



प्रस्तुत शोध सामान्य हिंदी पर औपचारिक व्याकरणिक तरीकों को लागू करता है। इसका उद्देश्य हिंदी वाक्य विन्यास और हिंदी शब्दार्थ को समझने के लिए वैज्ञानिक आधार लाना है।

इसमें प्रस्तावित सभी अनुभाग अलग-अलग विषय चर्चा के रूप में प्रकाशित किये गए हैं। अतः यदि यह संक्षिप्त विवरण आपको हिंदी व्याकरण को एक भिन्न तरीके से विश्लेषित करने की तरफ आकर्षित करता है तो कृपापूर्वक पूर्ण-लेखों को भी पढ़ें। ये शोध-लेख दिए गए ईमेल से आप प्राप्त कर सकते हैं।

विवेक त्रिपाठी
sopan.tripathi@gmail.com



मूल हिंदी के व्याकरण की प्रकृति का वर्णन करने के लिए कुछ वाक्य-विन्यास और शब्दार्थ-नियम प्रस्तावित किये गए हैं। हिंदी व्याकरण की जटिलताओं को समझने के लिए एक मूल नमूना L_{IH} नाम की भाषा बना कर समझाया गया है।

L_{IH} एक कृत्रिम भाषा है जिसका ऐसा नामकरण इसकी आधार भाषाओं की वजह से किया गया है। L_{IH} में 'I' कथन तर्क को और 'H' हिंदी भाषा का प्रतिनिधित्व करता है। इसके बाद L_{IH} का कई बार परिवर्धन किया गया है। परिवर्धित भाषाएँ (L_{2H} , L_{3H} और L_{4H}) को विभिन्न कथन-श्रेणियों के हिसाब से बाँधा गया है।

प्रमुख बिंदु

कथन-श्रेणियों का कुछ उदाहरण ऐसा हो सकता है जैसे कि (१) राम चलता है, (२) बच्चे चलते हैं, (३) राम चलता है और बच्चे चले जाते हैं, (४) राम चला था और बच्चे चलेंगे आदि।

इन सभी वाक्यों में एक समानता और एक अंतर है। समानता यह है कि ये सभी वाक्य 'धोषणात्मक प्रकृति' (affirmative nature) के हैं।

अंतर यह है कि इन सभी वाक्यों की क्रिया-संरचना अलग-अलग है। इसलिए नंबर (१) जैसे वाक्य L_{IH} भाषा से, (२) जैसे वाक्य L_{2H} से, (३) जैसे वाक्य L_{3H} से और (४) जैसे वाक्य L_{4H} से समझे जा सकते हैं।

हमारा मुख्य सैद्धांतिक चिंतन 'तार्किक वाक्यविन्यास' और 'भाषाई वाक्यविन्यास' के बीच के संबंध को समझना है। सभी व्याकरणिक नियमों को हमने सॉफ्टवेयर के माध्यम से कार्यान्वित भी किया है। उम्मीद है कि इस अध्ययन से हिंदी के लिए एक नया भाषाई ढाँचा विकसित होगा।



हिंदी कैसे वाक्य-संरचना करती है?

"हिंदी कैसे वाक्य-संरचना करती है?" लेख एक आधुनिक दृष्टिकोण को पेश करता है जिससे हमें हिंदी की वाक्य संरचना एवं अर्थ संरचना को समझने का एक मौका प्राप्त होता है।

भाषा विज्ञान की दृष्टि से वाक्य संरचना को 'सिंटेक्स (syntax)' और अर्थ संरचना को 'सिमेंटिक्स (semantics)' भी कहा जाता है। इस पद्धति से विश्व की लागभग सभी भाषाओं में अनुसंधान एवं अध्ययन किया गया है। अतः अनेकों विद्वानों में फर्जिनेंड डी सौस्युर, नोम चोमस्की, रिचर्ड मोटेयु, बारबरा पार्टी आदि का नाम विशेष रूप से आता है।

