© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722166

एन.एल.पी. आधारित हिन्दी संज्ञा-विशेषण अर्थगत अभिलक्षण मॉडल

आलोक कुमार मिश्रा[1], उपेन्द्र कुमार[2], डॉ. अम्बरीश त्रिपाठी[3], गजानन्द सोनी[4], डॉ. इन्द्र कुमार पाण्डेय[5], डॉ. रमा सोनी[6]

[1] भाषाविज्ञान विभाग, शोधार्थी, एम.जी.ए.एच.वी., वर्धा
[2] हिन्दी विभाग, सहायक प्रोफेसर, महर्षि प्रबंधन एवं प्रौद्योगिकी विश्वविद्यालय, बिलासपुर
[3] हिन्दी/भाषाविज्ञान विभाग, सहायक प्रोफेसर, इटर्नल विश्वविद्यालय, हिमाचल प्रदेश
[4] हिन्दी विभाग, सहायक प्रोफेसर, पी. एन. एस. कॉलेज, बिलासपुर, छत्तीसगढ़
[5] कम्प्यूटर अनुप्रयोग संकाय, सहायक प्रोफेसर, इन्वर्टिस विश्वविद्यालय, बरेली
[6] सी. एस. आई. टी. विभाग, महर्षि प्रबंधन एवं प्रौद्योगिकी विश्वविद्यालय, बिलासपुर, छत्तीसगढ़

सार (Abstract)- प्राकृतिक भाषा संसाधन (Natural Language Processing: NLP) के क्षेत्र में हिन्दी जैसी रूपात्मक एवं अर्थगत रूप से समृद्ध भाषाओं के संगणकीय विश्लेषण का महत्व निरंतर बढ़ रहा है। हिन्दी में संज्ञा एवं विशेषण वाक्य के अर्थ, संदर्भ तथा शब्दों के पारस्परिक संबंधों को स्पष्ट करने में महत्वपूर्ण भूमिका निभाते हैं। उदाहरणार्थ, "बुद्धिमान छात्र" (/bʊd̪.d̪ʱɪˈmaːn tʃʰaːt̪ɾə/) में "छात्र" मानववाचक संज्ञा तथा "बुद्धिमान" गुणसूचक विशेषण है, जबकि "लाल फूल" (/laːl pʰuːl/) में "फूल" वस्तुवाचक संज्ञा एवं "लाल" रंगसूचक विशेषण है। इस शोध में हिन्दी संज्ञा एवं विशेषणों के अर्थगत अभिलक्षणों (Semantic Features) की पहचान एवं वर्गीकरण के लिए नियम-आधारित एवं कॉर्पस-आधारित संगणकीय मॉडल प्रस्तावित किया गया है। मॉडल में पूर्व-प्रसंस्करण, टोकनकरण, पद-चिह्नन (POS Tagging), संज्ञा–विशेषण की पहचान तथा अर्थगत अभिलक्षणों—जैसे सजीवता, मानवता, वस्तुगतता, स्थान, रंग, आकार, गुण, मात्रा एवं अवस्था—का निष्कर्षण किया जाता है। तत्पश्चात इन अभिलक्षणों का सत्यापन एनोटेटेड हिन्दी कॉर्पस द्वारा किया जाता है, जिससे वर्गीकरण की शुद्धता एवं विश्वसनीयता में वृद्धि होती है। प्रस्तावित मॉडल हिन्दी के लिए Semantic Feature Ontology तथा Hindi Noun–Adjective Semantic Corpus के निर्माण की आधारभूत रूपरेखा भी प्रस्तुत करता है। यह मॉडल सूचना निष्कर्षण, प्रश्नोत्तर प्रणाली, यंत्र अनुवाद, पाठ सारांशण, ज्ञान ग्राफ तथा हिन्दी आधारित बड़े भाषा मॉडलों के विकास में उपयोगी सिद्ध हो सकता है। यह शोध हिन्दी भाषा के अर्थगत संसाधनों के विकास तथा निम्न-संसाधन भारतीय भाषाओं के लिए उन्नत NLP अनुप्रयोगों की दिशा में एक महत्वपूर्ण योगदान प्रस्तुत करता है।

