Journal of Advances and Scholarly Research In Allied Education
Vol.23, Issue No. 2 April-2026, ISSN 2230-7540

डिजिटल युग में हिंदी भाषा एवं उसकी बोलियों का संरक्षण और संवर्धन

डॉ. अरविन्द सिंह तेजावत*

सहायक आचार्य, हिंदी विभाग, हरियाणा केंद्रीय विश्वविद्यालय, महेंद्रगढ़, हरियाणा, भारत

office.tejawat@gmail.com

सार: डिजिटल प्रौद्योगिकी ने भाषाओं के अस्तित्व एवं विकास की शर्तों को मूलभूत रूप से परिवर्तित कर दिया है। जो भाषाएँ डिजिटल माध्यमों में पर्याप्त उपस्थिति नहीं बना पातीं, वे धीरे-धीरे प्रकार्यात्मक संकुचन की ओर बढ़ती हैं—इसी परिघटना को भाषाविज्ञान में "डिजिटल भाषा-मृत्यु" कहा गया है। प्रस्तुत शोध-पत्र इसी परिप्रेक्ष्य में हिंदी एवं उसकी बोलियों की स्थिति का परीक्षण करता है। अध्ययन में सर्वप्रथम डिजिटल भाषा-विभाजन की अवधारणा का विवेचन करते हुए हिंदी की डिजिटल यात्रा का कालक्रमिक विश्लेषण प्रस्तुत किया गया है—आइ.एस.सी.आइ.आइ. से यूनिकोड तक और स्थिर संगणक से मोबाइल तक। तत्पश्चात भाषा-प्रौद्योगिकी के प्रमुख क्षेत्रों—मशीन अनुवाद, प्रकाशिक वर्ण-अभिज्ञान, वाक्-संश्लेषण एवं वाक्-अभिज्ञान, संगणकीय कोश एवं संग्रह-निर्माण—में हिंदी की उपलब्धियों एवं कमियों का आकलन किया गया है। अध्ययन का एक प्रमुख निष्कर्ष यह है कि हिंदी की डिजिटल स्थिति यद्यपि संतोषजनक रूप से सुदृढ़ हुई है, तथापि उसकी बोलियाँ—ब्रजभाषा, अवधी, बुंदेली, बघेली, कन्नौजी एवं अन्य—डिजिटल परिदृश्य में लगभग अनुपस्थित हैं। इस असमानता के दीर्घकालिक परिणाम गंभीर हैं, क्योंकि हिंदी की समृद्धि का मूल स्रोत उसकी बोलियाँ ही रही हैं। अंत में अभिलेखीकरण, संग्रह-निर्माण, समुदाय-आधारित सामग्री-सृजन एवं नीतिगत हस्तक्षेप संबंधी सुझाव प्रस्तुत किए गए हैं।

मुख्य शब्द: डिजिटल भाषा-विभाजन, यूनिकोड, भाषा-प्रौद्योगिकी, मशीन अनुवाद, भाषिक संग्रह, बोली-अभिलेखीकरण, संकटग्रस्त भाषाएँ, डिजिटल भाषा-मृत्यु।

1. प्रस्तावना

किसी भाषा का भविष्य अब केवल उसके बोलने वालों की संख्या पर निर्भर नहीं रहा। बीसवीं शताब्दी तक भाषिक जीवंतता का प्रमुख मानदंड वक्ता-संख्या एवं अंतःपीढ़ीगत संचरण था; इक्कीसवीं शताब्दी में इसमें एक नया एवं निर्णायक आयाम जुड़ गया है—डिजिटल उपस्थिति। जो भाषा संगणक एवं मोबाइल उपकरणों पर सहजता से लिखी, खोजी, अनूदित एवं संसाधित नहीं की जा सकती, वह क्रमशः शिक्षा, व्यवसाय एवं सार्वजनिक विमर्श के क्षेत्रों से बाहर होती जाती है, भले ही उसके वक्ताओं की संख्या बड़ी हो [3], [6], [9]।

आंद्राश कोर्नाई ने अपने बहुचर्चित अध्ययन में यह दर्शाया कि विश्व की सात सहस्र से अधिक भाषाओं में से केवल कुछ सौ ही डिजिटल रूप से जीवंत हैं, और शेष के लिए "डिजिटल आरोहण" की संभावना क्षीण होती जा रही है [6]। इस पृष्ठभूमि में प्रस्तुत अध्ययन दो प्रश्नों का परीक्षण करता है: प्रथम, हिंदी की डिजिटल स्थिति वर्तमान में कैसी है और उसमें कौन-सी कमियाँ शेष हैं; द्वितीय, हिंदी की बोलियों की

डिजिटल स्थिति क्या है और उनके संरक्षण के लिए किस प्रकार के हस्तक्षेप अपेक्षित हैं। पद्धति की दृष्टि से यह अध्ययन विश्लेषणात्मक है तथा प्रौद्योगिकी-प्रतिवेदनों, संस्थागत दस्तावेज़ों एवं द्वितीयक शोध-साहित्य पर आधारित है।

