वरूप जैसे – रूपसाधक एवं व्युत्पादक दोनों के लिए बनाया गया है।<sup>[7]</sup> मार्फ़ोलाजिकल एनलाइजर फॉर हिंदी- ए रूल बेस्ड इंफ्लेमेंटेशन में अंकिता अग्रवाल, प्रमिला, शशिपाल सिंह, अजय कुमार एवं हेमंत दरबारी के अनुसार रूपिमिक विश्लेषण प्राकृतिक भाषा संसाधन का एक महत्वपूर्ण भाग है। इसके साथ मशीनी अनुवाद का कार्य बहुत आसान हो गया है। इस शोध-पत्र में हम हिंदी भाषा हेतु एक रूपिमिक विश्लेषक के स्वरूप पर ध्यान केंद्रित करते हैं। रूपिमिक विश्लेषक एक हिंदी वाक्य को शब्द के इनपुट के रूप में लेता है और इसके मूल शब्दों के साथ अपनी आवश्यक विशेषताओं को उत्पन्न करने के लिए इसका विश्लेषण किया जाता है। इन गुणों को निम्न श्रेणियों में विभक्त किया गया है, जैसे- वाक, लिंग, संख्या और व्यक्ति का ही एक स्वरूप है। यह प्रणाली रूपसाधक एवं व्युत्पादक दोनों पर कार्य करती है। यह नियम आधारित प्रक्रिया पर कार्य करती है। धारा 3 में वर्णित हिंदी के विभिन्न शब्द संरचनाओं के लिए हमने लगभग सभी संभावित नियमों को शामिल किया है, चाहे वह रूपसाधक पर या व्युत्पादक पर हो।

उपरोक्त विश्लेषक निर्माण प्रक्रिया में अधिकतर नियम आधारित प्रक्रिया के साथ कार्पस आधारित प्रक्रम का उपयोग किया गया है जो शब्द निर्माण प्रक्रिया में उपसर्ग एवं प्रत्यय के साथ शब्दों के स्वरूपों का स्वरूपित संयोजन प्रस्तुत करता है। साथ ही

संज्ञा के भिन्न स्वरूप जैसे – संज्ञा समास, आवृत्तिक संज्ञा एवं कई ऐसे स्वरूप हैं जोकि विश्लेषक की भूमिका में यथावत प्रस्तुत होते हैं परंतु रूपीम के स्तर पर यह मात्र केवल समास है।

### विश्लेषण प्रक्रिया

विश्लेषक प्रक्रिया के अंतर्गत मैथिली में व्युत्पादन आधारित रूपिमिक विश्लेषक का निर्माण शब्द संरचना के जटिल से जटिलतम स्वरूप को सरल अवस्था में विश्लेषित कर शब्दों के समस्त स्तर एवं बहुमुखी रूपीम अवस्था को प्रस्तुत किया गया है। जैसा कि उपरोक्त वर्णित है कि इस विश्लेषक के निर्माण में मैथिली भाषा को संस्कृत के अनुरूप विश्लेषित करने की प्रक्रिया पर बल दिया गया है। इसी क्रम में सर्वप्रथम मैथिली के शब्द वर्ग में व्युत्पादन की दृष्टि से संज्ञा, सर्वनाम, विशेषण एवं क्रिया-विशेषण को चुना गया है। उसके उपरांत मैथिली उपसर्ग एवं प्रत्यय को चिह्नित कर संकलित किया गया है। उपसर्ग एवं प्रत्यय की संकलन प्रक्रिया में संस्कृत के आगत उपसर्ग को एवं मैथिली के (Cognate) उपसर्ग को भी शामिल किया गया है जिनकी संख्या मिलकर लगभग 29 है। प्रत्यय के अंतर्गत मैथिली में कृदंत एवं तद्धित को मिलाकर लगभग 72 के करीब हैं। इसके अतिरिक्त संज्ञा के मूल शब्द (तद्धितांत) के रूप में एवं क्रिया के धातुरूप का भी संकलन किया गया है। इसी क्रम में आगे बढ़ते

