Natural Languages Understanding and Generation

Abstract

The objective of this paper is to dive into the topic of Natural Languages Understanding and Generation (NLU and NLG), as well as provide examples of its relevance in Human-Computer Interaction. Multiple scholarly sources were used to gather the information presented in this paper, however, some external resources such as educational videos are referenced as well. The result is a short paper outlining the areas that makeup NLG & NLU in easy to understand, high-level explanations followed by examples of real-life usage in Human-Computer applications.

Keywords: Natural Language Processing, Natural Language Generation, Machine Learning

1.0 Introduction and Background

Natural Language Understanding (NLU) and Natural Language Generation (NLG) are two pieces of the puzzle when it comes to Human-Computer Interaction via conversation. To understand exactly how these two processes take part in Human-Computer interaction, we must first understand what Natural Language even means, and then what each piece of the process (NLU & NLP) entail. To do this, we will take a look at each section in an easy to understand, high-level overview, and then we'll combine our newfound knowledge to understand how it all comes together.

1.1 Natural Language

Though its exact definition may vary depending on who you ask, in simple terms, a natural language is a language that has spawned or evolved from humans (Lyons 1991). As an example, all world languages, dialects, and variations of said languages constitute natural languages. In contrast, a Constructed Language is one that has been purposefully and thoughtfully put together, instead of having been developed through natural human interaction. Relevant examples of Constructed Languages include Programming Languages. This clear distinction is where humans and computers diverge.

1.1.1 The Difference Between Humans and Computers

To humans, understanding a conversation that involves a natural language seems quite trivial, however, that is far from reality. When listening and understanding even the simplest of phrases, there is a lot of information that we are subconsciously aware of. This information not only includes the structure of the language being spoken and how it conveys meaning but more importantly, the context surrounding the phrase is being spoken (Winograd 1973). This context and inherent knowledge allow us to parse communication that strays from any strict rules within a language. It is what allows us to communicate despite using slang, speaking in different dialects, and other anomalies we might inject into our speech.

While we have certainly created Constructed Languages to instruct computers, they inherently lack all the subconscious knowledge we possess that is required to understand Natural Languages. Therein lies the problem of Natural Language Processing. How do we teach things, which lack this crucial knowledge, to understand how we humans communicate?

1.2 Natural Language Processing & Understanding

Though the two terms are often used interchangeably, Natural Language Processing and Natural Language Understanding are not exactly the same thing. Natural Language Understanding, or NLU, is a sub-topic of Natural Language Processing that relates to machine learning comprehension. This is a very commercialized technology as it deals with voice recognition, text categorization, question & answer simulation, and archiving. John McCartney first used the term artificial intelligence when he used natural language to show how a computer could interpret algebraic word problems. Further down the road, IBM used machine learning and Natural Language Understanding to classify text. There is a broad scope of applications for Natural Language Understanding. On the simplistic end, we have simple database queries and on the more complex side, we have voice translation applications or text characterization of emails. The complexity of an NLU program is measured in two ways: breadth and depth. Breadth is measured by the size of the vocabulary in the system, while depth is measured by the fluency of the system (imagine a novice English as a Second Language student as compared to a native speaker.) As the Patom Theory States: “To have a meaningful conversation with machines is only possible when we match every word to the correct meaning based on the meanings of the other words in the sentence – just like a 3-year-old does without guesswork."

1.2.1 Distinction from Speech Recognition

Speech recognition is often lumped in with Natural Language Processing and Understanding, however, it is important to understand that the two processes are not indeed the same. Speech recognition deals with the problem of turning human speech into data that can be interpreted by a machine (Reddy 1976), while Natural Language Processing and Understanding deal with the problem of actually processing that data and using it to teach a machine how to *understand* that data. It is not always the case that NLP and NLU systems will take data from a speech recognition system. One could type out such data by hand if they wished. However, using the two in conjunction allows us to create a smooth and intuitive system for Human-Computer interaction in which a use speaks to a speech recognition system, the system parses the human’s speech and turns it into something that can be easily read by an NLP system, which can then use its techniques to try and decipher the meaning and intention of the human’s command.

