

भारतीय भाषा आणि सांस्कृतिक वारसा जतन करण्यात नैसर्गिक भाषा प्रक्रियेची (NPL) भूमिका

डॉ. सारिका राजेंद्र पाटील

सहाय्यक प्राध्यापिका, मराठी विभाग, कै. म. दि. सिसोदे कला, वाणिज्य आणि विज्ञान महाविद्यालय, नरडाणा,
जिल्हा धुळे.

Email: sarikapatil07583@gmail.com

सारांश:

प्रस्तुत संशोधनाचा उद्देश भारताच्या भाषिक व सांस्कृतिक वारशाच्या संरक्षणात नैसर्गिक भाषा प्रक्रियेची (Natural Language Processing) संकल्पनात्मक, तांत्रिक व कार्यात्मक भूमिका चिकित्सकपणे तपासणे हा आहे. भारतातील बहुभाषिकता, मौखिक परंपरा, हस्तलिखित संपदा आणि स्थानिक ज्ञानप्रणाली या सांस्कृतिक सातत्याच्या प्रमुख आधारस्तंभ असून, जागतिकीकरण, इंग्रजीचे वर्चस्व, डिजिटल दरी आणि भाष्ययुक्त डिजिटल पाठसंग्रहांचा अभाव या कारणांमुळे अनेक भाषा असुरक्षित स्थितीत पोहोचल्या आहेत, हे अध्ययनातून अधोरेखित होते. यंत्रानुवाद, ध्वनी-ओळख, अक्षर-ओळख आणि पाठसंग्रह विश्लेषण यांसारख्या भाषा तंत्रज्ञान साधनांच्या माध्यमातून हस्तलिखितांचे संगणक-वाचनीय रूपांतरण, मौखिक परंपरांचे संरचित दस्तऐवजीकरण आणि लुप्त होऊ पाहत असलेल्या भाषांचे संवर्धन साध्य होऊ शकते, अशी मांडणी करण्यात आली आहे. तथापि, लिपी विविधता, बोलीसमृद्धता, तांत्रिक पायाभूत सुविधांची विषमता आणि संसाधनांचे असमान वितरण ही अंमलबजावणीतील प्रमुख आव्हाने असल्याचे विश्लेषण दर्शविते. परिणामी, भाषावैज्ञानिक संवेदनशीलता, समुदायसहभाग आणि समतोल धोरणात्मक पाठबळ यांच्या समन्वयातूनच नैसर्गिक भाषा प्रक्रिया भारतीय भाषिक वारशाच्या दीर्घकालीन जतनास प्रभावीपणे योगदान देऊ शकते असे निष्कर्षात मांडले आहे.

मुख्य शब्द: नैसर्गिक भाषा प्रक्रिया, भाषिक वारसा, बहुभाषिकता, मौखिक परंपरा, यंत्रानुवाद, हस्तलिखिते

► *Corresponding Author: सारिका राजेंद्र पाटील*

१) प्रस्तावना:

भाषिक आणि सांस्कृतिक वारसा म्हणजे एखाद्या समाजाची सामूहिक स्मृती होय, जी त्याच्या भाषा, साहित्य, मौखिक परंपरा, लिपी, लोककथा, विधी, श्रद्धा आणि ज्ञानपरंपरा यांद्वारे व्यक्त होते. भाषा ही केवळ संवादाचे साधन नसून ती मूल्यव्यवस्था, विचारपद्धती आणि ऐतिहासिक जाणिवेचे संवाहन करणारी शक्ती आहे. जी. एन. देवी (१९९२) यांनी नमूद केल्याप्रमाणे भारतीय साहित्यिक व भाषिक परंपरांमध्ये सखोल सांस्कृतिक सातत्य आढळते; परंतु त्या जिवंत भाषिक समुदायांपासून दूर गेल्यास त्यांची असुरक्षितता वाढते. पुढे गणेश एन. देवी (२०१३) यांनी दाखवून दिले आहे की भारतातील अनेक भाषा प्रामुख्याने मौखिक परंपरेवर आधारलेल्या आहेत; त्यामुळे त्यांचे जतन करणे तातडीचे आणि आव्हानात्मक ठरते.