### उदाहरण

तालिका सं. 2

क्र. सं.	नाम/धातु शब्द	मैथिली उदाहरण	रूपिमिक विश्लेषण
1.	सेक (sEk)	मैथिली- माँ रोटी सेकैत छेथी। WX- mA rotI skeit CethI. हिंदी- माँ रोटी सेकती है।	सेकैत = सेक+इत (लिंग मुक्त)/N
नोट: उपरोक्त उदाहरण में यह देखा गया है कि 'रोटी' स्त्रीलिंग में होते हुए भी मैथिली में 'सेकैत' को लिंग मुक्त दर्शाया गया है। और 'छेथी' जो स्त्रीवाची के साथ सम्मान के भाव को प्रस्तुत करती है, 'माँ' के लिए दर्शाया गया है। लेकिन यदि 'रोटी' को 'हाथ' शब्द से स्थानांतरित कर दे तो भी 'सेक' के प्रारूप में लिंग को लेके कोई परिवर्तन संभव नहीं है। तो इस प्रकार N <sub>m</sub> या N <sub>f</sub> के टैग को स्वरूपित किया गया है।			
2.	लिख (liK)	मैथिली- ओ अपने पोथी पर लिख लेता। WX- O apnA poTI par liK letA. हिंदी- वो अपनी पुस्तिका पर लिख लेंगे।	लिख= लिख+0 (लिंग मुक्त)/v लिखा= लिख+आ/c1 लिखवा=लिख+वा/c2
नोट: उपरोक्त उदाहरण में शब्द 'लिख'/'लिखा'/'लिखवा' तीनों ही सकर्मक क्रिया के प्रेरणार्थक रूप में माने जाते हैं। परंतु जब भी 'अब' प्रत्यय इनके साथ जुड़ते हैं तो संज्ञा के रूप को धरण कर लेते हैं। फिर इनके साथ रूपिमिक स्तर पर संज्ञा के गुण संलग्न हो जाते हैं।			
3.	राज काज (rAj kAj)	मैथिली- अहाँ के अपने राज-काज स फुरसत नै अछि। WX- ahAn ke apne rAj kAj sM fursat nE eiCI. हिंदी- आपको अपने ही राज काज से फुरसत नहीं है।	राज काज (Work)/N_C
	खाएत पिबैत (KaEt pibEt)	मैथिली- हम सभ खाएत पिबैत घर जाएब। WX- ham saB KaEt pibEt Gar jAeb. हिंदी- हम लोग खाते पीते घर जाएंगे।	खाएत पिबैत (Enjoy)/ V_C
नोट: उपरोक्त उदाहरण में 'राज काज' को एक समास के रूप में विश्लेषित किया गया है जिसमें शाब्दिक अर्थ में 'राज' एवं 'काज' दोनों के अर्थ अस्तित्व में हैं एवं साथ ही मुहावरे के अर्थ में भी उपयोग किया जाता है। इसको N_C या V_C के टैग द्वारा दर्शाया गया है।			
4.	घरक घर (ghark ghar)	मैथिली- घरक घर तबाह भ' गेल। WX- Gark Gar tabAh Ba gel. हिंदी- घर के घर बर्बाद हो गए।	घरक घर (Home)/RD_C
	चाह ताह (cAh tAh)	मैथिली- चाह ताह बनत कि आय? WX- cAh tAh banat ki Aya. हिंदी- चाय वाय बनेगा क्या आज?	चाह ताह (Tea)/RD_Eco
5.	नोट: उपरोक्त उदाहरण में 'घरक-घर' को पूर्ण आवृत्तिक माना गया है। एवं 'चाह ताह' को आंशिक आवृत्तिक शब्द के रूप में दर्शाया गया है। इसको RD_C एवं RD_Eco के टैग से दर्शाया गया है।		
6.	मुह ताकब	मैथिली- अहां सभ मुह ताकैत रहि ज्याब। WX- ahAn saB muhEn tAkeIt rahi jAyab. हिंदी- आपलोग मुंह ही देखते रह जाएंगे।	मुह ताकब (for nothing)/MWE_IP
7.	नोट: उपरोक्त उदाहरण में 'मुह ताकब' को पूर्ण आवृत्तिक माना गया है। एवं उक्त शब्दों के अर्थों के अनुरूप MWE_IP का टैग निर्धारित किया गया है।		

