

paper is organized as follows. Section 2 deals with the first objective mentioning the various important terminologies of NLP and NLG. Section 3 deals with the history of NLP, applications of NLP and a walkthrough of the recent developments. Datasets used in NLP and various approaches are presented in Section 4, and Section 5 is written on evaluation metrics and challenges involved in NLP. Finally, a conclusion is presented in Section 6.

2 Components of NLP

NLP can be classified into two parts i.e., *Natural Language Understanding* and *Natural Language Generation* which evolves the task to understand and generate the text. Figure 1 presents the broad classification of NLP. The objective of this section is to discuss the Natural Language Understanding (Linguistic) (NLU) and the Natural Language Generation (NLG).

2.1 NLU

NLU enables machines to understand natural language and analyze it by extracting concepts, entities, emotion, keywords etc. It is used in customer care applications to understand the problems reported by customers either verbally or in writing. Linguistics is the science which involves the meaning of language, language context and various forms of the language. So, it is important to understand various important terminologies of NLP and different levels of NLP. We next discuss some of the commonly used terminologies in different levels of NLP.

a) Phonology

Phonology is the part of Linguistics which refers to the systematic arrangement of sound. The term phonology comes from Ancient Greek in which the term phono means voice or sound and the suffix -logy refers to word or speech. In 1993 Nikolai Trubetzkoy stated that

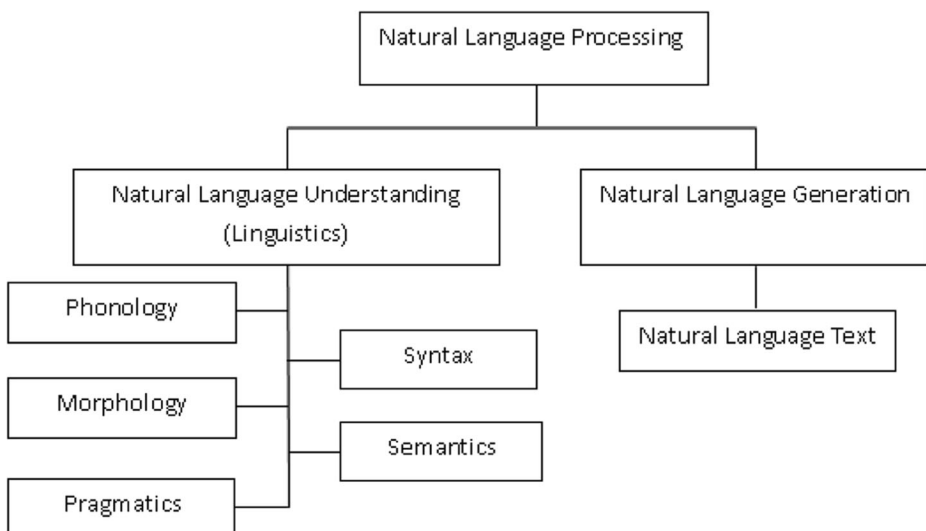


Fig. 1 Broad classification of NLP