Customise appearance

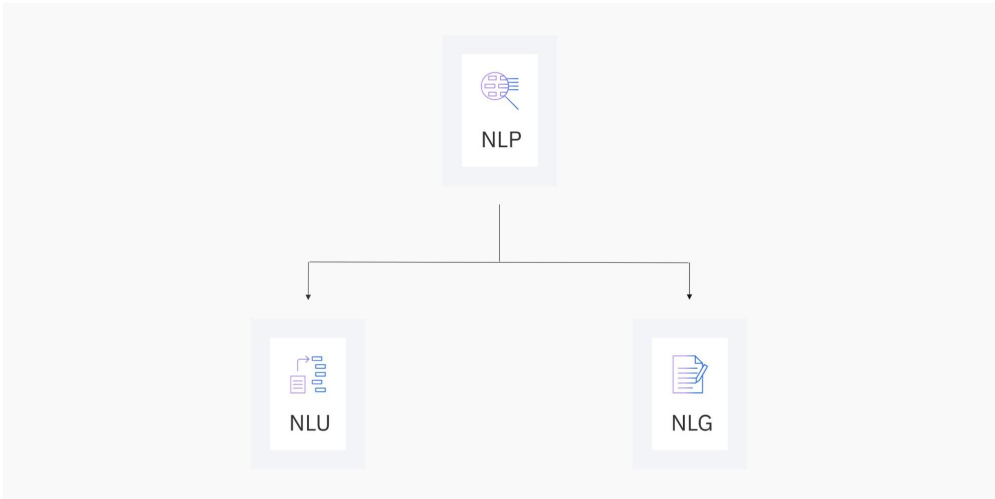# NLP vs. NLU vs. NLG: the differences between three natural language processing concepts - Watson Blog

By Eda Kavlakoglu | 5 minute read | November 12, 2020

While natural language processing (NLP), natural language understanding (NLU), and natural language generation (NLG) are all related topics, they are distinct ones. At a high level, NLU and NLG are just components of NLP. Given how they intersect, they are commonly confused within conversation, but in this post, we'll define each term individually and summarize their differences to clarify any ambiguities.
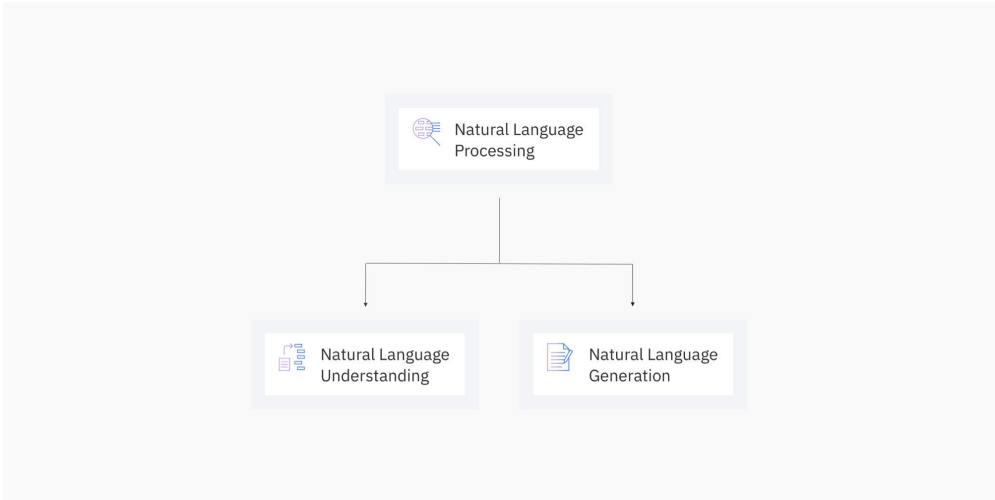
What is natural language processing?

Natural language processing, which evolved from computational linguistics, uses methods from various disciplines, such as computer science, artificial intelligence, linguistics, and data science, to enable computers to understand human language in both written and verbal forms. While computational linguistics has more of a focus on aspects of language, natural language processing emphasizes its use of machine learning and deep learning techniques to complete tasks, like language translation or question answering. Natural language processing works by taking unstructured data and converting it into a structured data format. It does this through the identification of named entities (a process called named entity recognition) and identification of word patterns, using methods like tokenization, stemming, and lemmatization, which examine the root forms of words. For example, the suffix -ed on a word, like called, indicates past tense, but it has the same base infinitive (to call) as the present tense verb calling.