2. डिजिटल भाषा-विभाजन की अवधारणा

"डिजिटल विभाजन" की चर्चा प्रायः उपकरण एवं संयोजकता की उपलब्धता तक सीमित रहती है—अर्थात् किसके पास संगणक एवं इंटरनेट है और किसके पास नहीं। किंतु इसका एक दूसरा एवं अधिक सूक्ष्म आयाम भाषिक है: उपकरण एवं संयोजकता उपलब्ध होने पर भी यदि उपयोगकर्ता की भाषा में सामग्री, अंतरापृष्ठ एवं सेवाएँ उपलब्ध न हों, तो वह वस्तुतः बाहर ही रह जाता है। यूनेस्को के लिए किए गए अध्ययन में जॉन पाओलिलो ने इसी को "भाषिक डिजिटल विभाजन" कहा और यह रेखांकित किया कि इंटरनेट पर भाषाओं का वितरण वास्तविक जनसंख्या-वितरण से अत्यधिक विषम है [9]।

डेविड क्रिस्टल ने भाषा-मृत्यु के अध्ययन में यह प्रतिपादित किया कि भाषा का लोप अचानक नहीं होता, अपितु प्रकार्यात्मक संकुचन की क्रमिक प्रक्रिया से होता है—भाषा पहले सार्वजनिक क्षेत्रों से हटती है, फिर शिक्षा से, फिर घरेलू व्यवहार से, और अंततः लुप्त हो जाती है [3]। डिजिटल युग में यह प्रक्रिया त्वरित हो गई है, क्योंकि सार्वजनिक विमर्श का बड़ा भाग अब डिजिटल मंचों पर स्थानांतरित हो चुका है। जोशुआ फ़िशमैन का "भाषा-परिवर्तन के प्रत्यावर्तन" संबंधी प्रतिरूप इस संदर्भ में विशेष रूप से उपयोगी है, क्योंकि वह हस्तक्षेप के स्तरों को क्रमबद्ध रूप से निर्धारित करता है [5]।

3. हिंदी की डिजिटल यात्रा

हिंदी की डिजिटल यात्रा को चार चरणों में विभाजित किया जा सकता है, जिनका संक्षिप्त विवरण निम्नलिखित सारणी में प्रस्तुत है।

सारणी 1: हिंदी की डिजिटल यात्रा का कालक्रम

चरण	प्रमुख विकास	सीमाएँ
प्रारंभिक चरण	विरासती फ़ॉन्ट-आधारित प्रणालियाँ; प्रत्येक फ़ॉन्ट का अपना संकेत-विन्यास	पाठ विनिमेय नहीं; खोज एवं छँटाई असंभव

चरण	प्रमुख विकास	सीमाएँ
मानकीकरण चरण	आइ.एस.सी.आइ.आइ. मानक तथा इनस्क्रिप्ट कुंजीपटल का विकास	सीमित अंगीकरण; अंतर्राष्ट्रीय समर्थन का अभाव
यूनिकोड चरण	देवनागरी का यूनिकोड में समावेश; ओपनटाइप फ़ॉन्ट	विरासती सामग्री का रूपांतरण अपूर्ण
मोबाइल एवं संजाल चरण	स्मार्टफ़ोन कुंजीपटल, लिप्यंतरण-आधारित निवेश, आवाज़-आधारित खोज	रोमन लिपि का बढ़ता प्रचलन; कोश-संसाधनों की कमी

यूनिकोड मानक का अंगीकरण हिंदी के डिजिटल इतिहास का निर्णायक मोड़ था। इससे पूर्व प्रत्येक प्रकाशन-संस्थान अपने निजी फ़ॉन्ट-संकेतन का प्रयोग करता था, जिससे एक प्रणाली में तैयार पाठ दूसरी में अपठनीय हो जाता था। यूनिकोड ने न केवल विनिमेयता प्रदान की, अपितु खोज, छँटाई, वर्तनी-परीक्षण एवं स्वचालित संसाधन को भी संभव बनाया। मोबाइल चरण में लिप्यंतरण-आधारित निवेश-पद्धति ने हिंदी टंकण की सीखने-संबंधी बाधा को लगभग समाप्त कर दिया, जिसका परिणाम हिंदी सामग्री के विस्फोटक विस्तार के रूप में सामने आया [7], [11], [14]।

4. भाषा-प्रौद्योगिकी में हिंदी की स्थिति

भारत सरकार के इलेक्ट्रॉनिक्स एवं सूचना प्रौद्योगिकी मंत्रालय के अंतर्गत भारतीय भाषाओं हेतु प्रौद्योगिकी विकास कार्यक्रम इस क्षेत्र की प्रमुख संस्थागत पहल है, जिसके अंतर्गत संग्रह-निर्माण, उपकरण-विकास एवं संसाधन-वितरण का कार्य किया जाता है [7]। प्रमुख क्षेत्रों की स्थिति इस प्रकार है:

- **मशीन अनुवाद:** अंग्रेज़ी-हिंदी अनुवाद-प्रणालियाँ अपेक्षाकृत परिपक्व हैं और तंत्रिका-आधारित प्रतिरूपों के आगमन से गुणवत्ता में उल्लेखनीय सुधार हुआ है; तथापि विधिक एवं तकनीकी पाठों में शुद्धता अभी भी सीमित है [2]।
- **प्रकाशिक वर्ण-अभिज्ञान:** मुद्रित देवनागरी के लिए स्थिति संतोषजनक है, किंतु हस्तलिखित सामग्री एवं पुरानी पांडुलिपियों के लिए त्रुटि-दर उच्च बनी हुई है।