भारत हा जगातील सर्वाधिक बहुभाषिक देशांपैकी एक आहे. भारताच्या २०११ च्या जनगणनेनुसार देशात शेकडो भाषा व बोली प्रचलित आहेत. ही बहुभाषिकता सामाजिक विविधता आणि सांस्कृतिक समृद्धीचे प्रतीक आहे. ब्रज बी. कचरू (१९८३) यांनी स्पष्ट केले आहे की भारतातील भाषिक वातावरण व्यक्तीची ओळख, सामाजिक भान आणि ज्ञाननिर्मिती यांना आकार देते. भारतीय बहुभाषिक परंपरा प्रादेशिक साहित्य, लोकपरंपरा आणि स्थानिक ज्ञानप्रणाली यांना आधार देते. ती शिक्षण आणि प्रशासनात सर्वसमावेशक सहभागास चालना देते तसेच सांस्कृतिक लोकशाही अधिक बळकट करते. म्हणूनच भाषिक वैविध्याचे संरक्षण करणे हे भारताच्या सांस्कृतिक वारशाच्या जतनाशी निगडित आहे.

तथापि, आज भारतीय भाषा अनेक गंभीर आव्हानांना सामोऱ्या जात आहेत. उच्च शिक्षण, तंत्रविकास आणि प्रशासकीय क्षेत्रांत इंग्रजीच्या वाढत्या वर्चस्वामुळे भाषिक स्तरभेद निर्माण झाला आहे. इंग्रजी जागतिक संपर्कासाठी उपयुक्त असली तरी ती अनेक प्रादेशिक व आदिवासी भाषांना दुय्यम स्थानावर ढकलते. याशिवाय डिजिटल दरीमुळे ही समस्या अधिक तीव्र होते, कारण अनेक भारतीय भाषांमध्ये संगणकीय माहितीभांडार, प्रमाणित लिपी, अक्षररचना आणि जाव्यावर उपलब्ध साधनांची कमतरता आहे. युनेस्को (अहवालानुसार) भारतातील अनेक भाषा पिढ्यान्पिढ्या हस्तांतरण कमी झाल्यामुळे लुप्त होण्याच्या धोक्यात असल्याचे स्पष्ट केले आहे. एखादी भाषा दुर्बल झाल्यास तिच्याशी निगडित मौखिक परंपरा, लोकगीते आणि स्थानिक ज्ञानसंपदा देखील लयास जाण्याचा धोका निर्माण होतो.

या पार्श्वभूमीवर नैसर्गिक भाषा प्रक्रिया हे एक परिवर्तनकारी साधन म्हणून पुढे येते. नैसर्गिक भाषा प्रक्रिया म्हणजे संगणकीय तंत्रांच्या साहाय्याने मानवी भाषेचे विश्लेषण, आकलन आणि निर्मिती करण्याची पद्धत होय. भारतात भारतीय भाषांसाठी तंत्रविकास कार्यक्रम आणि विविध शासकीय उपक्रम भाषिक माहितीभांडार निर्मिती, यंत्रानुवाद, वाणी ओळख प्रणाली आणि लिपी ओळख साधने विकसित करण्यासाठी कार्यरत आहेत. या तंत्रज्ञानामुळे हस्तलिखितांचे अंकन, मौखिक परंपरांचे दस्तऐवजीकरण आणि प्रादेशिक साहित्याची व्यापक उपलब्धता शक्य होते. भारतीय संगणकीय भाषाविज्ञानातील अभ्यासकांच्या मतानुसार भाषा तंत्रज्ञान परंपरा आणि आधुनिक माध्यम यांतील दरी कमी करण्यास सहाय्यक ठरते.

सदर अध्ययनाचा उद्देश भारताच्या भाषिक आणि सांस्कृतिक वारशाच्या संरक्षणात नैसर्गिक भाषा प्रक्रियेची भूमिका तपासणे हा आहे. डिजिटल साधने भारतीय भाषांच्या जतन, सुलभता आणि पुनरुज्जीवनात कशाप्रकारे योगदान देतात तसेच त्यासमोरील अडचणी कोणत्या आहेत, याचे विश्लेषण या अध्ययनात केले जाईल. डिजिटल मानवविद्या आणि भाषा तंत्रज्ञान यांचा समन्वय साधून भारताच्या बहुभाषिक भविष्यासाठी सर्वसमावेशक व सांस्कृतिकदृष्ट्या संवेदनशील तांत्रिक आराखड्याची आवश्यकता या अध्ययनातून अधोरेखित करण्यात येते.

