

प्राकृतिक भाषा संसाधन के लिए नियमाधारित मराठी विशेषण अभिज्ञानक

डॉ. राहुल म्हैसकर, सहायक प्राध्यापक भाषाशास्त्र विभाग, डेक्कन कॉलेज अभिमत विश्वाविद्यालय, येरवडा, पुणे

सारांश :

भाषा की संरचना ही व्याकरण होता है, जो नियमों से संयोजित होता है। मराठी व्याकरण के विशेषण का प्रभाव संपूर्ण वाक्य पर पड़ता है। सामान्य व्याकरण मनुष्य की बौद्धिक क्षमता एवं विश्लेषण सृजन शक्ति को ध्यान में रखकर रचे जाते हैं, जबकि संगणकीय व्याकरण संगणक को केन्द्र में रखकर (जैन, 1995: 113)। प्राकृतिक भाषा संसाधन के लिए विशेषण की पहचान संगणक द्वारा करने के लिए विश्लेषणात्मक अध्ययन करना आवश्यक होता है। मराठी के विशेषण अभिज्ञान के लिए विशेषण विश्लेषण को रूपविश्लेषणात्मक आधार और संरचनात्मक आधार और स्थान परिवर्तन तथा रूपपरिवर्तन के कारण होने वाला परिवर्तन यहाँ प्रस्तुत किया गया है। उदा. यदि धातु को वा, पट प्रत्यय लगे हैं तो उन्हें विशेषण समझा जाए। जैसे. पाचवा, पाचपट, ई। संगणक में यह नियम तथा विशेषण पहचान से संबंधित विश्लेषण के नियम एवं प्रत्यय आदि डाटाबेस और प्रोग्रामिंग के माध्यम से दिए गए हैं। इसके आधार पर ही संगणक द्वारा विशेषण की पहचान होती है।

संकेत शब्द: प्राकृतिक भाषा संसाधन, मराठी, विशेषण.

प्रस्तावना :

किसी भी भाषा की एक संरचना होती है जिसके संरचक होते हैं, वे संरचक एक निश्चित क्रम में व्यवस्थित होकर वाक्य बनाते हैं और इसी अनुक्रम में भाषा का विकास होता है। जैसे - संज्ञा, सर्वनाम, विशेषण, क्रिया और अव्यय आदि संबंध निर्धारित करते हैं। यह तत्व संरचकों को किसी न किसी नियम के द्वारा नियंत्रित करते हैं। यह संरचना ही व्याकरण होता है, जो नियमों से संयोजित होता है। मराठी भाषा की संरचना में मराठी व्याकरण के विशेषण का प्रभाव संपूर्ण वाक्य पर पड़ता है। विशेषण को मुख्यतः संज्ञा या सर्वनाम के विशेषक के रूप में देखा जाता है। लेकिन विशेषण का भी विशेषण होता है तथा कभी कभी वह अव्यय या क्रियाविशेषण के रूप में भी हो सकता है (जैन, 1995: 97)।

प्राकृतिक भाषा संसाधन के लिए विशेषण की पहचान संगणक द्वारा करने के लिए विश्लेषणात्मक अध्ययन करना अधिक आवश्यक हो जाता है। अतः संगणकीय व्याकरण को भी संक्षिप्त में स्पष्ट करना आवश्यक है। संगणकीय व्याकरण का प्रयोग विशेष अर्थों में क्रम में होता है, वह क्रम भी सुनिश्चित होता है। प्राक्कलन की भाषा के प्रतीकों के गठन एवं क्रम की सुनिश्चितता संगणकीय व्याकरण को गढ़ते हैं। दूसरी ओर विभिन्न प्राकृतिक भाषाओं की संरचनाओं के विश्लेषण और सृजन के लिए संगणक को जिन नियमों से होकर गुजरना पड़ता है, वे नियम संगणकीय व्याकरण को गढ़ते हैं। सामान्य व्याकरण मनुष्य की बौद्धिक क्षमता एवं विश्लेषण सृजन शक्ति को ध्यान में रखकर रचे जाते हैं, जबकि संगणकीय व्याकरण संगणक को केन्द्र में रखकर (जैन, 1995: 113)। यहाँ भी मनुष्य की विश्लेषण एवं सृजन क्षमता को ध्यान में रखना चाहिए और ऐसे नियम गढ़े जाने चाहिए जो मशीन को संसाधित करने में सहज हो और उसमें मनुष्य की सृजनशिलता तथा विश्लेषणात्मकता सम्मिलित हो।

