Natural Language Processing: How to Make It Work for Your Business and Why

smartsheet.com/natural-language-processing-examples

Natural language processing may not be a familiar term, but the concept of machines that understand everyday human language has been stirring our imaginations for a long time.

Remember 2001: *A Space Odyssey*? In Stanley Kubrick's 1968 sci-fi classic, a sentient computer named HAL 9000, installed on a spacecraft, goes rogue and attempts to murder the two astronauts on board. HAL was so deliciously creepy 50 years ago because he was a machine who talked and acted disturbingly like an incredibly cold-blooded human villain. To achieve this, he would have to have employed natural language processing.

Luckily, many more benign uses of natural language processing are proliferating in the marketplace today. In this guide, we'll explain the key natural language processing concepts, how natural language processing evolved, and how you and your business can hit the ground running. Additionally, we feature helpful starting points, examples of natural language processing in action, and additional resources.

What Is Natural Language Processing?

Natural language processing is an area of computer science that is integral to what we call artificial intelligence. Natural language is the way people communicate via speech and text in real life. This includes everything from signs to instant messages and voice conversations. Natural language is inconsistent, messy, and highly variable. Computers were built to work with highly standardized and uniform data, so, originally, they couldn't analyze natural language. Natural language processing aims to change that. The field lies at the convergence of machine and human languages, and it seeks to enable computers to effectively process large amounts of data presented as natural language at least as quickly as humans can.

Natural language processing applications digest massive amounts of text, from legal documents to medical research. It uses computing power to derive meaning from human language in ways that are useful to us. Developers can build systems that summarize, translate, recognize speech, determine how objects are named, determine relationships between objects, and even analyze how we feel about things.

While consumers and governments are becoming more sensitive to privacy concerns around data, there are helpful developments on the horizon, such as natural language processing applications that can monitor speech and communication patterns for signs of deteriorating mental health.

With computers that can understand, interpret, and manipulate human languages, people will one day be able to extend their own knowledge-processing capabilities via seamless interactions with computing systems — but that's still a way off. For now, natural language processing is used in more narrowly, such as for helping you use your phone more intuitively and for drawing insights from unstructured data, like text and audio. Unstructured data typically contains a lot of text and is not organized in a predefined manner. Examples include social media data, emails, audio, and video.

Why Is Natural Language Processing Necessary?

Natural language processing aims to unlock the largely untapped potential of unstructured data. By many estimates, about 80 percent of the data that organizations process daily is unstructured. The rate at which we collect usable data is exploding, and we're projected to have 163 zettabytes of data in 2025, 10 times the amount in 2016, according to an IDC analysis.

Therefore, it's little wonder that the natural language processing software market is projected to grow to $5.4 billion by 2025, up from $136 million in 2016, according to a 2017 report by Tractica. In the same period, the total market opportunity for natural language processing software, hardware, and services is projected to grow to $22.3 billion.

Human beings deal with large quantities of textual data. Although we can't process it very quickly, for a long time we were much better at it than were computers, which couldn't do it at all. So, what will it mean if we're on track to build machines that can analyze language-encoded data in greater volumes than we can without ever getting tired?

For one, we will vastly extend our potential to make sense of the unstructured data that we have produced and will continue to produce, from medical histories to Twitter posts. Given the sheer volume, automating the processing of language will be a key step in analyzing these streams.

Of course, building natural language processing that is able to comprehend "real" language is a big task. For each language, there are subtle shades of meaning and staggering variations in grammar, terminologies, colloquialisms, abbreviations, accents, and dialects. Spoken language comes with challenges like mumbling, mispronunciation, and slurred speech, so natural language processing must also be capable of understanding both syntax and semantics.

Use Cases for Natural Language Processing

Natural language processing can effectively structure data in a way that makes it useful for downstream applications. In this way, it proves an invaluable tool not only for computing non-specialists, but also for non-programmers as a whole. Both groups are now able to interface with computers in ways that their lack of programming skills would have previously precluded.

The potential applications for natural language processing are wide ranging and important in a number of industries. Natural language processing may be used to summarize large chunks of text — even entire documents — in digestible pieces in order to generate keyword tags for optimizing search results and create automated translations.