हिंदी में एक विस्तृत और यथासंभव 'सिंटेक्स और सिमेंटिक्स' को समझने के लिए एक पूर्ण ढाँचा तैयार करना ही इस शोध का प्रयास है। परन्तु हम ऐसा करना ही क्यूँ चाहते हैं, इसको समझने के लिए मैं आप से कुछ उद्हारण साझा करना चाहता हूँ।

उदाहरण २:

- राम चलता है
- राम चला जाता है
- राम चलता हुआ जाता है
- राम चल करके जाता है
- राम चलते चलते जाता है

इन सभी वाक्यों को हमारी आम व्याकरणिक दृष्टि (जिसमें संज्ञा, सर्वनाम आदि रूप में वाक्यांशों को समझा जाता है) से भी देखा जा सकता है। अन्य दृष्टि यह हो सकती है कि इन सभी वाक्यों को सिर्फ ($X + Y$) के रूप में तोड़ दिया जाए। अब यहाँ 'X' वाक्य का वह घटक हो सकता है जो हर वाक्य में एक जैसा है और 'Y' वह घटक जो हर वाक्य में बदल रहा है। अतः यहाँ $X =$ 'राम चल' और $Y =$ 'ता है / आ जाता है / ता हुआ जाता है / करके जाता है / ते चलते जाता है' आदि हो सकता है। इसी तरह की वाक्य विन्यास की प्रक्रिया चारों परिवर्धित भाषाओं को बनाने में लगाई गई है। फिर कंप्यूटर को यह प्रक्रिया गणित के रूप में समझाई गयी। इसे समझने के बाद कंप्यूटर के लिए ऐसे वाक्य या इस तरह के अन्य वाक्यों को बनाने की प्रणाली पता चली।

शोध-प्रविधि:

X या Y घटक हिंदी वाक्यों में किस प्रकार होने चाहिए? - इसकी संकल्पना "सिटेक्स के नियमों" द्वारा की जाती है। जैसे कि अगर मैं एक वाक्य जैसे कि "राम चलता है" का निर्माण करना चाहता हूँ, तो मैं यह कहूँगा कि "राम" एन.पी कैटेगरी का एक रूपिम है और "चलता है" वी.पी कैटेगरी का। यह एन.पी कैटेगरी हमारे लिए X के स्थान पे आ सकती है और वी.पी कैटेगरी Y के स्थान पर। यह संकल्पना आधुनिक काल में विश्व की कई भाषाओं के लिए की गयी है, जिसे हम "ट्रांसफॉर्मेशनल ग्रामर [Transformational Grammar (TG)]" के नाम से जानते हैं। इस ग्रामर (अर्थात् व्याकरण) में सबसे बड़ा नाम "नोम चॉस्ट्सी" का माना जाता है। यह विषय यदि आपको आकर्षित करता है, तो इसे विस्तार में इन्टरनेट द्वारा जरुर पढ़िए।