मराठी के विशेषण अभिज्ञान में भी स्वाभाविक भाषा की भाषायी सर्जनात्मकता अनुसार विशेषणों का स्थान-परिवर्तन होता रहता है। संरचना में प्रयोग के साथ स्थान, स्थिति और संदर्भ के अनुसार जिन नियमों से वह संयोजित होता है उसी के अनुरूप वह अभिज्ञानक बनाने की बात यहाँ हो रही है। इसलिए विशेषण विश्लेषण की बात रूपविश्लेषणात्मक आधार और संरचनात्मक आधार पर की जा रही है। साथ ही स्थान परिवर्तन तथा रूपपरिवर्तन के कारण होने वाला परिवर्तन भी इसमें शामिल है।

प्रथमतः विशेषण-विश्लेषण से संबंधित सभी नियमों और जानकारीयों को बताया जाएगा जो विशेषण अभिज्ञान के लिए सहयोगी होगी। रूपविश्लेषणात्मक आधार के अंतर्गत प्रत्यय के अनुसार विशेषण का विश्लेषण प्रस्तुत किया जाएगा। संरचनात्मक आधार पर विशेषण के विभिन्न स्थान और क्रम कैसे कार्य करता है और उसके पहचान से संबंधित नियम क्या होंगे, इसका विश्लेषण प्रस्तुत किया जाएगा।

साहित्य की समीक्षा

वैयाकरणों ने विशेषण की कई प्रकार की व्याख्याएँ प्रस्तुत की हैं। उन सब का सारांश सामान्यतः यही है की संज्ञा की विशेषता बतलाने वाले शब्द विशेषण कहलाते हैं।

उदा.

एक काळी गाय ताजे गवत खात आहे.

उपर्युक्त वाक्य में गाय के बारे में विशेष जानकारी प्रस्तुत की गई है।

कितनी गाय? एक, कैसी गाय? काळी, कैसा गवत? ताजे

उपर्युक्त वाक्यों में गाय के बारे में जो विशेष जानकारी दी गई है, यही विशेषताएँ विशेषण को पहचानने में मनुष्य को मदद करती हैं। विशेषण अभिज्ञानक के लिए भी यही विशेषताएँ महत्वपूर्ण हैं। संगणक को यह जानकारी समझा दी जाती है की कौन-कौन से विशेषण किस-किस रूप में और किन किन स्थानों पर प्रयुक्त होते हैं, जिससे संगणक उन विशेषणों को पहचानता है। जैसे - एक वाक्य है 'ती चरणारी गाय गाभण आहे'। इस वाक्य में 'ती चरणारी' तथा 'गाभण' यह दोनों 'गाय' के विशेषण हैं, परंतु एक 'गाय' संज्ञा के पहले प्रयुक्त हुआ है और दूसरा संज्ञा के बाद। विशेषण अभिज्ञान के लिए संगणक को यह भी बताना होगा की विशेषण संज्ञा के पहले तथा बाद में भी प्रयुक्त होता है। अन्यथा वह बाद में प्रयुक्त विशेषण को क्रियाविशेषण बना देगा, क्योंकि बाद में प्रयुक्त विशेषण क्रिया के पहले आता है। जैसे - 'गाभण आहे' में 'आहे' क्रिया है। लेकिन 'गाभण' शब्द 'गाय' के बारे में विशेष जानकारी देता है, इसलिए यह विशेषण है, क्योंकि यह 'आहे' क्रिया के बारे में विशेष जानकारी प्रस्तुत नहीं कर रहा।

