Annamalai University
Centre of Advanced Study in Linguistics

PG. Diploma in Natural Language Processing (Part Time)
Programme Code: LLIN14

Curriculum
(For students admitted from the academic year 2019-2020)

Programme Outcomes

PO1:	Apply the knowledge of language fundamentals and various literatures in society, computers, psychology, cognitive science and medicine.
PO2:	Formulate, solve and analyze complex problems in variety of domains that constitute the core of language and literature knowledge, including familiarity with diverse questions of interest in the areas of (and interfaces between) structures of language and aesthetics of literature.
PO3:	Apply the acquired knowledge for analyzing language and writing in appropriate genres and modes for a variety of purposes and audiences and provide solutions to societal and environmental contexts for problems related to language change, policy and planning.
PO4:	Design and conduct research, analyse and interpret data to provide valid conclusions in the field of literature and in the descriptive as well as applied language studies.
PO5:	Select and apply appropriate modern theories and techniques including cognitive, psychological, biological, cultural, and social factors for language study and research.
PO6:	Gain exposure to attain knowledge and understand interdisciplinary and multidisciplinary linguistic and literary approaches.
PO7:	Acquire professional and intellectual integrity, code of conduct and ethics on communicational practices, understanding responsibilities and norms for sustainable development of society.
PO8:	Interact with the specific linguistic community and with society at large, through critical conversations and prepare, organize, and deliver their work to the public through speaking and writing.
PO9:	Understand the aesthetic and scientific concepts of language and demonstrate the knowledge as a skilled person in teams and multidisciplinary tasks in their profession.
PO10:	Appreciate the need for self-preparation and life-long learning independently in the broadest context of language challenges in the context of multilingualism and globalization.

Programme Specific Outcomes

PSO1:	By Studying this course, the students will understand how to incorporate technology in Natural Language Processing
PSO2:	To make the students understand language teaching and learning through technology
PSO3:	To understand how to compile electronic dictionary
PSO4:	To understand the concept of artificial intelligence.
PSO5:	To import the knowledge for instructing computer through programming languages.

PAPER – 1 COMPUTER APPLICATIONS TO LANGUAGE STUDIES

Learning Objectives of the course Computer Applications to Language Studies

- To get exposure on fundamentals of computers and the basics of linguistics
- To know the applications of computers in linguistics and language studies
- To understand various generations of computer,
- To have sound knowledge on types and anatomy of computer
- To equip programming languages and applications of computer

UNIT – 1 Introduction to Computers:

Computer generations; Types of computers- main frame, mini and personal computers; Hardware- computer anatomy; input and output devices, central processing UNIT -- memories; Rom, Ram; software; programming Languages; Databases.

UNIT -2 Introduction to linguistics:

Definition of Language and Linguistics; Levels in Language analysis- Phonetics, Phonology, Morphology, Syntax, Semantics and Discourse; Theoretical Vs applied.

UNIT -3 Word processing and DTP (Desk Top Publishing):

Text processing; Translation; Lexicography; Language Learning and Teaching; Speech signal processing; Tools for Linguistic analysis- Sorting, Indexing, Frequency counting, KWIC (Key Word In Context).

UNIT – 4 Tools and Techniques for Computer Assisted Text Processing:

Preparing Text for computer analysis; Reusing text and encoding standard; Packages for Description of texts; Databases; Text Retrieval System; Object Oriented DBMS and Hypertext.

UNIT - 5 Introduction to Computational Linguistics:

Natural Language Processing (NLP); Issues and problems in NLP; Application of Natural Language Understanding (NLU); Evaluating Natural Language Understanding Systems; Representation and Understanding; Organization of Natural Language Understanding Systems.

Text Books:

1. Allen, J. *Natural Language Understanding*, The Benjamin Company. 1995,
2. Burnard, L. *'Tools and Techniques for Computer –Assisted Text processing'* in *Computers and Written Texts* (ed.) Christopher S. Butler, Oxford: Blackwell. 1992,
3. Rajaraman, D. et al., *Computer Primer*, New Delhi: Printice Hall of India Pvt. Ltd. 1986,
4. Richard, S. *Morphology and Computation*, Cambridge: MA: MIT Press. 1992,
5. Simon Garrod and Martin J. Pickering *Issues in Language Processing, in Language Processing* (ed.) Simon Garrod and Martin J. Pickering, University of Glasgow, UK: Psychology Press,1999

Supplementary Readings:

1. Hockett, C.F. 1973, *A Course in Modern Linguistics*, New Delhi: Oxford and IBH Publishing Co.
2. Verma, S.K. and Krishnaswamy, M. 1989, *Modern Linguistics*, Oxford university Press.
3. Leech Geofferey, Steven Fligelstone "*Computers and Corpus Analysis" in Computers and Written Texts* (Ed.) Christopher S Butler, Blackwell: U.K. 1992;

4. Leech Geofferey, *"Corpus Annotation Schemes" in Literary and Linguistic Computing*, Vol. 8, No. 4, OUP: U.K. 1993**,**
5. Mc Naught, John, *"User Needs for Textual Corpora in Natural Language Processing" in Literary and Linguistic Computing,* Vol. 8, No. 4, OUP: U.K. 1993,

Outcome:

After completion of this course, the students will be able to:

CO:1. Describe fundamentals of computers and the basics of linguistics
CO.2. Understand various applications of computers in linguistics and language studies.
CO.3.Explore generations, types and anatomy of computer
CO.4.Explain natural language processing, understanding, and evaluating systems
CO.5. Differentiate natural languages and programming languages

Outcome Mapping

Course Code:							**Course Title:**								
COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	√	√	√	√		√	√		√	√	✓		✓		✓
CO2	√		√	√	√		√	√		√		✓	✓	✓	✓
CO3	√	√	√		√	√	√	√	√	√	✓	✓			✓
CO4	√	√		√		√	√		√			✓	✓	✓	
CO5	√		√	√	√	√		√	√	√	✓		✓	✓	

PAPER – 2 CORPUS LINGUISTICS

Learning Objectives

- **To** quantitative study of language
- To use of language data in machine-readable
- To understand generations of corpus
- To investigate types of corpus,
- To have sound knowledge on text processing techniques

UNIT -1 Goals and Methods of the Corpus-based approach:

Definition of corpus; corpus creation; corpora vs. collection; corpus-based approach – characteristics, association patterns in language use, role of quantitative analysis, comparison with other approaches; Corpus Based Investigations of Language Use; Corpora of written and spoken English; Corpora for Indian Languages;

UNIT – 2 Corpus Design:

Corpus generation; Types of Corpora - Primary, Secondary, Parallel Corpora; Issues in Corpus design and Corpus Development- Data sampling, Size of Corpus, Method of data collection;

UNIT -3 Corpus Annotation:

Definition; Annotated vs. raw corpora; Levels of annotation – Phonetic/phonemic, grammatical, syntactic, semantic tagging; Criteria for annotation; Methodologies of Annotation – Automatic processing, Interactive processing, Machine-aided manual input; Characteristic of tagged corpora; Tagging of Indian Language corpora.

UNIT – 4 Corpus analyzing Tools:

Different text processing techniques - Frequency count, Concordance, Collocation, Key Word In Context (KWIC), morphological processing, parsing, and Lemmatization; UNIT -of analysis in corpus-based studies

UNIT – 5 Applications of Corpus Research:

Use of corpora in Natural Language Processing, Linguistics, Lexicography, Language teaching, Speech technology, Machine aided Translation

Text Books:

1. Dash, N. S. 2005, *Corpus Linguistics and Language* Technology, New Delhi: Mittal Publications.
2. Leech Geofferey, Steven Fligelstone 1992; "*Computers and Corpus Analysis" in Computers and Written Texts* (Ed.) Christopher S Butler, Blackwell: U.K.
3. Leech Geofferey, 1993**,** *"Corpus Annotation Schemes" in Literary and Linguistic Computing*, Vol. 8, No. 4, OUP: U.K.
4. Mc Naught, John, 1993, *"User Needs for Textual Corpora in Natural Language Processing" in Literary and Linguistic Computing,* Vol. 8, No. 4, OUP: U.K.
5. Murison-Bowie, 1996, *"Linguistic Corpora and Language Teaching", in Annual Review of Applied Linguistics* 16, pp. 182 -199.

