ज्ञान शांति मैत्री

महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा

हिंदी शिक्षण अधिगम केंद्र

एवं

प्रौद्योगिकी अध्ययन केंद्र एवं कंप्यूटेशनल भाषा विज्ञान विभाग

के संयुक्त तत्वावधान में आयोजित त्रिदिवसीय राष्ट्रीय संगोष्ठी

दिनांक : 17-19 अगस्त, 2016

प्रतिवेदन/रिपोर्ट

संगोष्ठी

कार्यक्रम विवरण	कार्यक्रम का स्वरूप	कार्यक्रम स्थल	अवधि	आयोजक
प्राकृतिक भाषा संसाधन : हिंदी एवं अन्य भारतीय भाषाओं के नवीन पक्षों के संदर्भ में	राष्ट्रीय संगोष्ठी	वर्धा, महाराष्ट्र	17-19 अगस्त, 2016	हिंदी शिक्षण अधिगम केंद्र, म.गां.अं.हिं.वि., वर्धा तथा प्रौद्योगिकी अध्ययन केंद्र एवं कंप्यूटेशनल भाषा विज्ञान विभाग, म.गां.अं.हिं.वि., वर्धा

'प्राकृतिक भाषा संसाधन : हिंदी एवं अन्य भारतीय भाषाओं के नवीन पक्षों के संदर्भ में' विषय पर आयोजित राष्ट्रीय संगोष्ठी (दिनांक 17-19 अगस्त, 2016)

आधुनिक युग विज्ञान और प्रौद्योगिकी का युग है। कोई भी देश इसके प्रभाव से अछूता नहीं है। किसी भी भाषा को लेकर प्रौद्योगिकीय प्रणालियों का निर्माण एक चुनौतीपूर्ण कार्य है। पिछले दो दशकों में प्राकृतिक भाषाओं को समझने और प्रजनित करने के लिए आवश्यक कंप्यूटर प्रणालियों के विकास में काफी तेजी आई है। इसलिए आज इस पर बात करना अनिवार्य है। हिंदी शिक्षण अधिगम केंद्र इसके महत्व को ध्यान में रखते हुए प्रौद्योगिकी अध्ययन केंद्र एवं कंप्यूटेशनल भाषा विज्ञान विभाग के सहयोग से 'प्राकृतिक भाषा संसाधन : हिंदी एवं अन्य भारतीय भाषाओं के नवीन पक्षों के संदर्भ में' विषय पर तीन दिवसीय राष्ट्रीय संगोष्ठी का आयोजन किया।

संगोष्ठी का उद्घाटन करते हुए मा. कुलपति प्रो.गिरीश्वर मिश्र ने अपने उद्बोधन में कहा कि भाषा के कारण ही मानव सभ्यता आदिकाल से आधुनिक काल तक की यात्रा कर सकी है। मानव सभ्यता के विकास क्रम में अनेक साधनों और संसाधनों का आविष्कार और निर्माण हुआ है जिसमें 'कंप्यूटर' डिजिटल माध्यमों के विकास का उन्नत बिंदु है और यह मानव जीवन के प्रत्येक क्षेत्र को किसी न किसी रूप में प्रभावित किया है, भाषा जगत भी इससे अछूता नहीं है। कंप्यूटर की अद्वितीय क्षमता के

कारण इसके माध्यम से मशीनी अनुवाद, सूचना प्रत्याययन, प्रकाशिक अक्षर अभिज्ञान एवं कृत्रिम बुद्धि आदि क्षेत्रों में हिंदी और अन्य भारतीय भाषाओं के अनुप्रयोग हेतु कार्य किया जा रहा है। यह संगोष्ठी भी उसी दिशा में बढ़ाया गया एक महत्वपूर्ण कदम है जिसका शोध के क्षेत्र में दूरगामी प्रभाव होगा। इस त्रिदिवसीय राष्ट्रीय संगोष्ठी में 12 बाह्य विशेषज्ञों ने हिस्सा लिया और अपनी अंतर्दृष्टि से संगोष्ठी को सार्थक बनाने में योग दिया। इसके अतिरिक्त इसमें 108 प्रतिभागियों ने सहभागिता की और अपना शोध पत्र प्रस्तुत किया।