मराठी में कही वैयाकरणों ने विशेषण के पहचान से संबंधित जानकारी प्रस्तुत की है। जैसे देशमुख (1990), बाळंबे (2006) तथा भागवत (2004), ई. किंतु यह प्राकृतिक भाषा संसाधन के लिए उपयुक्त नहीं है। प्राकृतिक भाषा संसाधन के लिए शब्द कोटी विश्लेषक के द्वारा व्याकरणिक कोटियों का अभिज्ञान और विश्लेषण किया जाता है। अंग्रेजी के लिए यह कार्य विशेष रूप से स्टैनफर्ड टैगर के रूप में किया जाता है। हिंदी के लिए नियमाधारित शब्द कोटी विश्लेषक को नवनित गर्ग, विशाल गोयल और सुमन प्रीत (2012) द्वारा प्रस्तुत किया गया है। पल्लवी बागुल (2014) तथा शरवरी गोवीलकर (2015) ने मराठी के लिए नियमाधारित शब्द कोटी विश्लेषक को प्रस्तुत किया है। ज्योती सिंह (2013) द्वारा सांख्यिकीय प्रणाली की उपयोग करते हुए शब्द कोटी विश्लेषक को प्रस्तुत किया गया है।

शोध प्रविधि -

प्राकृतिक भाषा संसाधन के लिए सर्वप्रथम महत्वपूर्ण होता है उस भाषा का कॉर्पस, इस कार्य के लिए सर्वप्रथम मराठी के शब्दों को उनकी व्याकरणिक कोटी के साथ डाटाबेस में संपादित किया गया। तत्पश्चात् विशेषण बनने से संबंधित नियमों को गढ़ा गया और उन नियमों को प्राक्कलन की भाषा द्वारा संगणक को समझाया गया है। इस प्रक्रिया को फ्लोचार्ट के द्वारा दर्शाया गया है।

विशेषण की विशेषताओं की पहचान :

विशेषण वह विकारी शब्द होता है जो संज्ञा या सर्वनाम के बारे में विशेष जानकारी देकर उनके बारे में अतिरिक्त सूचना प्रदान करता है। विशेषण से संबंधित पहचान हेतु तथा विशेषण में जो गुण होते हैं उन गुणों या विशेषताओं को आगे स्पष्ट किया जा रहा है।

1) विशेषण और विशेष्य संबंध :

लड्ड माणूस, क्रूर वाघ तथा तुटकी खुर्ची में माणूस, वाघ, खुर्ची संज्ञाएँ हैं। उनके साथ प्रयुक्त लड्ड, क्रूर और तुटकी विशेषण हैं। संज्ञाएँ विशेषणों के विशेष्य कहे जाते हैं। अर्थात् माणूस, वाघ, और खुर्ची यह संज्ञाएँ विशेष्य हैं।

गाय - संज्ञा, काळी गाय, एक काळी गाय, एक चरणारी काळी गाय, एक चरणारी सुरती काळी गाय

गाय संज्ञा के साथ सारे विशेषण जुड़े हैं, जैसे-जैसे गाय की व्याप्ति मर्यादित होती गई है वैसे-वैसे संज्ञा का विस्तार भी होता गया है।

2) विशेषण विकारी होते हैं।

विशेषण का कोई लिंग, वचन नहीं होता। संज्ञा और सर्वनाम के अनुसार उसे लिंग और वचन प्राप्त होते हैं। जैसे : काळी माती, काळा माणूस, काळे कपडे. यहाँ काळा रंग विशेषण है, पहले में माती संज्ञा स्त्रीलिंग है तथा एकवचन में है इसलिए 'काळी' विशेषण भी स्त्रीलिंग और एकवचन में है, जबकि दुसरे में संज्ञा एकवचन, पुलिंग होने के कारण विशेषण 'काळा' भी एकवचन पुलिंग में है, लेकिन तीसरे में संज्ञा का नपुंसकलिंग, बहुवचन में है इसलिए विशेषण भी नपुंसकलिंग और बहुवचन है। अर्थात् संज्ञा-सर्वनाम के अनुसार विशेषण का लिंग और वचन निर्धारित होता है, इसलिए विशेषण विकारी है।

3) पुर्वविशेषण और परविशेषण :

एक काळी गाय गाभण आहे.

तो स्त्री बिमार आहे.