२) संशोधनाची उद्दिष्टे:

१. भारताच्या भाषिक व सांस्कृतिक वारशाच्या संदर्भात नैसर्गिक भाषा प्रक्रियेची संकल्पनात्मक व कार्यात्मक भूमिका स्पष्ट करणे.
२. भारतीय हस्तलिखिते, मौखिक परंपरा व लुप्तप्राय भाषांच्या जतनासाठी भाषा तंत्रज्ञानाच्या उपयोजनांचे विश्लेषण करणे.
३. भारतीय भाषांसाठी नैसर्गिक भाषा प्रक्रियेपुढील तांत्रिक, संरचनात्मक व सामाजिक आव्हाने मांडणे.
४. भारतीय भाषा आणि बोली भाषा संवर्धनासाठी तांत्रिकदृष्टिकोनातून दिशा निर्देशित करणे.

३) संकल्पनात्मक चौकट:

नैसर्गिक भाषा प्रक्रिया (Natural Language Process) ही संगणकशास्त्र व भाषाविज्ञान यांची संयुक्त अभ्यासशाखा असून ती संगणकीय प्रणालींना मानवी भाषा समजणे, विश्लेषित करणे आणि निर्माण करणे सक्षम करते. डॅनियल आणि जेम्स (२००९) यांच्या मतानुसार, ती नियमाधिष्ठित व सांख्यिकीय नमुन्यांवर आधारित पद्धतींनी लेखी व वाचिक भाषेचे प्रक्रियाकरण करते; संशोधनदृष्ट्या ती व्यापक प्रमाणावर भाषिक रचना व वापराच्या पद्धतशीर अभ्यासासाठी महत्त्वपूर्ण साधन ठरते.

या क्षेत्रातील प्रमुख साधनांमध्ये यंत्रानुवाद, उच्चार-ओळख, पाठसंग्रह विश्लेषण आणि अक्षर-ओळख प्रणाली यांचा समावेश होतो, ज्यांद्वारे भाषांतर, बोलीभाषेचे लेखनरूपांतरण, आशयनिर्धारण आणि छापील किंवा हस्तलिखित ग्रंथांचे डिजिटल रूपांतरण शक्य होते. भारत सरकारच्या भारतीय भाषांचा तांत्रिक विकास कार्यक्रम (TDIL) अंतर्गत भारतीय भाषा व लिपींसाठी अशा साधनांचा विकास करण्यात आला असून दीर्घकालीन जतनासाठी हे अत्यावश्यक आहे.

भारतीय संदर्भात विविध भाषा-कुटुंबे, लिपीप्रणाली आणि बोलीवैविध्य असल्याने नैसर्गिक भाषा प्रक्रियेसमोर विशेष आव्हाने उभी राहतात. मागील जनगणनेनुसार, भारतात शेकडो मातृभाषांची नोंद असून, पुरेशा प्रमाणित डिजिटल पाठसंग्रहांच्या अभावामुळे संशोधनात भाषिक दस्तऐवजीकरण व पाठसंग्रह निर्मितीला प्राधान्य देणे आवश्यक ठरते.

भारतीय भाषिक व सांस्कृतिक वारसा मौखिक परंपरा, लोकसाहित्य, अभिजात ग्रंथ आणि हस्तलिखितांमध्ये संरक्षित आहे; गणेश देवी (२०१३) यांनी अनेक भाषा मौखिक समुदायांवर अवलंबून असल्याचे अधोरेखित केले आहे. या

ज्ञानसंपदेचे डिजिटल रूपांतरण व शास्त्रीय नोंदणी न झाल्यास ती लुप्त होण्याचा धोका निर्माण होतो, म्हणून भाषावैज्ञानिक अभ्यास आणि तांत्रिक साहाय्य यांचा समन्वय आवश्यक आहे.

भाषा ही सांस्कृतिक ओळख आणि सामूहिक स्मृतीची प्रमुख वाहक आहे, म्हणून बी. कचरू (१९८३) लिहितात कि, ती सामाजिक अनुभव आणि सांस्कृतिक रूपांतरणाचे प्रतिबिंब दर्शविते. त्यामुळे भाषिक वारशाचे संरक्षण हे केवळ भाषावैज्ञानिक कार्य नसून सांस्कृतिक जबाबदारी आहे. त्यामुळे नैसर्गिक भाषा प्रक्रिया पारंपरिक ज्ञान आणि आधुनिक डिजिटल युग यांतील दुवा ठरू शकते.