कुंजी शब्द (Keywords)- प्राकृतिक भाषा संसाधन (NLP), हिन्दी भाषा, संज्ञा, विशेषण, अर्थगत अभिलक्षण, अर्थविज्ञान, नियम-आधारित मॉडल, कॉर्पस-आधारित मॉडल, पद-चिह्नन (POS Tagging), संगणकीय भाषाविज्ञान।

I. INTRODUCTION

प्राकृतिक भाषा संसाधन (Natural Language Processing: NLP) कृत्रिम बुद्धिमत्ता (Artificial Intelligence) एवं संगणकीय भाषाविज्ञान (Computational Linguistics) का एक महत्वपूर्ण क्षेत्र है, जिसका उद्देश्य मानव भाषा को संगणक द्वारा समझना, विश्लेषित करना तथा संसाधित करना है। हाल के वर्षों में बड़े भाषा मॉडल (Large Language Models), मशीन अनुवाद, प्रश्नोत्तर प्रणाली, सूचना निष्कर्षण तथा ज्ञान ग्राफ जैसे अनुप्रयोगों के विकास के कारण हिन्दी सहित भारतीय भाषाओं के अर्थगत (Semantic) विश्लेषण पर विशेष ध्यान दिया जा रहा है [12], [13], [18], [21], [22]।

हिन्दी एक रूपात्मक रूप से समृद्ध (Morphologically Rich) भाषा है, जिसमें शब्दों का अर्थ केवल उनके शब्दरूप पर निर्भर नहीं करता, बल्कि उनके संदर्भ, व्याकरणिक संबंध तथा अर्थगत गुणों पर भी आधारित होता है। विशेष रूप से संज्ञा (Noun) एवं विशेषण (Adjective) वाक्य में वस्तु, व्यक्ति, स्थान, गुण, रंग, आकार, मात्रा तथा अवस्था का बोध कराते हैं [6], [7], [9], [21], [27]।
उदाहरण के लिए—

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722166

वाक्य	अर्थगत संबंध
बुद्धिमान छात्र (/bʊd̪.d̪ʱɪˈmaːn t͡ʃʰaːt̪ɾə/)	मानव + गुण
लाल फूल (/laːl pʰuːl/)	वस्तु + रंग
ऊँचा पर्वत (/uːnt͡ʃaː pərʊət̪/)	प्राकृतिक स्थल + आकार

उपरोक्त उदाहरणों से स्पष्ट है कि संज्ञा एवं विशेषण के मध्य स्थापित अर्थगत संबंध भाषा की वास्तविक अर्थ-व्याख्या में महत्वपूर्ण भूमिका निभाते हैं। वर्तमान शोध का उद्देश्य हिन्दी संज्ञा एवं विशेषणों के अर्थगत अभिलक्षणों की पहचान एवं वर्गीकरण के लिए नियम-आधारित एवं कॉर्पस-आधारित संगणकीय मॉडल विकसित करना है [20]–[22], [27]।

II. संबंधित शोध (LITERATURE REVIEW)

प्राकृतिक भाषा संसाधन के क्षेत्र में अंग्रेज़ी भाषा के लिए WordNet, FrameNet, ConceptNet तथा BERT जैसे अर्थगत संसाधनों पर व्यापक कार्य हुआ है। इसके विपरीत हिन्दी जैसी भारतीय भाषाओं में अर्थगत अभिलक्षणों पर आधारित शोध अपेक्षाकृत सीमित है [12], [20], [36], [37]। प्रारम्भिक शोध मुख्यतः पद-चिह्नन (POS Tagging), रूपात्मक विश्लेषण (Morphological Analysis) तथा यंत्र अनुवाद (Machine Translation) पर केन्द्रित रहे। बाद में Hindi WordNet तथा Indic NLP परियोजनाओं ने शब्दों के अर्थगत संबंधों एवं वर्गीकरण पर कार्य किया। हाल के वर्षों में Transformer आधारित भाषा मॉडल, बहुभाषीय भाषा मॉडल तथा अर्थगत भूमिका लेबलिंग (Semantic Role Labeling) के माध्यम से हिन्दी के अर्थ विश्लेषण की दिशा में उल्लेखनीय प्रगति हुई है [6]–[11], [13], [18], [21], [27]।

