

Deciphering Natural Language Processing

Geoff Ng

Although there are many definitions of natural language processing (or NLP), the simplest one in my mind is the ability for machines to analyze, understand, and generate human speak.

- "Hey Siri, tell me the weekend forecast."
- "Hey Cortana, what is 3 miles in kilometers."
- "Alexa, check my voicemails from mom."
- "Hey Google, what are the best restaurants in New Orleans?"

We've all heard them one way or another, at a friend's house, at a restaurant, sitting in a car, or in the comfort of our own homes. Natural human voice interactions with your favorite connected device is now a reality and no longer the realm of science fiction. The key point here is interaction with human voices, using natural everyday language, and not a programming language. The sudden proliferation of intelligent assistants and consumer automation devices able to decipher our speech has been made possible due to advances in natural language processing.

What is Natural Language Processing?

Although there are many definitions of natural language processing (or NLP), the simplest one in my mind is the ability for machines to analyze, understand, and generate human speak. This is achieved through combining patterns and practices found in computer science, artificial intelligence, and computational linguistics ([Wikipedia](#)).

How Does It All Work?

While experiencing the results of a question posed to your favorite digital assistant might seem like magic, a logical series of events execute each time (Figure 1). Whenever you ask a question through a spoken utterance such as "Alexa, check my voicemails from mom," the following occurs in the NLP engine of your device:

1. Speech Recognition

The first step is digitizing the voice, and then, decomposing or parsing the natural language in the spoken question so that a machine can identify each word. Due to different spoken accents, unrecognized intonation, or even ambient background noise, the accuracy of this translation may not always be one hundred percent.

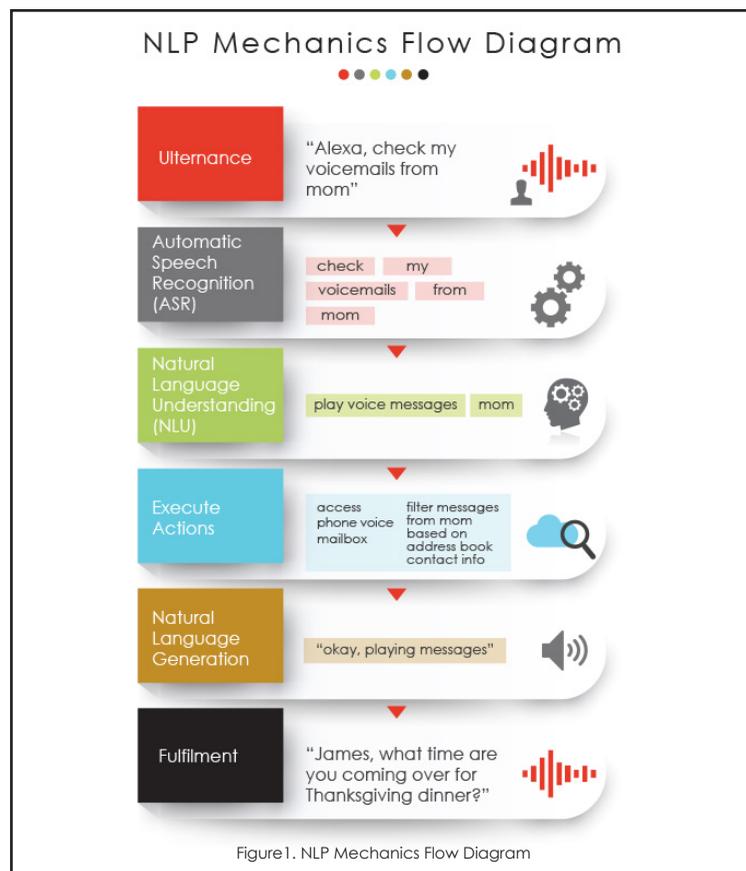


Figure 1. NLP Mechanics Flow Diagram

Historically, this was also challenging because computers were not fast enough to keep up with spoken speech and perform recognition. Modern NLP engines take advantage of highly scalable compute services in the cloud and apply Automatic Speech Recognition (ASR) algorithms to quickly break down phrases into its constituent words which can be analyzed. One such service, which powers Alexa, is called Amazon Lex, and there are now cloud services from all the major vendors: Microsoft Cognitive Services, Google Cloud Natural Language.