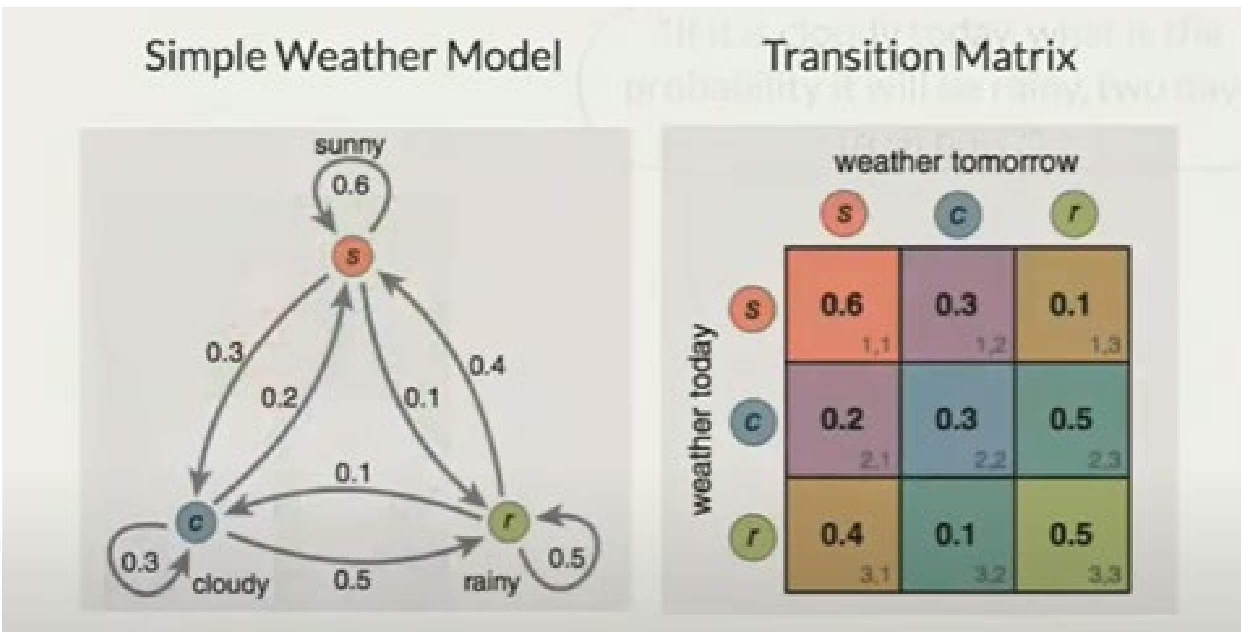
1.3 Natural Language Generation

What is a natural language generation (NLG)? It is a software process that produces structured data in the form of natural language. As noted in Stent et al. (2005), a good generator usually depends on several factors: adequacy, fluency, readability, and variation. And with it, it summarizes the data from analysts to automatically write reports that would be fitted to the audience. In the attempts to mimic human speech, NLG systems used different methods to adapt their writing style, tone, and structure according to the audience. An example is if you are using a chat box and asked your smartphone “Where is the closest restaurant?” the NLG will take in the data that it received, break it down, and reply back like a person by saying “There are fourteen different kinds of a restaurant near you” or “There is none near you at this moment.”

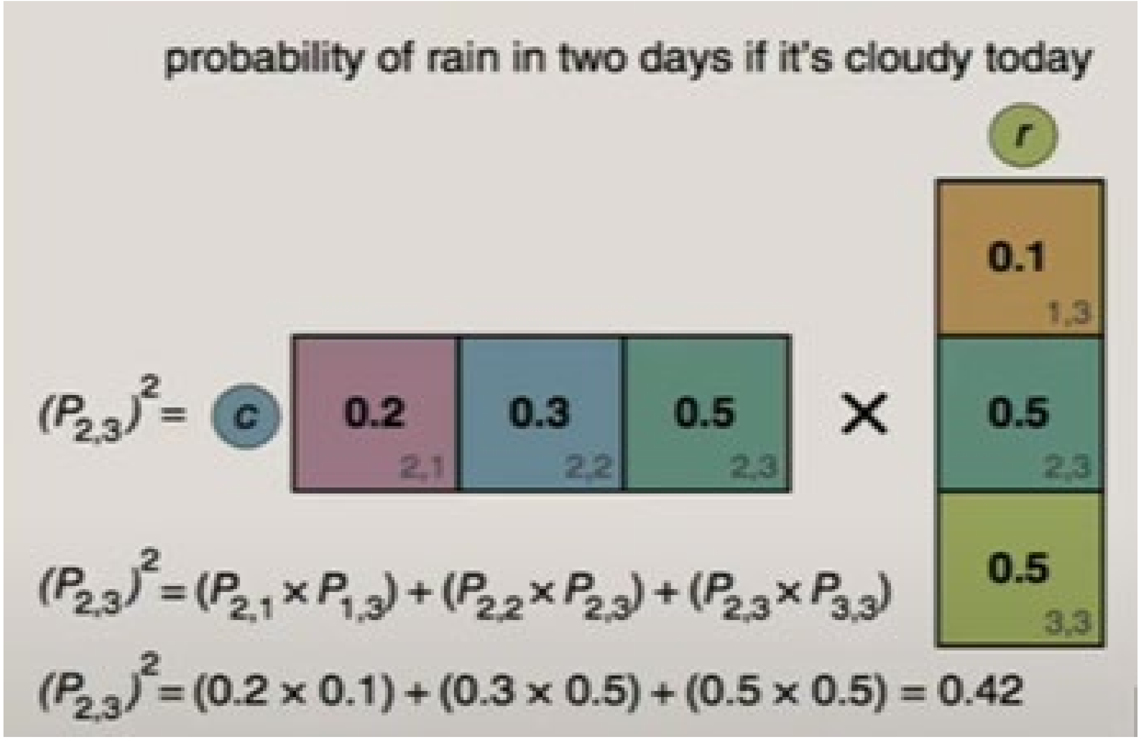
There are two major approaches to language generation: using templates and dynamic creation of documents. The template-based system is an old natural language generating system that has a predefined structure that automatically fills in gaps that have a limit because it uses data from a database table entry to fill in the gaps. The dynamics approach, on the other hand, is a more advanced generating system that creates sentences from representations of the desired linguistic structure. It does not need the developer to explicitly write code for every boundary case and it also allows the system to linguistically “optimize” sentences in several ways.