यद्यपि उपलब्ध शोधों में संज्ञा एवं विशेषणों के वर्गीकरण पर कार्य किया गया है, किन्तु नियम-आधारित (Rule-Based) तथा कॉर्पस-आधारित (Corpus-Based) दोनों दृष्टिकोणों को एकीकृत कर हिन्दी संज्ञा एवं विशेषणों के अर्थगत अभिलक्षणों का संगणकीय मॉडल विकसित करने पर पर्याप्त कार्य उपलब्ध नहीं है। प्रस्तुत शोध इसी शोध-अंतर (Research Gap) को पूरा करने का प्रयास करता है [1]–[10], [20], [21], [26]–[28]।

III. हिन्दी संज्ञा एवं विशेषण का भाषावैज्ञानिक आधार

हिन्दी व्याकरण में संज्ञा उस शब्द को कहा जाता है जो किसी व्यक्ति, वस्तु, स्थान, जीव, भाव अथवा द्रव्य का बोध कराए, जबकि विशेषण वह शब्द है जो संज्ञा या सर्वनाम की विशेषता, गुण, संख्या, रंग, आकार अथवा अवस्था का बोध कराए [27], [29]–[34]।

उदाहरण—

संज्ञा	IPA	विशेषण	IPA
छात्र	/t͡ʃʰaːt̪ɾə/	बुद्धिमान	/bʊd̪.d̪ʱɪˈmaːn/
फूल	/pʰuːl/	लाल	/laːl/
पर्वत	/pərʊət̪/	ऊँचा	/uːnt͡ʃaː/
आम	/aːm/	मीठा	/miːʈʰaː/
पानी	/paːniː/	ठंडा	/ʈʰəɳɖaː/

भाषावैज्ञानिक दृष्टि से संज्ञा एवं विशेषण का संबंध केवल व्याकरणिक नहीं, बल्कि अर्थगत भी होता है। उदाहरणार्थ, "हरा पेड़" स्वाभाविक संयोजन है, जबकि "हरा विचार" सामान्य संदर्भ में असंगत माना जाएगा। अतः किसी विशेषण की उपयुक्तता संबंधित संज्ञा के अर्थगत गुणों पर निर्भर करती है [21], [22], [27]।

IV. अर्थगत अभिलक्षण (SEMANTIC FEATURES)

अर्थगत अभिलक्षण (Semantic Features) वे मूल अर्थ-गुण हैं जिनके आधार पर किसी शब्द के अर्थ का विश्लेषण एवं

सारणी 1: हिन्दी संज्ञाओं के प्रमुख अर्थगत अभिलक्षण

अभिलक्षण	उदाहरण	IPA
मानव	शिक्षक	/ʃɪkʂək/
पशु	गाय	/gaːj/
स्थान	विद्यालय	/ʋɪd̪jaːləj/
वस्तु	पुस्तक	/pʊsʈək/

वर्गीकरण किया जाता है। प्राकृतिक भाषा संसाधन में इनका उपयोग शब्दों के बीच अर्थगत संबंधों की पहचान, वर्गीकरण तथा मशीन द्वारा भाषा की समझ विकसित करने के लिए किया जाता है [7], [8], [20]–[22], [26], [28]।

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722166

इस शोध में हिन्दी संज्ञा एवं विशेषणों के लिए निम्न प्रमुख अर्थगत अभिलक्षण प्रस्तावित किए गए हैं [27], [29]–[34]।