४) भाषिक वारशाच्या जतनात नैसर्गिक भाषा प्रक्रियेची भूमिका:

हस्तलिखिते आणि अभिजात ग्रंथांचे डिजिटल रूपांतरण:

भारत हा जगातील सर्वाधिक समृद्ध हस्तलिखित परंपरा असलेला देश मानला जातो. संस्कृत, प्राकृत, फारसी, मराठी, तमिळ तसेच इतर अनेक भारतीय भाषांमध्ये निर्मित हस्तलिखितांमध्ये तत्त्वज्ञान, धर्म, न्यायशास्त्र, आयुर्वेद, गणित, खगोलशास्त्र, साहित्य आणि समाजविचार यांसंबंधीची अमूल्य ज्ञानसंपदा संरक्षित आहे. सन २००३ मध्ये सांस्कृतिक मंत्रालयांतर्गत स्थापन झालेल्या राष्ट्रीय हस्तलिखित अभियानाने देशभर विखुरलेल्या या संपदेचे शोधकार्य, सूचीकरण आणि प्राथमिक जतनकार्य केले आहे. तथापि, केवळ सूचीकरणाने किंवा भौतिक जतनाने संशोधनात्मक उपयोगीता सिद्ध होत नाही. हस्तलिखितांचे छायाचित्रण किंवा स्कॅनिंग ही प्रक्रिया प्रारंभीची असली तरी ती अंतिम उद्दिष्ट नाही; कारण प्रतिमा स्वरूपात उपलब्ध असलेली सामग्री शोधसुलभ, संपादनयोग्य आणि विश्लेषणक्षम होत नाही.

यासाठी नैसर्गिक भाषा प्रक्रियेवर आधारित अक्षर-ओळख प्रणालींचा वापर अपरिहार्य ठरतो. अशा प्रणाली छापील किंवा हस्तलिखित प्रतिमांचे संगणक-वाचनीय मजकुरात रूपांतर करतात, ज्यामुळे त्या सामग्रीचे अनुलेखन, संपादन, तुलना आणि शोधकार्य शक्य होते. तथापि, भारतीय लिपींची विविधता या प्रक्रियेस मोठे आव्हान निर्माण करते. मोडी, ग्रंथ, शारदा, तसेच देवनागरीच्या प्रादेशिक रूपभेदांसारख्या लिपींच्या संगणकीय ओळखसाठी स्वतंत्र नमुना-विकास आवश्यक आहे. माधव देशपांडे (१९९३) यांनी भारतीय भाषांच्या आणि लिपींच्या ऐतिहासिक स्तररचनेचा विचार करतांना त्यांतील कालानुक्रमिक परिवर्तन व सामाजिक संदर्भ अधोरेखित केले आहेत. ही ऐतिहासिक गुंतागुंत स्वयंचलित ओळख तंत्रासाठी तांत्रिक आणि भाषावैज्ञानिक अशा दोन्ही स्तरांवर आव्हानात्मक ठरते.

म्हणूनच हस्तलिखितांचे डिजिटल रूपांतरण ही केवळ तांत्रिक प्रक्रिया नसून ती भाषाशास्त्र, पाठसमीक्षा आणि संगणकीय पद्धती यांच्या समन्वयातून साध्य होणारी बहुआयामी संशोधनक्रिया आहे. प्रामाणिकरणाच्या नावाखाली लिपी विविधता नष्ट होऊ नये, तसेच मूळ पाठवैशिष्ट्ये, संहिताशैली आणि भाषिक वैशिष्ट्ये जतन राहावीत याची काळजी घेणे आवश्यक आहे. आधुनिक प्रवेशसुलभता आणि ऐतिहासिक प्रामाणिकता यांचा समतोल साधणे हेच या प्रक्रियेचे प्रमुख ध्येय असले पाहिजे.