The first algorithm for the natural language generation is called the Markov chain. It predicts the next word in the sentence that the audience uses and finds the relationship between each word to calculate the next word. It is a mathematical system that transitions from one state to another according to certain probabilistic rules. From what the Brilliant said, “The characteristic of a Markov chain is that no matter how the process arrived at its present state, the possible future states are fixed.” This means that transitioning to any state is dependent on the current state and time elapsed. The Markov chain is mostly used in economics, game theory, communication theory, and finance.

From a mathematical perspective, a good example of a Markov chain is to figure out the probability of the weather forecast that the Full stack Academy created.



Looking at the weather model on the right, the state is the weather while the arrows are the transition probability. Using the weather model, it can create a transition matrix on the right of the picture. If the user wants to know what the probability of raining in two days if it is cloudy today, the Markov property will calculate the probability.



For a Markov chain to be related to NLG, the computer can track a person browsing data and create a Markov chain of state transitions on their application. Depending on how long and how much the person uses the application or website will let the computer know what the person will like and predict what the person will look for in the future. For example, if the user constantly searches for a random Pokémon on google. The computer will calculate the transition probability based on the user's behavior and current state.

A convolutional neural network (convNet/ CNN) is an algorithm that takes in an input image and can differentiate one from the other. An image is a matrix of pixels, so CNN can capture the spatial and temporal dependencies in an image through relevant filters. It has a better performance for the image dataset because the network can be trained to understand the image better. The ConvNet role is to reduce the image process without losing features. When the CNN takes in an input image and scans through every single portion of the image until the entire image is traversed. The purpose of the convolution operation is to extract high-level features such as edges from the image. There are multiple convolutional layers as the first layer is responsible for capturing the edges, color, gradient orientation, etc. With multiple layers, it will give CNN a network that understands the images in the dataset.

The pooling layer is just like the convolutional layer as it is responsible for reducing the size of the convolved feature. which has two types: "Max Pooling" and "Average Pooling." Max Pooling returns the maximum value from the portion of the image while the average pooling returns the average of all the values from the image. The max-pooling does a better performance than the average pooling because it can perform a noise suppressant. A noise suppressant discards the noisy activations and dimensionality reduction.

Recurrent neural networks (RNN) are models that store the previous words encountered in its memory and calculate the probability of the next word. For each word in the storage, the model assigns a probability based on the previous word, chooses the word, and stores it. This model is ideal for language generation because it can remember the background of the conversation at any time. The recurrent neural networks can take in one or more input from the user and produce one or more output. The thing about RNN is that the output is not just influenced by the current input, but also based on the context before the input or output of the user previously which is named "hidden." This shows that even if the user inputs the same vectors, the RNN will produce a different output base on the previous inputs. However, as the length of the sequence increases, the model moves cannot store words that were encountered remotely in the sentence and makes predictions based on only the most recent word. Because of this limitation, RNNs are unable to produce a consistent long sentence.

To stop the problem of long-range dependencies, the Long short-term memory (LSTM) was introduced by Hochreiter & Schmidhuber in 1997. The LSTM has four parts: the unit, the input door, the output door, and the forgotten door. The first part of LSTM is to decide what information to throw away from the cell state. The "forget gate layer" is the sigmoid that decides if the cell state should be kept or get thrown away. The second part is to store new information in the cell state which has two parts. The first is called a sigmoid layer named "input gate layer" that decides what values will update. Then a tanh layer will create a vector of new candidate values that could be added into the state. The last part of the LSTM is to decide what is going into the output. The output is based on the cell state and through tanh, the cell state is multiplied by the output of the sigmoid gate so there will only be outputs the user needs. This allows the network to find only relevant information, but the LSTM is only capable of storing a few hundred words.

Later, in 2017, Google introduced a new model called "self-attention mechanism." One example of the Transformer for language generation is the OpenAI which learns how to predict the next word in a sentence by focusing on words that were previously seen in the model. From what Sciforce said, "Transformer consists of a stack of encoders for processing inputs of any length and another set of decoders to output the generated sentences." Unlike LSTM, Transformer performs only a small number of steps, while applying a self-attention mechanism that simulates all words in a sentence. Looking into the structure, the attention is defined as the inner product of the query/giver and key/receiver divided by the square root of its dimensions.