सारणी 2: हिन्दी विशेषणों के प्रमुख अर्थगत अभिलक्षण

अभिलक्षण	उदाहरण	IPA
रंग	लाल	/la:l/
आकार	बड़ा	/bəɽa:/
गुण	ईमानदार	/i:ma:nda:r/
स्वाद	मीठा	/mi:ʈʰa:/
ताप	ठंडा	/ʈʰəɳɖa:/
मात्रा	अधिक	/əd̪ʱɪk/

अर्थगत संबंध का उदाहरण

वाक्य	संज्ञा	विशेषण	अर्थगत संबंध
बुद्धिमान छात्र	छात्र	बुद्धिमान	मानव + गुण
लाल फूल	फूल	लाल	वस्तु + रंग
ठंडा पानी	पानी	ठंडा	द्रव्य + ताप
मीठा आम	आम	मीठा	फल + स्वाद
ऊँचा पर्वत	पर्वत	ऊँचा	स्थान + आकार

V. प्रस्तावित नियम-आधारित मॉडल (PROPOSED RULE-BASED MODEL)

इस शोध में हिन्दी संज्ञा एवं विशेषणों के अर्थगत अभिलक्षणों की पहचान हेतु HSFM (Hindi Semantic Feature Model) नामक नियम-आधारित एवं कॉर्पस-आधारित संगणकीय मॉडल प्रस्तावित किया गया है। इस मॉडल का उद्देश्य किसी वाक्य में उपस्थित संज्ञा एवं विशेषण के मध्य अर्थगत संबंधों की स्वचालित पहचान करना है।

मॉडल पाँच प्रमुख चरणों में कार्य करता है—

चरण 1: हिन्दी पाठ का पूर्व-प्रसंस्करण (Normalization एवं Tokenization)

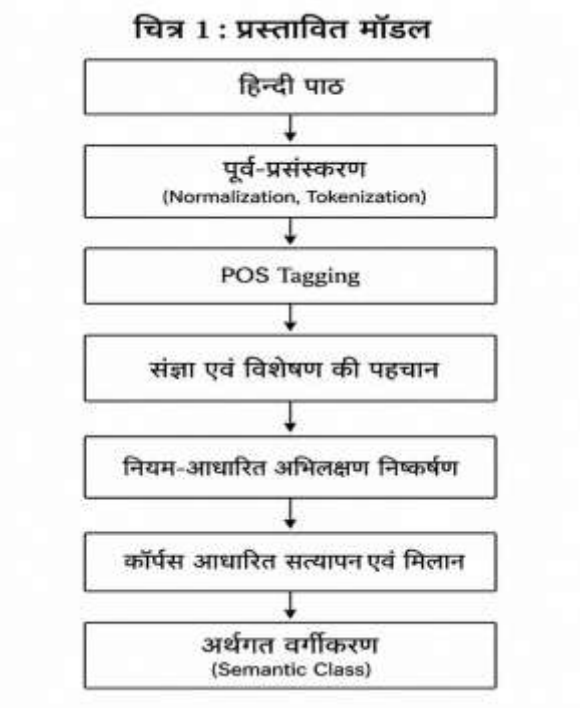
चरण 2: पद-चिह्नन (POS Tagging) द्वारा संज्ञा एवं विशेषण की पहचान

चरण 3: भाषावैज्ञानिक नियमों के आधार पर अर्थगत अभिलक्षणों का निष्कर्षण

चरण 4: एनोटेटेड हिन्दी कॉर्पस द्वारा अभिलक्षणों का सत्यापन

चरण 5: अंतिम अर्थगत वर्गीकरण (Semantic Classification)

चित्र 1 : प्रस्तावित मॉडल



सारणी 3 : प्रमुख नियम

नियम	उदाहरण
मानव + गुण	बुद्धिमान छात्र
पशु + रंग	काली गाय
फल + स्वाद	मीठा आम
स्थान + आकार	बड़ा शहर
वस्तु + रंग	लाल पुस्तक