लुप्त होऊ पाहत असलेल्या भाषांचे संरक्षण:

भारतातील अनेक भाषा सामाजिक, आर्थिक आणि शैक्षणिक दबावांमुळे झपाट्याने लुप्तप्राय अवस्थेकडे वाटचाल करीत आहेत. भाषेचा न्हास हा केवळ शब्दसंग्रहाचा लोप नसून एका विशिष्ट समुदायाच्या सांस्कृतिक स्मृती, ज्ञानप्रणाली आणि विश्वदृष्टीचा लोप होय. अन्विता अब्बी (२००६) यांनी अंदमानी भाषांच्या संदर्भात दाखवून दिले आहे की भाषासोबत पर्यावरणीय ज्ञान, लोकपरंपरा आणि विशिष्ट सांस्कृतिक ओळख देखील लुप्त होते. या पार्श्वभूमीवर नैसर्गिक भाषा प्रक्रियेवर आधारित उच्चारण ते लिपी प्रणाली उरलेल्या भाषिकांच्या कथनांचे, संभाषणांचे आणि गीत-परंपरांचे दस्तऐवजीकरण करण्यास साहाय्यक ठरतात.

तथापि, तंत्रज्ञान हे साहाय्यक साधन असून भाषासंवर्धनाची खरी शक्ती समुदायातच निहित असते. डिजिटल शब्दकोश, भाषिक पाठसंग्रह आणि ध्वनिसंग्रह तयार करतांना स्थानिक समुदायांचा सहभाग, त्यांची संमती आणि सांस्कृतिक स्वामित्वाचा आदर अनिवार्य आहे. दस्तऐवजीकरणात केवळ शब्दरूपांची नोंद पुरेशी नसून त्यामागील सांस्कृतिक संदर्भ, वाक्प्रचार, आख्यायिका आणि मौखिक इतिहास यांचा समावेश आवश्यक आहे. नैसर्गिक भाषा प्रक्रिया साधने माहितीचे वर्गीकरण, अनुलेखन आणि विश्लेषण सुलभ करतात; मात्र संशोधन प्रक्रियेत नैतिक जबाबदारी आणि सामाजिक संवेदनशीलता या तत्वांचे पालन तितकेच आवश्यक आहे.

मौखिक परंपरांचे दस्तऐवजीकरण:

भारतीय सांस्कृतिक परंपरेत मौखिक साहित्याला विशेष स्थान आहे. लोककथा, लोकगीते, म्हणी, आख्यायिका आणि धार्मिक कथनपरंपरा या अनेक समुदायांच्या सांस्कृतिक जीवनाचा अविभाज्य भाग आहेत. रामानुजन (१९९१) यांनी

स्पष्ट केले आहे की लोककथा केवळ मनोरंजनाचे साधन नसून त्या समाजाच्या नैतिक मूल्यव्यवस्था, विनोदबुद्धी आणि श्रद्धासंस्थेचे प्रतिबिंब असतात. मौखिक परंपरा स्मृतीच्या आधारे पिढ्यांपिढ्या प्रवाहित होत असल्याने त्यांचे अस्तित्व तुलनेने अस्थिर असते.

नैसर्गिक भाषा प्रक्रियेच्या साहाय्याने अशा कथांचे ध्वनिमुद्रण, लिप्यंतरण आणि अनुलेखन केल्यास त्या संशोधनासाठी स्थिर अभिलेखांत रूपांतरित होतात. तथापि, मौखिक सादरीकरणातील स्वर, लय, उच्चार, हावभाव आणि श्रोत्यांशी होणारा संवाद हे घटक केवळ लेखी स्वरूपात पूर्णपणे व्यक्त होत नाहीत. त्यामुळे लिप्यंतरणासोबत ध्वनिमुद्रण, दृश्य नोंदी आणि संदर्भमाहितीचे जतन करणे आवश्यक आहे. पाठसंग्रह विश्लेषणाच्या साहाय्याने पुनरावृत्त कथानमुन्यांचे आणि आशयसूत्रांचे तुलनात्मक अध्ययन करता येते; परंतु डिजिटल संग्रहणाने परंपरांना स्थिर व अपरिवर्तनीय रूप देऊ नये. मौखिक साहित्य स्वभावतः गतिशील असते आणि त्याच्या परिवर्तनशीलतेला संशोधनात स्थान दिले गेले पाहिजे.

यंत्रानुवाद आणि ज्ञानसुलभता:

भारताच्या बहुभाषिक संरचनेत ज्ञानसुलभतेचा प्रश्न अत्यंत महत्त्वाचा आहे. कचरू (१९८६) यांनी भाषिक स्तररचना आणि सत्तासंबंध यांतील परस्परसंबंध अधोरेखित करताना दाखवून दिले आहे की विशिष्ट भाषांचे वर्चस्व ज्ञानप्रवेशावर प्रभाव टाकते. उच्च शिक्षण, न्यायव्यवस्था आणि प्रशासनात एखाद्या भाषेचे प्राबल्य असल्यास इतर भाषिक समुदायांना माहितीपासून वंचित राहावे लागू शकते. या संदर्भात नैसर्गिक भाषा प्रक्रियेवर आधारित यंत्रानुवाद प्रणाली विविध भारतीय भाषांमध्ये शैक्षणिक, प्रशासकीय आणि कायदेशीर माहिती उपलब्ध करून ज्ञानलोकशाहीला चालना देऊ शकतात.