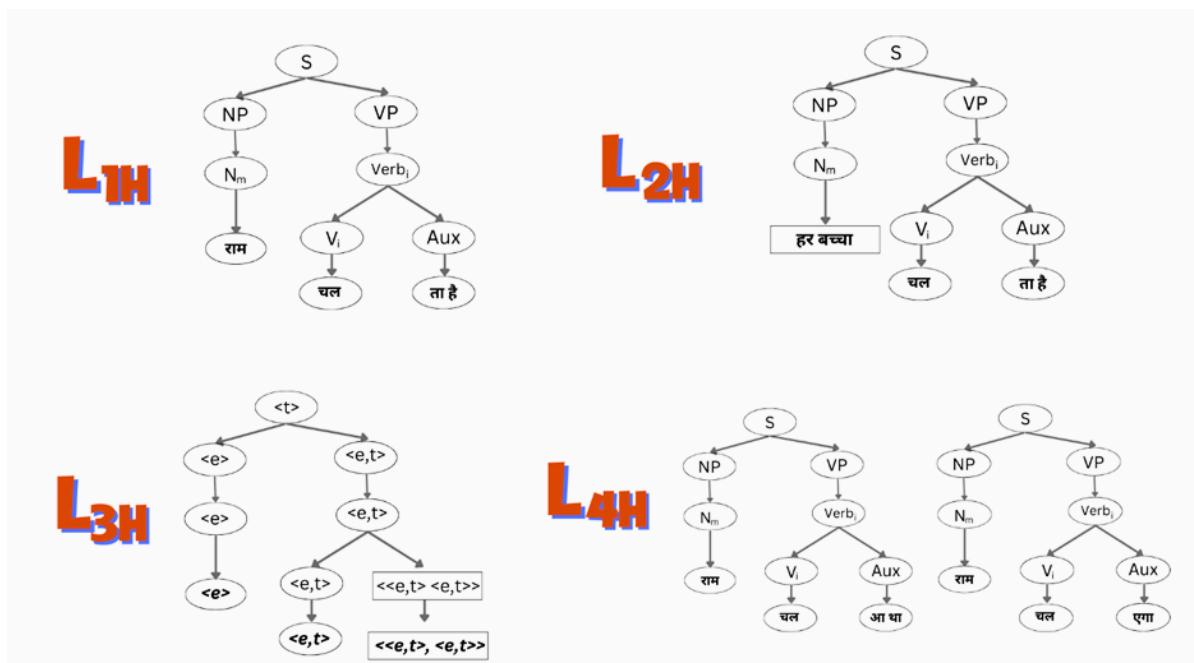
अब आते हैं अगली अवधारणा पर - क्योंकि हम हिंदी के वाक्य को किसी पूर्व-तय व्याकरणिक इकाई (जैसे कि संज्ञा, सर्वनाम आदि) के अनुसार नहीं तोड़ रहे हैं, बल्कि अपने स्व-सिद्धांत (कोई भी सिद्धांत जो आप अपने विमर्श में रख सकते हैं) के हिसाब से तोड़ रहे, इसलिए आप X और Y के गुण भी खुद से ही निर्धारित कर सकते हैं। यह भाषा को बनाने का यानी कि वाक्यों की संरचना करने का बड़ा ही अनोखा दृष्टिकोण है। इसी संकल्पना का प्रयोग करके इस शोध में हिंदी के वाक्यों की संरचना की गई है। उम्मीद है कि अब आपको "हिंदी अपनी वाक्य संरचना (हमारे शोध अनुसार) कैसे कर रही है?" यह समझ आ गया होगा।

जैसा कि मैंने कहा कि X और Y स्थापित व्याकरणिक इकाई नहीं हैं, इसलिए ये हिंदी व्याकरण के गुण के हिसाब से बंधे भी नहीं हैं। हम गुणों का निर्धारण स्वयं कर सकते हैं। हमने X और Y के लिए जो गुण निर्धारित किये हैं, वो हैं "तर्क शास्त्र" के। | आप चाहे तो किसी और विषय के गुण भी रख सकते हैं।

गुण को निर्धारित करने के लिए हमारा प्रयोजन "तर्क-शास्त्र" इसलिए था क्योंकि हम चाहते थे कि वाक्य संरचना के बाद जो हिंदी वाक्य बने, उनसे हम तर्क भी कर सकें। जैसे कि आगर हमने कहा कि "राम चलता है" जगत में होने वाला एक सत्य वाक्य है, तो इसका अर्थ ये भी है कि "राम बैठा नहीं है" क्यूंकि उसके अन्दर चलने का गुण पहले से विद्यमान था।