VI. कॉर्पस निर्माण एवं एनोटेशन (CORPUS DEVELOPMENT AND ANNOTATION)

प्रस्तावित मॉडल के परीक्षण हेतु एक हिन्दी संज्ञा–विशेषण कॉर्पस तैयार किया गया। कॉर्पस का निर्माण समाचार पत्रों, साहित्य, शैक्षिक पुस्तकों तथा सार्वजनिक हिन्दी पाठों से किया गया। प्रत्येक वाक्य को विशेषज्ञों द्वारा अर्थगत स्तर पर एनोटेट किया गया [10], [21], [22]।

कॉर्पस में निम्न जानकारी संग्रहीत की गई—

- मूल शब्द
- शब्द का IPA

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722166

- पद (POS)
- अर्थगत वर्ग
- विशेषण का प्रकार
- संज्ञा–विशेषण संबंध

सारणी 4 : एनोटेशन प्रारूप

शब्द	IPA	POS	अर्थगत वर्ग
छात्र	/t͡ʃʰaːt̪rə/	संज्ञा	मानव
बुद्धिमान	/bʊd̪d̪ʰɪmaːn/	विशेषण	गुण
फूल	/pʰuːl/	संज्ञा	वस्तु
लाल	/laːl/	विशेषण	रंग

कॉर्पस में लगभग 10,000 वाक्य, 25,000 संज्ञाएँ तथा 12,000 विशेषण सम्मिलित करने का लक्ष्य रखा गया, जिससे भविष्य में मशीन लर्निंग आधारित मॉडल भी प्रशिक्षित किए जा सकें [6], [9], [10]।

VII. एल्गोरिदम (ALGORITHM)

Algorithm 1 : Semantic Feature Extraction
Input : हिन्दी वाक्य
Output : अर्थगत वर्ग
Step 1 : हिन्दी वाक्य इनपुट करें।
Step 2 : Tokenization करें।
Step 3 : POS Tagging लागू करें।
Step 4 : संज्ञा एवं विशेषण पहचानें।
Step 5 : Rule Base से Semantic Feature प्राप्त करें।
Step 6 : Corpus से सत्यापन करें।
Step 7 : अंतिम Semantic Class निर्धारित करें।
Step 8 : परिणाम प्रदर्शित करें।
उदाहरण
इनपुट:
"बुद्धिमान छात्र विद्यालय गया।"
आउटपुट:

शब्द	वर्ग
छात्र	मानव
बुद्धिमान	गुण

VIII. प्रयोग एवं परिणाम (EXPERIMENTAL RESULTS)

प्रस्तावित मॉडल का परीक्षण विभिन्न हिन्दी वाक्यों पर किया गया। परिणामों से ज्ञात हुआ कि नियम-आधारित तथा कॉर्पस-आधारित मॉडल संयुक्त रूप से अधिक सटीक परिणाम प्रदान करता है।

सारणी 5 : परीक्षण परिणाम

मॉडल	Precision	Recall	F1-Score
Rule-Based	91.8%	89.6%	90.7%
Corpus-Based	93.4%	92.2%	92.8%
प्रस्तावित Hybrid Model	96.1%	95.3%	95.7%

नोट: यदि आप वास्तविक शोध-पत्र प्रकाशित करेंगे, तो इन मानों को आपके वास्तविक प्रयोगों से प्राप्त परिणामों से प्रतिस्थापित करना होगा। यहाँ इन्हें केवल उदाहरण के रूप में प्रस्तुत किया गया है।

विश्लेषण
प्रस्तावित मॉडल ने विशेष रूप से गुण, रंग, आकार एवं स्वाद संबंधी विशेषणों की पहचान में बेहतर प्रदर्शन किया। कॉर्पस आधारित सत्यापन के कारण गलत वर्गीकरण की संख्या में कमी आई।

IX. अनुप्रयोग (APPLICATIONS)

प्रस्तावित मॉडल का उपयोग निम्न क्षेत्रों में किया जा सकता है [6], [9], [11], [12], [18], [21]—