राजनाथ भट्ट के शोध पत्र का शीर्षक था- **भाषा-प्रौद्योगिकी : इस शताब्दी की प्रमुख आवश्यकता**। उन्होंने कहा कि भाषा-प्रौद्योगिकी का महत्व हर क्षेत्र में बढ़ रहा है। यह सूचना-क्रांति और ज्ञान-विस्तार का काल है। किसी भी व्यवसाय से

संगोष्ठी के उद्घाटन अवसर पर अध्यक्षता करते हुए मा. कुलपति प्रो. गिरीश्वर मिश्र, मा. प्रतिकुलपति प्रो. आनंद वर्धन शर्मा, संबोधन करते हुए भाषा विद्यापीठ के प्रो. विजय कुमार कौल, तत्कालीन निदेशक प्रो. अरबिंद कुमार झा, तत्कालीन कुलसचिव डॉ. राजेन्द्र प्रसाद मिश्र सहित अन्य अतिथिगण।

जुड़ा व्यक्ति आधुनिक ज्ञान से लाभान्वित होने के साथ-साथ सूचना-क्रांति का भरपूर उपयोग करने की इच्छा रखता है। इंटरनेट, ई-मेल, एस-एम-एस जैसे जन-उपयोगी साधन और तकनीक हर व्यक्ति इस्तेमाल करना चाहता है। इसके लिए यदि भाषा अथवा लिपि बाधक बन जाए, तो स्थिति गंभीर हो जाती है। इस दिशा में भाषा-प्रौद्योगिकी की भूमिका और अधिक हो जाती है। उन्होंने बताया कि देवनागरी इस देश की एक प्रमुख लिपि है, जिसमें संपर्क-भाषा हिंदी, इतिहास-ज्ञान-दर्शन की भाषाएँ संस्कृत, पालि, प्राकृत, कई प्रदेशों की राज्य-भाषाएँ जैसे-मराठी, कोंकणी, अन्य राज्यों की भाषाएँ, उपभाषाएँ जैसे-डोगरी, पहाड़ी, राजस्थानी, भोजपुरी, छत्तीसगढ़ी, मालवी, मैथिली, मगही, अवधी, ब्रज, कन्नौजी, खड़ी-बोली, हरियाणवी, कश्मीरी आदि लिखी-पढ़ी जाती हैं। देवनागरी का 'अनुकूलन' करने से अन्य भाषाएँ [लिखित एवं अलिखित] भी इसी लिपि में लिखी-पढ़ी जा सकती हैं। इसके लिए आवश्यक है कि भाषा-प्रौद्योगिकी का हर प्रकार से विकास और प्रसार किया जाए। भाषा-प्रौद्योगिकी विद्यापीठ की स्थापना करने से निम्नलिखित स्तरों पर कार्य किया जा सकता है-

1-देवनागरी लिपि का 'अनुकूलन' – इससे भारत की भाषाएँ/ उपभाषाएँ व बोलियाँ इस लिपि में लिखी जा सकेंगी। इस शोध-अध्ययन में कंप्यूटर एवं मोबाईल निर्माता अहम भूमिका निभा सकते हैं। ब्राह्मी-प्रसूत दक्षिण-एशिया की बोलियाँ इस नवीन-देवनागरी में सहज ही लिखी जा सकेंगी। इसके लिए ब्राह्मी लिपि के अध्ययन केंद्र/ विभाग स्थापित किए जा सकते हैं।

2-मानक हिंदी और अन्य भारतीय भाषा कोश- मानक हिंदी और हिंदी उप-भाषा कोश निर्मित होने से कोश निर्माण, शोध, खरीद आदि की संस्कृति उभर सकती है। इसके लिए अलग-अलग आकार-प्रकार के द्वि-भाषिक/द्वि-बोली/द्वि-

पारिभाषिक शब्दकोश तैयार हो सकते हैं। ऐसे शब्दकोश हिंदी के प्रसार के साथ-साथ हिंदी से इतर प्रदेशों के भाषा-भाषियों को हिंदी जानने के लिए प्रेरित कर सकते हैं।

3-संगणकीय भाषाविज्ञान विभाग- इसमें प्राचीन एवं अर्वाचीन भाषा वैज्ञानिक सिद्धांत प्रस्तुत होंगे ताकि शिक्षार्थियों को दोनों प्रकार से जानकार बनाया जा सके।