The square root acts as a balance since longer sentences result in a much larger inner product. Once the user inputs a word, the output is being calculated by the sum of attention multiplied by the information from the value of each word. Query and key are the two variances that construct the relationships, while the value summarizes all the relationships to make the output, but to create multiple outputs, multi-head was created. It is a feature that creates multiple attention matrices and it simply doubles the query, key, and value. Once it does that, it will calculate the attention matrix independently. In the table below, there is a comparison time complexity of self-attention, LSTM, and CNN. The self-attention time complexity is 4 times faster than RNN while 1.5 faster than CNN.

	Time Complexity
Self-Attention	O (length^2 * dim) = 4 * 10^9

RNN (LSTM)	O (length * dim^2) = 16 * 10^9
Convolutional Neural Network (CNN)	O (length * dim^2 * kernel-width) = 6 * 10^9
length = 1000, dim = 1000, kernel-width = 3	

1.4 Use Cases in Human-Computer Interaction

While many people look at NLU/NLG as tools, the reality is that they are interfaces much in the same way a GUI or CLI are interfaces. These innovations can not only enhance the interaction between human and computer, but they can enable users who might not have previously been able to interact with computers to have an avenue with which they can interact.

1.4.1 Use Cases of Natural Language Understanding

Natural Language Understanding is key in enabling humans to interact with computers via voice. When used in conjunction with Voice Recognition, it can allow direct voice commands to be interpreted by a computer. Because of this, these technologies can be leveraged to solve a plethora of problems that non-technically literate people may have as well as people with disabilities. When a base model smartphone has voice recognition technology it means that a special device for the aforementioned people is not necessary. Accessibility is a huge benefit of implementing voice recognition.

An example of NLP and Voice Recognition working in conjunction to help those with disabilities can be seen in a paper by Tatale et al., where a system is described, implemented, and tested for the educational aid of disabled children, with a heavy emphasis on Natural Language Processing as the backbone for understanding and generating the objects that are to be displayed to a student in a virtual reality space. This system allows for students with learning disabilities to interact with a computer through voice in order to create and interact with objects in 3D space. The system takes a voice input query, which is then passed to an NLP engine for processing. Once processed, the system renders out the query in 3D space. figure 1.4.1.1 shows a high-level diagram of the system.