हुए समास के अंतर्गत निम्न वर्गों को भी शामिल किया है एवं चरण दर चरण विश्लेषण को निम्न प्रकार से दर्शाया गया है-

तालिका सं. 1

क्र. सं.	विवरण	टैग	उप-टैग
1.	संज्ञा धातु (N_root)	NN	N <sub>m</sub> /f N <sub>c</sub>
2.	क्रिया धातु (V_root)	VM	V <sub>c</sub>
3.	समास (Compounds)	C	N <sub>c</sub> V <sub>c</sub>
4.	आवृत्तिक शब्द (Reduplication)	RD	RD <sub>c</sub> RD <sub>Eco</sub>
5.	एम.डबल्यू.ई. (MWE)	MWE	MWE_IP

उपरोक्त तालिका सं. 1 में समस्त वर्गों एवं उपवर्ग के टैग से दर्शाये गए हैं। विश्लेषण प्रक्रिया में यह रूपिम व्यवस्था के व्युत्पादन के आधार पर विश्लेषण एवं संबंधित टैग को उदाहरण के साथ प्रस्तुत किया गया है। सर्वप्रथम संज्ञा को जोकि संज्ञा-पुरुषवाची एम संज्ञा-स्त्रीवाची के रूप में विभक्त है-

### निष्कर्ष एवं परिणाम

मैथिली अनुवाद हेतु शाब्दिक स्तर पर विश्लेषण की प्रक्रिया में व्युत्पादन के स्वरूप को उपरोक्त में निम्न स्तरों पर दर्शाया गया है। संज्ञा के समस्त विश्लेषित रूप को अलग-अलग टैग के साथ प्रस्तुत है। अभी इस प्रणाली प्रक्रिया का निर्माण अग्रसर है एवं नियम आधारित एवं कार्पस आधारित स्वरूपों के आधार पर व्युत्पादित हो रहा है। परिणाम के दृष्टिकोण से अभी केवल संज्ञा एवं विशेषण वर्ग में भिन्न स्वरूप के सभी आधारों को प्रस्तुत कर रहा है। साथ ही समास एवं अन्य संज्ञा संबंधित को पूर्णतः खोजने में सक्षम है। नियम आधारित एवं कार्पस वर्गीकरण के आधार पर 68% से 72% तक रूपिम का विश्लेषण करने में सक्षम है। साथ ही सामासिक पद एवं पुनोक्ति शब्दों को प्रतिशत में देखें तो यह प्रतिशत सम्पूर्ण रूप से 73.28 % के लगभग प्राप्त हुआ है।