संगोष्ठी में स्वागत वक्तव्य देते हुए तत्कालीन निदेशक प्रो. अरबिंद कुमार झा तथा मंचासीन पदाधिकारी एवं अतिथिगण।

4-भाषा-प्रौद्योगिकी- इस विभाग के अंतर्गत मानव भाषा की विशेषताओं का मशीनी अध्ययन हो सकेगा। ऐसे विभाग में शब्द-निर्माण, शब्द-संरचना, वाक्य-संरचना, वाक्य-निर्माण सरीखे मुद्दों की बारीकियों को मशीन द्वारा विश्लेषित किये जाने की पद्धति समझी-समझाई जा सकेगी।

5-मशीनी अनुवाद विभाग- मशीनी अनुवाद इस युग की आवश्यकता बन चुका है। फ़िल्म, साहित्य, चिकित्सा, दर्शन, इतिहास, समाचार, विधि, हर विषय से संबंधित जानकारी पूरे विश्व में प्रचारित-प्रसारित करने का एकमात्र उपाय अनुवाद है। त्वरित अनुवाद मशीन द्वारा ही संभव हो सकता है। मशीन द्वारा अनुदित सामग्री को न्यूनतम सम्पादन के पश्चात प्रसारित किया जा सकता है। इस दिशा में ऐसा विभाग स्थापित करने की आवश्याकता है जहां शोधार्थी मशीनी-अनुवाद को अधिक से अधिक सटीक बनाने की ओर अग्रसर रहें।

संगणकीय भाषा विज्ञान के शोधार्थी और महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय के प्राध्यापक **डॉ. धनजी प्रसाद** ने '**हिंदी में योजक संदिग्धार्थकता और विसंदिग्धीकरण**' विषय पर अपना शोध पत्र प्रस्तुत किया। उन्होंने कहा कि किसी एक ही शब्द की रचना के किसी वाक्य में प्रयुक्त होने पर एक से अधिक शब्दवर्ग या टैग या अर्थ होने की स्थिति को संदिग्धार्थकता कहते हैं, जैसे- अब मुझे सोना चाहिए। इस वाक्य में 'सोना' शब्द के दो अर्थ हैं- 'एक मूल्यवान धातु = संज्ञा' और 'नींद में सोने की क्रिया = क्रिया'। अत: इसमें सदिग्धार्थकता है। प्रत्येक भाषा में कुछ शब्द ऐसे होते हैं जिनका प्रयोग एक से अधिक शब्दवर्गों के लिए होता हैं। ऐसे शब्दों को संदिग्धार्थक शब्द कहा जाता है। उन्होंने कहा कि संदिग्धार्थकता दो प्रकार की होती है। 1-शाब्दिक संदिग्धार्थकता। 2-संरचनात्मक संदिग्धार्थकता।

शाब्दिक संदिग्धार्थकता की स्थिति में वाक्य में प्रयुक्त शब्द के ही दो शब्दवर्ग होते हैं। इस कारण टैगिंग के दौरान उसे दो टैग प्राप्त होते हैं। जैसे – 'सोना के दो शब्दवर्ग हैं- 'संज्ञा' और 'क्रिया' । अत: इसे दो टैग दिए जाएँगे- '**NN** और **VB**'। इसी प्रकार 'आम' शब्द– के भी दो शब्दवर्ग हैं – 'संज्ञा' और 'विशेषण'। अत: इसे भी दो टैग दिए जाएँगे - '**NN**' और '**JJ**'।

इसी प्रकार संरचनात्मक संदिग्धार्थकता में एक से अधिक अर्थों की स्थिति प्राप्त होती है। यह स्थिति वाक्य स्तर पर होती है। ऐसे वाक्यों में आए पदबंध मूलत: किसी अन्य वाक्य के रूपांतरण होते हैं जो पदबंध के रूप में वाक्य में प्रयुक्त होते हैं। ऐसे वाक्यों का चॉमस्कीय विसंदिग्धीकरण बाह्य संरचना और आंतरिक संरचना की संकल्पना का प्रयोग करते हुए सरलता से किया जा सकता है। धनजी प्रसाद ने इसे समझने के लिए एक उदाहरण दिया - मैंने दौड़ते हुए शेर को देखा। इस वाक्य में 'दौड़ते हुए शेर' एक संदिग्धार्थक संरचना है। इसके विसंदिग्धीकरण के लिए हमें इसकी आंतरिक संरचना में आए दोनों वाक्यों को अलग-अलग व्यवस्थित करना होगा। यथा –