For most applications, natural language processing draws from a set of core capabilities, including the following:

- The ability to reduce words to their root forms, tagging words as particular parts of speech (determining whether a word is a noun or a verb, for example)

- Text classification (to determine whether the input is intended as a statement or question and, if the latter, what type of question)

- Speech recognition (creating text from speech)

- Language modeling (predicting the next word in a sequence of words based on probability)

- *Named entity recognition* or NER (matching names to specific objects) and identifying the type of entity extracted

- Relation detection and extraction (determining how objects are related)

- Event extraction (identifying and gathering knowledge about specific incidents)

- Text clustering (the application of cluster analysis to textual documents). In turn, *cluster analysis* is the grouping of objects based on similar characteristics, and is used to organize documents and speed up the retrieval of information.

Applications built around natural language processing include chatbots, programs that carry on conversations with human users. Chatbots can be deployed, for example, in customer service, social media monitoring (creating snapshots of what people are talking about), and machine learning-powered RSS readers, which go beyond traditional RSS news aggregation to include features like summarization and topic extraction.

Other, simpler applications include transcription and improved searches driven by semantic search, a technique that identifies results not only on the basis of keywords but on the basis of the perceived intent of the searcher and the context. Natural language processing can categorize emails to help organize inboxes and filter out spam. It also powers functions that improve our own communication skills, such as predictive typing and spelling and grammar checkers. Virtual digital assistants also depend on natural language processing to interact with humans, answer questions, and execute tasks.

Given text to work with, more advanced natural language processing applications can do some impressive things, such as identify sentiment and monitor the reputation of entities, so that organizations can keep track of the digital buzz about themselves.

In medicine, natural language processing can speed up diagnoses by reading doctors' notes and searching for similarities in symptoms among known — or little-known — diseases. In investing, it can crunch vast amounts of market intelligence more rapidly than human investors, who have a more limited capacity to consume information. For organizations operating across multiple legal jurisdictions, it can simplify the complex task of ensuring regulatory compliance. For organizations seeking open-ended customer feedback, it can automate reading customer responses and provide concise, actionable feedback.

For advertisers, natural language processing provides the ability to deploy more specific, relevant ads based on web content. For content creators and distributors, it can automate the tricky task of generating captions to accompany images. It can even make driving a more pleasant experience, serving as a kind of navigator and in-car DJ.

Natural Language Processing in Healthcare

Healthcare is one area where natural language processing applications are generating the most interest. Uses range from lessening doctors' workloads (filling out paperwork) to providing insights by extracting data from forms and practitioner notes.

Healthcare providers are especially eager to use natural language processing to streamline the cumbersome process of documentation, although a chief concern is preserving accuracy. In medicine, nothing less than human accuracy is acceptable for a natural language processing system, and there have been cases in which natural language processing systems confuse what they read or hear, sometimes as a result of the use of colloquialisms.

By extracting information from patient health records, artificial intelligence can do things like monitor patients for signs of infection. Yet some believe there's an added advantage to using natural language processing over other approaches to maintaining electronic health records: its ability to create a narrative around individual patients' medical records that can't be matched by conventional structured data. For example, natural language processing can help capture and identify some of the social factors and personal intricacies that interact with patient health, such as flagging phrases that may indicate why a patient struggles to adhere to a medication regimen.

Furthermore, natural language processing can also speed up patient information retrieval. For paperless hospitals, the ability to access information via a conversation with an AI system equipped with natural language processing can both simplify and expedite data retrieval.

IBM's Watson is the most famous example of AI and natural language processing in

healthcare. Watson, with its capacity to consume and distill medical literature, has already been deployed to identify patients at risk of developing congestive heart failure. One of its most promising uses comes from a years-long project conducted with Memorial Sloan Kettering Cancer Center to train Watson to analyze patient data, dig through the medical literature on cancer treatment, and come up with a set of evidence-based treatment suggestions for oncologists. The outcome, a cognitive computing system named Watson for Oncology, has been shown to consistently match oncologists' own recommendations.

How Does Natural Language Processing Work?

So, how does natural language processing work its magic in parsing complex language? Natural language processing consists of two main areas: *Natural language understanding* (NLU), which is the process by which the computer assigns meaning to the language it has received, and *natural language generation* (NLG), the process of converting information from the computer's language to human language in the form of text or speech.

