1 Introduction

A language can be defined as a set of rules or set of symbols where symbols are combined and used for conveying information or broadcasting the information. Since all the users may not be well-versed in machine specific language, **N**atural Language Processing (NLP) caters those users who do not have enough time to learn new languages or get perfection in it. In fact, NLP is a tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages. It came into existence to ease the user's work and to satisfy the wish to communicate with the computer in natural language, and can be classified into two parts i.e. *Natural Language Understanding* or Linguistics and *Natural Language Generation* which evolves the task to understand and generate the text. **L**inguistics is the science of language which includes *Phonology* that refers to sound, *Morphology* word formation, *Syntax* sentence structure, *Semantics* syntax and *Pragmatics* which refers to understanding. Noah Chomsky, one of the first linguists of twelfth century that started syntactic theories, marked a unique position in the field of theoretical linguistics because he revolutionized the area of syntax (Chomsky, 1965) [23]. Further, Natural Language Generation (NLG) is the process of producing phrases, sentences and paragraphs that are meaningful from an internal representation. The first objective of this paper is to give insights of the various important terminologies of NLP and NLG.

In the existing literature, most of the work in NLP is conducted by computer scientists while various other professionals have also shown interest such as linguistics, psychologists, and philosophers etc. One of the most interesting aspects of NLP is that it adds up to the knowledge of human language. The field of NLP is related with different theories and techniques that deal with the problem of natural language of communicating with the computers. Few of the researched tasks of NLP are Automatic Summarization (*Automatic summarization* produces an understandable summary of a set of text and provides summaries or detailed information of text of a known type), Co-Reference Resolution (*Co-reference resolution* refers to a sentence or larger set of text that determines all words which refer to the same object), Discourse Analysis (*Discourse analysis* refers to the task of identifying the discourse structure of connected text i.e. the study of text in relation to social context),Machine Translation (*Machine translation* refers to automatic translation of text from one language to another),Morphological Segmentation (*Morphological segmentation* refers to breaking words into individual meaning-bearing morphemes), Named Entity Recognition (*Named entity recognition* (NER) is used for information extraction to recognized name entities and then classify them to different classes), Optical Character Recognition (*Optical character recognition* (OCR) is used for automatic text recognition by translating printed and handwritten text into machine-readable format), Part Of Speech Tagging (*Part of speech tagging* describes a sentence, determines the part of speech for each word) etc. Some of these tasks have direct real-world applications such as Machine translation, Named entity recognition, Optical character recognition etc. Though NLP tasks are obviously very closely interwoven but they are used frequently, for convenience. Some of the tasks such as automatic summarization, co-reference analysis etc. act as subtasks that are used in solving larger tasks. Nowadays NLP is in the talks because of various applications and recent developments although in the late 1940s the term wasn't even in existence. So, it will be interesting to know about the history of NLP, the progress so far has been made and some of the ongoing projects by making use of NLP. The second objective of this paper focus on these aspects. The third objective of this paper is on datasets, approaches, evaluation metrics and involved challenges in NLP. The rest of this