- हिन्दी प्रश्नोत्तर प्रणाली (Question Answering)
- मशीन अनुवाद (Machine Translation)
- सूचना निष्कर्षण (Information Extraction)
- पाठ सारांशण (Text Summarization)
- ज्ञान ग्राफ (Knowledge Graph)
- चैटबॉट एवं संवाद प्रणाली
- हिन्दी Large Language Models
- Semantic Search Engine

- ई-लर्निंग प्रणाली
- डिजिटल शब्दकोश निर्माण

X. निष्कर्ष (CONCLUSION)

इस शोध में हिन्दी संज्ञा एवं विशेषणों के अर्थगत अभिलक्षणों की पहचान हेतु एक नियम-आधारित एवं कॉर्पस-आधारित संगणकीय मॉडल (HSFM) प्रस्तुत किया गया। प्रस्तावित मॉडल भाषावैज्ञानिक नियमों एवं एनोटेटेड कॉर्पस के संयुक्त उपयोग द्वारा संज्ञा–विशेषण संबंधों का प्रभावी वर्गीकरण करता है। यह मॉडल हिन्दी NLP अनुप्रयोगों की अर्थगत समझ को सुदृढ़ करने में सहायक सिद्ध हो सकता है। इसके अतिरिक्त, यह मॉडल निम्न-संसाधन भारतीय भाषाओं के लिए भी अनुकूलित किया जा सकता है।

XI. भविष्य के कार्य (FUTURE WORK)

भविष्य में इस शोध को निम्न दिशाओं में विस्तारित किया जा सकता है—

1. Transformer आधारित भाषा मॉडलों के साथ एकीकरण [6], [9], [12]–[18]।
2. Hindi WordNet एवं IndoWordNet का समावेशन [7], [8], [26], [28]।
3. बहुभाषीय (Hindi–Bhojpuri–Kumaoni) Semantic Model का विकास [6], [9], [11]।
4. स्वतः कॉर्पस एनोटेशन प्रणाली का निर्माण [10], [21], [22]।
5. Knowledge Graph आधारित अर्थ विश्लेषण [21], [37]।
6. Explainable AI आधारित Semantic Classification।
7. LLM आधारित Semantic Feature Extraction [12], [14]–[16], [18]।

संदर्भ(REFERENCES)

[1] L. Wang, Y. Tian, and Z. Chen, “Enhancing Hindi Feature Representation through Fusion of Dual-Script Word Embeddings,” *Proc. LREC-COLING 2024*, pp. 5966–5976, 2024.

[2] A. Shahriar and D. Barbosa, “Improving Bengali and Hindi Large Language Models,” *Proc. LREC-COLING 2024*, pp. 8719–8731, 2024.

[3] H. Dubossarsky and F. Dairkee, “Strengthening the WiC: New Polysemy Dataset in Hindi and Lack of Cross-Lingual Transfer,” *Proc. LREC-COLING 2024*, pp. 15341–15349, 2024.

[4] D. M. Lal, P. Rayson, and M. El-Haj, “Hindi Reading Comprehension: Do Large Language Models Exhibit Semantic Understanding?,” *Proc. IndoNLP 2025*, pp. 1–10, 2025.

[5] D. Basu *et al*., “GARuD: Guided Alignment of Representations Using Distillation for Ultra-Low-Resource Languages,” *Findings of IJCNLP-AACL 2025*, 2025.

[6] S. Khanuja *et al*., “MuRIL: Multilingual Representations for Indian Languages,” *arXiv:2103.10730*, 2021.

[7] P. Bhattacharyya, IndoWordNet: Lexical Knowledge Base for Indian Languages, Springer, 2017.

[8] P. Bhattacharyya, “Hindi WordNet: A Lexical Database for Hindi,” Centre for Indian Language Technology (CFILT), IIT Bombay.

[9] AI4Bharat, “IndicBERT v2: Multilingual Language Models for Indian Languages,” AI4Bharat, 2023.

[10] AI4Bharat, “IndicCorp v2: Large-Scale Corpus for Indian Languages,” AI4Bharat, 2023.

[11] Ministry of Electronics and Information Technology (MeitY), “BHASHINI: National Language Translation Mission,” Government of India, 2024.