- **वाक्-प्रौद्योगिकी:** हिंदी के लिए वाक्-संश्लेषण एवं वाक्-अभिज्ञान प्रणालियाँ उपलब्ध हैं, किंतु उनका प्रशिक्षण प्रायः मानक हिंदी पर हुआ है, अतः क्षेत्रीय उच्चारण-भेदों पर उनका निष्पादन दुर्बल रहता है।

- **संग्रह एवं कोश:** हिंदी के लिए अनेक भाषिक संग्रह विकसित हुए हैं, तथापि उनकी मात्रा एवं अनुशीर्षण-गुणवत्ता प्रमुख विश्व-भाषाओं की तुलना में अत्यंत सीमित है [4], [7]।

इस आकलन का सार यह है कि हिंदी "डिजिटल रूप से जीवंत" भाषाओं की श्रेणी में तो सम्मिलित हो चुकी है, किंतु "डिजिटल रूप से समृद्ध" भाषाओं की श्रेणी में नहीं। इस अंतराल का मुख्य कारण अनुशीर्षित एवं मुक्त रूप से उपलब्ध भाषिक संसाधनों की कमी है, जो आधुनिक भाषा-प्रौद्योगिकी की मूलभूत आवश्यकता है [2], [4], [6]।

5. बोलियों की डिजिटल स्थिति

हिंदी की डिजिटल प्रगति की चर्चा प्रायः इस तथ्य को ढक देती है कि यह प्रगति लगभग पूर्णतः मानक खड़ी बोली तक सीमित है। ब्रजभाषा, अवधी, बुंदेली, बघेली, कन्नौजी, हरियाणवी एवं छत्तीसगढ़ी—इन सभी की डिजिटल उपस्थिति नगण्य है। यह स्थिति विशेष रूप से विडंबनापूर्ण है, क्योंकि हिंदी का शास्त्रीय साहित्य-भंडार मुख्यतः इन्हीं बोलियों में रचा गया—सूरदास एवं बिहारी ब्रजभाषा में, तुलसीदास एवं जायसी अवधी में [1], [12], [13]।

सारणी 2: प्रमुख हिंदी बोलियों की डिजिटल उपस्थिति का आकलन

बोली		लिखित सामग्री	कुंजीपटल/निवेश-समर्थन	संग्रह/भाषा-प्रौद्योगिकी संसाधन
ब्रजभाषा		शास्त्रीय साहित्य उपलब्ध, समकालीन न्यून	देवनागरी के माध्यम से अप्रत्यक्ष	लगभग अनुपस्थित
अवधी		शास्त्रीय साहित्य समृद्ध,	पृथक् समर्थन नहीं	लगभग अनुपस्थित

बोली		**लिखित सामग्री**	**कुंजीपटल/निवेश-समर्थन**	**संग्रह/भाषा-प्रौद्योगिकी संसाधन**
		डिजिटलीकरण आंशिक		
भोजपुरी		लोकगीत एवं दृश्य-सामग्री प्रचुर	आंशिक	प्रारंभिक स्तर पर
बुंदेली/बघेली		अत्यल्प	पृथक् समर्थन नहीं	अनुपस्थित
छत्तीसगढ़ी		राज्य-प्रायोजित सामग्री बढ़ती हुई	आंशिक	प्रारंभिक स्तर पर
कन्नौजी		नगण्य	पृथक् समर्थन नहीं	अनुपस्थित

सारणी से यह स्पष्ट है कि जिन बोलियों को राज्य-स्तरीय संस्थागत समर्थन प्राप्त हुआ अथवा जिनका सुदृढ़ लोकप्रिय संस्कृति-आधार रहा, उनकी डिजिटल स्थिति अपेक्षाकृत बेहतर है। इसके विपरीत जो बोलियाँ केवल शास्त्रीय साहित्य की भाषा रह गईं और जिनका समकालीन जन-प्रयोग संकुचित हुआ, वे डिजिटल परिदृश्य से लगभग अनुपस्थित हैं [1], [13], [15]।

यहाँ एक पद्धतिगत स्पष्टीकरण आवश्यक है। संकटग्रस्तता का आकलन करते समय वक्ता-संख्या को एकमात्र मानदंड मानना भ्रामक है। यूनेस्को के भाषा-संकट संबंधी ढाँचे में अंतःपीढ़ीगत संचरण, नए क्षेत्रों में प्रयोग तथा शैक्षिक सामग्री की उपलब्धता—इन सभी को स्वतंत्र मानदंड माना गया है [8]। इस दृष्टि से देखने पर अनेक हिंदी बोलियाँ, जिनके वक्ताओं की संख्या करोड़ों में है, फिर भी प्रकार्यात्मक संकुचन की स्थिति में हैं, क्योंकि नई पीढ़ी उन्हें घरेलू परिवेश से बाहर प्रयोग नहीं करती और डिजिटल क्षेत्र में उनका कोई स्थान नहीं है [3], [5], [8]।

6. हिंदी की डिजिटल उपस्थिति: आँकड़ात्मक चित्र

गुणात्मक विवेचन के साथ-साथ मात्रात्मक चित्र भी आवश्यक है, यद्यपि इस क्षेत्र में विश्वसनीय आँकड़े सीमित हैं और विभिन्न स्रोतों के अनुमानों में पर्याप्त भिन्नता मिलती है। केपीएमजी एवं गूगल के संयुक्त अध्ययन में यह अनुमान प्रस्तुत किया गया था कि भारतीय भाषाओं के इंटरनेट उपयोगकर्ताओं की संख्या अंग्रेज़ी उपयोगकर्ताओं से आगे निकल जाएगी और इस वृद्धि में हिंदी की भागीदारी सर्वाधिक होगी [14]। यह प्रवृत्ति दो कारकों से प्रेरित रही है—सस्ती आँकड़ा-सेवा की उपलब्धता तथा स्मार्टफ़ोन का ग्रामीण क्षेत्रों तक विस्तार।