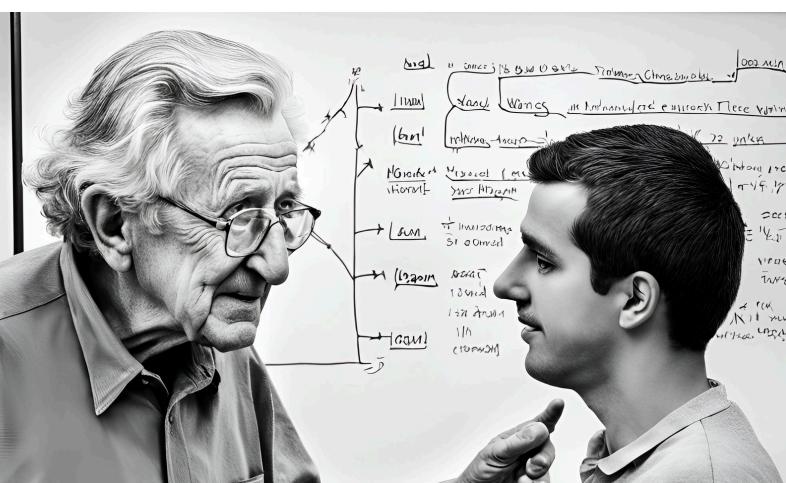
अतः "राम चलता है" का सत्य होना "राम बैठा नहीं है" इसको भी स्वयं सिद्ध कर देता है। इसी प्रकार यदि "राम चलता है" की सत्यता स्थापित है, तो यह ये भी निर्धारित कर सकता है कि "राम जाता है" क्योंकि जाने वाला व्यक्ति चलते हुए ही जाता है। इन्हीं गुणों को हमने अपने शोध में "सिमेटिक्स" कहा है। इस प्रकार की संकल्पना का प्रयोग रिचर्ड मॉटेंगु द्वारा "मॉटेंगु ग्रामर [Montague Grammar (MG)]" के तहत किया गया है।

इन्हीं "सिटेक्स और सिमेटिक्स के नियमों" का हमने गूढ़ता से अध्ययन करके 4 प्रकार की भाषाएँ बनाई हैं, जिनके अन्दर वाक्य संरचना की शक्ति होने के साथ-साथ, अर्थ-निर्धारण के लिए तर्क का भी प्रयोग किया जा सकता है। ये 4 भाषाओं की संरचना कुछ इस प्रकार दिखाई देती है:



MG और TG क्रमशः 'द स्कूल ऑफ फिलोसोफी' और 'द स्कूल ऑफ लिंग्विस्टिक्स' के शोधकर्ताओं द्वारा अध्ययन के दो प्रमुख शास्त्रीय क्षेत्र रहे हैं। यह शोध हिंदी भाषा का विश्लेषण करने और उसके औपचारिक समक्ष को विकसित करने पर केंद्रित है।

हिंदी के लिए जो ढाँचा विकसित किया गया है वह MG और TG का मिश्रण है। सफल वाक्य-विन्यास और इस तरह की शब्दार्थ-प्रक्रिया से हिंदी के प्राकृतिक भाषा प्रसंस्करण [Natural Language Processing (NLP)] में आशाजनक योगदान देने के लिए यह एक प्रयास है।



जिन भाषा-विद्वानों को परिवर्धित भाषाओं के बीच का मूल अंतर जानना है, उनके लिए मैं यह कहना चाहता हूँ कि -