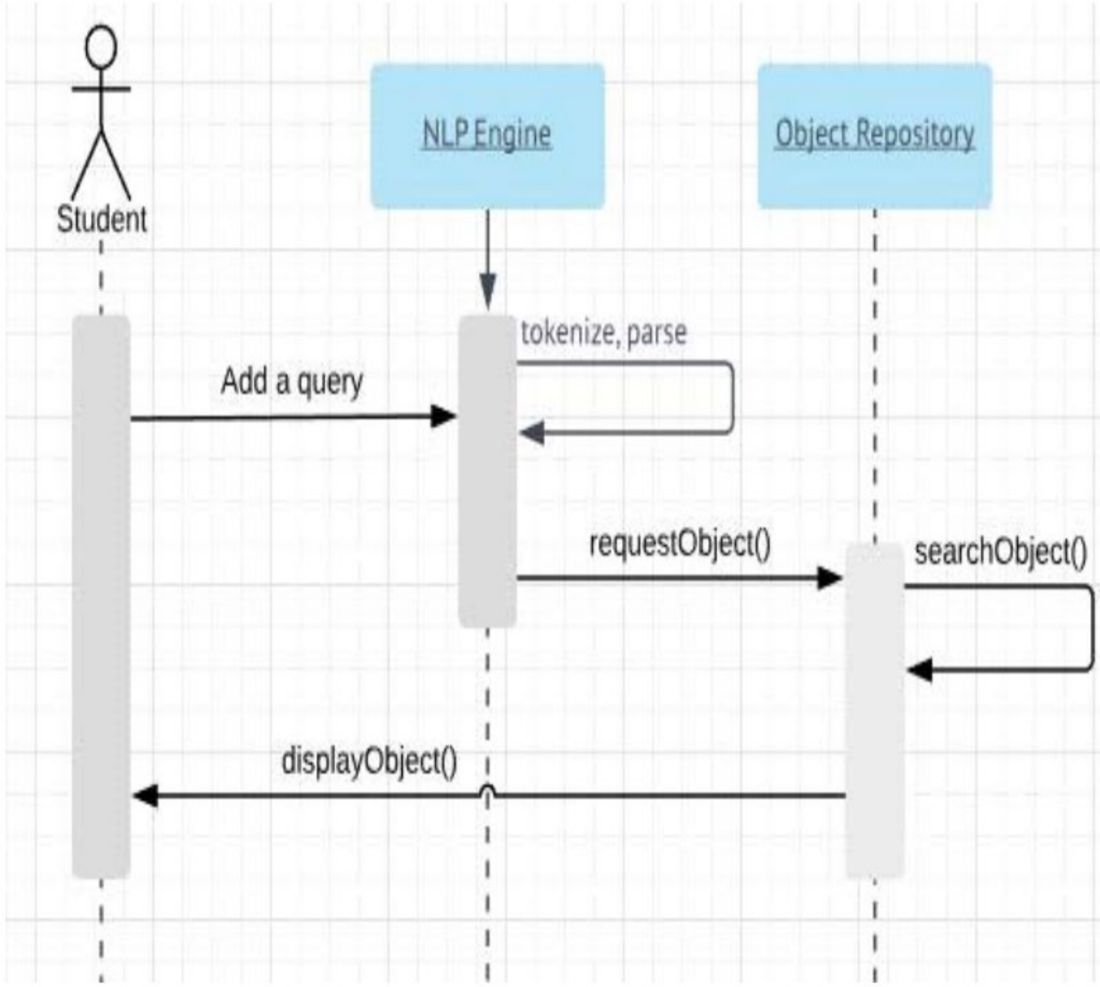


Figure 1.4.1.1; Tatale et al.

To evaluate the results, Tatale et al used a number of syntax rules as shown in Figure 1.4.1 & 1.4.2 in which the following abbreviations were used to compose said rules:

- NNP - Proper Noun
- VBG - Verb
- NN - Noun
- JJ - Adjective
- PRP - Pronoun

The results can be seen in Figures 1.4.3 and 1.4.4, where the system correctly generated the proper geometry and text based on the user's voice query.

TABLE I: SYNTAX RULE FOR HUMANOID ACTION MODULE

Sr. No.	Syntax Rule	Example
1.	NNP VBG	Mohan is dancing
2.	(NNP VBG) + (NNP VBG)	Mohan is dancing, and Maya is running
3.	(NNP VBG) + (VBG PRP)	Mohan is dancing and running is being performed by me
4.	(NNP VBG) + (VBG NNP)	Maya is running, and dancing is being performed by Mohan
5.	(NNP VBG) + (NNP VBG) + (PRP VBG)	Akshay is kicking, and Mohan is dancing. On the other hand, I am punching
6.	(PRP VBG) + (NNP VBG) + (NNP VBG) + (VBG NNP)	I am running, and Mohan is dancing. On the other hand, Akshay is punching and rolling is being performed by dadi

Figure 1.4.1; Tatale et al.

TABLE II: SYNTAX RULE FOR DYNAMIC OBJECT GENERATION MODULE

Sr. No.	Syntax Rule	Example
1.	NN JJ JJ	The apple is red in color, and it is big
2.	NN JJ	The apple is on the right/left side
3.	NN	Show me a zebra/computer/spoon etc.
4.	JJ JJ NN JJ	Show me a big red apple which is on the left side
5.	JJ NN JJ	Show me a big apple which is close/closer to me
6.	(NN) + (NN)	Show me an apple and an elephant
7.	(NN) + (JJ NN)	Show me an apple and a big brown table
8.	(JJ NN JJ) + (JJ JJ NN JJ)	Show me a red apple which is on the left side of the big brown table

Figure 1.4.2; Tatale et al.

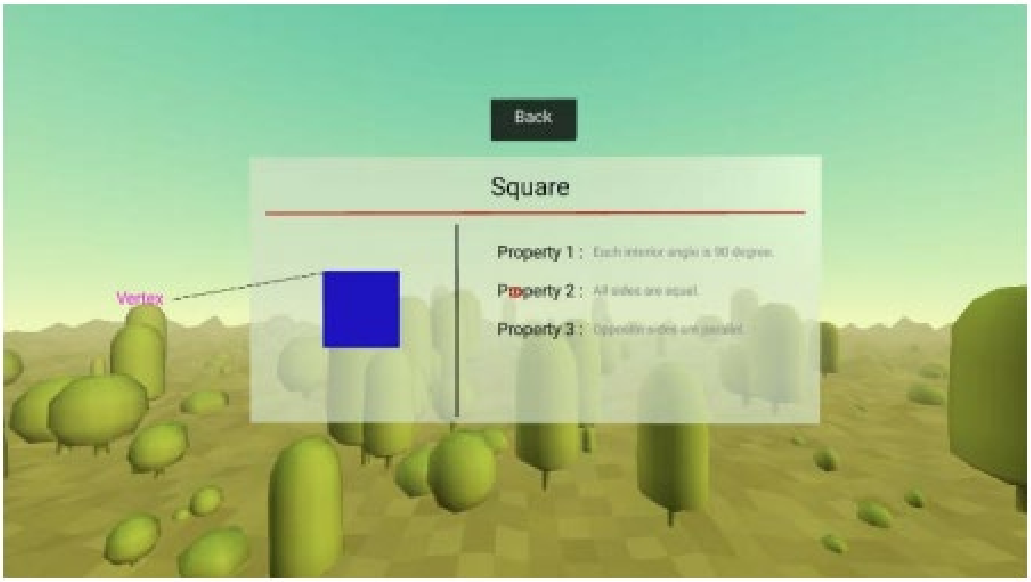


Figure 1.4.3 Courses - Shapes being taught in a 3D world; Tatale et al.