सारणी 3: हिंदी की डिजिटल उपस्थिति के प्रमुख आयाम

आयाम	वर्तमान स्थिति	प्रमुख अंतराल
उपयोगकर्ता-आधार	भारतीय भाषाओं में सर्वाधिक; तीव्र वृद्धि	ग्रामीण-नगरीय एवं लैंगिक विषमता
सामग्री-निर्माण	समाचार, मनोरंजन एवं सामाजिक माध्यम में प्रचुर	शैक्षिक एवं शोध-सामग्री अत्यल्प
अंतरापृष्ठ-समर्थन	प्रमुख प्रचालन-तंत्र एवं अनुप्रयोगों में उपलब्ध	विशेषीकृत व्यावसायिक सॉफ़्टवेयर में सीमित
निवेश-पद्धति	लिप्यंतरण एवं वाक्-निवेश सुगम	देवनागरी टंकण-दक्षता का ह्रास
खोज एवं सूचना-प्राप्ति	हिंदी प्रश्नों का समर्थन उपलब्ध	प्रामाणिक हिंदी स्रोतों की कमी
भाषा-प्रौद्योगिकी संसाधन	संग्रह एवं उपकरण विकसित	अनुशीर्षित एवं मुक्त संसाधनों का अभाव

इन आँकड़ों की व्याख्या में सावधानी अपेक्षित है। उपयोगकर्ता-संख्या में वृद्धि स्वतः भाषिक सशक्तीकरण का संकेत नहीं है। यदि उपयोगकर्ता हिंदी में केवल उपभोग करता है—मनोरंजन देखता है, संदेश पढ़ता है—किंतु सृजन नहीं करता, तो भाषा की डिजिटल स्थिति सतही ही रहती है। भाषा-

प्रौद्योगिकी की दृष्टि से निर्णायक प्रश्न यह है कि उस भाषा में कितनी मौलिक, संरचित एवं पुनःप्रयोज्य सामग्री उत्पन्न हो रही है। इस कसौटी पर हिंदी की स्थिति उसकी उपयोगकर्ता-संख्या की तुलना में कहीं दुर्बल है [2], [6], [9]।

7. संस्थागत प्रयास एवं उनकी सीमाएँ

इस क्षेत्र में अनेक संस्थागत पहलें विद्यमान हैं। भारतीय भाषा संस्थान की भारतीय भाषाओं हेतु भाषिक आँकड़ा संघ पहल ने विभिन्न भाषाओं के लिए भाषिक संग्रह विकसित किए हैं। भारतवाणी परियोजना ने बहुभाषिक कोश एवं ज्ञान-सामग्री को एक मंच पर लाने का प्रयास किया। राष्ट्रीय पांडुलिपि मिशन ने पांडुलिपियों के सूचीकरण एवं आंशिक डिजिटलीकरण का व्यापक कार्य किया। इसके अतिरिक्त विकिपीडिया एवं विकिस्रोत जैसी स्वयंसेवी परियोजनाओं ने कतिपय भारतीय भाषाओं में उल्लेखनीय सामग्री-निर्माण किया है [7], [11], [14]।

इन प्रयासों की सीमाएँ भी स्पष्ट हैं। प्रथम, अधिकांश परियोजनाएँ अनुसूचित भाषाओं पर केंद्रित हैं, जिससे गैर-अनुसूचित बोलियाँ स्वतः बाहर रह जाती हैं। द्वितीय, अभिलेखीकरण एवं जीवंतीकरण के बीच का अंतर प्रायः उपेक्षित रहता है—किसी बोली की रिकॉर्डिंग संग्रहालय में सुरक्षित रख देना उसे जीवित रखने के समान नहीं है। तृतीय, परियोजनाएँ प्रायः अल्पकालिक अनुदान-चक्रों पर निर्भर होती हैं, जिससे दीर्घकालिक अनुरक्षण कठिन हो जाता है। चतुर्थ, समुदाय की भागीदारी प्रायः सीमित रहती है, जबकि भाषा-जीवंतीकरण संबंधी अनुसंधान यह दर्शाते हैं कि समुदाय-नेतृत्व के बिना कोई भी प्रयास टिकाऊ नहीं होता [5], [10], [15]।

संस्थागत प्रयासों के मूल्यांकन में एक और पक्ष विचारणीय है—प्रयासों का अभिविन्यास। अधिकांश परियोजनाएँ "ऊपर से नीचे" की पद्धति पर आधारित हैं, जिनमें विशेषज्ञ-संस्थान सामग्री का निर्माण करते हैं और समुदाय उसका प्राप्तकर्ता मात्र रहता है। भाषा-जीवंतीकरण संबंधी अंतर्राष्ट्रीय अनुभव यह दर्शाता है कि इस पद्धति की उपयोगिता अभिलेखीकरण तक तो है, किंतु जीवंतीकरण के लिए वह अपर्याप्त है। जिन प्रयासों में समुदाय स्वयं सामग्री-निर्माता की भूमिका में आया—चाहे वह स्थानीय विकिपीडिया हो, लोकगीत-संग्रह हो अथवा बोली में पॉडकास्ट—वहाँ परिणाम अधिक स्थायी रहे [5], [10]।

