मराठी विशेषण पदबंध चिन्हक

"Marathi Adjective Phrase Marker"

महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय में एम.फिल. कंप्युटेशनल भाषाविज्ञान उपाधि के पाठ्यक्रम संबंधी आवश्यकता की आंशिक परिपूर्ति हेतु प्रस्तुत

लघु शोध-प्रबंध

शोधार्थी

प्रसंजीत जगदीश ताकसांडे

एम.फिल. कंप्युटेशनल भाषाविज्ञान

सत्र:2013-14

शोध-निर्देशक

डॉ. विजय कौल

अधिष्ठाता

कंप्युटेशनल भाषाविज्ञान विभाग

सह-शोध-निर्देशक

जगदीप सिंह दांगी

रीडर

कंप्युटेशनल भाषाविज्ञान विभाग

कंप्युटेशनल भाषाविज्ञान विभाग, भाषा विद्यापीठ

महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय

(संसद द्वारा पारित अधिनियम 1997 अंतर्गत, क्रमांक 3 के अंतर्गत स्थापित)

पोस्ट बॉक्स सं. –16, गांधी हिल्स - वर्धा ,442005, (महाराष्ट्र)

फोन :07152-251661, फैक्स :07152-230903

ई मेल-hindiuni_wda@sancharnet.in,info@hindivishwa.org

वेबसाइट:www.hindivishwa.org

अनुक्रमणिका

- आवरण पृष्ट ----------1
- घोषणा पत्र ----------2
- प्रमाणपत्र ----------3-5
- आभार ----------6-7
- अनुक्रमणिका ----------8-9
- भूमिका 10-11

❖ **प्रथम अध्याय** :- 12-36

1) कंप्युटेशनल भाषाविज्ञान एवं पाठ संसाधन 13-22

1.1) कंप्युटेशनल भाषाविज्ञान 13

1.1.1) परिचय 13-19

1.1.2) अनुप्रयोगिक पक्ष 19-22

1.1.3) उपयोग एवं महत्व

1.2) पाठ संसाधन 23-36

1.2.1) मशीनी अनुवाद 23-28

1.2.2) टॅगर 29-32

1.2.3) पार्सर 33-36

❖ **द्वितीय अध्याय** :- 37-67

2) मराठी विशेषण पदबंध संरचना 38-55

2.1) मराठी वाक्य संरचना 56-61

2.2) मराठी पदबंध संरचना 62-65

2.3) मराठी विशेषण पदबंध संरचना का संगणकीय प्रारूप 66-67

2.3.1) टॅगिंग नियमन

2.3.2) विशेषण पदबंध प्रारूपम

❖ **तृतीय अध्याय :-** **68-77**

3) डाटाबेस प्रबंधन प्रणाली 69-77

3.1) डाटाबेस निर्माण एवं प्रबंधन

3.1.1) डाटाबेस स्वरूप 69-74

3.1.2) डाटाबेस निर्माण प्रक्रिया 75-77

3.1.3) डाटाबेस प्रबंधन

❖ **चतुर्थ अध्याय :-** **78-99**

4) प्रणाली निर्माण प्रक्रिया एवं प्रारूप

4.1) प्रणाली संरचना 79-83

4.1.1) कलनविधि एवं प्रवाह तालिका 83-87

4.1.1.2) कार्यप्रणाली एवं डाटाप्रवाह 88-93

4.2) अंतरापृष्ठ निर्माण

4.2.1) अंतरापृष्ठ उद्देश निर्माण 96

4.2.2) उद्देश के कार्य एवं अंतरसंबंध 97

4.2.3) कोड 98-99

- उपसंहार 100
- सीमाएं 101
- संदर्भ ग्रंथ सूची 102

भूमिका:-

भाषा समाज को देखने का माध्यम है। हम समाज को इसी माध्यम से देखते है। एक व्यक्ति का समाज में क्या स्थान है। यह उस व्यक्ति की भाषा ही बताती है। व्यक्ति की भाषा से पहचान होती है। भाषा मनुष्य को एक भाषाई समुदाय से जोड़ने का काम करती है, तो अन्य भाषाई समुदाय से तोड़ने का भी। मनुष्य भाषा के द्वारा ही समाज को देखता है। संसार में 2000 से भी जादा भाषाएँ बोली जाती है। भारत के संदर्भ में देखें तो 600 के आस-पास बोलिया है। भारत एक बहुभाषी देश है। जिसमें तकरीबन छब्बीस भाषाएँ संविधान द्वारा मान्य है।