मैंने शेर को देखा, शेर दौड़ रहा था।
मैंने शेर को देखा, मैं दौड़ रहा था।

सत्येन्द्र कुमार अवस्थी ने अपने शोध पत्र- **स्वचालित शब्द-टैगिंग में संदिग्धार्थकता विषयक समस्याएँ** में बताया कि टैगिंग एक रोचक एवं चुनौतीपूर्ण प्रक्रिया है। टैगिंग से तात्पर्य, वाक्यांश में प्रयुक्त शब्दों को उनके पदविभाग के आधार पर अंकित करना है जैसे- राम/**NNP** घर/**NN** जाता/**VM** है/**VAUX** ।/**PUNC**। हिंदी की अपनी अलग प्रकृति तथा भिन्न संरचना है। इसकी अपनी संदिग्धार्थक जटिलताएँ हैं। संदिग्धार्थकता से तात्पर्य ऐसे शब्दों से हैं जिनका एक से अधिक प्रकार्य व अर्थ होता है।

हिंदी भाषा प्रौद्योगिकी विभाग के शोध छात्र **नितिन कुमार जानबाजी रामटेक** ने विसंदिग्धीकरण के इस मसले को मराठी के संदर्भ में देखा। उनके शोध पत्र का शीर्षक था- समनामी शब्दों का विसंदिग्धीकरण। **प्रवेश कुमार द्विवेदी** का विषय था- **पाठ संसाधन में संदिग्धार्थकता की प्रकृति**। उन्होंने कहा कि भाषा अभियांत्रिकी के प्रमुख दो लक्ष्य हैं: पहला प्राकृतिक भाषा का विश्लेषण, प्रक्रम एवं संश्लेषण करना तथा दूसरा प्राकृतिक भाषा को संगणक के समझने योग्य बनाना। इन दोनों ही परिस्थिति की समस्याओं में से एक प्रमुख समस्या संदिग्धार्थकता की समस्या है। मानवीय भाषिक अभिव्यक्ति के अंतर्गत संदिग्धार्थकता की समस्या कम होती है क्योंकि उसके पास संदर्भगत एवं सांसारिक ज्ञान का भण्डार होता है, जबकि संगणक का ज्ञान मानव द्वारा प्रदत्त ज्ञान तक ही सीमित होता है। जिस कारण संगणक में संदिग्धता की समस्या पर्याप्त मात्रा में होती है। चूंकि

संगोष्ठी में उपस्थित विश्वविद्यालय के पदाधिकारी, प्राध्यापकगण तथा प्रतिभागीगण।

संगणक में विश्व की समस्त भाषाओं के शब्दों को भण्डारित अथवा विस्तृत कॉर्पस का संकलन कर उसका विश्लेषण, प्रक्रम एवं संश्लेषण कर समझने योग्य बनाया जाता है। जहां पर एक शब्द, पद, पदबंध एवं वाक्य के अर्थ एकाधिक होते हैं, वहाँ यह संदिग्धता आ जाती है कि संगणक किस अर्थ को ग्रहण करे। यहाँ पर संगणक के लिए ज्यादा आवश्यक हो जाता है कि भाषा के

संगोष्ठी में उद्बोधन देते हुए भाषा विद्यापीठ के प्रो. हनुमान प्रसाद शुक्ल, मंचासीन कुलपति महोदय तथा अतिथिगण।