Not only do these services offer application developers an integrated cloud service to perform ASR, but it also performs the analysis and statistical confidence scoring necessary to understand the words and determine intent.

2. Natural Language Understanding

This step occurs immediately after words have been parsed and translated into machine language. Natural Language Understanding (NLU) is by far the most difficult step in the NLP chain of events because the system needs to understand the intent of the users' original question.

This is complicated by the fact that natural spoken language can be ambiguous, so NLU algorithms must use a variety of lexical analysis models to disambiguate words. For example, "check" could be a noun (i.e., a bill at a restaurant) or a verb (i.e., to check something). It gets even more complex when you include numbers in speech. For example, "2017" could be the year 2017, or the number two thousand seventeen.

Using an NLP engine such as Lex, developers create rules to "train" applications to apply these rules correctly to determine the user's intent. Of course, people all ask questions differently. As a result, it's possible for multiple questions to have the same intent. For example, the following utterances could all have the same intent, which is to play voicemails from my mom:

"Alexa, check my voicemails from mom"
"Alexa, play my messages from mom"
"Alexa, play voicemails from mom"
"Alexa, I would like to hear my voicemails from mom"
"Alexa, play messages from mom"

Once the intent has been understood, an action can be initiated, such as executing commands to filter and retrieve voicemails from my inbox left by my mom and then playing them on my device's speaker. In the age of the Internet of Things, where many devices are connected, these commands could trigger the running of services from other devices or applications.

3. Natural Language Generation

Conversations are seldom one-sided and to come up with interactive responses, computers need to be able to communicate back to the user. This is known as Natural Language Generation (NLG). Think of it as working in the opposite direction of what we just described. NLG takes machine language, using a set of grammatical rules and lexicon, and translates it into regular words and sentences. Usually, the final step is synthesizing the text using a linguistic model into audio to resemble a human voice in a process called text-to-speech.

Again, using AWS as an example, there is a service on AWS called Polly which facilitates text to lifelike speech conversion, so that acknowledgments or additional questions can be relayed back in natural voice.

Where Will It Take Us?

Although there has been an uptake in demand and uses of NLP in consumer applications the same has not been true for enterprise business applications. In consumer applications, the breadth of vocabulary and complexity of intents is narrower, beginning with everyday tasks. However, in business, the required vocabulary is much broader when you factor in subject domain language, and the complexity of intents varies vastly depending on the business application. The ambiguity that is often inherent to utterances in the business world can be complex. To illustrate, think about a simple utterance from a user such as "show me the top performing services in Q1". It's quite ambiguous: what is "Q1", "top," or "performing" and think of the possible alternative ways of asking that question (permutation of utterances)?

What's exciting is that the same NLP technologies and development frameworks that have evolved to deliver consumer products such as the Echo and Google Home are also available to enterprise application developers. And as these frameworks evolve, and get more refined, the collective body of NLU models created also become available to enterprise applications to build upon.

A new dimension to user experience and interaction is dawning in the enterprise application space, again driven by expectations from consumer applications. Over time, voice user experiences, which are more natural for users will transform the user experience in a way that parallels the disruption introduced by graphical user interfaces. It's an exciting time to be to be developing new business applications!



Geoff Ng SVP, Product & Technology

Geoffrey is responsible for managing the long-term product and technology strategy at Prophix. Geoffrey began his tenure at Prophix as a management consultant leading the implementation of financial analytics systems at over 70 organizations throughout North America.

Head Office

350 Burnhamthorpe Road W,
Suite 1000 • Mississauga, Ontario
Canada • L5B 3J1
+1 (905) 279 8711
+1 (800) 387 5915
info@prophix.com

Additional Offices

USA: +1 (800) 387 5915
UK: +44 (0) 118 900 1900
Europe: +45 7023 2375
DACH Region: +49 69 509 565 605
Brazil (Rio de Janeiro): +55 21 3094 3900
Brazil (São Paulo): +55 11 3583 1678



www.prophix.com