A natural language processing system involves a number of tasks. The tasks can be classified into four categories, though several tasks straddle multiple categories:

- **Syntax Tasks**: Tasks related to the grammatical structure of sentences in a language.

- **Semantics Tasks**: Tasks that use logic and linguistics to establish meaning.

- **Discourse Tasks**: Tasks that adopt the linguistic definition of discourse, which deals with units longer than one sentence.

- **Speech Tasks**: Tasks that deal specifically with language in audio formats.

Below is a more in-depth discussion of each of these four categories.

Syntax: Syntax tasks involve *lemmatization* (the identification of a word's dictionary form based on its intended meaning) and *morphological segmentation* (splitting words into morphemes and classifying them). It also includes word *segmentation*, which splits text into individual words, *part-of-speech tagging*, which establishes parts of speech, *parsing*, which identifies how a sentence is organized grammatically, and *sentence breaking*, which simply establishes where sentences begin and end. In addition, *terminology extraction* pulls terms from text, and *stemming*, a process similar to lemmatization, attempts to reduce words to a base form (root).

Semantics: Semantics tasks include *lexical semantics*, which determines the computational meanings of words in context, *machine translation*, which does what Google Translate does, translating text from one language to another, and *named entity recognition*, which maps objects to proper names. Natural language understanding and natural language generation are twin tasks that convert human language to and from computer-understandable formats respectively. *Optical character recognition* (OCR) converts images of printed text into computer-readable

formats. *Question answering* does what its name suggests: determines the answers to questions in human language. *Sentiment analysis*, which we touched upon earlier, assesses emotions. Also included among semantics tasks are *word sense disambiguation*, which decides the intended meaning of a word with multiple possible meanings, *relationship extraction*, which establishes relationships between objects, *recognizing textual entailment*, which deals with how fragments of text affect each other's truth or negation, and *topic segmentation*, which breaks texts down into topical fragments.

Discourse: Discourse tasks include *discourse analysis*, which establishes the role that sentences play in larger blocks of text with reference to each other, *coreference resolution*, which determines which words (or "mentions") refer to the same objects, and *automatic summarization*.

Speech: Speech tasks involve the two opposing processes of *speech recognition* and *text-to-speech*, which convert speech to and from text respectively; the former is much more challenging for natural language processing systems. Speech recognition includes a sub-task called *speech segmentation*, which separates speech into sequences of intelligible words.

If you are interested in further details, this guide contains codes for the <u>top 10 common natural language processing tasks.</u> For programmers looking to implement elements of text processing, the following are open-source tools for natural language processing tasks:

- **Natural Language Toolkit (NLTK)**: This provides text-processing libraries for tasks like classification, tokenization, stemming, tagging, parsing, and semantic reasoning.

- **Stanford's CoreNLP Suite**: This performs part-of-speech tagging, named entity recognition, parsing, coreference resolution, and sentiment analysis, among other things. The system was also designed from the start to work with multiple languages, overcoming hurdles of differing grammar and syntax.

- **Apache OpenNLP**: This also does tokenization, sentence segmentation, part-of-speech tagging, named entity extraction, chunking, parsing, language detection, and coreference resolution.

- **MALLET**: From the University of Massachusetts Amherst, this is a more advanced set of tools for document classification, cluster analysis, topic modeling, and information extraction.

Challenges in Natural Language Processing

Natural language processing is still maturing. Speech recognition has proven to be the most difficult task to execute accurately because of humans' tendency to speak imperfectly, run words into one another, mumble, and disregard grammar — and that's

before we even begin to discuss confounding factors like accents. All of these factors, of course, impact computers' natural language understanding.

Even in text-based communication, however, interactions with natural language processing systems can be frustrating. AI natural language processing capabilities look great in advertisements and carefully orchestrated demonstrations — remember Mark Zuckerberg's viral introduction to Jarvis? But, they're underwhelming in the real world. Facebook, for example, reported last year that its virtual assistant named M, deployed on its Messenger platform, could fulfil less than a third of user requests without human agents intervening. One of its weaknesses proved to be its inability to fully understand the nuances of natural language. Facebook said in early 2018 that it would discontinue M.

