महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा

सूचना एवं भाषा अभियांत्रिकी केंद्र(&)

भाषाविज्ञान एवं भाषा प्रौद्योगिकी विभाग

हिंदी मिश्र वाक्यों के लिए टैग्ड डेटाबेस निर्माण

(Creation of Tagged Database for Hindi Complex Sentences)

(अंतरराष्ट्रीय कार्यशाला, *12-17 फ़रवरी 2018*)

प्रशिक्षक

प्रो. तत्याना ओ.

(Prof. Tatiana Oranskaia)

संयोजक

डॉ. शमीम फ़ातमा

डॉ. धनजी प्रसाद

किसी भी भाषा के सर्वांगीण विकास के लिए आवश्यक है कि उसके सभी पक्षों का वस्तुनिष्ठ विश्लेषण किया जाए। साथ ही वर्तमान संदर्भों में विश्लेषण से प्राप्त ज्ञान का डेटाबेस के रूप में संचय भी उतना ही महत्वपूर्ण है। इसी क्रम में इस कार्यशाला का आयोजन किया जा रहा है जिसका उद्देश्य मानव भाषा एवं उसके अधीनस्थ उपकरणों में सूचना पदानुक्रम की समस्या को हल करना है। यह कार्यशाला क्रिया विशेषण उपवाक्यों वाले संयुक्त एवं मिश्रित वाक्यों का टैग्ड डेटाबेस तैयार करने पर केंद्रित है। यह कार्यशाला एक बड़ी परियोजना का हिस्सा है जिसका उद्देश्य अंतः उपवाक्यों की निर्भरता एवं अर्थ विज्ञान एवं रूप/आकृति के संबंध के दृष्टिकोण से उनकी संरचना का गहन विश्लेषण करना है।

तैयार किया गया डेटाबेस इन्टरनेट पर उपलब्ध होगा जिससे वाक्य विज्ञान एवं पाठ्य संरचना के टाईपोलॉजिकल अनुसंधान के लिए संसाधित खंडित (segmented) डेटा प्राप्त होगा।

कार्यशाला का मुख्य कार्य निम्न मानदंडों के आधार पर हिंदी पाठ्य के वाक्यों की वाक्यविन्यास्मिक टैगिंग करना होगा।

- पाठ्य की वाक्यों में पारसिंग
- अंतः उपवाक्यों के अर्थों की संदिग्धार्ता को दूर करना
- अंतः वाक्यों के अर्थों की संदिग्धार्ता को दूर करना
- वाक्यों/उपवाक्यों के संयोजन के प्रकार

मूल हिंदी पाठों के संसाधन का उद्देश्य उपवाक्यों और क्रियाविशेषण अर्थ वाले उपवाक्यों /वाक्यों कुल संख्या का आंकड़ा/गणना का पता लगाना है। कंप्यूटेशनल भाषाविज्ञान के साथ साथ वर्णात्मक और विश्लेष्णात्मक भाषा विज्ञान के क्षेत्र में होने वाले अनुसंधान में भी निर्मित डेटा का प्रभावी ढंग से उपयोग किया जा सकता है। ये डेटा उपवाक्यों और उपवाक्यों के संभाव्यता अनुक्रम की मॉडलिंग में योगदान दे सकते है।

कार्यक्रम स्थल : कक्ष संख्या-02, भाषा विद्यापीठ, म.गा.अं.हिं.वि., वर्धा

तिथि : 12-17 फ़रवरी 2018

The workshop deals with the problem of information hierarchy in human language and subordinating devices. Its overall objective is to create a tagged data base of complex and compound Hindi sentences with adverbial clauses. The workshop is a part of a larger project aiming at providing an amount of data sufficient for an in-depth analysis of inter-clausal dependencies and their structuring from the point of view of relationship between semantics and form.

The data base shall be available on Internet and provide processed segmented Hindi data for typological research in the fields of syntax and text structure.