पूर्ण किंवा संदर्भविरहित अनुवादामुळे आशयभ्रंश होऊ शकतो आणि सांस्कृतिक सूक्ष्मता हरवू शकते. म्हणूनच यंत्रानुवाद प्रणाली विकसित करताना विविध, संतुलित आणि संदर्भसमृद्ध पाठसंग्रहांचा वापर आवश्यक आहे. नैसर्गिक भाषा प्रक्रिया नियोजनबद्ध विकसित झाल्यास भाषिक वैविध्याचा सन्मान राखत सर्वसमावेशक ज्ञानप्रवेश आणि सांस्कृतिक समतोल यांना बळकटी देऊ शकते.

५) नैसर्गिक भाषा प्रक्रियेपुढील आव्हाने आणि मर्यादा:

भारतीय भाषांच्या संदर्भात नैसर्गिक भाषा प्रक्रियेच्या वापरासमोर उभे राहणारे सर्वात मूलभूत आव्हान म्हणजे भाष्ययुक्त आणि विश्वसनीय डिजिटल पाठसंग्रहांचा अभाव होय. कोणतीही संगणकीय प्रणाली प्रभावीपणे कार्य करण्यासाठी तिला व्याकरणचिन्हे, अर्थनिर्देश, संदर्भ आणि वापरपरिस्थिती यांसह संरचित माहितीची आवश्यकता असते. परंतु बहुतेक भारतीय भाषांमध्ये, विशेषतः प्रादेशिक आणि आदिवासी भाषांमध्ये, अशी समृद्ध व प्रमाणित माहिती उपलब्ध नाही. परिणामी, विकसित होणारी भाषिक नमुना-प्रणाली अपुरी ठरते किंवा ती प्रभावी वर्चस्व असलेल्या इंग्रजी व हिंदीसारख्या भाषांकडे झुकते. यामुळे भाषिक समतोल राखण्याऐवजी विद्यमान असमानता अधिक दृढ होण्याचा धोका निर्माण होतो.

दुसरे महत्त्वाचे आव्हान म्हणजे भारतीय भाषांची लिपीवैविध्यपूर्ण आणि बोलीसमृद्ध रचना. देवनागरी, तमिळ, बंगाली, गुरुमुखी इत्यादी विविध लिपींमध्ये लेखनसंहितेची स्वतंत्र वैशिष्ट्ये, संयुक्ताक्षरांची रचना, मात्रा-चिन्हे आणि शब्दक्रमातील लवचिकता आढळते. या वैशिष्ट्यांमुळे स्वयंचलित ओळख आणि विश्लेषण प्रक्रियेची गुंतागुंत वाढते. त्याचबरोबर एकाच भाषेच्या अनेक प्रादेशिक बोली अस्तित्वात आहेत; उच्चार, शब्दसंपदा आणि वाक्यरचना जिल्हानिहाय बदलते. प्रमाणभाषेवर आधारित प्रणाली स्थानिक बोलरूपे अचूकपणे समजू शकत नाही, ज्यामुळे तिची अचूकता आणि सर्वसमावेशकता मर्यादित राहते.

तिसरे आव्हान म्हणजे तांत्रिक साधनसामग्रीतील विषमता. ग्रामीण व आदिवासी भागांमध्ये डिजिटल पायाभूत सुविधा, स्थिर इंटरनेटसुविधा आणि तांत्रिक प्रशिक्षणाचा अभाव जाणवतो. परिणामी, ज्या समुदायांच्या भाषांचे संरक्षण अत्यावश्यक आहे, त्यांच्यापर्यंत विकसित साधने पोहोचत नाहीत. याशिवाय आर्थिक सहाय्य व संशोधनाचे केंद्र प्रामुख्याने व्यापक प्रमाणावर बोलल्या जाणाऱ्या भाषांवर राहते. सर्वसमावेशक धोरणात्मक पाठबळ आणि समसमान तांत्रिक प्रवेशाविना नैसर्गिक भाषा प्रक्रियेवर आधारित उपक्रम भाषिक विषमता कमी करण्याऐवजी अनवधानाने ती वाढवू शकतात.