Part of the problem with natural language understanding is that intended meanings can be difficult to infer. The contextual awareness that is second nature for people can be an insurmountable obstacle for a machine trying to understand humans. Ambiguity, which is rarely a problem when we converse with other humans, can cripple our efforts to communicate with machines.

How to Integrate Natural Language Processing into Your Business

You may be eager to leverage natural language technology in your business. Before you embark, however, do a reality check and make sure your organization is ready. Introducing this new technology requires a culture willing to embrace change, an ability to handle workflow disruption, and the resources (time, staff, and money) to manage new IT initiatives.

You'll also need a very clear idea of how you want to deploy natural language processing systems. Businesses have typically had the most success by starting with a targeted and clearly defined application. Once you have worked out any kinks, you can build on that success and expand into other areas.

IC-Infographic-of-Natural-Language-Processing-Applications-c.jpg

How Natural Language Processing Is Applied

Machine Translation

Natural language processing breaks language barriers with automatic translation but still struggles to preserve meaning, handle idioms.

Blocking Spam

Filters use natural language processing to keep out junk emails by extracting meaning and comparing frequency of certain words to non-spam emails. Also seek to make sure wanted emails are not filtered.

Detecting Emotion

Natural language processing analyzes sentiment to gauge emotional state, take the pulse of consumer moods and form review snapshots.

Information Extraction

Natural language processing helps power rapid algorithmic trading in financial markets by extracting key information from deal news and making it usable by trading programs.

Question Answering

Search engines historically have struggled to provide relevant answers to questions posed in natural language. Most people prefer to seek information this way, so natural language processing is being applied to the challenge.

Chatbots

Computer programs that can simulate human conversation are used to help people perform simple tasks or ask routine questions.

Here are some popular starting points:

- **Customer Service**: Chatbots can add to your customer service capabilities by answering routine questions and handling simple requests. They can help make the quality of customer service more consistent and free up agents to focus on more complicated needs. Chatbots have the advantage of being able to work 24/7. The

are also cost effective and never lose their tempers. Chatbots work best in businesses that offer a single type of product or service, such as airlines or florists. On the downside, however, they have limitations and can frustrate customers, especially those who are highly emotional or have nuanced questions. Therefore, you should not expect chatbots to replace humans.

Chatbots can be deployed internally as well to answer questions and perform tasks. This enables users who do not have programming experience to use organization systems with self-service, which pull answers from a number of sources, such as relational databases, RESTful APIs, and search engine results.

- **Sentiment Analysis**: Sentiment analysis uses natural language processing to provide structured, quantifiable data on how people feel. It usually searches for emotional content by combining speech analysis of things like customer emails with monitoring social media posts about your business. This action automates and speeds up the process of customer feedback collection and analysis, which allows brand managers to be much more responsive to the ebbs and flows of customer opinion.

  Using natural language processing this way, managers can track the impact of customer outreach efforts efficiently. But perhaps the most compelling aspect of sentiment analysis is the ability to see not just what people think about your brand, but also what they think of your competitors.

- **Information Extraction**: Using natural language processing systems to extract information enables you to quickly gather and collate relevant information. *Text mining* uses algorithms to discover meaningful information, trends, and patterns in unstructured text. Specific tasks include entity extraction, fact extraction, relationship extraction, text categorization, and clustering. (*Clustering* is the technique of organizing a collection of documents by finding documents that are similar or related to one another.)

  Information extraction can improve business decision making because it makes a vast quantity of information accessible and analyzable that wouldn't be accessible manually. In business transactions where speed is key — such as stock trading decisions — the increased breadth and depth of information that can be digested amounts to a big advantage.

  Information extraction is also part of sentiment analysis.

- **Semantic Search**: The last application involves on-site semantic search, which is a type of smart online search powered by natural language processing technology. Unlike keyword searches, which generate search results based on keyword matching, semantic searches are able to identify what search queries actually mean.

  This capability generates more relevant results, gradually eliminating those that

users aren't interested in and mitigating the effect of misspellings. It helps customers find more valuable answers to their questions and decreases the likelihood that they will leave websites without finding what they are looking for. Search results are also a valuable source of data: Among other things, they can tell you what customers are looking for and why they're looking for it. This data can then be used to personalize the on-site experience, perhaps by offering product recommendations based on customer search and browsing habits.