Supplementary Readings:

1. Allen, J. *Natural Language Understanding*, The Benjamin Company. 1995,
2. Ganesan, M. *Scheme for Tagging Large Corpus in Indian Languages" in Language and Technology* (Ed.) B.B. Rajapurohit, Mysore: CIIL. 1994,
3. Hockett, C.F. 1973, *A Course in Modern Linguistics*, New Delhi: Oxford and IBH Publishing Co.
4. Richard, S. *Morphology and Computation*, Cambridge: MA: MIT Press. 1992,
5. Simon Garrod and Martin J. Pickering *Issues in Language Processing, in Language Processing* (ed.) Simon Garrod and Martin J. Pickering, University of Glasgow, UK: Psychology Press,1999

Outcome:

After completion of this course, the students will be able to:

CO 1: Know the importance of quantitative study of language

CO 2: Explain use of language data in machine-readable form in natural language processing

CO 3:Create a corpus and compare with other approaches

CO 4:Differentiate the types of corpus, text processing techniques and corpus generations

CO5:Describe the use of corpora in various fields like Linguistics, Lexicography, Language teaching, Speech technology, Machine aided Translation

Outcome Mapping

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	√	√	√	√	√		√		√	√	✓		✓		✓
CO2	√	√		√	√		√	√		√		✓	✓	✓	✓
CO3	√	√	√		√	√			√		✓	✓			✓
CO4	√		√	√		√	√	√	√	√	✓	✓	✓	✓	
CO5	√	√			√	√	√	√		√			✓	✓	✓

PAPER-3 NATURAL LANGUAGE PARSING

Learning Objectives of the course Natural Language Parsing

- To understand the system of language
- To imbibe the knowledge on various levels of language
- To acquire organization of language components internally and externally
- To know how to analyze language components
- To study the parsing of smallest elements in the larger components

UNIT -1 Morphological Analysis:

Introduction to Morphology; morph, morpheme and allomorph; free vs. bound morph; Morphological Processes - Inflectional, Derivational and Compositional Morphology; Automatic Morphological Analysis; Representation of Morphological information; MRD (Machine Readable Dictionary) for stems, suffixes; A Simple Morphological parsing scheme

UNIT – 2 Finite State Automata:

Morphology and Finite State Transducers – Finite State Morphological Parsing; Word structure rules- context free rules- features and categories- word structure formalism; Unification and its varieties; role of word grammar.

UNIT – 3 Grammar and Parsers:

Definition; Parsing in traditional grammar; Parsing in formal Linguistics; Parsing in Artificial Intelligence Grammar and sentence structure; Top – Down Parser; Bottom-up Chart parser; Transition Network Grammars; Top-Down Chart Parsing; Augmented Transition Networks ATN); Recursive Transition Network (RTN).

UNIT - 4 Features and Unification:

Feature Structures; Unification of Feature Structures; Features Structures in the Grammar; Agreement; Head Features; Subcategorization; Long-Distance Dependencies; Implementing Unification.

UNIT – 5 ***Lexicalized and Probabilistic Parsing:***

Probabilistic Context-Free Grammars; Learning PCFC Probabilities; Problems with PCFC; Probabilistic Lexicalized CFGs; Dependency Grammars; Categorical Grammar; Human Parsing.

Text Books:

1. Allen, J. 1995, *Natural Language Understanding*, The Benjamin Company.
2. Jurafsky, Daniel and James H. Martin, 2002, *Speech and Language Processing*, New Delhi: Pearson Education
3. King, M. 1998, *Parsing Natural Languages* (ed.) New York: Academic Press.
4. Ritchie, D. et al., 1992, *Computational Morphology,* England: MIT.
5. Patten 1992, *'Tools and Techniques for Computer –Assisted Text processing' in Computers and Written Texts* (ed.) Christopher S. Butler, Blackwell: Oxford.

Supplementary Readings:

1. Allen, J. *Natural Language Understanding*, The Benjamin Company. 1995,
2. Ganesan, M. *Scheme for Tagging Large Corpus in Indian Languages" in Language and Technology* (Ed.) B.B. Rajapurohit, Mysore: CIIL. 1994,
3. Hockett, C.F. 1973, *A Course in Modern Linguistics*, New Delhi: Oxford and IBH Publishing Co.
4. Richard, S. *Morphology and Computation*, Cambridge: MA: MIT Press. 1992,
5. Simon Garrod and Martin J. Pickering *Issues in Language Processing, in Language*
6. *Processing* (ed.) Simon Garrod and Martin J. Pickering, University of Glasgow, UK: Psychology Press,1999

Outcome

After completion of this course, the students will be able to:

CO1:Explain different levels of natural language parsing such as morphological, syntactic, semantic and lexical levels

CO2:Differentiate the traditional and computational grammar and its applications

CO3:Create automatic morphological analyzer and parser.