While a number of NLP algorithms exist, different approaches tend to be used for different types of language tasks. For example, hidden Markov chains tend to be used for part-of-speech tagging. Recurrent neural networks help to generate the appropriate sequence of text. N-grams, a simple language model (LM), assign probabilities to sentences or phrases to predict the accuracy of a response. These techniques work together to support popular technology such as chatbots, or speech recognition products like Amazon's Alexa or Apple's Siri. However, its application has been broader than that, affecting other industries such as education and healthcare.

What is natural language understanding?

Natural language understanding is a subset of natural language processing, which uses syntactic and semantic analysis of text and speech to determine the meaning of a sentence. Syntax refers to the grammatical structure of a sentence, while semantics alludes to its intended meaning. NLU also establishes a relevant ontology: a data structure which specifies the relationships between words and phrases. While humans naturally do this in conversation, the combination of these analyses is required for a machine to understand the intended meaning of different texts.

Our ability to distinguish between homonyms and homophones illustrates the nuances of language well. For example, let's take the following two sentences:

1. Alice is swimming against the current.
2. The current version of the report is in the folder.

In the first sentence, the word, current is a noun. The verb that precedes it, swimming, provides additional context to the reader, allowing us to conclude that we are referring to the flow of water in the ocean. The second sentence uses the word current, but as an adjective. The noun it describes, version, denotes multiple iterations of a report, enabling us to determine that we are referring to the most up-to-date status of a file.

These approaches are also commonly used in data mining to understand consumer attitudes. In particular, sentiment analysis enables brands to monitor their customer feedback more closely, allowing them to cluster positive and negative social media comments and track net promoter scores. By reviewing comments with negative sentiment, companies are able to identify and address potential problem areas within their products or services more quickly.

What is natural language generation?

Natural language generation is another subset of natural language processing. While natural language understanding focuses on computer reading comprehension, natural language generation enables computers to write. NLG is the process of producing a human language text response based on some data input. This text can also be converted into a speech format through text-to-speech services.

NLG also encompasses text summarization capabilities that generate summaries from in-put documents while maintaining the integrity of the information. Extractive summarization is the AI innovation powering Key Point Analysis used in That's Debatable.

Initially, NLG systems used templates to generate text. Based on some data or query, an NLG system would fill in the blank, like a game

of Mad Libs. But over time, natural language generation systems have evolved with the application of hidden Markov chains, recurrent neural networks, and transformers, enabling more dynamic text generation in real time.

As with NLU, NLG applications need to consider language rules based on morphology, lexicons, syntax and semantics to make choices on how to phrase responses appropriately. They tackle this in three stages:

- Text planning: During this stage, general content is formulated and ordered in a logical manner.
- Sentence planning: This stage considers punctuation and text flow, breaking out content into paragraphs and sentences and incorporating pronouns or conjunctions where appropriate.
- Realization: This stage accounts for grammatical accuracy, ensuring that rules around punctation and conjugations are followed. For example, the past tense of the verb *run* is *ran*, not *runned*.

NLP vs NLU vs. NLG summary

- **Natural language processing (NLP)** seeks to convert unstructured language data into a structured data format to enable machines to understand speech and text and formulate relevant, contextual responses. Its subtopics include natural language processing and natural language generation.
- **Natural language understanding (NLU)** focuses on machine reading comprehension through grammar and context, enabling it to determine the intended meaning of a sentence.
- **Natural language generation (NLG)** focuses on text generation, or the construction of text in English or other languages, by a machine and based on a given dataset.

Infuse your data for AI

Natural language processing and its subsets have numerous practical applications within today's world, like healthcare diagnoses or online customer service.

Explore some of the latest NLP research at IBM or take a look at some of IBM's product offerings, like Watson Natural Language Understanding. Its text analytics service offers insight into categories, concepts, entities, keywords, relationships, sentiment, and syntax from your textual data to help you respond to user needs quickly and efficiently. Help your business get on the right track to analyze and infuse your data at scale for AI.

Explore some of the latest NLP research at IBM or take a look at some of IBM's product offerings, like Watson Natural Language Understanding. Its text analytics service offers insight into categories, concepts, entities, keywords, relationships, sentiment, and syntax from your textual data to help you respond to user needs quickly and efficiently. Help your business get on the right track to analyze and infuse your data at scale for AI.