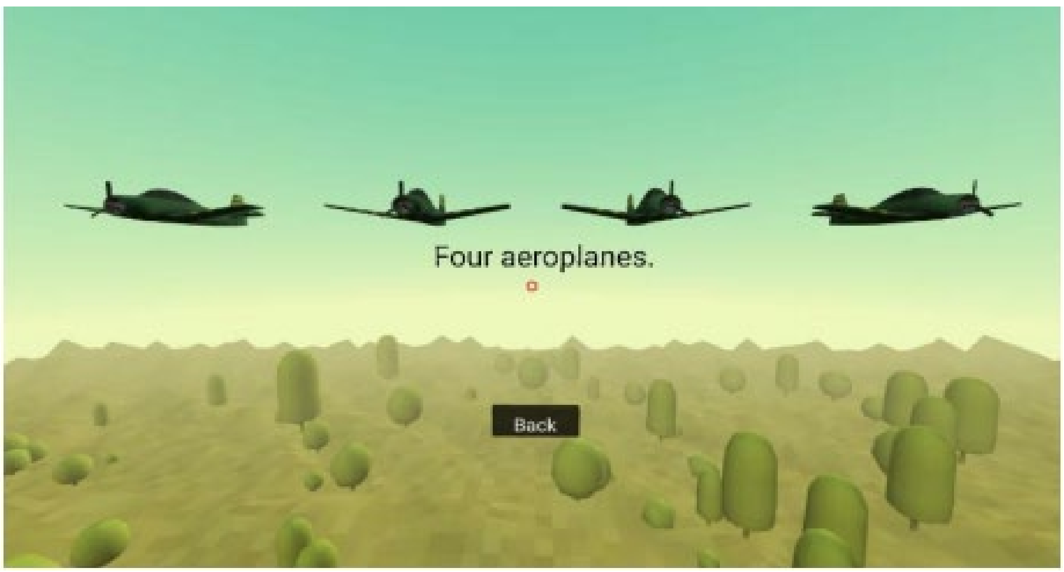


Figure 1.4.4 Courses - Quantity being taught in 3D world; Tatale et al.

NLP also eliminates the need for software to hardcode multiple translations of its text, as it can instead use NLP to handle all translations in the cloud. A paper by Padro, L., & Turmo, J. shows this in effect by detailing a system called "TextServer", a cloud based solution for text related problems, which not only offers multilinguality, but also scalability, replicability, and ease of use. Figure 1.4.5 shows a table of the distribution of services for each TextServer service across 15 languages, which gives an insight into the power and usefulness of a service like this.

	as	ca	cs	cy	de	en	es	fr	gl	hr	it	nb	pt	ru	sl
language Identification	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Phonetic Transcription			X			X	X								
Tokenization&Splitting	X	X		X	X	X	X	X	X	X	X	X	X	X	X
Morphological Analysis	X	X		X	X	X	X	X	X		X	X	X	X	X
PoS tagging	X	X		X	X	X	X	X	X		X	X	X	X	X
WSD		X				X	X	X	X		X		X		X
NERC		X				X	X						X		
Chunking	X	X				X	X		X				X		
Dependency parsing	X	X			X	X	X		X	X					X
SRL		X			X	X	X								
Coreference resolution						X	X								
Semantic graph generation		X				X	X								

TABLE I. DISTRIBUTION OF LANGUAGES FOR EACH TEXTSERVER SERVICE

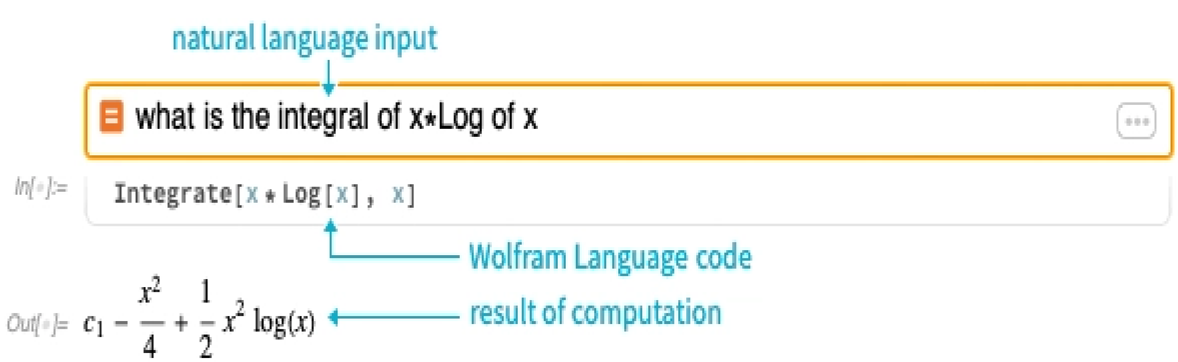
Figure 1.4.5 DISTRIBUTION OF LANGUAGES FOR EACH TEXTSERVER SERVICE; Padro, L., & Turmo, J

Another very important use for Natural Language Understanding is voice recognition in automobiles. Thousands of people die every year due to cell phone use while operating a car. Using voice recognition to render a phone call, navigate to an address, or to change a song can save countless lives. This application allows for hands-free navigation of several interfaces. As Natural Language Understanding improves, so does the quality and swiftness of voice recognition. Because of this, technologies become more intuitive and consequently more powerful.