इस संदर्भ में यह उल्लेखनीय है कि हिंदी बोलियों के क्षेत्र में सर्वाधिक जीवंत डिजिटल गतिविधि किसी शासकीय परियोजना से नहीं, अपितु स्वतःस्फूर्त लोकप्रिय संस्कृति से उत्पन्न हुई है। भोजपुरी की डिजिटल उपस्थिति का आधार मुख्यतः उसका संगीत एवं दृश्य-सामग्री उद्योग रहा है, न कि कोई संरक्षण-

कार्यक्रम। इसी प्रकार हरियाणवी एवं राजस्थानी सामग्री का विस्तार भी लोक-संगीत एवं हास्य-सामग्री के माध्यम से हुआ। यह अनुभव एक महत्वपूर्ण नीतिगत संकेत देता है: बोलियों के डिजिटल संरक्षण की सर्वाधिक प्रभावी रणनीति संभवतः उन्हें संरक्षण का विषय बनाना नहीं, अपितु उनके समकालीन सृजनात्मक प्रयोग को सुगम बनाना है [10], [13], [15]।

तथापि इस स्वतःस्फूर्त प्रक्रिया की भी अपनी सीमा है। लोकप्रिय संस्कृति से उत्पन्न सामग्री प्रायः असंरचित एवं अननुशीर्षित होती है, अतः वह भाषा-प्रौद्योगिकी के लिए सीधे उपयोगी नहीं होती। अतः अपेक्षित यह है कि संस्थागत प्रयास इस स्वतःस्फूर्त सामग्री का प्रतिस्थापन करने के स्थान पर उसका अनुशीर्षण, संरचनीकरण एवं संग्रहण करें—अर्थात् दोनों धाराएँ पूरक बनें, प्रतिस्पर्धी नहीं [4], [7], [10]।

8. लिपि, मानकीकरण एवं तकनीकी समस्याएँ

देवनागरी की संरचनात्मक विशेषताएँ उसे ध्वन्यात्मक दृष्टि से अत्यंत सुसंगत बनाती हैं, किंतु डिजिटल संसाधन की दृष्टि से कतिपय विशिष्ट कठिनाइयाँ भी उत्पन्न करती हैं। संयुक्ताक्षरों का निर्माण, मात्राओं की स्थिति-निर्भरता, अनुस्वार एवं पंचमाक्षर के बीच विकल्प, नुक्ता का ऐच्छिक प्रयोग तथा एक ही उच्चारण के लिए एकाधिक संकेतन-विकल्पों की उपलब्धता—ये सभी पाठ-सामान्यीकरण की समस्या उत्पन्न करते हैं [7], [11]।

व्यावहारिक स्तर पर इसका परिणाम यह होता है कि दृष्टिगत रूप से समान दिखने वाले दो पाठ संकेतन-स्तर पर भिन्न हो सकते हैं, जिससे खोज-परिणाम अपूर्ण रहते हैं और संग्रह-निर्माण में अशुद्धियाँ आती हैं। यूनिकोड मानक इन विकल्पों को वैध मानता है, अतः समाधान मानक के स्तर पर नहीं, अपितु सामान्यीकरण-प्रक्रिया के स्तर पर खोजना होगा। हिंदी के लिए एक सर्वस्वीकृत पाठ-सामान्यीकरण मानक का अभाव आज भी एक महत्वपूर्ण तकनीकी रिक्ति है [4], [7]।

एक अन्य गंभीर समस्या विरासती सामग्री की है। यूनिकोड-पूर्व काल में तैयार विपुल हिंदी सामग्री—समाचार-पत्रों के अभिलेख, शासकीय दस्तावेज़, शोध-पत्रिकाएँ एवं प्रकाशन-गृहों की फ़ाइलें—निजी फ़ॉन्ट-संकेतन में बंद पड़ी है। इस सामग्री का यूनिकोड-रूपांतरण तकनीकी रूप से संभव है, किंतु प्रत्येक फ़ॉन्ट के लिए पृथक् मानचित्रण अपेक्षित होने के कारण यह श्रमसाध्य है। इस दिशा में समन्वित राष्ट्रीय प्रयास के अभाव में हिंदी का एक बड़ा बौद्धिक भंडार व्यावहारिक रूप से अनुपलब्ध बना हुआ है [7], [11], [14]।

रोमन लिपि का बढ़ता प्रचलन एक भिन्न प्रकार की चुनौती प्रस्तुत करता है। सामाजिक माध्यमों पर हिंदी का बड़ा भाग रोमन में लिखा जा रहा है, और इस लेखन की कोई मानक वर्तनी नहीं है—एक ही शब्द अनेक रूपों में लिखा जाता है। इससे दो परिणाम निकलते हैं: प्रथम, यह सामग्री संगणकीय विश्लेषण के लिए अत्यंत कठिन हो जाती है; द्वितीय, देवनागरी की पठन-लेखन दक्षता पर दीर्घकालिक प्रभाव की आशंका उत्पन्न होती है। यद्यपि लिपि-परिवर्तन को स्वतः भाषा-क्षय नहीं माना जा सकता, तथापि हिंदी के संदर्भ में देवनागरी का सांस्कृतिक एवं संवैधानिक महत्व इसे एक विचारणीय प्रश्न बनाता है [3], [11], [13]।