CO4:Describe morphophonemic changes and morphotactic rules and sandhi rules

CO5:Unify the structural features with grammar and implement the unification

Outcome Mapping

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	√	√	√	√	√		√	√	√		✓		✓		✓
CO2	√	√	√		√	√	√	√		√	✓	✓		✓	✓
CO3	√	√		√	√		√		√	√	✓	✓			✓
CO4	√		√	√		√			√			✓	✓	✓	
CO5	√		√		√	√	√	√		√	✓		✓	✓	

PAPER – 4 ARTIFICIAL INTELLIGENCE

Learning Objectives

- **To** understand the unambiguous Representations
- To equip knowledge representations to the systems
- To know meaning structure of language
- To get exposure on semantic network and frames
- To acquire knowledge on lexical relations

UNIT – 1 Representing Meaning:

Computational Desiderata for Representations; Verifiability; Unambiguous Representations; Canonical Form; Inference and Variables; Expressiveness; Meaning Structure of Language; Predicate-Argument Structure; First order Predicate Calculus; Elements of FOPC; The Semantics of FOPC; Variables and Quantifiers; Inference.

UNIT – 2 Some Linguistically Relevant Concepts:

Categories; Events; Representing Time; Aspect; Representing Beliefs; Pitfalls; Related Representational Approaches; Alternative Approaches to Meaning; Meaning as Action; Meaning as Truth; Semantic Networks; Frames

UNIT – 3 Semantic Analysis:

Syntax-Driven Semantic Analysis; Semantic Augmentations to Context-Free Grammar Rules; Quantifier Scooping and the Translation of Complex-Terms; Attachments for a Fragment of English; Sentences; Noun Phrases; Verb

Phrases; Prepositional Phrases; Idioms and Compositionality; Robust Semantic Analysis; Semantic Grammars; Information Extraction.

UNIT – 4 Lexical Semantics:

Relations among Lexemes and their Senses; Homonymy; Polysemy; Synonymy; Hyponymy; WordNet: A Database of Lexical relations; The Internal Structure of Words; Thematic Roles; Selectional Restrictions; Primitive Decomposition: Semantic Fields.

UNIT – 5 Word Sense Disambiguation and Information Retrieval:

Selectional Restriction-Based Disambiguation; Limitations of Selectional Restrictions; Robust Word Sense Disambiguation; Machine Learning Approaches; Dictionary-Based Approaches; Informational Retrieval; The Vector Space Model; Term Weighting; Term Selection and Creation; Homonymy, Polysemy and Creation.

Text books:

1. Allen, J. 1995, *Natural Language Understanding*, The Benjamin Company.
2. Jurafsky, Daniel and James H. Martin, 2002, *Speech and Language Processing*, New Delhi: Pearson Education
3. King, M. 1998, *Parsing Natural Languages* (ed.) New York: Academic Press.
4. Ritchie, D. et al., 1992, *Computational Morphology,* England: MIT.
5. Patten 1992, *'Tools and Techniques for Computer –Assisted Text processing' in Computers and Written Texts* (ed.) Christopher S. Butler, Blackwell: Oxford.

Supplementary Readings:

1. Hockett, C.F. 1973, *A Course in Modern Linguistics*, New Delhi: Oxford and IBH Publishing Co.
2. Verma, S.K. and Krishnaswamy, M. 1989, *Modern Linguistics*, Oxford university Press.
3. Leech Geofferey, Steven Fligelstone "*Computers and Corpus Analysis" in Computers and Written Texts* (Ed.) Christopher S Butler, Blackwell: U.K. 1992;
4. Leech Geofferey, *"Corpus Annotation Schemes" in Literary and Linguistic Computing*, Vol. 8, No. 4, OUP: U.K. 1993**,**
5. Mc Naught, John, *"User Needs for Textual Corpora in Natural Language Processing" in Literary and Linguistic Computing,* Vol. 8, No. 4, OUP: U.K. 1993,

Outcome

After completion of this course, the students will be able to:

CO1:Explain the concept of meanings and their representation
CO2:Differentiate various methods of semantic analysis
CO3:Distinguish the word sense relations and word sense disambiguation
CO4:Describe machine learning approaches; dictionary-based approaches
CO5:Create semantic networks and frames

Outcome Mapping

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	√	√	√	√		√	√	√		√	✓	✓		✓	✓
CO2	√		√		√		√		√			✓	✓		✓
CO3		√		√		√	√	√		√	✓	✓			✓
CO4	√	√	√	√	√			√	√		✓	✓		✓	
CO5	√		√		√	√	√		√	√	✓		✓	✓	

****