सामन्यात: भाषा के वर्गिकरण का वैज्ञानिक प्रयास भाषा की प्रकृति की समानता और असमानता के आधार पर किया गया है। भाषाओं में मिलने वाली समानताएँ दो प्रकार की होती हैं। पहली समानता बाह्य संरचनात्मक होती है एवं दूसरी आंतरिक। जब भाषाओं को उनकी आकृति अर्थात् बाह्य समानता के आधार पर अलग-अलग वर्गों में वर्गीकृत किया जाता है तब उसे आकृतिमूलक वर्गिकरण कहते है। और जब भाषाओं को उनकी आंतरिक संरचना के आधार पर विभिन्न समूहों में बताया जाता तब इस प्रकार के वर्गिकरण को पारिवारिक या ऐतिहासिक वर्गिकरण कहा जाता है। भाषाओं में आंतरिक समानता से युक्त भाषा को एक ही परिवार का माना जाता है। इस प्रकार के वर्गिकरण को पारिवारिक वर्गिकरण कहा जाता है। इस पद्धति के अनुसार उन भाषाओं को ही एक वर्ग में रखा जाता है जिन भाषाओं में ऐतिहासिक संबंध रहता है। संसार की समस्त भाषाओं को भाषा परिवारों में बाँटा गया है। जिनका पारिवारिक वर्गिकरण इस प्रकार किया गया है। 1. यूरेशीय खण्ड, 2. अफ्रीका खण्ड 3. प्रशांत महासागरीय खण्ड 4. अमेरिका खण्ड। यूरेशीय खण्ड की भाषा परिवार में भारोपीय परिवार की भाषाओं का समावेश होता है। मराठी भाषा भारोपीय परिवार की भाषा है। जिसके मूल में संस्कृत है। मराठी भाषा का जन्म संस्कृत भाषा से हुआ है, संस्कृत इसकी जननी भाषा है। मराठी भाषा की शब्द संपदा संस्कृत से ही प्राप्त हुई है। मराठी भाषा के जादातर शब्द संस्कृत भाषा से ही आए हुए है। प्रस्तुत शोध में मराठी भाषा के विशेषण पदबंध को कंप्यूटर के माध्यम से चिन्हित करने का कार्य किया जाएगा साथ ही मराठी विशेषण पदबंध का विश्लेषण भी कंप्यूटर के द्वारा किया जाएगा।

भारतीय भाषाओं में मशीनी अनुवाद के लिए यह बहुत ही महत्त्वपूर्ण कार्य हो सकता है। साथ ही प्रकृतिक भाषा संसाधन के लिए भी यह कार्य उपयोगी सिद्ध हो सकता है। वर्तमान समय में मेरे संज्ञान में अभी तक विकसित रूप में कोई कार्य.....

प्रस्तावित शोध के माध्यम से मराठी भाषा के वाक्यों में निहित विशेषण पदबंध को चिन्हित किया जाएगा। उस स्थान विशेष पर उनके व्याकरणिक प्रकार्य को बताना। विशेषण पदबंध के प्रयोग की शुद्धता की जाँच करना और संज्ञा विशेषण और विशेषण पदबंध को रंगीत करेंगे जिससे पदबन्धो को पहचानने में आसानी होगी। यानि मुख्य रूप से शोध का उद्देश्य यह है कि कंप्यूटर के माध्यम से मराठी विशेषण पदबंध को पहचान/चिन्हित करना उसके के पच्छात विशेषण पदबंध को विश्लेषण कर के विशेषण पदबंध में जुड़े प्रत्यय और विभक्ति चिह्न या परसर्गों की पहचान करना।

'भारतीय समाजात सभ्यता आहे.' मराठी भाषा के इस वाक्य में 'भारतीय' शब्द 'भारत + इय' दो रूपिमों से मिलकर बना है, तथा इसमें 'भारत' मूल शब्द या प्रातिपदिक है और 'इय' प्रत्यय लगा है। 'भारत' संज्ञा को 'इय' प्रत्यय लगने से वह विशेषण बन गया है तथा यह 'इय' प्रत्यय व्युत्पादक प्रत्यय की श्रेणी में आता है। उपर्युक्त शोध कार्य में यह बताने का प्रयास किया गया है कि मराठी भाषा में विशेषण/विशेषण पदबन्धो के शब्दों को कंप्यूटर की सहायता से पहचानना/चिन्हित करना (Marathi adjective phrase marker), यह कंप्यूटर की सहायता से स्वचलित प्रणाली (Automatic System) के द्वारा /चिन्हित करना। इससे मराठी को प्राकृतिक भाषा (Natural Language) के रूप में संसाधन (Processing) करने में आसानी हो जाएगी साथ ही मराठी के विशेषण पदबन्धो को कंप्यूटर की सहायता से कॉर्पस निर्माण (Corpus Development) के कार्य में प्रयोग किया जाएगा। मशीनी अनुवाद के लिए यह एक उपयोगी कार्य सिद्ध होगा। प्रस्तावित शोध में विश्लेषणात्मक प्रविधि का उपयोग किया जाएगा। इसमें मराठी विशेषण पदबन्धो से संबंधित स्थितियों और नियमों का गलती और प्रयत्न (Try and Error) रूप से विवेचन किया जाएगा। इसके बाद इन्ही के आधार पर "मराठी विशेषण पदबन्ध चीन्हक" का विकास किया जाएगा।