9. संरक्षण का संवैधानिक एवं नीतिगत आधार

बोलियों के डिजिटल संरक्षण की चर्चा को संवैधानिक एवं नीतिगत ढाँचे से जोड़े बिना वह केवल तकनीकी विमर्श बनकर रह जाती है। भारतीय संविधान में इस विषय से संबद्ध कतिपय प्रत्यक्ष एवं परोक्ष प्रावधान विद्यमान हैं। अनुच्छेद 29(1) नागरिकों के किसी भी वर्ग को अपनी विशिष्ट भाषा, लिपि एवं संस्कृति के संरक्षण का अधिकार प्रदान करता है—यह एक मौलिक अधिकार है और भाषिक समुदायों के लिए सर्वाधिक सुदृढ़ संवैधानिक आधार प्रस्तुत करता है। अनुच्छेद 350क राज्यों को भाषाई अल्पसंख्यक वर्गों के बालकों हेतु प्राथमिक स्तर पर मातृभाषा में शिक्षा की सुविधा देने का निदेश देता है, जो बोली-आधारित प्रारंभिक शिक्षा-सामग्री के विकास का आधार बन सकता है [12], [13]।

सारणी 4: बोली-संरक्षण से संबद्ध प्रमुख संवैधानिक एवं नीतिगत प्रावधान

प्रावधान	सारभूत उपबंध	डिजिटल संरक्षण की दृष्टि से प्रासंगिकता
अनुच्छेद 29(1)	विशिष्ट भाषा, लिपि एवं संस्कृति के संरक्षण का अधिकार	बोली-समुदायों हेतु मौलिक अधिकार का आधार
अनुच्छेद 350	किसी भी भाषा में अभ्यावेदन प्रस्तुत करने का अधिकार	बहुभाषिक डिजिटल शासन-सेवाओं का आधार
अनुच्छेद 350क	प्राथमिक स्तर पर मातृभाषा में शिक्षा की सुविधा	बोली-आधारित डिजिटल शिक्षण-सामग्री

प्रावधान	सारभूत उपबंध	डिजिटल संरक्षण की दृष्टि से प्रासंगिकता
अनुच्छेद 351	हिंदी द्वारा अन्य भाषा-रूपों से ग्रहण	बोलियों को स्रोत मानने का संवैधानिक निदेश
आठवीं अनुसूची	बाईस भाषाओं को विशेष स्थान	संसाधन-आवंटन में अनुसूचित भाषाओं की वरीयता

इस ढाँचे का परीक्षण करने पर एक संरचनात्मक विसंगति स्पष्ट होती है। संवैधानिक स्तर पर अनुच्छेद 29(1) एवं 351 बोलियों के लिए पर्याप्त आधार प्रस्तुत करते हैं, किंतु संसाधन-आवंटन की व्यावहारिक व्यवस्था आठवीं अनुसूची पर केंद्रित है। परिणामस्वरूप जो भाषा-रूप अनुसूची में नहीं हैं—और हिंदी की अधिकांश बोलियाँ इसी श्रेणी में आती हैं—वे संस्थागत सहायता से प्रायः वंचित रह जाते हैं। यह विसंगति डिजिटल क्षेत्र में विशेष रूप से तीव्र हो जाती है, क्योंकि भाषा-प्रौद्योगिकी संसाधनों के निर्माण में व्यय अधिक होता है और वह प्रायः शासकीय वित्तपोषण पर निर्भर रहता है [8], [12], [15]।

इस विसंगति के समाधान के लिए दो भिन्न दृष्टिकोण संभव हैं। प्रथम, अधिक भाषा-रूपों को आठवीं अनुसूची में सम्मिलित किया जाए—किंतु यह मार्ग राजनीतिक रूप से जटिल है और इससे "भाषा बनाम बोली" का विवाद और तीव्र होता है। द्वितीय, संसाधन-आवंटन को अनुसूची-आधारित मानदंड से पृथक् करके भाषिक आवश्यकता एवं संकटग्रस्तता के आधार पर पुनर्गठित किया जाए। द्वितीय मार्ग तकनीकी दृष्टि से अधिक व्यावहारिक प्रतीत होता है, क्योंकि वह अस्मिता-संबंधी विवाद में प्रविष्ट हुए बिना संरक्षण-कार्य को आगे बढ़ा सकता है [5], [8], [10]।

10. उभरती चुनौतियाँ

- **रोमन लिपि का बढ़ता प्रचलन:** सामाजिक माध्यमों पर हिंदी का बड़ा भाग रोमन लिपि में लिखा जा रहा है, जिससे देवनागरी की पठन-लेखन दक्षता पर दीर्घकालिक प्रभाव की आशंका है [11], [14]।
- **संसाधन-असंतुलन:** भाषा-प्रौद्योगिकी के आधुनिक प्रतिरूप विशाल आँकड़ा-राशि पर निर्भर हैं; जिन भाषा-रूपों के आँकड़े नहीं हैं, वे प्रौद्योगिकी से स्वतः बाहर हो जाते हैं।

