

# Applications Associated With Morphological Analysis And Generation In Natural Language Processing

Neha Yadav

**Abstract:** Natural Language Processing is one of the most developing fields in research area. In most of the applications related to the Natural Language Processing, findings of the Morphological Analysis and Morphological Generation can be considered very important. As morphological study is the technique to recognise a word and its output can be used on later on stages .Keeping in view this importance, this paper describes how Morphological Analysis and Morphological Generation can be proved as an important part of various Natural Language Processing fields such as Spell checker, Machine Translation etc.

**Keywords:** Natural Language Processing, Morphological Analysis, Morphological Generation, Spell checker, Machine Translation

## INTRODUCTION

Morphological study is one of the branch of linguistic which is used for study of structure of words[1]. It identifies how a word is formed using combination of various morphemes [2]. A morpheme is considered a basic unit of the word. The smallest part of the word with grammatical function and meaning is considered as the morpheme. Morpheme is categorized in 2 types: Free morpheme and Bound morpheme. A free morpheme can constitute a word by itself. For example: Bus,Cycle etc. While a bound morpheme can not stand alone but attaches itself to a free morpheme to form a word. Examples of bound morpheme are: ing, un etc. Morphology can be categorized in 2 parts: Inflectional Morphology and Derivational Morphology. Both of these types have their own significance in various areas related to the Natural Language Processing.

## APPLICATIONS OF MORPHOLOGICAL ANALYSIS AND GENERATION IN NATURAL LANGUAGE PROCESSING FIELDS

### A. Text to Speech Synthesis

These days people are largely dependent on technology for their daily basis tasks. Various technology mediums such as computers, mobiles are used for fulfillment of daily need. But when we talk about those people who are not well aware of interface of these technical Medias and disabled people, they face a lot of difficulty in interfacing of these mediums. There arises the need of Text to Speech Synthesis [3].In case of Bengali language, Morphological information plays an important role in determining the pronunciation of a homograph [3]. Morphological analysis can be used to reduce the size of lexicon also. Because here one needs to store only the root word and various inflections need not to be stored.

As search time also depends on the size of lexicon, so it will affect the search time also[3]. Schwa deletion is one of the significant problems faced during Text to Speech synthesis. At some places, schwa is considered while at some places, schwa sound needs to be ignored. There is no proper clarification regarding existence of this sound. In case of the compound words, Morphological Analyzer can be used to segregate a compound word into basic form. Phonemes can be used for these basic root forms and later on these phonemes can be combined using Morphological Generator to generate the result[4].

### B. Machine Translation

In this technical era, amount of digital data is increasing day by day. People belonging to different communities want to interact with this data. But as the data may be stored using different languages, people might face problem while interfacing with that data. One of the prominent solutions for this problem might be Machine Translation[5]. For a few languages Machine translators have been developed which produce a remarkable result. There are a few other languages for which no such efficient translators have been developed but work is going on in this direction. In case of the machine translation in between two languages, database of words play a significant rule. In lack of the morphological analysis, we need to store all word forms of a particular word enhancing the size of lexicon. It will create wastage of memory and as large will be the size of the database, it will take more time to search a particular word[6].Another important benefit can be that, Morphological analysis provides the information of the word such as number, gender etc. This information can be used in target language to generate the correct form of the word.

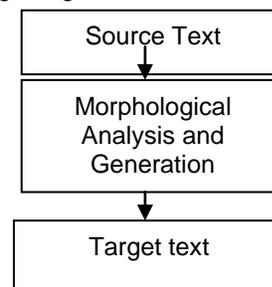


Fig 1. Morphological study in machine translation

- Neha Yadav
- Department of Computer Science, Central University of Haryana, Mahendergarh, India
- [Yadavneha2511@gmail.com](mailto:Yadavneha2511@gmail.com)

### C. SPELL CHECKER

While writing a over a document, there arise a possibility of misspell a word. A Spell checker is an application that is used to identify whether a word has been spelled correctly or not. Sometimes misspelled words show the carelessness of one's personality. A spell checker might be an individual unit or it might be part of a larger application such as Microsoft office word application [7]. Spell checker functionality can be divided into two parts: Spell check error detection and Spell check error correction. Spell check error detection phase only detects the error while Spell check error correction will provide some suggestions also to correct the error detected by Spell check error detection phase. For morphologically rich language, it is more suitable to use morphological based spell checker. Because for a morphologically rich language, there will be large number of inflections and it will take too much space to store all forms, In that case morphology can reduce the size. One more advantage associated with morphology based spell checker is that it can handle the Name Entity problem[8]. If any new word is found which is not stored already in the lexicon, then that can be included in the particular paradigm of the lexicon.