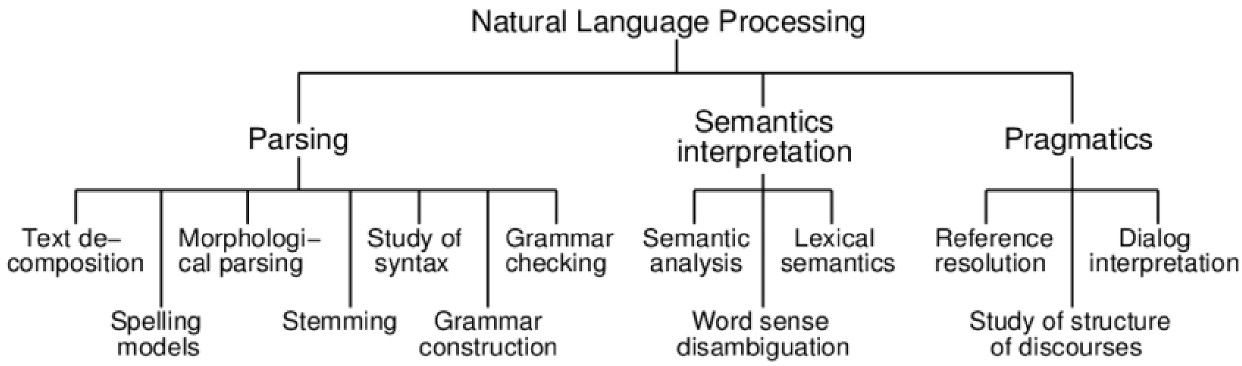
1.5 Project Summary

Natural Language Understanding is “the comprehension by computers of the structure and meaning of human language (e.g., English, Spanish, Japanese), allowing users to interact with the computer using natural sentences”. (source) This technology is a subcategory of Artificial Intelligence. NLU can take an input such as voice or text and interpret it in a computing language then create an output of that computer language into a language that can be understood by humans. The most direct example of this is Google Translate.

Another very commonly used instance of NLU is Siri or Cortana. The goal of Natural Language Understanding is to seamlessly translate the human vernacular using computer programs. Essentially, NLU in a perfect world would be able to translate, parse, understand and respond to a given human input. For example, Siri would be able to tell you the answer to a question regardless of language, tone or slang used. When someone asks their phone, “How hot is it outside?” Natural Language Understanding allows for the program to realize that the person is asking for the temperature, without explicitly inquiring, “How many degrees is it today?”. This can also be applicable with calculators such as Wolfram Alpha.



Natural Language Understanding and Natural Language Processing are two branches of AI that are often used interchangeably however they have distinct differences. Natural Language Understanding is a subset of Natural Language Processing. Natural Language Understanding is related to error management in regard to dialects, slang, mispronunciations and phrasing of languages in human interaction. Think of the first time you heard the phrase, “It’s raining cats and dogs”. It doesn’t make a lot of sense to a non-english speaker. A huge roadblock of Natural Language Processing is mispronunciations and NLU seeks to minimize the consequences of these errors.



Conversely, Natural Language Processing is the broader term used to describe the process of taking text, voice, other human inputs and data and returning an output that can be understood by humans. Some common Natural Language Processing applications include Text-to-Speech translations, Optical Character Recognition, Part of Speech Tagging, Lexical Semantics, and Question Answering. Optical character recognition relates to matching photos to words or people. Lexical Semantics refers to recognizing slang and euphemisms in context of a sentence or conversation. For instance, when a user inputs the word “match”, is she referring to a soccer game or a potential dating partner? Natural Language Processing helps computer programs understand these discrete semantics.

Extended Resources

1. https://www.youtube.com/watch?v=fOvTtapxa9c

A quick video that summarizes the need for Natural Language Processing and gives a high-level overview of how some systems implement it. The video also goes slightly into the history of NLP and describes speech recognition's importance in the process of NLP

---

2. https://www.youtube.com/watch?v=MNvT5JekDpg

In this video, two Googlers talk about Natural Language Generation and how it is implemented at Google in applications like Google Assistant. The video is good for understanding the applications of NLG, though more specifically, Machine Learning backed NLG.