६) भविष्यकालीन दिशा:

भारतात नैसर्गिक भाषा प्रक्रिया तंत्रज्ञानाचे भविष्य मोठ्या प्रमाणावर कृत्रिम बुद्धिमत्तेवर आधारित भाषिक नमुन्यांच्या विकासावर अवलंबून आहे. सध्या प्रगत भाषिक साधने मुख्यतः इंग्रजी आणि काही जागतिक भाषांपुरती मर्यादित आहेत; त्यामुळे मराठी, तमिळ, बंगाली, गुजराती तसेच आदिवासी भाषांसाठी मोठ्या प्रमाणावर दर्जेदार भाषिक माहिती-संकलनाची तातडीची गरज आहे. स्थानिक भाषिक माहितीवर प्रशिक्षित नमुने यांत्रिक भाषांतर, उच्चार ओळख आणि लेखननिर्मिती अधिक अचूक करू शकतात. योग्य नियोजनाने विकसित झाल्यास हे तंत्रज्ञान भाषिक विविधतेचे संरक्षण करत शिक्षण, संशोधन आणि प्रशासनाला मातृभाषेत बळ देऊ शकते.

भारत २०४७ या विकसित राष्ट्राच्या दृष्टीकोनात भाषा तंत्रज्ञानाची भूमिका अत्यंत महत्त्वाची ठरते. विकसित भारताची संकल्पना केवळ काही प्रभावी भाषांपुरती मर्यादित राहू शकत नाही; शासनव्यवस्था, ऑनलाईन शिक्षण, आरोग्य सेवा आणि न्यायव्यवस्था नागरिकांच्या मातृभाषेत उपलब्ध असणे आवश्यक आहे. नैसर्गिक भाषा प्रक्रिया साधनांमुळे ज्ञान सर्वसामान्यांपर्यंत त्यांच्या समजणाऱ्या भाषेत पोहोचू शकते. त्यामुळे हे तंत्रज्ञान केवळ नवोपक्रम नसून सामाजिक न्याय आणि सांस्कृतिक सातत्य यांचे प्रभावी माध्यम आहे.

भविष्यातील प्रगतीसाठी धोरणात्मक पाठबळ अत्यावश्यक आहे. शासन आणि शैक्षणिक संस्था यांनी भाषिक माहिती-संकलन, मुक्त स्रोत साधनांची निर्मिती आणि भारतीय भाषांतील संगणकीय प्रशिक्षण यासाठी गुंतवणूक करावी. हस्तलिखितांचे डिजिटायझेशन, लुप्तप्राय भाषांचे दस्तऐवजीकरण आणि बहुभाषिक डिजिटल व्यासपीठांची उभारणी यांना स्पष्ट धोरणात्मक दिशा देणे गरजेचे आहे. राष्ट्रीय नियोजन आणि निधी केवळ प्रमुख भाषांपुरता मर्यादित न राहता लहान व आदिवासी भाषांनाही समप्रमाणात मिळाला पाहिजे; अन्यथा तांत्रिक प्रगती असमान राहिल.

शेवटी, भाषातज्ञ आणि तंत्रज्ञ यांचे मिळून सहकार्य अत्यंत आवश्यक आहे. भाषा ही केवळ माहिती नसून ती इतिहास, संस्कृती आणि सामाजिक अर्थ यांची वाहक असते. तंत्रज्ञ साधने तयार करतात, परंतु व्याकरण, बोलीभेद आणि सांस्कृतिक संदर्भ यांचे सखोल ज्ञान भाषातज्ञांकडे असते. दोन्ही घटक एकत्र कार्य केल्यास भाषा प्रक्रिया प्रणाली अधिक अचूक आणि सांस्कृतिकदृष्ट्या संवेदनशील ठरतील.