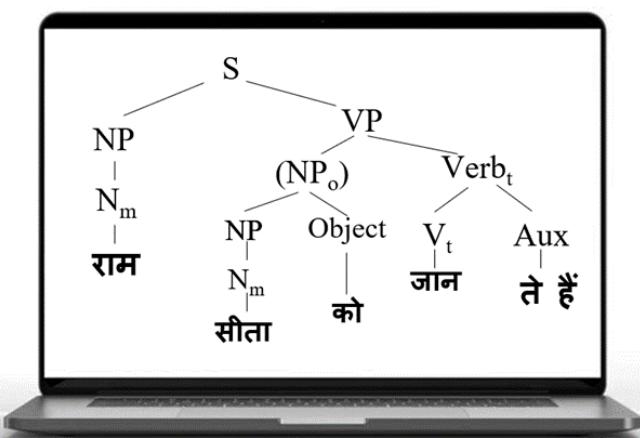
ऊपर बनाई गई चारों परिवर्धित भाषाएँ अपनी तार्किक भाषाओं के समरूपी हैं। इस शोध में सभी परिवर्धित भाषाओं को एक सॉफ्टवेयर के माध्यम से प्रदर्शित भी किया गया है। जहाँ L_{1H} भाषा मुख्य रूप से संदर्भात्मक संज्ञा के लिए वाक्यात्मक नियम विकसित करने पर ध्यान केंद्रित करती है, वहाँ L_{2H} गैर-संदर्भात्मक संज्ञाओं के लिए परिमाणीकरण (जिसे अंग्रेजी में quantification कहा जाता है) का उपयोग करती है। L_{3H} भाषाई अभिव्यक्तियों को उनके शब्दार्थ-प्रकारों के आधार (जिन्हें Type कहा जाता है; मूल टाइप केवल दो प्रकार के होते हैं - 'इ टाइप' और 'टी टाइप') पर वर्गीकृत करती है, और L_{4H} हिंदी के भूत और भविष्य काल के लिए "समय अवधि" के तर्क का प्रयोग करती है।



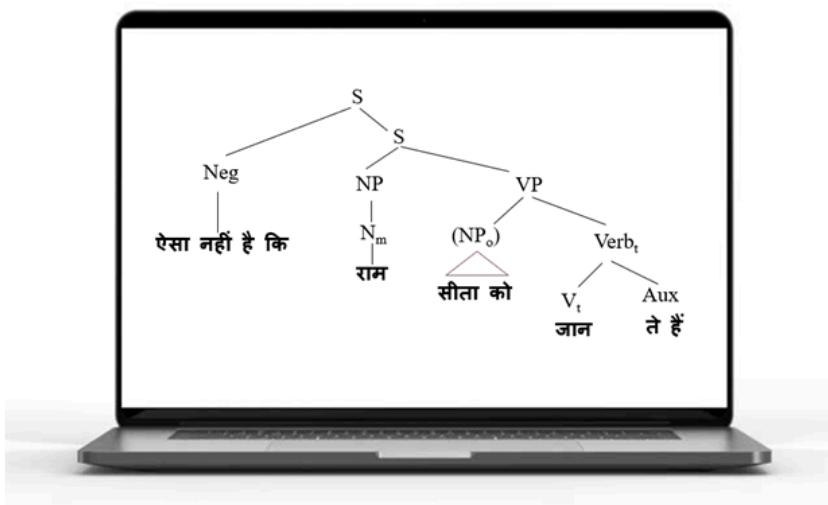
नवीनता एवं उपलब्धि

इस शोध अध्ययन से हिन्दी की वाक्य रचना एवं शब्दार्थ की कार्यप्रणाली की वैज्ञानिक समझ प्राप्त हुई है। परिवर्धित भाषाएँ जिनके भाषाई व्यवहार अलग हैं, वे मूल समाज में बोले जाने वाली हिंदी की वैज्ञानिक रूप से समझने के लिए एक संपूर्ण रूपरेखा प्रदान करते हैं। टाइप प्रणाली के प्रयोग से हिंदी के संयुक्त क्रियाओं के व्यवहार को समझने का एक वैकल्पिक दृष्टिकोण मिलता है। हिंदी में इस योगदान को कंप्यूटर हेतु क्रियान्वित करने के लिए हमने टी.एच.टी (THT) नामक पार्सर भी बनाया है जो हिंदी वाक्यों का भाषाई वाक्यविन्यास या उन्हें शाखाओं के रूप में तोड़ कर वाक्य को समझने का एक सरल तरीका बताता है। यह पार्सर छोटे या बड़े किसी भी हिंदी वाक्य को पार्स (वर्गीकरण) कर सकता है जो हिंदी के वाक्यविन्यास नियमों के अनुसार ही चलता है। हमने पाठकों की सुलभता हेतु सॉफ्टवेर से बनाये गए कुछ वाक्यों का वर्गीकरण नीचे दिया है।