### संदर्भ सूची

1. टीना बोगल, मिरियम बट्ट, एनेटे हौतीली, एंड सेबस्तीन सुगलर, “डब्लपिंग ए फाइनाइट स्टेट मोर्फोलोजिकल एनलाइजर फॉर उर्दू एवं हिंदी”, फाइनाइट स्टेट मेथड्स एंड नेचुरल लैंग्वेज प्रोसेसिंग, पेज 86, 2007.
2. विशाल गोयल, गुरप्रीत सिंह लेहल, “हिंदी मोर्फोलोजिकल एनालिसर एंड जनरेटर”, फ्रंट इंटरनेशनल कॉन्फ्रेंस ऑन एमर्जिंग ट्रेड्स इन इंजीनियरिंग टेक्नोलोजी, यूएसए, पेज 1156-1159, 2008.
3. नीरज अश्वनी एवं रॉबर्ट गाइज़ोस्वास, “डब्लपिंग मोर्फोलोजिकल एनलाइजर फॉर साउथ एशियन लैंग्वेजेस: एक्सपेरिमेंटिंग विथ द हिंदी एंड गुजराती लैंग्वेजेस”, प्रोसिडिंग्स ऑफ द सेवेन्थ इंटरनेशनल कॉन्फ्रेंस ऑन लैंग्वेजेस रिसोर्सेस एंड एवाल्यूसन, वालेटा, माल्टा, पेज 811-815, मई, 2010.
4. निखिल कानहुप्रती, अभिलाषा इनुमेला एंड दीप्ति मिश्रा शर्मा “हिंदी डेरिवेशनल मोर्फोलोजिकल एनलाइजर”, प्रोसिडिंग्स ऑफ द ट्वेन्थ मीटिंग ऑफ द स्पेशल इंटररेस्ट ग्रुप ऑन कंप्यूटेशनल मोर्फोलोजी एंड फोनोलोजी, कनाडा, पेज. 10-16, जून 2012.
5. पवन दीप सिंह, अर्चना कोरे, रेखा सुगंधी, गौरव आर्या एंड स्नेहा जादव “हिंदी मोर्फोलोजिकल एनालिसिस एवं इंफ्लेक्सनल जनरेटर फॉर इंग्लिश टू हिंदी ट्रांसलेसन”, आईजेआईटी, 2008.
6. रस्तोगी मयूरी एवं खन्ना पूजा “डेवलपमेंट ऑफ मार्फोलोजिकल एनलाइजर हिंदी”, आईजेसीए, 2014
7. अग्रवाल अंकिता, प्रमिला, सिंह शशिपाल, कुमार अजय एवं दरबारी हेमंत “मार्फोलोजिकल एनलाइजर फॉर हिंदी- ए रूल बेस्ड इंप्लेमेंटेशन”, इन प्रोसिडिंग्स ऑफ इंटरनेशनल जर्नल ऑफ एडवांस कंप्यूटर रिसर्च (आईएसएसएन (प्रिंट): 2249-7277 ISSN (ऑनलाइन): 2277-7970) वाल्यूम-4 नवंबर-1 इशू – 14, मार्च-2014.
8. सिंह रेणु, ओझा अतुल, एंड झा गिरीश नाथ, “क्लासिफिकेशन एंड आईडेंटिफिकेशन ऑफ रेडूप्लिकेशन मल्टी-वर्ड एक्सप्रेसन इन हिंदी” 2016.
9. अक्षर भारती, राजीव संगल, एस.एम. बेंद्रे, ऐश्वर्या, पवन कुमार, “अनरीजर्वड इमप्रूवमेंट ऑफ मोर्फोलोजिकल अनलायजर फॉर इनफ्लेक्सनली रिच लैंग्वेजस” इन प्रोसिडिंग्स ऑफ 6 एनएलपी पेसिफिक रिम सिंपोसियम, 27-30 नवंबर 2001, टोकियो.
10. दीपक कुमार मल्लाड़ी, प्रशांत मेनन, “स्टैटिस्टिकल मोर्फोलोजिकल अनलायजर फॉर हिंदी” इन प्रोसिडिंग्स ऑफ 6 इंटरनेशनल जोईंट कॉन्फ्रेंस ऑन नेचुरल लैंग्वेज प्रोसेसिंग 2013.
11. अदिति मूले, एट. अल. “मोर्फोलोजिकल अनलायजर फॉर ए गिवेन टेक्स्ट ऑफ मराठी लैंग्वेज” इंटरनेशनल जर्नल ऑफ कंप्यूटर साइंस एंड कम्यूनिकेशन नेटवर्क भाग-4 (1), 13-17, 2014.
12. <http://sampark.iit.ac.in/hindimorph/web/restapi.php/indic/morphclient>
13. <http://www.cfilt.iitb.ac.in/~ankitb/ma>
14. <http://sanskrit.jnu.ac.in/morph/analyze.jsp>