



# ARTIFICIAL INTELLIGENCE - NATURAL LANGUAGE PROCESSING ITS RISE AND THEIR APPLICATIONS

**R Maheswari<sup>1</sup>, S Sunitha<sup>2</sup>, S Krishnaveni<sup>3</sup>, M Krishna Santhi<sup>4</sup>**

Assistant Professor – Computer Science, BBCIT, O.U, Hyderabad, Telengana, India<sup>1</sup>

Associate Professor – Computer Science, BBCIT, O.U, Hyderabad, Telengana, India<sup>2</sup>

Associate Professor – Computer Science, BBCIT, O.U, Hyderabad, Telengana, India<sup>3</sup>

Associate Professor– Computer Science, BBCIT, O.U, Hyderabad, Telengana, India<sup>4</sup>

**Abstract:** The Natural Language Processing (NLP) is a sub domain of Artificial Intelligence (AI). NLP is playing a vital role in AI, which is used to bridge the gap between human and machine. AI is a platform for learning outcomes, in that NLP is used to solve problems like conversion of one human language to another human language. NLP spreads in all domains like Questions generation, Question Answering, Evaluation of descriptive answers and knowledge graph completion. The surge of modern NLP is accredited to the evolution of a simple model, perceptions. Including of perceptions was not just a second order with techniques altogether or boosting, but rather exponential if not asymptotic, with the advent of deep neural networks. The influence of NLP has made modernized advancements into real-world applications, i.e. chatbots conversion, real-time translations, hate speech, or forged news detection. Natural Language Processing is highly influenced on human lives, so it is working on many applications like Translator, Search Autocorrect and Autocomplete, Social Media Monitoring, Hiring and Recruitment Survey Analysis, Targeted Advertising ,Voice Assistants Grammar Checkers, Chatbots, End Notes ,Email Filtering, etc.. In this article, we shall start by exploring some machine learning algorithms to give solutions for NLP approaches. The different algorithms Bayesian Networks, Maximum Entropy, Linear Regression, Logistic Regression, Decision Tree, SVM, Naive Bayes, kNN, K-Means, Random Forest.

**Keywords:** NLP, AI, Machine Language Algorithms, Translator

## I. INTRODUCTION

Natural language processing (NLP) is an area in computer science and artificial intelligence apprehensive with the interaction between computers and humans in natural language. The ultimate plan of NLP is to help computers understand language as well. It is the leading force behind things like virtual assistants, speech recognition, sentiment analysis, automatic text summarization, machine translation and much more.

This paper covers the fundamentals of natural language processing, plunge into few of its techniques and also learn how NLP has benefited recent advances in deep learning. Natural language processing (NLP) is the convergence of computer science, linguistics and machine learning.

The dicipline states out on communication of computers and humans in natural language and NLP is all about making computers understand and generate human language. Applications of NLP techniques encompass voice assistants like Amazon's Alexa and Apple's Siri, but also things like machine translation and text-filtering.

NLP has deliberately included from recent advances in machine learning, especially from deep learning techniques. The field is divided into the three parts: Speech Recognition- The translation of spoken language into text. Natural Language Understanding- The computer's potentiality to understand the voice specified. Natural Language Generation- The generation of natural language by a computer. Understanding human language is considered a strenuous task due to its complexity. For example, there is an unbounded number of dissimilar ways to arrange words in a sentence. Also, words can have several meanings and contextual information is necessary to correctly interpret sentences. Every language is more or less eccentric and vague. Just take a look at the adherents newspaper highlights "The Pope's baby steps on gays." This sentence clearly has two very different interpretations, which is a pretty good example of the challenges in NLP.



## II. APPLICATIONS: LANGUAGE TRANSLATION

One of the signs of evading on Spanish homework is that syntactically, it's a mess. Various languages do not permit for straight translation and have different orders for sentence structure, which converting services are being used to give over. But, they've come a long way. With NLP, online translators can translate languages more perfect and display grammatically-perfect results. This is limitlessly helpful when trying to communicate with another person in another language. Not only that, but when translating from other language to one's own, tools also recognize the language depending on input text given and translate it.

### 1. Digital phone calls :

We all might listen in mobile phones this particular announcement that "this call may be recorded for training purposes," but rarely do we ponder what actually happens. Turns out, these recordings may be used for training purposes, if a customer is dissatisfied, but most of the time, they all are stored into the database for an NLP system to learn from and improve in the future. Automated systems direct customer calls to a service representative or online chatbots, which respond to customer requests with helpful information. This is a NLP enactment that many companies, including large telecommunications provides have been put to use. NLP also allows computer-generated language to suitable nearest to the voice of a human. Phone calls to arranging appointments like an oil change or haircut can be atomized, as proofs by this video showing Google Assistant making a hair appointment.

### 2. Predictive text:

Things like autocorrect, autocomplete, and predictive text are so commonplace on our smartphones that we take them for granted. Autocomplete and predictive text are searched for similarities to search engines in that they forecast things said based on what is typed there, finishing the word or suggesting a relevant one. And autocorrect will sometimes even change words so that the complete message seems more meaningful. They also absorb from us. Predictive text will personalize itself to your personal language quirks the prolonged you use it. This makes for fun assessment where individuals will share entire sentences made up entirely of predictive text on their phones. The results are surprisingly personal and enlightening; they've even been highlighted by several media outlets.

NLP libraries for Indian Languages

iNLTK (Natural Language Toolkit for Indic Languages)

iNLTK provides support for various NLP applications in Indic languages. The languages supported are Hindi (hi), Punjabi (pa), Sanskrit (sa), Gujarati (gu), Kannada (kn), Malayalam (ml), Nepali (ne), Odia (or), Marathi (mr), Bengali (bn), Tamil (ta), Urdu (ur), English (en).

iNLTK is like the NLTK Python package. It provides the feature for NLP tasks such as tokenisation and vector embedding for input text with an easy API interface.

One has to first install;

```
pip install torch==1.3.1+cpu -f https://download.pytorch.org/whl/torch_stable.html
```

Then next is installing iNLTK using pip:

```
pip install inltk
```

Indic NLP Library:

The Indian languages have some difficulties which come from sharing a lot of similarity in terms of script, phonology, language syntax, etc., and this library provides a general solution.

Indic NLP Library provides functionalities like text normalisation, script normalisation, tokenisation, word segmentation, romanisation, indicisation, script conversion, transliteration and translation.

Languages supported:

•Indo-aryan:

Assamese (asm), Bengali (ben), Gujarati (guj), Hindi/Urdu (hin/urd), Marathi (mar), Nepali (nep), Odiaa (ori), Punjabi (pan).

•Dravidian:

Sindhi (snd), Sinhala (sin), Sanskrit (san), Konkani (kok), Kannada (kan), Malayalam (mal), Teugu (tel), Tami (tam).

•Others:English (eng).

Tasks handled:

- It handles bilingual tasks like Script conversions for languages mentioned above except Urdu and English.
- Monolingual tasks:

•This language supports languages like Konkani, Sindhi, Telugu and some others which aren't supported by iNLTK library.

•Transliteration amongst the 18 above mentioned languages.



•Translation amongst ten languages.

The library needs Python 2.7+, Indic NLP Resources (only for some modules) and Morfessor 2.0 Python Library.

Installation:

```
pip install indic-nlp-library
```

Nextstep, download the resources folder which contains the models for different languages.

```
# download the resource
```

```
git clone https://github.com/anoopkunchukuttan/indic_nlp_resources.git
```

StanfordNLP:

StanfordNLP contains tools which can be used to convert a string containing human language text into lists of words and sentences. This library converts the human language texts into lists to generate base forms of those words, parts of speech and morphological features, and also to give a syntactic structure dependency parse. This Syntactic structure dependency parse is designed to be parallel among more than 70 languages using the Universal Dependencies formalism.

The language inherits additional functionality from CoreNLP Java package such as constituency parsing, linguistic pattern matching and conference resolution.

The modules are built on top of PyTorch, and the package is a combination of software based on Stanford entry in the CoNLL 2018 Shared Task on Universal Dependency Parsing and Java Stanford CoreNLP software.

StanfordNLP offers features like:

- Easy Native Python Implementation.
- Complete neural network pipeline for better and easy text analytics which includes multi-word token (MWT) expansion, tokenisation, parts-of-speech (POS), lemmatisation, morphological features tagging and dependency parsing.
- Stable Python interface to CoreNLP.
- The neural network model has support for 53 human languages featured in 73 treebanks.

Install using pip, pip install stanfordnlp

### Top datasets for NLP (Indian languages)

**Semantic Relations from Wikipedia:** [Contains](#) automatically extracted semantic relations from multilingual Wikipedia corpus.

**HC Corpora (Old Newspapers):** This [dataset](#) is a subset of HC Corpora newspapers containing around 16,806,041 sentences and paragraphs in 67 languages including Hindi.

**Sentiment Lexicons for 81 Languages:** This [dataset](#) contains positive and negative sentiment lexicons for 81 languages which also includes Hindi.

**IIT Bombay English-Hindi Parallel Corpus:** This dataset contains parallel corpus for English-Hindi and monolingual Hindi corpus. This [dataset](#) was developed at the Center for Indian Language Technology.

**Indic Languages Multilingual Parallel Corpus:** This [parallel corpus](#) covers 7 Indic languages (in addition to English) like Bengali, Hindi, Malayalam, Tamil, Telugu, Sinhalese, Urdu.

**Microsoft Speech Corpus (Indian languages)(Audio dataset):** This [corpus](#) [contains](#) conversational, phrasal training and test data for Telugu, Gujarati and Tamil.

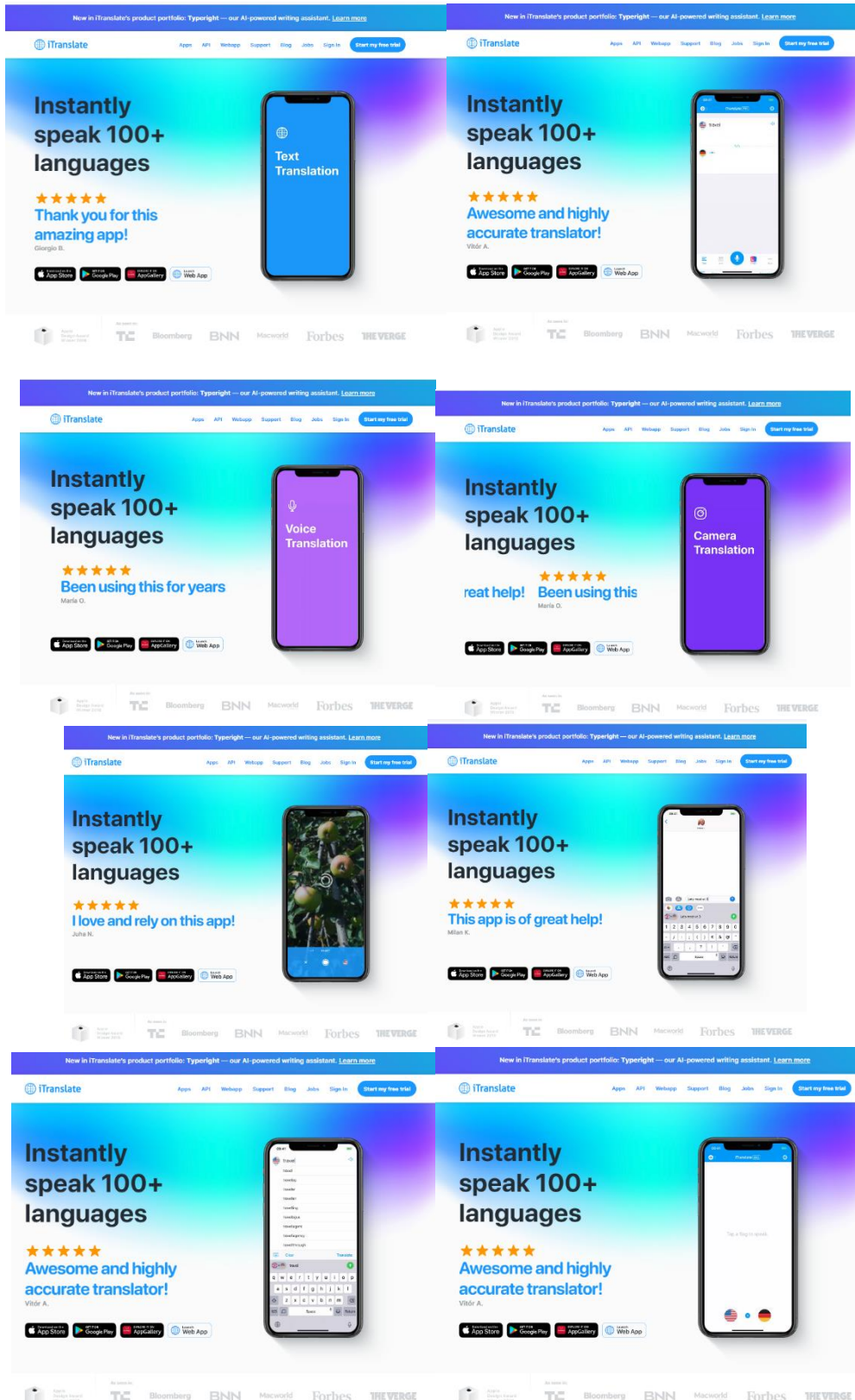
**Hindi Speech Recognition Corpus(Audio Dataset):** This is a [corpus collected](#) in India consisting of voices of 200 different speakers from different regions of the country. It also contains 100 pairs of daily spontaneous conversational speech data.

Fig. 1 A sample graph

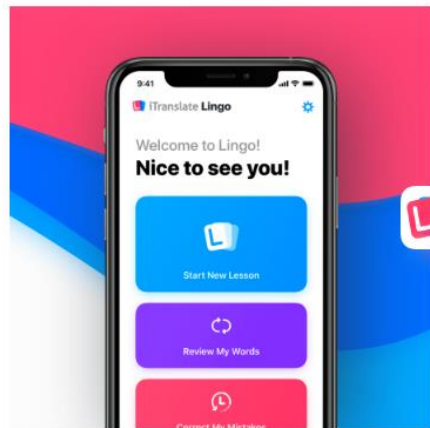


A. Figures:

1.iTranslate







**iTranslate Lingo**

Welcome to Lingo!  
**Nice to see you!**

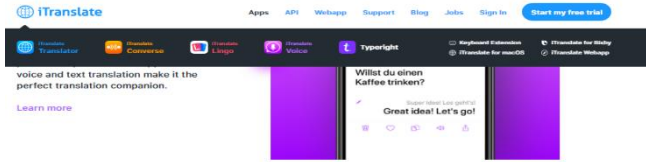
Start New Lesson

Review My Words

Correct My Mistakes

The easiest way to learn a language. Choose from 14 languages and begin learning in a fun, modern and playful way in just 5 minutes a day.

[Learn more](#)



**Typerright by iTranslate**

Compose mistake-free, clear and simple writing with Typerright's AI-powered writing assistant.

[Visit typerright.co](#)

## B. References

- <https://www.itranslate.com/>
- <https://www.analyticsvidhya.com/blog/2020/07/top-10-applications-of-natural-language-processing-nlp/>
- <https://www.mygreatlearning.com/blog/trending-natural-language-processing-applications/>
- <https://arxiv.org/pdf/1807.02200.pdf>

## III. CONCLUSION

The Natural Language Processing (NLP) is a sub domain of Artificial Intelligence (AI).

The surge of modern NLP is accredited to the evolution of a simple model Natural Language Processing is highly influenced on human lives, so it is working on many applications like Translator, Search Autocorrect and Auto complete, Social Media Monitoring, Hiring and Recruitment Survey Analysis, Targeted Advertising, Voice Assistants Grammar Checkers, Chat bots, End Notes, Email Filtering, etc.. In this article, we shall start by exploring some machine learning algorithms to give solutions for NLP approaches. The main contribution of this work is a demonstration of the usefulness for musicological research of applying systematic linguistic processing techniques

Moreover, we envision that the combination of knowledge extracted from text with knowledge extracted from other data modalities (e.g., audio signals or music scores) would be a further step in the construction of real-world music understanding systems. And the main point for advancement in this paper is to demonstrate the ability to create a voice alert drones that identifies the voice of threat and alerts the nearest concerned officials.

## IV. REFERENCES

- [1]. <https://www.itranslate.com/>
- [2]. <https://www.analyticsvidhya.com/blog/2020/07/top-10-applications-of-natural-language-processing-nlp/>
- [3]. <https://www.mygreatlearning.com/blog/trending-natural-language-processing-applications/>