प्रत्येक स्तर पर एकाधिक अर्थ को प्रकट करने वाली समस्याओं को रेखांकित कर उसके समाधान हेतु नियमों का प्रतिपादन एवं निरूपण किया जाय। उन्होंने कहा कि जब भाषा को संगणक के साथ जोड़ा जाता है तो यह कई मायनों में भिन्न स्वरूप ग्रहण करती है। भाषा आधारित मानकीकरण और व्याकरणिक स्वरूप भाषा अभियांत्रिकी के प्रयोग हेतु आवश्यक हो जाते हैं। जहां भाषा में मानवीय व्यवहार चिंतन प्रक्रियाओं के जरिए संदर्भों से अर्थ ग्रहण करते हैं वहां प्रौद्योगिकी में भाषा का प्रयोग करने हेतु मानकीकरण की कोडिंग-डिकोडिंग की जाती है। चूंकि मशीन संदर्भ और विचारगत प्रक्रियाओं से भिन्न कार्य करती है, अतः अभियांत्रिकी का संबंध संदर्भों से इतर व्याकरणिक और मानकीकरण आधारित होता है। वह भाषा के जरिए व्यक्त किए गए किसी वाक्य को किसी एक अर्थ के रूप में ही ग्रहण कर पाती है। जिसकी ग्रहणीयता उसकी कोडिंग पर निर्भर करती है।

राजीव रंजन राय ने अपना शोध- **'लिंग्विस्टिक लैंडस्केप और भारतीय डायस्पोरा में पूर्वज भाषाओं की गतिशीलता'** विषय पर प्रस्तुत करते हुए बताया कि भारतीय डायस्पोरा उन प्रवासी भारतीय समुदायों से मिलकर बना है, जिन्होंने इतिहास के विभिन्न काल खंडों में क्रमिक रूप में अलग-अलग व्यवस्थाओं के अंतर्गत विश्व के अनेक स्थानों के लिए प्रस्थान किया। प्राचीन व्यापारिक और धार्मिक समूहों, औपनिवेशिक काल में सिद्ध दोष, अनुबंधित श्रमिकों 'कुली', गिरमिटिया, कंगनी, मैस्त्री से लेकर स्वतंत्र भारत के 'प्रतिभा-पलायन' से जुड़े समूह और आधुनिक वैश्विक प्रबंधकों व तकनीकी विशेषज्ञों द्वारा विश्व के अधिकांश देशों में भारतीय भाषा और संस्कृति सजीव रूप में उपस्थित है। डायस्पोरा में भारतीयों ने अत्यंत कठिन संघर्ष के साथ अपने कौशल और सृजनात्मक क्षमता के बल पर अपने नए गंतव्य देशों की सामाजिक दशाओं से सामंजस्य बनाते हुए अपनी पहचान को संरक्षित रखा है। भारतीय डायस्पोरा का विकास भारतीय भाषाओं/बोलियों और सांस्कृतिक मान्यताओं को संरक्षित रखने में सफल हुआ है। अकादमिक डायस्पोरा विमर्श के अंतर्गत प्रतिकूल माहौल में स्वसंस्कृति का पुनरुत्पादन और स्वभूमि से संबंध बनाए रखने की भावना को डायस्पोरा की अवधारणा का मुख्य लक्षण माना गया है। प्रवासी भारतीयों में डायस्पोरा समुदायों की उत्पत्ति और एक समुदाय के रूप में उनका विकास स्वत: स्फूर्त (सुई जेनेरिस)

न होकर मेज़बान देशों की सामान्य ऐतिहासिक प्रक्रियाओं से गहराई से जुड़ा है। इन देशों में हुए राजनीतिक, आर्थिक और सामाजिक परिवर्तन भारतीय प्रवासियों को डायस्पोरा समुदाय बनाने में सदा प्रभावित करते रहे हैं। भारत की सांस्कृतिक विविधता की ही भांति विदेशों में भी भारतीय समुदाय अत्यंत विविधतापूर्ण स्थिति में है। इनके द्वारा प्रदर्शित विविधता में क्षेत्रीयता, धर्म और भाषा प्रमुख निर्धारक हैं। जिसके मूल में सांस्कृतिक अस्मिता को केंद्रीय तत्व के रूप में माना जा सकता है। इन प्रवासी समुदायों द्वारा क्षेत्रीय और भाषिक आधार पर अपनी अस्मिता को विदेशी भूमि पर बनाए रखा गया है। बहुभाषिक सांस्कृतिक स्थिति में डायस्पोरा समुदायों में भाषा सर्वाधिक सन्निकट और प्रत्यक्ष पहचान के साथ ही सांस्कृतिक परिवर्तन या संरक्षण की सर्वाधिक संवेदी सूचक है।

