

Certificate Course in Computational Linguistics

Centre for Endangered Languages (CEL) **GGV, Bilaspur**

Objectives

The objectives of the course are to provide students with a broad overview of the field and prepare them for further study in computational linguistics and language technologies. This course aims at enabling the students to understand and evaluate the approaches of language processing in relation to various linguistic, computational issues, and applications. It also aims at analyzing the text, context, sentiment or other affective qualities of the people speaking endangered languages. The course may help to facilitate human interaction with machines and software and make the textual and other resources of the internet readily available in multiple languages.

Expected Outcomes:


1. It will enhance the knowledge base of the students regarding the status of endangered linguistic groups (Tribal communities) in Chhattisgarh.
2. It will enhance the skill levels of the students to pursue research on endangered languages in their future career.
3. The students having knowledge of computational linguistics can find employment in Government and international agencies engaged in preserving the endangered languages of the small communities and their oral literature.

Medium of Instruction: The medium of instruction will be English.

Procedure: Each paper will carry 4 credits (Lecture-3, Practical/Tutorial-1). The students have to carry a fieldwork for seven days to study the endangered tribal languages of Chhattisgarh by staying in the community with the teachers.

Attendance: Students have to secure a minimum of 75 % attendance during the course as per University rule.

Students Intake: Maximum seats 20 nos.


विभागाध्यक्ष / H.O.D.
मानव विज्ञान एवं जनजातीय विकास विभाग
Department of Anthropology & T.D.
गुरु घासीदास विश्वविद्यालय, बिलासपुर (छ.ग.)
Guru Ghansidas Vishwavidyalaya, Bilaspur (C.G.)

Scheme of Examination

Sl. No.	Code	Title of Paper	Internal Marks	Semester Marks	Total	Minimum Passing Marks
1	CEL-CCCL-01	General Linguistics- An Introduction	30	70	100	40
2	CEL-CCCL-02	Applied Linguistics	30	70	100	40
3	CEL-CCCL-03	Computational Linguistics	30	70	100	40
4	CEL-CCCL-04	Fieldwork and Data Processing	30	70	100	40

The Paper aims at the phonetic study of a language. Apart from the theoretical background of the subject, the paper is also designed to enable the students to identify the phonetic specifications of endangered languages extension in the central region of India. This paper also focuses on the specified study of phonetic structure of a given language.

**Paper- CEL-CCCL-01.General Linguistics- An Introduction:
(30+70)=100**

Unit-1.Definition of Language, What is Linguistics, Scope of Linguistics and its relationship with other disciplines, Language and Society.

Unit-2.Phonetics and Phonology- Introduction, Area of Study,Characteristics, Phonetic Analysis, Speech units, Place of Articulation, Speech Classification and Features, International Phonetic Alphabet (IPA), Phoneme, Allophone, segmental and supra-segmental phonemes.

Unit-3.Morphology- Introduction, Area of Study, Morph, Morpheme and Allomorph, Morphological Units, Word, Word Formation, Word Analysis, Distributional Morphology.

Unit-4.Syntax, Semantics and Pragmatics- Introduction, Area of Study, Basic Types of Sentence, Units of Sentence, Sentence Patterns, Introduction to syntactical/Grammatical Theories (Phrase Structure, Case, Dependency based), Sentence and Meaning Relationship, Definition, Semantic Range (Lexical Level), Types of Meaning, Social and Cultural Meaning.

Unit-5.Discourse- Definition, Area of Study, Sentence-Discourse Relationship, Basic Methods of Discourse Analysis.

Note: This paper will be connected with the field work and the analysis part will be made by using computer.Both of these may be taken as a part of the project work.

One internal examination consists of 30 marks will be carried.

References:

Abercrombie,D.1967. Elements of General Phonetics. Edinburg: Edinburg University Press.

Ashby, Michael and John Maidment. 2005. Introducing Phonetics Science, Cambridge:Cambridge University Press.

Bhaskar Rao, Peri.1077.Practical Phonetics. Pune: Deccan College.

Roach,P.2001.Phonetics.Oxford University Press.

Paper- CEL-CCCL-02.Applied Linguistics:
(30+70)=100

Theoretical or structural linguistics deals with the generalizations of language use i.e. without taking the context into consideration, whereas applied linguistics deals with each and every aspects of language use, viz: psychological, social-cultural, geographical, biological, literary etc. The emergence of applied linguistics has a vast range of studies like language teaching, socio-linguistics, stylistics, translation studies, lexicography, computational linguistics etc. The paper introduces key areas of Applied Linguistics studies because these areas can be practically used as a tool for the paper, and the students will be able to collect data and analyse them accordingly as and when required.

Unit-1. Language Teaching- Mother tongue acquisitions, Second Language learning, contrastive analysis at lexical, syntactic and semantic level, multi lingual teaching in India and status of tribal languages .