उपर्युक्त वाक्यों में देखने पर पता चलता है कि रेखांकित विशेषण संज्ञा के पहले भी प्रयुक्त हुए हैं तथा संज्ञा के बाद में भी। सार्वनामिक विशेषण हमेशा संज्ञा के पहले ही प्रयुक्त होते हैं।

विशेषण के प्रकार

विशेषण उनके गुण, भाव तथा संख्या आदि के द्वारा व्याकरणिक संरचना में समाविष्ट होकर भाषा को सुंदरता प्रदान करते हैं। उनकी इन विशेषताओं के आधार पर उनकी पहचान की जा सकती है। गुणवाचक विशेषण गुण, अवगुण, रूप, रंग, दर्जा आदि को व्यक्त करते हैं। उदा. उंच झाड, रागीट शिक्षक, काळी माती, सुंदर स्त्री। संज्ञा को 'कितने' से प्रश्न पुछने पर जो जवाब मिलता है वह संख्यावाचक विशेषण है। उदा. दोन मुले, पाच छड्या, पंचतत्व, थोडा आहार, खूप प्रश्न, दोन-चार पुस्तक, अर्धा लिटर, पाव भोपळा. संख्यावाचक विशेषण में पट, घे, घी, ओं, हा, वा प्रत्यय जुडकर भी संख्यावाचक विशेषण बनते हैं। जैसे : पाचपट लोकं, दोघे मुलं, दोघी मुली, लाखों रूपये, दोन्ही साड्या, पाचही रत्न, पाचवां वर्ग। सार्वनामिक विशेषण मुख्यतः संज्ञा के पहले आता है। सर्वनाम \$ संज्ञा और संज्ञा \$ संज्ञा इस रूप में आते हैं। संज्ञा जब विशेषण रूप में आते हैं, तो उसमें कोई न कोई प्रत्यय जुडा होता है। जैसे : कट, लक - तेलकट पाणी, इक - शारीरिक स्थिति, सामाजिक, ईय - माननीय नेते, मननीय वातावरण, वान - बुद्धिवान मनुष्य, शीलवान स्त्री, मंत - श्रीमंत व्यक्ति, वंत - गुणवंत मुलगा, गरजवंत नौकर, लेले - पिकलेले फड, उघडलेले दार, आदि। धातुओं से बनने वाले विशेषणों को धातुसाधित विशेषण कहते हैं। उदा. वाहणारी नदी, उडनारे पाखरू, हसरे मूलं, हलणारे झाड, ई. धातुसाधित विशेषणों में अगर अंत्य वर्ण ण / न है तो -अरे प्रत्यय द्वारा विशेषण बनते हैं।

विशेषण हमेशा संज्ञा के पहले या बाद में ही आते हैं। तो कभी कभी संज्ञा या सर्वनाम भी विशेषण का कार्य करते हैं। तब उनमें कम से कम एक संज्ञा होती है और बाकी संज्ञाएँ विशेषण का रूप ले लेती हैं। दो या दो से अधिक विशेषण प्रयुक्त होने पर कभी कभी विशेषण भी किसी विशेषण की विशेष जानकारी प्रस्तुत करता है न कि संज्ञा की, ऐसे विशेषण ज्यादातर विशेषण के पहले प्रयुक्त होते हैं। परंतु इनका स्थान बदल देने से या यूँ कहे कि इन्हें विशेषण के पहले न लगाकर बाद में संज्ञा के पहले लगाने से वह विशेषण विशेषण की जानकारी न देकर संज्ञा की जानकारी देने लगता है। और इसके कारण वाक्य का अर्थ बदल जाता है तथा विशेषण कभी गुण को दर्शाता है तो कभी उसकी संख्या को दर्शाता है। जैसे : खूप सुंदर स्त्रीया, सुंदर खूप स्त्रीया इन वाक्यों में पहले वाक्य में खूप सुंदर विशेषण की जानकारी दे रहा है, तथा दूसरे वाक्य में खूप संज्ञा की संख्या की जानकारी दे रहा है।

1) संरचनात्मक आधार

संरचनात्मक आधार में विशेषणों के प्रयोग की स्थितियों के आधार पर विशेषण की जानकारी प्रस्तुत की जा रही है।

1. तो माणूस उंच आहे.

तो उंच माणूस आहे.

उंच हा माणूस आहे.