पिछले कुछ दशकों में सामाजि भाषा विज्ञान की नई शाखा के रूप में 'लिंग्विस्टिक लैंडस्केप' अध्ययनों का उभार हुआ है, जिसके अंतर्गत चरम विविधता युक्त सांस्कृतिक नगरों की बहुभाषिक स्थिति और उनसे जुड़ी प्रक्रियाओं को बेहतर तरीके से समझने का प्रयास किया जा रहा है। लिंग्विस्टिक लैंडस्केप अध्ययनों के अंतर्गत सार्वजनिक रूप से दृश्यमान लिखित भाषा और अभिव्यक्ति के अन्य स्वरूपों को विवेचित किया जा रहा है। भारतवंशीय समुदायों द्वारा अपनी पूर्वज भाषाओं को सजग तौर पर अपनी अगली पीढ़ियों तक पहुंचाने के प्रयास के साथ-साथ चरम विविधता की बहुभाषिक सांस्कृतिक स्थिति में सार्वजनिक पहचान के एक माध्यम के रूप में अभिव्यक्त किया जा रहा है।

प्रियंका श्रीवास्तव ने **'भारत में अनुवाद के द्वारा संवाद की परम्परा'** विषय पर अपना शोध पत्र प्रस्तुत करते हुए अनुवाद और संवाद परम्परा को स्पष्ट किया तथा भारत में अनुवाद के इतिहास की समीक्षा की।

आशीष कुमार ने **विज्ञान तथा प्रौद्योगिकी साहित्य और अनुवाद** विषय पर अपने शोध पत्र में कहा कि मनुष्य को एक दूसरे के करीब लाने, मानवीय जीवन को अधिक संपन्न और सुखी बनाने तथा सारी दुनिया को एक कर 'विश्वग्राम' बनाने में अनुवाद की महती भूमिका रही है। हम सभी जानते हैं कि भाषाओं की भिन्नताएँ मनुष्य को एक दूसरे से अलग करती हैं। उदाहरण स्वरूप यदि भारत में बहुत सारी भाषाएँ न होतीं तो राष्ट्रीय एकता की समस्या भी उस स्तर पर सामने न आती जितना कि वर्तमान में हमारे सामने दिखती है। इक्कीसवीं सदी के बाद का समय कंप्यूटर एवं सूचना प्रौद्योगिकी का है। पूरी दुनिया के किसी भी कोने तक पहुँचने में सूचना प्रौद्योगिकी के माध्यम से अनुवाद कारगर साबित हो रहा है। मूलभूत भाषा में उत्पन्न ज्ञान की अवधारणा और उसके विस्तार को अनुवाद के द्वारा ही सूक्ष्मता व गहनता से अंतरित कर दूसरी भाषाओं में उपलब्ध कराया जा सकता है। आज हमारे पास विज्ञान व प्रौद्योगिकी का जो प्रचुर ज्ञान उपलब्ध है उसमें अनुवाद की भूमिका महत्वपूर्ण है। अकादमिक दुनिया ने शुरू-शुरू में यह भ्रम लोगों के दिमाग में बैठा दिया था कि वैज्ञानिक एवं प्रौद्योगिकी साहित्य की संकल्पनाओं को विश्व की कुछ भाषाएं ही आम लोगों तक पहुंचा रही हैं किन्तु अब यह भ्रामक भूत लोगों के दिमाग से उतर रहा है और कई देश अपनी भाषाओं के महत्व को पहचानने लगे हैं। **हेमंत कुमार राय** ने '**बुन्देलखंडी कविताओं में अलंकार का प्रयोग**' विषय पर शोध पत्र प्रस्तुत किया। **शिशिर कुमार और सरोज कुमार झा** ने '**बनारस की पिजिन भाषा**' पर अपना अंतर्दृष्टिपूर्ण शोध पत्र प्रस्तुत किया।

यह राष्ट्रीय संगोष्ठी विषय-विशेषज्ञों, प्रतिभागियों की सक्रिय सहभागिता से बहुत अर्थपूर्ण बन गयी। उल्लेखनीय है कि हिंदी शिक्षण अधिगम केंद्र ऐसी गुणवत्तापूर्ण संगोष्ठियों के लिए प्रतिबद्ध है।