---

3. https://narrativescience.com/resource/blog/what-is-natural-language-generation

A brief article that breaks down Natural Language Generation, its difference with NLP, as well as potential future applications of this technology. The article also touches on other technologies' reliance on innovation in the field of Natural Language Generation for the implementation of user communication

---

4. https://www.forbes.com/sites/danwoods/2015/07/09/why-big-data-needs-natural-languag

This article briefly goes over the history of human language rendering before moving on to talk about the uses of Natural Language Generation in the field of Big Data. It also talks about the importance of context when trying to not only understand but also generate natural language and how semantic engines in this endeavor by breaking down the process into different types of analysis.

---

5. https://www.youtube.com/watch?v=4q3H_ZN01kk

In this video, a person will thoroughly explain about the Markov chain, what it is, how it was created, example, and use cases.

---

6. https://www.youtube.com/watch?v=5vcj8kSwBCY

In this video, a professor from Stanford University named Christopher Manning explains about transformers and self-attention.

---

7. https://www.youtube.com/watch?v=EzDgw4_gCVU

This video shows multiple use cases of NLP in the Healthcare industry, such as lowering patient waiting times by using NLP to process patient forms and using Named Entity Recognition alongside Deep Leaning technologies to aid with medical diagnosis

---

8. https://becominghuman.ai/8-thought-provoking-cases-of-nlp-and-text-mining-use-in-business-60bd8031c5b5

This article talks briefly about 8 different modern use cases for NLP. Some of the use cases on here are quite common, but a lot of them are different and very interesting. The article also does a good job of giving exact figures for the impact of NLP in these technologies

---

9. https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/

This fascinating article gives you a brief, yet detailed history of Natural Language Processing all the way back to the early 1900s. It breaks down the stages that NLP has gone through to get to where it is today, and really shows how much time, energy, and effort has been put into this particular field of Machine Learning.

---

10. https://lance-eliot.medium.com/natural-language-processing-nlp-for-in-car-voice-discussion-with-riders-of-driverless-cars-26f34b0248e8

This article talks about the many use cases for In-Car Voice recognition and understanding services. It also briefly talks about the drawbacks as well as impact on the safety of drivers that NLP/NLU has had in the automotive industry. The article is also heavily focused on telling the reader how self driving cars will utilize Natural Language Generation to communicate with its passengers.

---

11. https://www.synaptiq.ai/machine-learning-nlp-for-vehicle-book-matching/

This Case Study talks about a specific use of NLP and Machine Learning in the automotive industry to better understand the value of used cars. Because different people describe things in different ways, this article talks about how NLP was used to decipher different descriptions and help automate the process of used car appraisals.

---

12. https://analyticsindiamag.com/how-mercedes-benz-is-using-ai-nlp-to-give-driving-a-tech-makeover/

This article describes hope Mercedes-Benz is implementing NLP in their vehicles to enhance the user's experience in their automobiles by making their in house voice assistant, MBUX, have more features. One of the key features is that MBUX can create a better relationship with its driver by "routine to make personal predictions and offer appropriate recommendations".

---

13. https://www.the-learning-agency-lab.com/the-learning-curve/how-npl-will-change-education

This article touches on NLP in the context of education, and how it can improve learning for students in Behavior & Motion and Reading & Writing. It talks about what metrics can be used to measure NLP's impact in these areas, as well as the individual impact it has on students.

References

Joshi, A. K. (1991). Natural language processing. Science, 5025, 1242.

Lyons, John (1991). Natural Language and Universal Grammar. New York: Cambridge University Press. pp. 68–70. ISBN 978-0521246965.

T. Winograd, "Research in natural language," in Computer, vol. 6, no. 5, pp. 25-30, May 1973, doi: 10.1109/MC.1973.6536715.

Sciforce. (2019, July 4). A Comprehensive Guide to Natural Language Generation. Medium. https://medium.com/sciforce/a-comprehensive-guide-to-natural-language-generation-dd63a4b6e548

D. R. Reddy, "Speech recognition by machine: A review," in Proceedings of the IEEE, vol. 64, no. 4, pp. 501-531, April 1976, doi: 10.1109/PROC.1976.10158.

Gasic. M, Mrksic. N, Su. P, Vandyke. D, Wen. T, Young. S, "Semantically Conditioned

LSTM-based Natural Language Generation for Spoken Dialogue Systems," Retrieved August 26, 2016, from https://arxiv.org/pdf/1508.01745v2.pdf

Saha Sumit, "A Comprehensive Guide to Convolutional Neural Networks - the ELI5 way," Retrieved December 15, 2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Tatale, S., Bhirud, N., Parmar, R., & Pawar, S. (2019). Education using Virtual Reality for students with learning disabilities. 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Computing, Communication, Control And Automation (ICCUBEA), 2019 5th International Conference On, 1–7. https://doi.org/10.1109/ICCUBEA47591.2019.9128515

Padro, L., & Turmo, J. (2015). TextServer: Cloud-Based Multilingual Natural Language Processing. 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, Data Mining

Workshops (ICDMW), 2013 IEEE 13th International Conference On, 1636–1639.

https://doi.org/10.1109/ICDMW.2015.102