उंच माणूस हा आहे

2. पाच रूमालांचा रंग हिरवा आहे
हिरवा रंग पाच रूमालांचा आहे
रंग हिरवा पाच रूमालांचा आहे
रंग पाच रूमालांचा हिरवा आहे

उपर्युक्त सभी वाक्यों में विशेषण का स्थान बदलने पर भी वाक्य के अर्थ में कोई अंतर नहीं आया। प्रथम वाक्यों में सर्वनाम का स्थान संज्ञा के पहले जो अर्थ देता है वही अर्थ संज्ञा के बाद भी देता है। दूसरे वाक्यों में भी रूमाल की संख्या पहले हो या बाद में या उसका रंग पहले हो या बाद में अर्थ वही देता है। रंग 'हिरवा' विशेषण का विशेषण है इसलिए यह हमेशा वाक्य में कहीं भी प्रयुक्त हो 'हिरवा' को ही विशेष्य बनाएगा न की रूमाल को।

1. ते खूप चांगले लोकं आहेत.
ते चांगले खूप लोकं आहेत.
2. मी एका नंबरच्या खूर्चीवर बसतो.
मी नंबरच्या एका खूर्चीवर बसतो.

उपर्युक्त वाक्यों में विशेषण का क्रम या स्थान बदलने से वाक्य का अर्थ बदल रहा है। पहले वाक्यों में खूप अधिकता को दर्शाता है तो उसी के बाद वाले वाक्य में वह लोगों की संख्या को दर्शाता है। दूसरे वाक्यों में भी 'एक' पहले में नंबर को दर्शाता है तो द्वितीय में खूर्ची को दर्शाता है अर्थात् कभी कभी विशेषण का स्थान बदलने से वाक्य में विशेषण और विशेष्य संबंध तो कभी विशेषण का प्रकार बदल जाता है, और उसका जो मूल गुणधर्म है उसे छोड़कर संज्ञा के अनुसार परिवर्तित हो जाता है। यह कुछ विशेषणों के साथ ही होता है।

2) प्रत्यय के अनुसार

कुछ प्रत्यय विशेषण प्रत्यय से जुड़ने पर संज्ञा का कार्य करते हैं। जैसे - 'मानवीय', ता लगने से मानवीयता होगा। कुछ संज्ञा प्रत्यय से जुड़ने पर विशेषण का कार्य करते हैं। कुछ विशेषण ही प्रत्यय आदि लगकर अलग स्वरूप प्राप्त करके अलग-अलग तरह के विशेषण बनते हैं। विशेषणों में प्रत्यय जुड़कर जो संज्ञाएं बनती हैं उनका यहां कोई मतलब नहीं, लेकिन बाकी शब्द जो प्रत्यय जुड़ने से विशेषण बनते हैं, उन प्रत्ययों को यहां देखा जाएगा।

व्यंजन अंत्य - सूंदर, खूप, जास्त, लड्डू, गरम, थंड, सर्वांत.

इ - सुखी, दूखी, गुलाबी, काळी, थोडी.

ऊ - कडू, मऊ, ठेंगू, मोटू,

ओ - लाखो, करोडो, हजारो, षेकडो

वा - पाचवा, सहावा, शंभरावा, हजारवा

पट - पाचपट, शंभरपट, लाखपट,

पदरी - दोनपदरी, चारपदरी, नवपदरी,

झा - माझा, तुझा,

चा - त्याचा, तिचा, आमचा, तुमचा, त्यांचा, रमेषचा, कल्पनाचा. वरचा, खालचा, जवळचा, मागचा.

ला - आपला, पूढला, गल्लीतला, धाणीतला,

वाला - रिकशावाला, गाडीवाला

आरे, आरा, आरी, - ओरडणारा, बोलणारी, पकडणारे, गाणारा, उचलणारे.

लेले - पिकलेले, उघडलेले

कट - मातीकट, तेलकट, पाणकट,

इक - नागरिक, शारिरिक, भाषिक, मानसिक, सामाजिक,

ईय - माननीय, राजकीय,

वंत, वान - गुणवंत, गरजवंत, भाग्यवंत, गुणवान, गरजवान, भाग्यवान. शीलवान, बूद्धिवान, प्रतिभावान

विशेषण अभिज्ञान के नियम

नियम 1- विकारी विशेषण बदलते नहीं इसलिए ऐसे शब्दों का उनकी व्याकरणिक कोटी के अनुसार अभिज्ञान होगा। उदा. सुंदर

नियम 2- संज्ञा के पहले वाले शब्दों विशेषण के डाटाबेस से जोडकर देखा जाएगा ऐसे शब्द जादातर विशेषण होते हैं। उदा. उंच (विशेषण) माणूस (संज्ञा)