The focal task of the workshop will be syntactic tagging on the basis of following criteria:

- Parsing into sentences
- Disambiguation of inter-clausal meanings
- Disambiguation of inter-sentential meanings
- Types of clause combining

Processing of original Hindi texts will also aim at finding count data on total number of clauses and of clauses/sentences with adverbial meanings.

The results of research using the data can be deployed in descriptive and analytic as well as in computational linguistics. They may contribute to modelling of probabilistic sequences of clauses and sentences.

Venue: Room No.-02, School of Language, MGAHV, Wardha

Date: 12-17 February, 2018

कार्यक्रम की प्रस्तावित रूपरेखा

दिन	10:00 – 11:30	11:30 – 12:00	12:00-01:30	01:30-03:00	03:00–04:15	04:15-04:30	04:30 – 6:00
1.	**पंजीयन**	**उद्घाटन**	विषय का परिचय **Prof. Tatiana Oranskaia**		प्रशिक्षण **Prof. Tatiana Oranskaia**		परिचर्चा
	10:00 – 11:45	**11:45-12:00**	**12:00-01:15**	भोजन अंतराल		चाय अंतराल	
2.	Lecture Prof. Anil K. Pandey	चाय अंतराल	Lecture Dr. Shamim Fatma		Lecture Dr. Dhanji Prasad		टैगिंग एवं परिचर्चा
3.	वाक्य विश्लेषण		टैगिंग		वाक्य विश्लेषण		टैगिंग एवं परिचर्चा
4.	वाक्य विश्लेषण		टैगिंग		वाक्य विश्लेषण		टैगिंग एवं परिचर्चा
5.	वाक्य विश्लेषण		टैगिंग		वाक्य विश्लेषण		टैगिंग एवं परिचर्चा
6.	वाक्य विश्लेषण		टैगिंग		**समापन एवं प्रमाण पत्र वितरण**		

महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा

सूचना एवं भाषा अभियांत्रिकी केंद्र(&)
भाषाविज्ञान एवं भाषा प्रौद्योगिकी विभाग

हिंदी मिश्र वाक्यों के लिए टैग्ड डेटाबेस निर्माण

(Creation of Tagged Database for Hindi Complex Sentences)
(अंतरराष्ट्रीय कार्यशाला, 12-17 फ़रवरी 2018)

[पंजीकरण फॉर्म]

1. नाम (Name) : ....
2. कोर्स (Course) : ....
3. विभाग (Dept.) : ....
4. संपर्क नं. (Contact No.) : .... 5. ई-मेल (E-mail) : ....
6. पता (Address) : ....
....
7. पंजीयन शुल्क : रु. 200/-

नोट: कार्यशाला में सहभागिता के लिए 'वाक्यविज्ञान' (syntax) का प्राथमिक ज्ञान आवश्यक है।

(हस्ताक्षर, दिनांक सहित)

महात्मा गांधी अंतरराष्ट्रीय हिंदी विश्वविद्यालय, वर्धा

सूचना एवं भाषा अभियांत्रिकी केंद्र(&)
भाषाविज्ञान एवं भाषा प्रौद्योगिकी विभाग

हिंदी मिश्र वाक्यों के लिए टैग्ड डाटाबेस निर्माण

(Creation of Tagged Database for Hindi Complex Sentences)
(अंतरराष्ट्रीय कार्यशाला, 12-17 फरवरी 2018)
[पंजीकरण फॉर्म]

1. नाम (Name) : ....
2. कोर्स (Course) : ....
3. विभाग (Dept.) : ....
4. संपर्क नं. (Contact No.) : .... 5. ई-मेल (E-mail): ....
6. पता (Address) : ....
....
7. पंजीयन शुल्क : रु. 200/-

नोट: कार्यशाला में सहभागिता के लिए 'वाक्यविज्ञान' (syntax) का प्राथमिक ज्ञान आवश्यक है।

(हस्ताक्षर, दिनांक सहित)