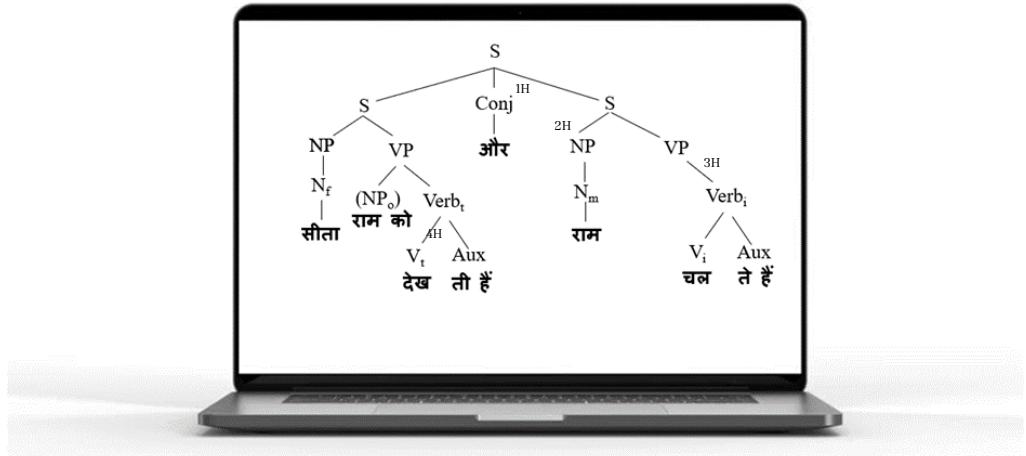
वाक्य १: राम सीता को जानते हैं।



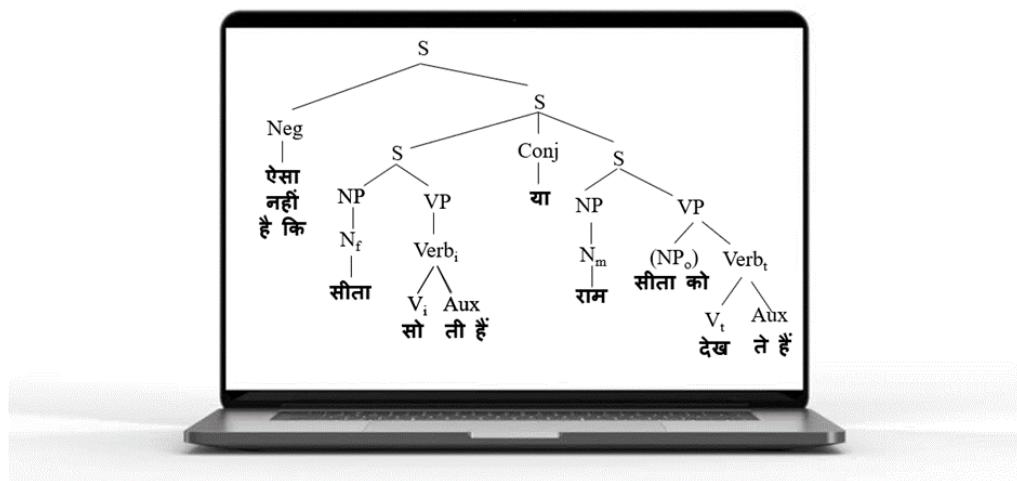
वाक्य २: राम सीता को नहीं जानते हैं। (इस वाक्य को following रूप से बदल के तोड़ा गया है।)



वाक्य ३: सीता राम को देखती हैं और राम चलते हैं।



वाक्य ४: सीता नहीं सोती हैं या राम सीता को देखते हैं (इस वाक्य को following रूप से बदल के तोड़ा गया है)।



वाक्य ५: राम चलते हैं या श्याम सोते हैं।