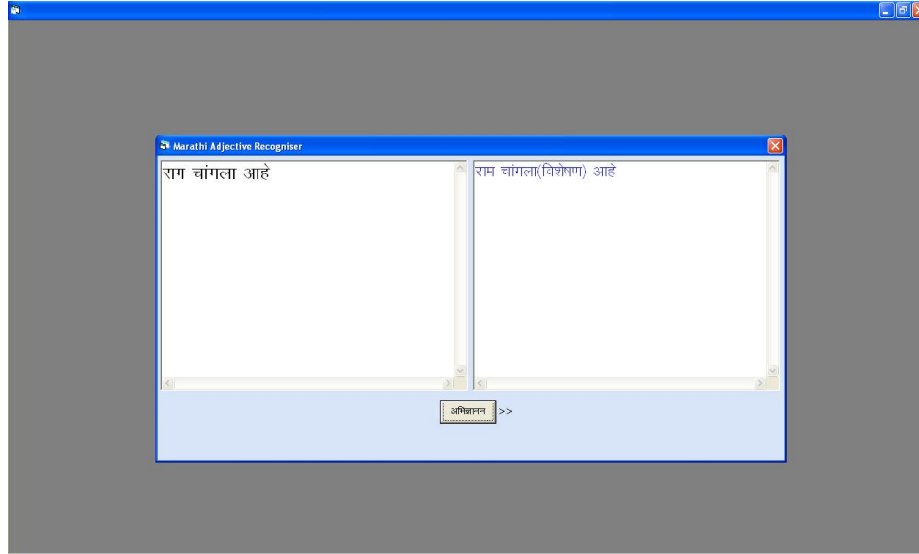
नियम 3- विशेषण के पहले वाले शब्दों विशेषण के डाटाबेस से जोडकर देखा जाएगा ऐसे शब्द जादातर विशेषण होते हैं। उदा. खूप (विशेषण) चांगले (विशेषण) लोकं (संज्ञा)

नियम 4- संज्ञा के बाद वाले शब्द अगर क्रिया, क्रियाविशेषण, संयोजक या कारक नहीं है तो उसे विशेषण समझा जाए। उदा. माणूस (संज्ञा) उंच (विशेषण) आहे (क्रिया)।

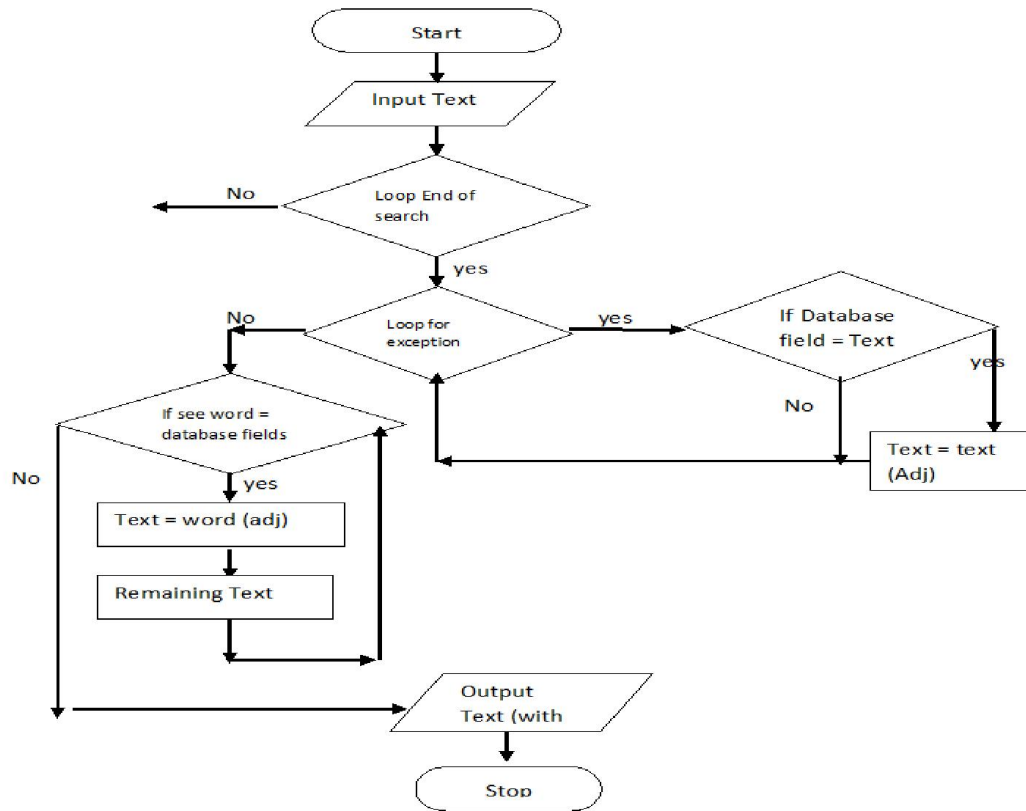
नियम 5- यदि धातु को इ, ऊ, ओ, वा, पट, पदरी, झा, चा, ला, वाला, कट, इक, ईय, वंत, वान, आरे, आरा, आरी, लेले प्रत्यय लगे हैं तो उन्हे विशेषण समझा जाए।