- **मानकीकरण का दबाव:** स्वतः-सुधार एवं वर्तनी-परीक्षण उपकरण क्षेत्रीय रूपों को "त्रुटि" के रूप में चिह्नित करते हैं, जिससे भाषिक विविधता पर सूक्ष्म किंतु सतत दबाव पड़ता है।
- **अभिलेखीकरण की तात्कालिकता:** बोलियों के प्रामाणिक वक्ता प्रायः वयोवृद्ध पीढ़ी में हैं; अभिलेखीकरण में विलंब का अर्थ अपूरणीय क्षति है [3], [8]।
- **सामग्री का असंतुलन:** हिंदी में मनोरंजन-सामग्री प्रचुर है, किंतु विज्ञान, प्रौद्योगिकी एवं शोध-सामग्री अत्यल्प, जिससे भाषा का प्रकार्यात्मक विस्तार अवरुद्ध होता है [2], [6]।

11. सुझाव

- प्रत्येक प्रमुख बोली के लिए मानकीकृत पद्धति से ध्वनि-अभिलेख एवं अनुशीर्षित पाठ-संग्रह तैयार किए जाएँ तथा उन्हें मुक्त अनुज्ञप्ति के अंतर्गत उपलब्ध कराया जाए।
- अभिलेखीकरण को जीवंतीकरण से जोड़ा जाए—बोलियों में समकालीन सामग्री-सृजन, पॉडकास्ट, लघु-चलचित्र एवं शैक्षिक सामग्री को प्रोत्साहन दिया जाए।
- विद्यालयी स्तर पर बोली-आधारित प्रारंभिक शिक्षा-सामग्री विकसित की जाए, जो अनुच्छेद 350क की भावना के अनुरूप भी है।
- समुदाय-आधारित डिजिटल परियोजनाओं—स्थानीय विकिपीडिया, लोक-साहित्य संग्रह, मौखिक इतिहास—को तकनीकी एवं वित्तीय सहायता दी जाए।
- विरासती फ़ॉन्ट में उपलब्ध विपुल हिंदी सामग्री के यूनिकोड-रूपांतरण हेतु राष्ट्रीय स्तर का अभियान चलाया जाए।
- हिंदी में मौलिक वैज्ञानिक एवं तकनीकी लेखन को प्रोत्साहित किया जाए, जिससे भाषा का प्रकार्यात्मक क्षेत्र विस्तृत हो।
- भाषा-प्रौद्योगिकी उपकरणों के अभिकल्प में बोलीगत विविधता को त्रुटि नहीं, अपितु वैध विकल्प के रूप में समाहित किया जाए।

12. अध्ययन की सीमाएँ एवं भावी अनुसंधान-दिशाएँ

प्रस्तुत अध्ययन की कतिपय सीमाएँ स्पष्ट रूप से स्वीकार्य हैं। प्रथम, यह अध्ययन द्वितीयक स्रोतों एवं प्रकाशित प्रतिवेदनों पर आधारित है; इसमें कोई मौलिक आँकड़ा-संग्रह अथवा क्षेत्र-सर्वेक्षण सम्मिलित

नहीं है। द्वितीय, डिजिटल उपस्थिति संबंधी आँकड़े तीव्र गति से परिवर्तित होते हैं, अतः किसी भी मात्रात्मक आकलन की वैधता सीमित कालावधि तक ही रहती है। तृतीय, सारणी 2 में प्रस्तुत बोलियों का आकलन उपलब्ध विवरणात्मक साक्ष्य पर आधारित है, किसी मानकीकृत मापन-पद्धति पर नहीं—अतः उसे संकेतात्मक माना जाना चाहिए, निर्णायक नहीं [8], [12]।

इन सीमाओं को देखते हुए भावी अनुसंधान की कुछ दिशाएँ विशेष रूप से उपयोगी होंगी। सर्वप्रथम, हिंदी बोलियों के लिए एक मानकीकृत डिजिटल-जीवंतता सूचकांक का विकास अपेक्षित है, जो सामग्री-मात्रा, उपकरण-समर्थन, अंतःपीढ़ीगत प्रयोग एवं नए क्षेत्रों में उपस्थिति—इन सभी को समाहित करे। द्वितीय, सामाजिक माध्यमों पर उपलब्ध रोमन-लिपि हिंदी का संगणकीय विश्लेषण एक विस्तृत एवं अल्प-अन्वेषित क्षेत्र है, जो समकालीन भाषिक परिवर्तन को समझने की दृष्टि से अत्यंत मूल्यवान है। तृतीय, बोली-वक्ता समुदायों के बीच भाषिक अभिवृत्ति का अध्ययन आवश्यक है, क्योंकि संरक्षण-नीति की सफलता अंततः समुदाय की स्वयं की अभिवृत्ति पर निर्भर करती है। चतुर्थ, विरासती फ़ॉन्ट-सामग्री के रूपांतरण हेतु स्वचालित मानचित्रण-पद्धतियों का विकास एक व्यावहारिक एवं तात्कालिक उपयोगिता वाला अनुसंधान-क्षेत्र है [4], [5], [10], [15]।

13. निष्कर्ष

प्रस्तुत अध्ययन से यह निष्कर्ष निकलता है कि डिजिटल युग हिंदी के लिए संकट नहीं, अपितु अवसर सिद्ध हुआ है। यूनिकोड मानक के अंगीकरण, मोबाइल क्रांति एवं लिप्यंतरण-आधारित निवेश-पद्धति ने हिंदी को उस डिजिटल जीवंतता तक पहुँचाया है, जो एक दशक पूर्व अनिश्चित प्रतीत होती थी। तथापि यह प्रगति असमान है—वह मानक खड़ी बोली तक केंद्रित है और हिंदी की बोलियाँ इस परिवर्तन से लगभग अछूती रह गई हैं।