७) निष्कर्ष:

उपरोक्त विश्लेषणावरून असा निष्कर्ष मांडता येतो की भारताचा भाषिक व सांस्कृतिक वारसा बहुभाषिकता, मौखिक परंपरा, लिपीवैविध्य आणि स्थानिक ज्ञानप्रणाली यांच्या आधारे उभा आहे; तथापि आधुनिक शिक्षणव्यवस्था, प्रशासकीय रचना, इंग्रजीचे वाढते वर्चस्व आणि डिजिटल दरी यांच्या परिणामस्वरूप अनेक भारतीय भाषा असुरक्षित स्थितीत पोहोचल्या आहेत. भाषेचा न्हास हा केवळ भाषिक रूपांचा लोप नसून सांस्कृतिक स्मृती, समुदायाधिष्ठित ज्ञान आणि ऐतिहासिक सातत्य यांच्या क्षीणतेशी निगडित आहे. म्हणून भाषिक वारशाचे संरक्षण हे केवळ भाषावैज्ञानिक कार्य न राहता व्यापक सांस्कृतिक आणि सामाजिक जबाबदारीचे स्वरूप धारण करते, ही बाब अध्ययनातून स्पष्ट होते.

या पार्श्वभूमीवर नैसर्गिक भाषा प्रक्रिया भारतीय भाषांच्या जतन, संवर्धन आणि ज्ञानसुलभतेसाठी परिणामकारक साधन म्हणून उदयास येते. हस्तलिखितांचे संगणक-वाचनीय रूपांतरण, मौखिक परंपरांचे ध्वनिमुद्रण व लिप्यंतरण, लुप्तप्राय भाषांचे दस्तऐवजीकरण आणि यंत्रानुवादाच्या साहाय्याने बहुभाषिक ज्ञानप्रवेश यांसारख्या उपयोजनांद्वारे परंपरा आणि आधुनिक डिजिटल माध्यम यांतील दरी कमी करण्याची क्षमता या तंत्रज्ञानात आढळते. तथापि, भाष्ययुक्त पाठसंग्रहांचा अभाव, लिपी व बोलीवैविध्य, तांत्रिक विषमता आणि संसाधनांचे असमान वितरण ही मर्यादा लक्षात घेतल्यास नियोजनबद्ध, संतुलित आणि संदर्भसंवेदनशील अंमलबजावणी अत्यावश्यक ठरते.

अध्ययनातून असे सूचित होते की नैसर्गिक भाषा प्रक्रिया ही स्वतंत्र तांत्रिक प्रक्रिया नसून भाषाविज्ञान, पाठसमीक्षा, समुदायसहभाग आणि धोरणात्मक पाठबळ यांच्या समन्वयातूनच परिणामकारक ठरू शकते. ऐतिहासिक प्रामाणिकता, सांस्कृतिक स्वामित्व आणि भाषिक विविधतेचा सन्मान राखत विकसित झालेली भाषा तंत्रज्ञान प्रणालीच ज्ञानलोकशाही, सामाजिक समावेशकता आणि सांस्कृतिक सातत्य यांना बळ देऊ शकते. म्हणूनच भारतीय बहुभाषिक भविष्यासाठी सर्वसमावेशक आणि सांस्कृतिकदृष्ट्या सजग तांत्रिक आराखडा उभारणे ही काळाची गरज असल्याचे या अध्ययनातून अधोरेखित होते.

८) संदर्भसूची:

१. Devy, Ganesh N. *After Amnesia: Tradition and Change in Indian Literary Criticism*. Orient Longman, 1992.
२. Devy, Ganesh N., editor. *The People's Linguistic Survey of India*. 50+ vols., Orient Blackswan, 2013–present.
३. Government of India. *Census of India 2011: Language Data*. Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, 2011.
४. Kachru, Braj B. *The Indianization of English: The English Language in India*. Oxford University Press, 1983.
५. Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing*. Pearson, 2009.
६. अब्बी, अन्विता. (२००६). ग्रेट अंडमानी भाषांचे लुप्त होत जाणारे स्वर : एक भाषावैज्ञानिक अभ्यास. नवी दिल्ली: ओरिएंट ब्लॅकस्वान.
७. देशपांडे, माधव म. (१९९३). संस्कृत आणि प्राकृत : समाजभाषावैज्ञानिक प्रश्नांचा अभ्यास. दिल्ली: मोतीलाल बनारसीदास.
८. राष्ट्रीय हस्तलिखित अभियान. (२००३). सांस्कृतिक मंत्रालय, भारत सरकार. अधिकृत अहवाल व सूचीपत्रे.
९. रामानुजन, अ. के. (१९९१). भारतातील लोककथा : सांस्कृतिक आणि साहित्यिक अभ्यास. न्यूयॉर्क: पॅथिऑन बुक्स.