संरचना (आर्टिटेक्चर)

मराठी विशेषण अभिज्ञानक हेतु कंप्यूटर की प्राक्कलन भाषा (विज्युअल बेसिक 7.0) में एक आर्टिटेक्चर अथवा संरचना का निर्माण किया गया है। यह वह विण्डो है जिसमें प्रयोक्ता को इनपुट के रूप में वाक्य को देना तथा उसका अभिज्ञानित स्वरूप को देखने का कार्य होता है अर्थात् किसी कार्य का जो मूर्त स्वरूप होता है वह यह है।



फ्लोचार्ट



निष्कर्ष :

निष्कर्षतः यही कहा जा सकता है कि भाषिक संरचना में विशेषण की पहचान करना बहुत कठिन कार्य है। लेकिन विशेषण के विश्लेषण से यह कार्य बहुत सरल हो जाता है। विशेषण की विशेषताएं, प्रकार इनके माध्यम से विशेषण का संरचनात्मक स्वरूप क्या है इसके बारे में पता चल जाता है। वहीं विशेषण विश्लेषण द्वारा रूपविश्लेषणात्मक आधार पर विशेषण किन किन रूपों में आता है तथा वह प्रत्यय आदि लगकर किस तरह से कार्य करता है इसके बारे में जानकारी प्राप्त होती है। इसलिए विशेषण का संरचनात्मक विश्लेषण भी यहां प्रस्तुत किया गया है, तथा उसकी पहचान से संबंधित शब्दों को दिया गया है। इससे विशेषण अभिज्ञान से संबंधित नियम गढ़े गये हैं। संगणक में यह नियम तथा विशेषण पहचान से संबंधित विश्लेषण के नियम एवं प्रत्यय आदि डाटाबेस और प्रोग्रामिंग के माध्यम से दिए गए हैं। इसके आधार पर ही संगणक द्वारा विशेषण की पहचान होती है।

संदर्भ सूची :

1. गर्ग, नवनित और अन्य (2012) रूल बेस्ड हिंदी पार्ट ऑफ स्पिच टैगिंग, प्रोसेडिंग ऑफ COLING-2012, 163.174.
2. गोवीलकर और अन्य (2015), पार्ट ऑफ स्पिच टैग फोर मराठी लैंग्वेज, इन्टरनेशनल जर्नल ऑफ कम्प्यूटर अप्लिकेशन्स (0975-8887) 119(18), 29-23.
3. जैन, वृषभ प्रसाद (1995), अनुवाद और मशीनी अनुवाद (प्रथम संस्करण), नई दिल्ली, सारांश प्रकाशन.
4. देशमुख, अम्बादास (1990), हिन्दी और मराठी की व्याकरणिक कोटियाँ, कानपूर, अतुल प्रकाशन
5. बागुल, पल्लवी और अन्य (2014), रूल बेस्ड पीओएस टैग फोर मराठी टेक्ट, इन्टरनेशनल जर्नल ऑफ कम्प्यूटर सायंस अँड इन्फार्मेशन टेक्नोलॉजीज़ 5 (2), 1322.1326.
6. बाळंबे, कै. मो. रा. (2006), सुगम मराठी व्याकरण लेखन, पुणे, नितीन प्रकाशन
7. भागवत. श्रीपाद, (2004), तुमचे - आमचे मराठी व्याकरण, लातूर, विद्याभारती प्रकाशन