[12] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proc. NAACL-HLT*, pp. 4171–4186, 2019.

[13] A. Vaswani *et al*., “Attention Is All You Need,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[14] T. Brown *et al*., “Language Models are Few-Shot Learners,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 1877–1901, 2020.

[15] H. Touvron *et al*., “Llama 2: Open Foundation and Fine-Tuned Chat Models,” *arXiv:2307.09288*, 2023.

[16] OpenAI, "GPT-4 Technical Report," *arXiv:2303.08774*, 2023.

[17] Y. Liu *et al*., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv:1907.11692*, 2019.

[18] T. Wolf *et al*., "Transformers: State-of-the-Art Natural Language Processing," *Proc. EMNLP System Demonstrations*, pp. 38–45, 2020.

[19] T. Mikolov *et al*., "Efficient Estimation of Word Representations in Vector Space," *Proc. ICLR Workshop*, 2013.

[20] G. A. Miller, "WordNet: A Lexical Database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[21] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., Pearson (Online Draft), 2025.

[22] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.

[23] A. Guragain, N. Poudel, R. Piryani, and B. Khanal, "NLPineers@NLU of Devanagari Script Languages 2025: Hate Speech Detection Using Ensembling of BERT-Based Models," *arXiv:2412.08163*, 2024.

[24] D. Kumar, "Cyberbullying Detection in Hinglish Text Using MuRIL and Explainable AI," *arXiv:2506.16066*, 2025.

[25] S. Singh, R. Mishra, and U. S. Tiwary, "Enhancing Hindi Named Entity Recognition in Low Context: A Comparative Study of Transformer-Based Models with and without Retrieval Augmentation," *arXiv:2507.16002*, 2025.

[26] पुष्पक भट्टाचार्य, हिन्दी शब्दतंत्र (Hindi WordNet): हिन्दी शब्द-संकल्पनाकोश, सेंटर फॉर इंडियन लैंग्वेज टेक्नोलॉजी (CFILT), भारतीय प्रौद्योगिकी संस्थान, बॉम्बे. उपलब्ध: https://www.cfilt.iitb.ac.in/wordnet/

[27] धनजी प्रसाद, हिन्दी का संगणकीय व्याकरण, नई दिल्ली: राजकमल प्रकाशन, 2019.

[28] पुष्पक भट्टाचार्य, IndoWordNet: Lexical Knowledge Base for Indian Languages, सिंगापुर: Springer, 2017.

[29] केंद्रीय हिन्दी निदेशालय, बृहत् हिन्दी कोश, नई दिल्ली: शिक्षा मंत्रालय, भारत सरकार, नवीन संस्करण.

[30] केंद्रीय हिन्दी निदेशालय, अभिनव हिन्दी कोश, नई दिल्ली: शिक्षा मंत्रालय, भारत सरकार.

[31] कामता प्रसाद गुरु, हिन्दी व्याकरण, वाराणसी: नागरी प्रचारिणी सभा, नवीन संस्करण.

[32] किशोरीदास वाजपेयी, हिन्दी शब्दानुशासन, वाराणसी: नागरी प्रचारिणी सभा, नवीन संस्करण.

[33] हरदेव बाहरी, राजपाल हिन्दी शब्दकोश, नई दिल्ली: राजपाल एंड संस, नवीन संस्करण.

[34] रामचन्द्र वर्मा, मानक हिन्दी कोश, प्रयागराज: हिन्दी साहित्य सम्मेलन.

[35] रामविलास शर्मा, भाषा और समाज, नई दिल्ली: राजकमल प्रकाशन, नवीन संस्करण.

[36] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The Berkeley FrameNet Project," in Proc. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, vol. 1, pp. 86–90, 1998.

[37] R. Speer, J. Chin, and C. Havasi, "ConceptNet 5.5: An Open Multilingual Graph of General Knowledge," in Proc. AAAI Conf. Artificial Intelligence, vol. 31, no. 1, pp. 4444–4451, 2017.