यह असमानता हिंदी के लिए भी हितकर नहीं है। ऐतिहासिक रूप से हिंदी की जीवंतता का स्रोत उसकी बोलियाँ रही हैं—शब्दावली, मुहावरे, लय एवं अभिव्यंजना-शक्ति उसे इन्हीं से प्राप्त हुई। यदि बोलियाँ क्षीण होंगी तो दीर्घकाल में मानक हिंदी भी निर्जीव एवं यांत्रिक होती जाएगी। अनुच्छेद 351 का वह निदेश, जिसमें हिंदी को अन्य भाषा-रूपों से ग्रहण करने को कहा गया है, इसी अंतर्दृष्टि पर आधारित था।

अंततः यह रेखांकित करना आवश्यक है कि डिजिटल संरक्षण का अर्थ केवल अभिलेखीकरण नहीं है। किसी बोली की ध्वनि-रिकॉर्डिंग को संरक्षित कर लेना उसे जीवित रखने के समान नहीं है—जीवंतता तभी है जब भाषा नई सामग्री रचती रहे, नई पीढ़ी उसे नए प्रयोजनों में प्रयुक्त करे और वह समकालीन

विमर्श में सहभागी बनी रहे। अतः नीति का लक्ष्य बोलियों को संग्रहालय में सुरक्षित करना नहीं, अपितु उन्हें डिजिटल सार्वजनिक क्षेत्र में सक्रिय भागीदार बनाना होना चाहिए। यही इस अध्ययन का केंद्रीय प्रतिपादन है।

संदर्भ सूची

1. ग्रियर्सन, जी. ए. (1903–1928). लिंग्विस्टिक सर्वे ऑफ़ इंडिया (खंड I–XI). कलकत्ता: ऑफिस ऑफ़ द सुपरिटेंडेंट ऑफ़ गवर्नमेंट प्रिंटिंग।
2. भट्टाचार्य, पी. (2015). मशीन ट्रांसलेशन. बोका रैटन: CRC प्रेस।
3. क्रिस्टल, डी. (2000). लैंग्वेज डेथ. कैम्ब्रिज: कैम्ब्रिज यूनिवर्सिटी प्रेस।
4. डैश, एन. एस. (2005). कॉर्पस लिंग्विस्टिक्स एंड लैंग्वेज टेक्नोलॉजी: विद रेफरेंस टू इंडियन लैंग्वेजेज. नई दिल्ली: मित्तल पब्लिकेशन्स।
5. फिशमैन, जे. ए. (1991). रिवर्सिंग लैंग्वेज शिफ्ट: थ्योरेटिकल एंड एम्पिरिकल फाउंडेशन्स ऑफ़ असिस्टेंस टू थ्रेटन्ड लैंग्वेजेज. क्लीवेडन: मल्टीलिंगुअल मैटर्स।
6. कोर्नाई, ए. (2013). "डिजिटल लैंग्वेज डेथ." PLOS ONE, 8(10), e77056.
7. झा, जी. एन. (2010). "द TDIL प्रोग्राम एंड द इंडियन लैंग्वेज कॉर्पोरा इनिशिएटिव." प्रोसीडिंग्स ऑफ़ द सेवंथ इंटरनेशनल कॉन्फ्रेंस ऑन लैंग्वेज रिसोर्सेज एंड इवैल्यूएशन (LREC 2010), वैलेटा, माल्टा।
8. मोसले, सी. (संपादक) (2010). एटलस ऑफ़ द वर्ल्ड्स लैंग्वेजेज इन डेंजर (तीसरा संस्करण). पेरिस: UNESCO पब्लिशिंग।
9. पाओलिलो, जे. (2005). मेज़रिंग लिंग्विस्टिक डाइवर्सिटी ऑन द इंटरनेट. पेरिस: UNESCO इंस्टीट्यूट फॉर स्टैटिस्टिक्स।
10. नेटल, डी., और रोमेन, एस. (2000). वैनिशिंग वॉयसेस: द एक्सटिंक्शन ऑफ़ द वर्ल्ड्स लैंग्वेजेज. ऑक्सफोर्ड: ऑक्सफोर्ड यूनिवर्सिटी प्रेस।
11. क्रिस्टल, डी. (2006). लैंग्वेज एंड द इंटरनेट (दूसरा संस्करण). कैम्ब्रिज: कैम्ब्रिज यूनिवर्सिटी प्रेस।

12. मासिका, सी. पी. (1991). द इंडो-आर्यन लैंग्वेजेज. कैम्ब्रिज: कैम्ब्रिज यूनिवर्सिटी प्रेस।

13. श्रीवास्तव, रवीन्द्रनाथ (1992). हिंदी भाषा का समाजशास्त्र. नई दिल्ली: राधाकृष्ण प्रकाशन।

14. KPMG इंडिया और Google (2017). भारतीय भाषाएँ: भारत के इंटरनेट को परिभाषित करना. मुंबई: KPMG इंडिया.

15. तिवारी, भोलानाथ (1966). *भाषाविज्ञान.* इलाहाबाद: किताब महल.