Semantic search can be very powerful for retail companies when combined with speech recognition. This is what Amazon's voice-controlled smart speaker Echo does. People who bought Echo increased their spending by 10 percent, a 2016 study by NPD Group found.

The History of Natural Language Processing

The concept of natural language processing is actually several hundred years old. The philosopher Descartes proposed a sort of machine translation that could relate words between languages. It wasn't until the mid-1930s, however, that the first patents for translating machines were recorded. One of these machines, designed by Georges Artsrouni, was simply an automatic dictionary. Another, invented by Peter Troyanskii, stretched toward grammatical understandings of language as well.

In 1950, when Alan Turing published his famous "Computing Machinery and Intelligence," the idea of natural language processing in its modern form emerged. Turing's article proposed what came to be called the *Turing test*: a method for judging whether a computer program could impersonate a human being in conversation successfully enough that a person could not tell whether they were talking to a human or a machine.

In 1954, the so-called "Georgetown experiment" was a landmark demonstration of machine translation by IBM and Georgetown University. The effort involved translating 60 sentences from Russian to English. But, the sentences themselves had been carefully chosen and did not constitute a representative sample of actual speech, so excitement proved premature.

Other natural language processing advances that seemed to work well in restricted conditions included the following: Daniel Bobrow's 1964 program STUDENT, which could solve simple algebra word problems; Joseph Weizenbaum's ELIZA program in the mid-1960s, which was designed, ironically, to show the superficiality of human-computer "conversations;" and Terry Winograd's SHRDLU computer program in the late 1960s, which facilitated manipulation of a virtual world of blocks. The 1970s saw the popular emergence of *chatterbots* (now called chatbots) as programmers began to write "conceptual ontologies" that essentially transformed real-world data into structured forms that computers could understand. These included Roger Schank's MARGIE (1975). Rollo Carpenter's Jabberwacky, which was developed in the 1980s and 90s, was a conversational chatterbot designed to be interesting and funny.

No discussion of the history of natural language processing would be complete without mentioning key figures who shaped its development.

- David Ferrucci, IBM Principal Investigator, led a team of researchers and engineers to develop the Watson computing system that won *Jeopardy*! in 2011.

- Dan Jurafsky, the co-author of *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*, also developed the first automatic system for a natural language process. That system is called *semantic role labeling*.

- Victor Yngve, a physicist with a passion for machine translation, created the first major programming language for text processing, known as COMIT.

- William Aaron Woods built one of the first question-answering systems for the NASA Manned Spacecraft Center, where it answered questions about the Apollo 11 moon rocks.

- Stephen Wolfram developed the famous computational knowledge engine Wolfram Alpha, based on natural language processing.

What Is Natural Language Processing in AI?

Artificial intelligence (AI) is a term coined by Stanford University researcher John McCarthy in 1956. AI describes computing systems that can think and learn much like people do. AI researchers attempt to build systems that replicate human thought processes and actions.

Machine learning is part of artificial intelligence, and machine learning algorithms revolutionized natural language processing In the late 1980s. In machine learning, computers use statistical methods to "learn" on their own by being exposed to new or different data without direct programming. Prior to machine learning, natural language processing systems were based on rules laboriously defined by people. But, increases in processing power and the decline of influential linguist Noam Chomsky's theories paved the way for smarter machines. Chomsky had discouraged the machine-learning approach of studying language in large samples of real texts.

Early machine-learning algorithms used decision trees, which really led to the same sort of hard if-then rules that had previously been written by hand. But, the field moved towards statistical modeling using techniques like Hidden Markov Models, which are statistical models that convert speech to text by performing mathematical calculations to determine what you said. This led to systems better equipped to understand language in natural forms. As statistical models became more advanced, thanks largely to IBM, early successes in machine translation proliferated.

Compared to the hard if-then rules created by handwritten coding — and by early iterations of machine learning — these statistical techniques allowed for soft probability-

based decision making, hence expressing the relative certainty of multiple possible answers.

These techniques relied on inferring grammar and syntax from large bodies of real-world texts, such as the documents produced by the Parliament of Canada and the European Union in multiple official languages.