### D. Search Engine

Search engine is an application program used to search a particular document over the internet using World Wide Web. Input is provided to the search engine, after that search engine provides results based on the input provided by user. A lot of search engines are available in the market, One of the such popular search engine is Google[9]. Morphological Analysis and Generation improves the result of the search engine. Output of the search engine depends on the input words[10]. Suppose a word is provided as an input in inflected form and this word is not present in lexicon of search engine, it will affect output of search engine. In this case, Morphological Analysis of that word is done and root form is find out and if that root word is present in the dictionary, better output will be provided by the search engine. For language such as Arabic, Morphological Analysis can be taken as incremental process taking the next sensitive word pattern there by increasing the recall[11].

### E. Information Retrieval

These days, a lot of are using internet. With the vast number of users, amount of data to be stored is also increasing day by day. There arises the need of effective technique for information retrieval .Information retrieval is the process of retrieval of information resources from a collection of information resources. Searching process can be based on metadata or it can be based on full-text indexing. Process of information retrieval begins with a query entered into a system by the user. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy. Most of the IR system, First compute a numeric score based on how well an object is related to the query , and then objects are ranked according to this value. After that, the top ranking objects are then shown to the user. In case of cross-lingual Information Retrieval, Morphological Analysis and Generation yield more effectiveness in getting output.

Morphological Analysis can be applied on the words which are not present in the dictionary, After getting the root word as an output of Morphological Analysis, that root word is searched in the dictionary[12]. For morphologically analyzed queries , MAP score is better than those without that analysis.

### CONCLUSION

After discussing all the applications of morphological analysis and morphological generation in the field of Natural Language Processing, it can be concluded that morphology is one of the dominant part of the literature for morphologically rich language, as it is providing the basic structure of the word. From the root form, various inflections are generated and these inflections are acting as the sub-units for sentences and paragraphs etc. All these applications of Morphological Analysis and Morphological Generation help to achieve better results in various processes. These applications are not limited to a particular language but can be included upto Hindi, Arabic, Marathi depending upon the particular field.

### REFERENCES

- [1]. Nikhil Kanuparthi, Abhilash Inumella, Dipti Mishra Sharma," Hindi Derivational Morphological Analyzer", Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON2012), pages 10–16, Montre'al, Canada, June 7, 2012. c 2012 Association for Computational Linguistic.
- [2]. Neha Yadav, "Existing Morphological Analyzer and Different approaches used" Proceedings of National Coference on Emerging Trends in Computer Science and Electronics Engineering(ETCSEE-2016)",pages 148-151, April 2016
- [3]. Shyamal Das Mandal, Somnath Chandra2 and Swaran Lata," Use of Part Of Speech (POS) and morphological information for resolving Multiple Pronunciations in Pronunciation Lexicon Specification (PLS) for Indian Languages – Bengali as a Case Study"
- [4]. Shikha Kabra,Ritika Agarwal, Neha Yadav," Rule Based Schwa deletion Algorithm for Text To Speech Synthesis in Hindi", adfa, p. 1, 2011. © Springer-Verlag Berlin Heidelberg 2011
- [5]. Sadik Bessou, Mohamed Touahria," Morphological Analysis and Generation for Machine Translation from and to Arabic" International Journal of Computer Applications (0975 – 8887) Volume 18– No.2, March 2011
- [6]. Vishal Goyal, Gurpreet Singh Lehal," Hindi Morphological Analyzer and Generator" First International Conference on Emerging Trends in Engineering and Technology 978-0-7695-3267-7/08 \$25.00 © 2008 IEEE DOI 10.1109/ICETET.2008.11

- [7]. Baljeet kaur , Harsharndeeep Singh,” Design and Implementation of HINSPELL -Hindi Spell Checker using Hybrid approach”, International Journal of scientific research and management (IJSRM) ,Volume 3,Issue 2,Pages 2058-2061,2015
- [8]. Veena Dixit, Satish Dethe, Rushikesh K. Joshi, “Design and Implementation of a Morphology-based Spellchecker for Marathi, an Indian Language”, Archives of Control Sciences Volume 15(LI), 2005 No. 3, pages 251–258
- [9]. Ayush Jain,” The Role and Importance of Search Engine and Search Engine Optimization” ,International Journal of Emerging Trends & Technology in Computer Science, Volume 2, Issue 3, May – June 2013
- [10]. Dr. S.K Dwivedi, Rjesh Kr. Goutam, Parul Rastogi,” Impact of language morphologies on Search Engines Performance for Hindi and English language” International Journal of Computer Science and Information Security, Vol. 8, No. 3, June 2010
- [11]. Mamoun Hattab, Bassam Haddad, Mustafa Yaseen, Asem Duraidi and A. Abu Shmais,” Addaall Arabic Search Engine: Improving Search based on Combination of Morphological Analysis and Generation Considering Semantic Patterns”
- [12]. H. Alizadeh, R. Fattahi, M. R. Davarpanah” Applying Natural Language Processing Techniques for Effective Persian- English Cross-Language Information Retrieval” , International Journal of Information Science and Management, Vol. 8, No. 2 July / December 2010