

Butler (1985) believes that analysis should be carried out on the level of graphology, phonology (the sound system), lexis (vocabulary), syntax (the combination of words in grammatical constructions), and semantics (meaning).

2.3.1. Graphological Analysis

A graphological analysis is considered one of the simplest matters in the analytical process. The computer can easily count individual letters and words appearing in a text and generate word-lists, indexes, concordances, or statistical information. Therefore, most computational work on texts has relied on analysis at this level.

2.3.2. Lexical Analysis

After generating a word-list, index, or concordance, lexical analysis takes place. As words have inflected forms, the computer has to decide the lemma for the given word, i.e. its origin. For example, the verb ‘to walk’ may appear in different forms, ‘walk’, ‘walking’, ‘walks’, etc. the base form of the word ‘walk’ is called lemma, the word one looks up from the dictionary. The activity of grouping the different lexemes of one lemma is called lemmatization in the science of computational linguistics. Lemmatization is therefore the algorithmic process of determining the lemma of a given word which is considered to be one of the most challenging tasks in MT.

2.3.3. Syntactic Analysis

Syntactic analysis is sought next in MT which requires the recognition of clause and phrase boundaries, and the classification of clauses (main or subordinate), phrases (nominal, verbal, prepositional, etc), words (as nouns, verbs, adjectives, or other parts of speech). Explicitly, such analysis is of an intricate nature which needs an automatic parsing system which is not 100% accurate. Nirenburg (1987) illustrates the importance of the knowledge of the syntactic structure by furnishing the following example: