NLP TERM REPORT

Natural Language Processing

Saranya M S (CS15D006) *

Correspondence:
cs15d006@cse.iitm.ac.in
Department of CSE, IITM,
Chennai, India
Full list of author information is
available at the end of the article
*Single contributor

Chapter 1

1 Overview of NLP

Languages born when human tried to communicate more information that can be conveyed by their sign languages and sounds. The languages evolve over a period of time. Though the vocabulary size of a language increases, the basic unit of sounds (phonemes) produced to form those vocabulary are confined. The basic unit of speech is called phones. The set of phones confined to a language is called phonemes. Once humans start to civilize and think, he analysed the nature and started to discover the science behind everything. His quest did not stop at the understanding, he also tried to apply it in his day to day life.

This analysing quest turned towards the language which is the basic mode of communication when he tried to make the system to understand the same. During 1950's this natural language processing starts to get more attention as it intersected with the another vast domain called artificial intelligence (AI).

The natural languages are very vast and un-restrictive in nature with many ambiguities This makes it tougher to come up some set of hand-written rules to feed to the systems. Two main problems that are needed to be faced while using hand-crafted rules are [1],

- Extracting meaning from the text (semantics) is hard.
- Hard to get even the hand-written rules to come up with human comprehensible, 'ungrammatical' phrases.

In recent years, the real time applications like

- Language Translation: translating the sentence from one language to the other,
- Language Identification: identifying a language from the speech, or from text,
- Voice commanding machines: detecting or understanding the emotion or context,
- Context based spell checking, etc

makes the natural language processing (NLP) a rigorous research field. NLP deals with text. We can describe NLP as the process of understanding and producing the understandable texts in any human language either as text or speech.

The NLP can be split into two components namely natural language understanding (NLU) and natural language generation (NLG) as shown in Figure 1

The NLU is the process of interpreting a text and converting it into some symbolic representation for machine understanding and usage. The NLG is the process

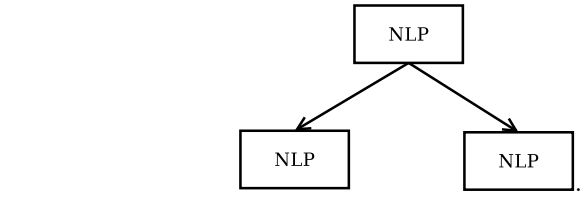

Figure 1: Components of NLP

Figure 2: Natural language understanding

Figure 3: Natural language generation

of taking some source or representation like discourse model of a language and generating a text or speech or embodied part of a document. Few examples for existing NLG are

- Traffic mapping: Interpreting the satellite road maps and giving voice instructions accordingly.
- Language translation: Converting text from one language to the other like *G*oogle translator
- Text to speech (speech synthesis) or speech to text conversion (speech recognition)

2 System Vs Human Intelligence

The general assumption is that NLU is harder problem than NLG. Because NLG does not require much precision as NLU. For example, one can compare automated flight take model as the NLG model where the choices and corresponding situations can be given in prior. It has to select one among the existing choices in according to the situation. But NLU is like trying to automate the aircraft landing where many unexpected situations can appear due to the evolving nature of a language.

Another reason for NLU is the ambiguity in the source. The text which serves as the source for the NLU component, exists in both structured and unstructured ways. Humans do interpretation even from unstructured form of resources due to the complicated and still unexplored nervous system. Expecting the same from system is nearly impossible until we find out a way to interpret all those stuffs. Interpreting those information are even hard to humans because **"we do many stuffs but do not know how it is being done?"**. This is because of the **non-conscious** knowledge in our mind. It is hard to represent something clearly without much understanding. And to make the systems work on it, it requires more precision as discussed before.

2.1 Structure Function Correspondence

The human thinking can be related to a black box, where the input is fed and output is obtained without bothering much about its internal functionality. But machines need a sequence of steps which converts the input to an output. The process is well structured. For the successful modelling of a scenario via a machine,

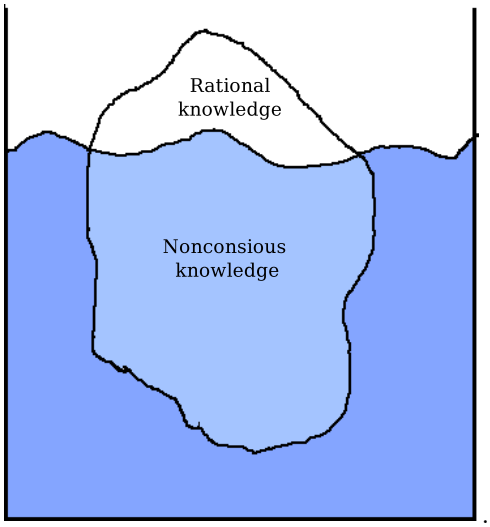

Figure 4: Conscious vs non-conscious knowledge

the whole process has to be structured well. The modelling will be successful if the structure function correspondence is high. To be more clear, the way in which the human brain reads a newspaper by selecting the required column or section out of all the paper and interpreting the information can be considered as a function. But the same is not so simple with the system. Making the machine to understand the semantics of a phrase is a harder problem. It requires the information to be in a more structured form like tables.

2.2 Model Diagnosis

The other main problem in the structure in structure-function correspondence is **"notion of a structure is not universal"**. It is based on the application and its context. The system model of a scenario need not to be perfect in its first attempt. The structure of the model can be updated iteratively by observing how far its behaviour is deviating from the empirical behaviour. The model is the abstract behaviour of the system and it can be incomplete. Given the observation of a system (real-time), the model simulates the behaviour of the system. The actual observations and predicted observations are compared to understand how far the model has deviated from the system as in 5.

Classical artificial intelligence (AI) can be used for better modelling of a system. The model of a system is developed by certain rules. The rule based systems may not be perfect due to certain level of uncertainty prevailing in the structure. In such cases "rules are padded with certain probabilities". The two ways in structuring a model are

- bottom-up-representation: the models are formulated from set of rules.
- top-down-representation: first model is formulated, then rules are derived from the models.

Top down approach ensures that one is much aware of what is needed and what is to be done in a much promising way.

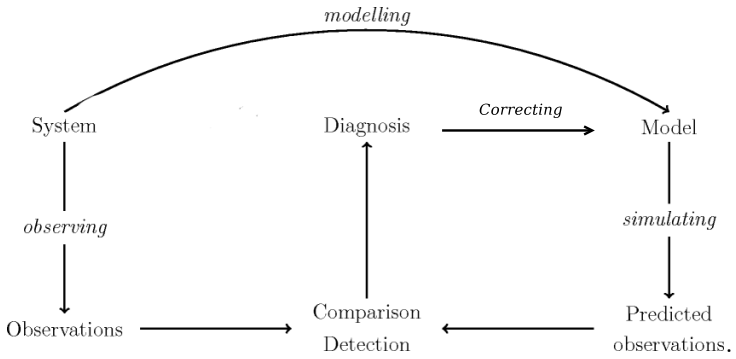

Figure 5: Model Diagnosis

2.3 Classical AI

The study of understanding the intelligent entities of human and an attempt to represent/model it, is called artificial intelligence (AI) [2]. AI helps to learn about ourselves and our thinking process. AI tries to address the functionality of a tiny brain. AI turned the philosophers study on how the inputs of sense organs are being remembered by this little brain into real experimental and theoretical discipline. The two broad sub-fields of AI are

- General purpose: logical reasoning, perception etc
- Application specific: chess games, poetry, disease diagnosis etc

Over the decades, many definitions have been evolved to describe AI. Though there are many definitions they vary in only two dimensions namely "reasoning" and "behaviour".

Thinks / Acts	Rationally	like Humans
Rationally	Act rationally & think rationally	Act rationally & think like human
like Humans	Act like human & think rationally	Act like human & think like human

The definitions are based on *"human"* performance and ideal concept of *" intelligence"* called as *"rationality"*. A human-centered approach is an empirical science with lot os hypothesis and experimental confirmation, while the rationalist approach is defined by the combination of mathematics and engineering.

2.4 AI system and Conventional system

The fundamental difference between the normal conventional system and AI system is **decision making**. In conventional system the process of decision making is encoded completely in prior with all possibilities and appropriate decisions at every possible conditions. Eg:- tree parsing

On the other hand, in AI system *"what to do"* is specified but "how to do" is not specified. That decision is to be made by the AI system.

3 Applications of NLP

NLP is the field that deals with most of our day-to-day life products like emails, web pages, tweets, social media, newspapers, scientific articles in many languages

across the world. It has the unavoidable usage right from spell check, grammar check to language translation, toy applications like automatic question answer , text comprehension etc.

Information extraction: Reading a mail or passage identifying the information like date, place and create a calendar entry automatically.

Sentiment analysis: Getting user feelings from the comments given for a product online. For example, when the consumer page of a camera which has lots of comments by various users about the products, analysing and splitting it into good and bad reviews can be done by sentiment analysis.

Machine translation: Converting a text from one language to another language. The best and simpler example for this is the *"google translator"*. Though it does not work upto the mark for some uncommon words or vernacular speech.

Text summarization: This produce abstract version of a document with user specific relevant information. Eg: Reading a documnet and producing the headlines, summary, outline etc. This summarization can be done on multiple documents also where a gist of the documents, or the stories that are common across the documents, or identifying the set of web pages with same content can be retrieved.

Parsing: Parsing is the process of representing a sentence in the form of a tree represented by the parts of speech of that sentence. Using grammar and parts of speech for parsing is a poor idea. As the grammar structure increases, the parsing tree structure will grow exponentially. If we restrict the grammar to control the parse structure it reduces the robustness nature of it. *Penn's tree bank* came with an idea of annotated parsing which reduces the complication of parsing by using annotated parse tree. The annotated parse tree contains the information of parts of speech , frequency of the words, and the distributional information.

IBM Watson System: This is a *"question-answer"* system, based on cognitive computing. It behaves like humans by observing the data fed to it, evaluating the data based on grammar and huge set of rules programmed into it followed by decision making. Do not need structured data. It can understand from unstructured data like blogs, newspapers, websites etc. Watson does not simply looks for a synonym or keywords alone as he search engines. It interprets the text like a person by breaking the sentence grammatically, relationally and structurally by understanding the context. Watson performs corpus collection which is overlooked by human to remove anything that is out of date. This process is called as "curating the content". The data is preprocessed and an analogy graph is created. Training is done with the help of machine learning to understand the human interpretation via question answer pairs. Watson keeps on reading in every Q&A process.

Synchronize-3-application: This applicationalso called as *"applink"*, allows the consumer's smart phones to synchronize with the vehicles thereby giving access to their smart phones seamlessly. Once synchronization is done, the user can utilize the smart phone completely in a hands free conversation mode while driving the vehicle.

Ring (reads text for blind): This is an innovative invention of a ring which can be worn on a finger to read the text on a page or book. This will be very useful for the blind people as it converts text to speech.

4 Critics of NLP in history

The specific branch of AI that deals with natural language processing is called computational linguistics. The work of Norwig on computational logic is criticized by Chomsky as *"a method of doing linguistics without involving language and linguistic science"*. But Norwig comments on Chomsky's work by saying it as a "notorious work in linguistics without any special regard for real-data. Chomsky wants to represent language with elegant theory which clearly surpasses the human erroneous nature and to get the simple structure lying under his complex brain structure. Meanwhile Norwig tries to represent everything by simplistic statistics. Chomsky commented on the statistical machine learning methods that mimics the behaviour of some real world scenarios, as to simulate the bee dance without understanding the reason behind that dance. But Norvig claims that why should we complicate the process instead of following statistical model which is simpler and efficient. Norwig argues that every word in a language occurs with certain frequency and that frequency can be learned from the humongous amount of data.

5 Morphology

The source of NLP is "TEXT". The basic element of text is character. A sequence of character giving a proper meaning is called as a word. The sequence of words forming conveying an information is called sentence. Sentences together form the paragraph also know as discourse. This structure can be linked with chemistry and it looks like the Figure 8.

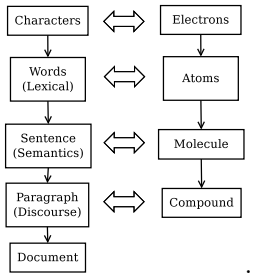

Figure 6: Morphology

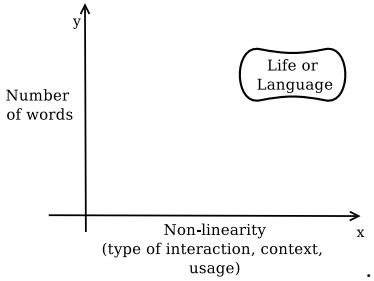

Figure 7: Words-Language relationship

The characters are similar to electrons which are the basic elements and do not convey any meaning when it is used alone. The combination of electrons leads to atoms and atoms are bonded to get a valid molecule. The combination is determined by the valency bond. Similarly the combination of words to make a meaningful sentence is determined by grammar and structure of the language.

The words evolve over a period of time. One word can be used to form the other by many process like *derivation* (Eg: educate –¿ education), *inflexion* (sing –¿ singing), etc. When words used in sentence, the role of it in the sentence is determined by *parts of speech*. The meaning of two individual words change completely when they come together. *Eg: ice-cream, honeymoon, pass-away ...*

Lexical analysis comes into picture in the formation of words from characters and semantics comes in to picture when words are combined to form a sentence.

Sentences together form a paragraph or passage. Analysis the paragraph is totally a different task known as *discourse analysis*.

Words evolve over a period of word, which leads to higher dimensionality problem when documents are processed on word basis. As the words of a language increases it is not necessary for the language also to grow in same rate. This is because, all the words in the language are not used in daily interactions or in all context base don the life-style. Thus the growth of words of language an d the language are non-linearly related.

5.1 Word Co-occurrence

The relation between two words is defined as word co-occurrences if both words are in association with each other in some sense. That is, if there is a document with a word heart surgery, it is more likely to find the word by-pass surgery or angiography in the same document. If the related word are found in same document, then they are referred to as *"first order co-occurrences"*. If the related works are found in two different documents then they are called as *"second order co-occurrences"*.

5.2 Ambiguity and its types

Analysing and understanding a text/speech segment is pretty tough for a machine. Humans understand it duet to the cognitive knowledge that acts as a very big database formed right from one's child hood. When some one asks you *Is there water in jug?* it means if the water is there we have to pass that jug to the person who asked that question. Such sentences are called **pragmatics**. But system will not understand pragmatics. The process of understanding a word in more than one way is referred to as **ambiguity**. Ambiguity can occur at various levels of NLP. Ambiguity could be Lexical, Syntactic, Semantic etc [3].

- **Syntactic ambiguity:** This ambiguity involves operators and quantifiers. *E*g: Every man loves a woman. Here there are two kinds of interpretations. The one is "For every man there is a woman", and the other is "there is one particular woman who is loved by every man". The scope of quantifiers are not clear and it can create ambiguity.
- **Semantic ambiguity:** This occurs when the meaning of the word itself is interpreted in more than one ways. *Eg: We saw her duck*. Duck can refer to the person's bird or to a motion he made. Semantic ambiguity happens when a sentence contains an ambiguous word or phrase.
- **Discourse analysis:** This analysis gives the information about the influence and relationship of across the sentences.This analysis includes *(i)Co-reference resolution* and *(ii) information extraction*. Co-reference identifies the phrases in a document that refers to same information. Information extraction locates specific piece of data from a document.

6 Types of Knowledge

Humans infer the language through the knowledge they acquire from there child-hood [4]. The process of a child getting the worldly knowledge from its care-takers and surroundings is termed as **"child knowledge acquisition**. Similar to this knowledge source, system needs knowledge to process the natural language. The

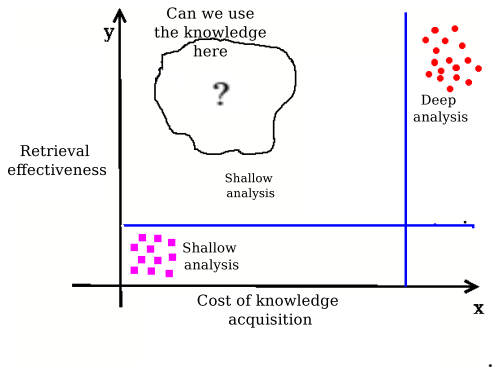

Figure 8: Shallow and deep linguistic analysis

knowledge that contributes the database for NLP can be divided into three types namely

- **Linguistic knowledge:** The dictionary is an example for linguistic knowledge from where one can get the common knowledge about the use of language. Explanations, meanings and synonyms of a word and their usage helps us in understanding the language in a proper way. *Eg: Oxford dictionary*
- **Introspective knowledge:** The knowledge within the given documents or reports are considered as introspective knowledge. The information like co-occurrences of words at different order (statistical co-occurrences) constitutes the introspective knowledge.
- **External knowledge:**The world knowledge or external knowledge can be divided under variety of various fields. This knowledge also includes the common information shared across the fields also. The effective representation of knowledge via human media such as pictures, diagrams, voice, knowledge drawn from textbooks, encyclopedia, glossaries etc. *Eg: Wikipedia*

7 Linguistic processing

Language processing needs theoretical and descriptive analysis. Based on the level of analysis the linguistic processing is classified into (i) deep and (ii) shallow processing.

- **Shallow linguistic processing (SLP):** SLP also known as machine learning approaches altered the fundamental processing of NLP. Many rapid and robust creation of tools have been developed for this SLP which requires lesser time and manual effort. So the SLP gained more interest than deep linguistic processing. This can also be termed as *knowledge light NLP*.
- **Deep linguistic processing (DLP):** DLP is concerned with more computational development. The grammars for DLP were manually developed maintained and are computationally expensive. But the cost can be trade off with the effective performance given by DLP than SLP.

Chapter 2

8 History of NLP

In 1948 the first NLP application was developed for searching a word in dictionary using a look-up system. In 1950 the first machine translation was implemented to translate Russian language to English. The NLP really took its intense phase of growth from 1957 after the syntactic structures proposed by Noam Chomsky. Many researches are started after that and still evolving. After the introduction of machine learning and artificial intelligence the research development in NLP took a vertical growth.

The unavoidable usage of internet and technologies, made NLP as an essential part of peoples every day life. Chatter-bots like PARRY, Recter, and Jabberwacky were written and developed during early times of NLP research. The IBM Watson discussed in Section 3 was invented in 2011 which is a question and answer (Q&A) application, which makes the user so transparent to the system that responds from the other end. It is a general purpose artificial intelligence based system, which can hold a human-level conversation.

An enormous quantity of preprogrammed knowledge concerning both the language and the domain under examination is the prerequisite of any natural language processing application. The increase in computer power made this transformation possible thereby allowing more research and programs to be written for NLP.

9 Why NLP is hard?

Sheer complexity of sentence structure of our text/spoken language makes the NLP problem harder to get implemented. The indicatives also known as *known facts*, and our rationally questioning nature of *"what if...?* in many cases increases complexity in formulating the rules for these kind of applications. Processing text in becoming more hard because of the various methods of formulating the text. For example it is rule to use periods (.) after the title like *Mr., Mrs.,* etc and at the end of the sentences. its is harder to find and differentiate between these two periods, wherein other unnecessary and unwanted periods are being inserted in many places due to lack of awareness which makes the **sentence boundary detection** problem more complicated.

Apart from the lack of common notation in writing a text information, the syntactic ambiguity because of the polysemous words makes it tough to understand and differentiate the words. The major and still unsolved problem is the context detection in the language. The meaning of the sentences vary depending on the intent of the speaker, listener, location, prior, time and social content as shown in some scenarios given below [5].

- Depends on the people present e.g. "How far is it?" (miles or km)
- Depends on the social context: "That was bad!"
- Depends on the location, e.g. "It is playing in upstairs" (It refers to song that is being played in upstairs.)
- Depends on the time of day, e.g. "Let's go eat" (eat may refer to breakfast, lunch or dinner)
- Depends on prior sentences: "The third one" (depends on what other two are?)

Apart from this, recognising the names of people, place, animal, object and differentiating the various slangs, jargon, sarcasm, spelling mistakes, humour, grammar mistakes etc makes the natural language processing the more challenging research problem.

10 Artificial Intelligence

Artificial intelligence can be literally defined as the intelligence possessed by machines or software. AI is the separate engineering field where the study deals with making intelligible machines and software. Intelligence in here describes the ability of the system that perceives the environment and takes the action such that it maximizes the chance of success in whatever application it work towards. AI has its own part from generalized applications to specific applications like gaming. The AI is used where there is the need for reasoning, planning and learning. Since NLP has all these in its field, AI is used widely in NLP. Humans have the ability to precisely describe their need and making the system to simulate it. But the range of precision clearly affects the performance of the system. As far as the rules are clear, systems will work fine with it.

AI leads to many latest and admirable inventions like smart phones, automatic vehicles, smart homes, auto-bots, language learning tools, hardware interactions etc. Many IVR based systems are developed for ubiquitous interaction of humans with the machines and electronics on the fly. This is made possible only because of the fusion of two most important and complicated fields by thte name AI nad NLP.

11 Conscious and non-conscious Knowledge

Right from the child hood we humans do many task with the help of our organs controlled by the only CPU in our head called brain. The tasks like understanding a language, picking up the required sound from crowded hall, summarizing the text document, inferring the image and storing the details, everything are done just like that without much knowledge of "actually how it is done?". These actions are done as an involuntary reaction for the instructions given by the brain. If we decode these set of actions and understand it, then it will be easier for us to instruct the machines to follow the same set of instructions to simulate the human behaviour.

But it is hard to explain something that one does not understand black and white. Those unexplainable knowledge is termed as *"non-conscious-knowledge"*. If one can able to answer a question clearly by description or diagram or by any means the same can be interpreted by system if the way of our interpretation is given to it as input. For example, if some one asks to define force, it can be defined empirically and mathematically. This is because of *conscious knowledge* which we obtain through our rational understanding. If someone asks to explain a color, it is impossible to explain unless one see many object with same color and classifies it internally with that class label "green". This happens because of the *non-conscious* knowledge that resides in our brain due to the humongous processed database collected right from our childhood.

12 Morphology

The smallest meaningful grammatical unit in a language is called *morpheme*. The study of morpheme sis called morphology. The morpheme and words are not same. Morpheme may or may not stand alone. But words always stand-alone [6]. There are different types of morphemes as described below:

- **Free morphemes:** The morphemes that are stand alone and act as words. Eg: dog, cat, town, city etc.
- **Bound morphemes:** Appears as a part of a word. Always occurs in conjunction with a root word as affixes, prefixes, and suffixes(tion, ing, ation etc).
- **Derivational morphemes:** The morpheme together with the standalone word called as root word, changes the meaning or parts of speech of that root word. *Eg: happy+ness=happiness. Happy (adjective), happiness(noun).*
- **Inflectional morphemes:** Modifies the verb's tense or a noun's number without affecting the meaning. *Eg: Sing+ing=singing. Root: sing, singing(current ongoing action).*

As discussed in Section 5, the character forms the basic of a written language from where the words, phrases, sentences and paragraphs are formed. In spoken language the same sequence can be compared with phonemes, syllables, words, sentences and paragraphs. Phonemes are like the basic elements. The combination of vowels (V) and consonants (C) as CV, VC, CVC, CCV, etc forms the syllables. The syllable structure of every language varies. The basic sound units of a language are termed as *"phones"* and the sounds that are confined to a language are called as *"phonemes"*.

The words of a language keep on evolving. The one brilliant idea followed or insisted by our ancestors are to keep the basic sound units, i.e., phonemes constant and evolving new words as per the requirement based on these phonemes. Apart from the semantic ambiguity or context difference, difference and complications starts to occur right from the word level of a language. There are different words with same meaning(synonyms), different or same spelling words with same pronunciation (homonyms), a word having more specific meaning than the other (hyponym), a metaphorical word, that used to refer the other word which actually points to the part of a whole (meronym) etc. Understanding and differentiating these words are not a trivial task.

These problems can be solved by having large number of examples for each word in the database and tagging them with descriptions called synsets. This helps us to resolve the syntactic ambiguity problem to a certain extent.

13 Ambiguity

Ambiguity in NLP occurs in all levels of analysis right from the morphology to discourse analysis. Coming up with a software that decides the meaning of a given text or speech without considering the context is like a lost battle. The basic definition of ambiguity can be given as *"the ability to understand the piece of text/information in more than one way"* [7]. The most common types of ambiguity from literature are

- **Lexical ambiguity:** More than one interpretation of a word. Eg: Flies (an action of flying / an insect)

- **Syntactic ambiguity:** Variation in the interpretation of a sentence due to the grammatical structure. Eg: He saw her duck (duck may be an action/a bird).
- **Semantic ambiguity:** Different interpretations of meaning based on the based on the meaning of words in the phrase when combined. Eg: *"Colourless green ideas sleep furiously"* is a sentence composed by Noam Chomsky in his book "Syntactic Structures" as an example of a grammatically correct, but semantically nonsensical statement.
- **Pragmatic ambiguity:** The context of the phrase or sentence gives completely different meaning for the sentence's actual meaning. This commonly happens when conversation happens between two entities in different domain without any common basis. Eg: *"superfluous hair remover"*. Here we are not sure whether the superfluous describes the process of hair remover or the noun remover. Because of the improper representation of scope of specifier this pragmatic ambiguity may occur.

14 Child knowledge acquisition

Language acquisition is the process of acquiring the capacity to understand and interpret a language, and to communicate in that language. This language based communication is meant especially for homosapiens [4]. Usually the language acquisition starts right from our childhood, from the surroundings and caretakers. There are two types of language acquisition in general for a child namely (i) first language acquisition and (ii) second language acquisition. To get expertise in a language, one need to now different range of tools right from phonology, morphology to extensive vocabulary in that language. Language can be written as text or vocalized as speech.

The general approaches of child knowledge acquisition are (i) social interaction and (ii) emergent-ism. The former one happens because of the interaction interaction of a child with the linguistic-knowledge-adults. The instructive corrections given to the kid by its care-givers helps the child to correct the language. This works as a black box with feed-back mechanism. The later one can be defined as the cognitive process that emerges by the biological pressures and environment. This can be related a=to an example for better understanding. If a person familiar with only the mother tongue is relocated to another place with completely different language, he will learn that new language quickly for his better survival in that environment. The theory of child and adult knowledge acquisition process is detailed in [8].

15 Other applications of NLP

- **Co-reference resolution:** This is the process of analysing the given text document or a chunk of text and identifying the words referring to same information. For examples, matching up the pronouns with the nouns that are referring to. It is also called as "bridging-relationships" while referring the structure of one thing in a sentence. *Eg: She peeped in to Rosy's house via the glass window.* In this sentence, the glass window is the structure of Rosy's house. It indirectly refers to Rosy's house.

- **Morphological segmentation:** The process of identifying the morphed words, separating it from the root word and classifying the morphemes is called morphological segmentation. For few languages like English which has simple morphology, all possible words can be listed instead of going for morphological segmentation. But this is not the case in highly agglutinated languages like Tamil, Turkish etc.
- **Named Entity Recognition (NER):** Given a stream of text, predicting the upcoming word or next word or completing the present typing the word is called NER. Forming a common way of writing the text to identify the starting of hte word will also fail in case of representing the name of a a person, place or animal. Handling this is also another difficult problem in NLP.
- **Optical Character Recognition (OCR):** Scanning text from a scanned image of a document/book. There is a problem of mis-interpreting two closely written characters. *Eg: Interpreting "cl" as "d".*
- **Sentence boundary disambiguation:** Usually the written texts are bounded to some rules like the usage of periods and punctuations for better readability and sense. But the same can also be used to represent abbreviation. To make the difference between the sentence period and abbreviation-al notations, and segmenting the sentence accordingly is called as sentence boundary disambiguation. *Eg: I woke up at 5 A.M. I went for a walk and met the C.E.O of a company in the park.* Here the difference between hte period after 5 A.M and the punctuation between A.M and C.E.O needs to be differentiated.

16 Summary and Conclusion

The Sections 1 to 7 in Chapter 1, gives the verbatim of the introduction of NLP taught in the first week classes. These chapters helped us to understand the basics and origin of NLP field and the current trending application of it. The NLP evolved when humans try to simulate the language behaviour through the system for faster and easier communication in this technological world. The mapping between the human knowledge and that of the systems (atificial intelligence) are discussed in Section 2. The general structure of any language is learnt from morphology discussed in Section 5. The type of knowledge required by the system to perform any NLP tasks is elaborated in Section 6 and the methods of language processing by the systems is discussed in Section 7.

The remaining Sections from 8 to 15 discusses the information I gained through the self-study suggested during class hours. Section 8 is just the output of a curiosity in digging the origin of NLP and its evolution in the current trend. Section 9 is the suggestion for defining natural language processing as a hard problem. Section 10, 11 and 13 is about my views on AI, conscious vs non-conscious knowledge and ambiguity. The term child knowledge acquisition mentioned in class and through the video lecture from TED talk made me to search and understand the process of language knowledge acquired by a child, which is explained in Section 14. Apart from the NLP applications discussed in class, few more applications are elaborated in Section 15 which caught my interest.

Acknowledgements
I sincerely acknowledge to Prof. Sutanu Chakraborti for teaching us the concepts of NLP and guiding us in a righteous way to develop the subject knowledge that we acquire through the NLP classes. The cited articles and web-pages are acknowledged by referring them in reference section.

References

1. Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W.: Natural language processing: an introduction. Journal of the American Medical Informatics Association : JAMIA **18**, 544–51 (2011). doi:10.1136/amiajnl-2011-000464
2. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall Press, Upper Saddle River, NJ, USA (2009)
3. Anjali M K, B.A.: Ambiguities in natural language processing. International Journal of Innovative Research in Computer and Communication Engineering **2**, 392–394 (2014)
4. Taylor, M., Esbensen, B.M., Bennett, R.T.: Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. Child Development **65**(6), 1581–1604 (1994)
5. problem in NLP, H.: http://nlp.abodit.com/home/nlp-is-hard
6. Morphology:: https://en.wikipedia.org/wiki/morpheme
7. Shemtov, H.: Ambiguity management in natural language generation. PhD thesis, Stanford, CA, USA (1997). AAI9901643
8. Deanna Kuhn, A.Z. Merce Garcia-Mila, Andersen, C.: Strategies of knowledge acquisition. MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT **60** (1995)