Unit-2.Sociolinguistics- Language and society, definition of speech community and basic speech communities in India (unilingual,bi-lingual, multilingual). Regional language and sociolect, process of standardization and modernization,registered varieties, sociology of language (address terms, taboo words, kinship terms, pronouns, naming and their usage),

Unit-3.Language Mapping, Questionnaire making, Pilot Survey (in fieldwork for projects), Language Planning.

Unit-4.Translation Studies / Lexicography- Introduction and definition, theory of translation and lexicography, Types of Translation, Process of Translation, Types of Dictionary, Process of Dictionary making,

Unit-5.The Role of Contrastive analysis in Translation work and Dictionary making

This paper aims to introduce the relationship between language and machine. It introduces the history, procedure and basic components of CL. It also deals with the analytical steps and the computer programming languages to prepare the students to analyze the data collected from the fieldwork related to EL. This paper also includes sociolinguistics, nature of language a given society possesses. In this context the sociology of Endangered Languages will be studied and data will be collected for more elaborate study of the language.

Note: One internal examination consists of 30 marks will be carried.

Reference Books:

- Alan Davies, 2007. An Introduction to Applied Linguistics: from Practice to Theory. (2nd Edition), Edinburg : Edinburg University Press.
- Abbi, Anvita. 2001. A Manual of Linguistic Field Work and Indian Language Structures, Munich: LuromEurop.

- Cattford, J. C., Linguistic Theory of Translation, Oxford University Press, London.
- Fasold, R. 1999. The Sociolinguistics of Language. London: Basil Blackwell.
- Guy Cook, 2003. Applied Linguistics, Oxford University Press.
- Hudson, R.A.1980. Sociolinguistics. Cambridge. Cambridge University Press.
- Machine Translation: Post,Present,Future. Chichester: Ellis Horwood Limited (Chapter-19)-Hutchins, W.J.:1986
- Nida, E.A., Language Structure and Translation, Stanford University Press.
- Nida and Taper, The Theory and Practice of Translation, Leiden, E.J.Brile
- Pandit, P.B. 1972. India as a Sociolinguistics Area, (Gune Memorial Lectures). Pune: University of Poona.
- Wardhaugh, R.1992. An Introduction to Socio Linguistics, Oxford.Blackwell.

Paper- CEL-CCCL-03. Computational Linguistics (CL):
(30+70)=100

Unit-1. Fundamentals of Computer - Introduction to Hardware and software, operating system (windows), database structure and management, programming language (Python).

Unit-2. Introduction, History of CL in Indian context, The interface between Language and Machine, Artificial Intelligence.

Unit-3. Natural Language Processing (NLP), Steps of CL (Morphological analyzer, POS tagging, Chunking, Parsing),

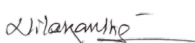
Unit-4. Applications of CL (Machine Translation, Search Engine (Google, Bing), Information Extraction. Approaches: Rule-based approach.

Unit-5. Computer lab will be utilized to introduce the software related to learning a programming language like python.

Note: Students have to appear for one internal examination carrying 30 marks.

Reference:

- Bird, S. and E. Klien. E. and Loper. 2009. Natural Language Processing with Python. O'reilly Media.
- Bolskov, I. A. and A. Gelbukh. 2004. Computational Linguistics: Models, Resources, Applications. Ciencia De La Computacion.
- David G. Hays. 1967. Introduction to Computational Linguistics. American Elsevier Publication Company, Inc. New York.
- Hausser, R.R. 2012. Foundations of Computational Linguistics: Human Computer communication in Natural language, Spinger.


विभागाध्यक्ष/H.O.D.
मानव विज्ञान एवं जनजातीय विकास विभाग
Department of Anthropology & T.D.
गुरु घासीदास विश्वविद्यालय, बिलासपुर (छ.ग.)
G.G. Ghastidas Vishwavidyalaya, Bilaspur (C.G.)


Paper- CEL-CCCL-04:Fieldwork and Data Processing
(30+70)=100

This paper is mainly based on the field work to be conducted for seven (07) days in the tribal areas and endangered tribal languages. The collected data will be categorized by different levels of linguistic studies like phonetic and socio-cultural. The student will present a different sort of linguistics with the help of computer. Reality of TL and in a combined way a dictionary or grammar or socio-cultural realities will be presented in a computerized and digitized form.

- Field work
- Making of word list and other lexical categories
- Contrastive study at all linguistic levels
- Electronic dictionary
- Machine translation of words and simple sentences
- Morphological analyzer and generator

This part of the course will train the students with the basics of programming language to enable the students to understand the computer knowledge, data structure and its function. It will also focus on the main software of their projects the students select/choose.

The fieldwork report will be presented in the centre before final publication of result. The internal/external examiner will evaluate the dissertation. 30 marks will be divided into two presentations.


विभागाध्यक्ष/H.O.D.
मानव विज्ञान एवं जनजातीय विकास विभाग
Department of Anthropology & T.D.
गुरु घासीदास विश्वविद्यालय, बिलासपुर (छ.ग.)
Guru Ghansidas Vishwavidyalaya, Bilaspur (C.G.)