More recent developments have tended to favor *semi-supervised* or *unsupervised learning* techniques — that is, the partial or complete use of non-annotated data sets instead of those where desired answers have been indicated. These techniques are facilitated by the availability of vast amounts of analyzable information — a phenomenon called *big data* — and increased computing power. *Deep learning* techniques in particular have achieved promising results in natural language processing tasks. Deep learning is a branch of machine learning in which algorithms are patterned after the structure of the human brain. These algorithms are called *artificial neural networks*.

Other approaches include *reinforcement learning*, which enables machines to create their own languages by giving them the ability to communicate and placing them in "worlds" where they must achieve goals that are most effectively pursued cooperatively. To achieve this approach, intelligent agents (autonomous artificial intelligence entities) develop grounded language. A *grounded language* is one in which our understanding of a word derives from our interactions with the physical world. (This is the opposite of a dictionary, which defines words in terms of other words.)

Four Approaches to Natural Language Processing

Stanford University's Percy Liang, a natural language processing expert, says there are four main approaches to natural language processing: distributional, frame-based, model-theoretical, and interactive learning.

To compare the four types of approaches, it helps to understand the three levels of linguistic analysis:

- **Syntax**: This deals with the grammatical structure of text.

- **Semantics**: This deals with what text is supposed to mean.

- **Pragmatics**: This relates to the purpose of the text.

Following is a description of the four main approaches:

> **Distributional Approaches**: These approaches involve the large-scale statistical tactics you see in machine learning. They rely on turning content into word vectors and perform strongly on tasks like part-of-speech tagging, dependency parsing, and semantic relatedness. (*Semantic relatedness* refers to tasks that don't need to understand what words mean but simply how the words are related to each other.)

While distributional approaches are flexible enough to be applied widely to texts of different types and lengths, they're weak at understanding semantics and pragmatics.

Frame-Based Approaches: These involve the use of frames, which are structures that represent what cognitive scientist Marvin Minsky called "stereotyped situations." An easy example of a frame is an accusation from the murder mystery board game Clue, where you accuse a specific character of committing a murder with a specific murder weapon in a specific room. For each accusation — that is, each *frame* — you have a murderer, a murder weapon, and a murder location. Regardless of how the accusation is phrased grammatically, it fits semantically into the same frame because it conveys the same information. However, frames call for supervision, and in some domains they must be created by experts. Moreover, since frames only detail specific situations, information outside the parameters of the frame cannot be analyzed, so frame-based approaches can be incomplete.

Model-Theoretical Approaches: These combine two concepts in linguistics: *model theory*, which is the idea that sentences refer to the real world, and *compositionality*, which is the idea that you can combine the meanings of the various parts of a sentence to deduce the whole meaning. Liang says this approach is like using language as a computer program. For example, to answer the question "Which is the cheapest model produced by the car maker XYZ?," the concepts of model and car maker XYZ would have to be identified, and a search would have to be created and filled with all models created by car maker XYZ. The resulting list of objects would have to be sorted by price, and the cheapest model returned as the answer. The amount of supervision required for model-theoretical approaches ranges from heavy to light. These methods are strong with semantics, can represent the full real world, and feature end-to-end processing. However, since their features must be engineered by hand, these approaches are limited in scope and need narrow use cases.

Interactive Learning Approaches: These approaches hold exciting promise for teaching natural language processing systems to understand language via interaction with human beings. To do this, a human instructs a computer to perform specific simple tasks using syntactically consistent instructions. The person then tells the computer what the outcome should be from carrying out each instruction. Liang, for example, created a modern-day version of Terry Winograd's SHRDLU, called SHRDLRN, which involves a world populated by Lego-style colored blocks that the computer must manipulate based on user instructions to achieve a specific end-state. With enough practice, a computer can learn to associate words with colors or positions — and this can be done in any language, so long as consistent syntactic forms are employed.

If you're interested in an extensive discussion of natural language processing, you can watch Percy Liang's full 91-minute talk here.

Google's Impact on Natural Language Processing

Google is a leader in natural language processing, and its use of these techniques show how far reaching their impact can be. Google's research focuses on wide-ranging algorithms that apply at scale, across languages and domains. Google deploys natural language processing across a number of its core technologies and services, including its search and translate functions and ads.

Let's use Google's famous search engine as an example. The company applies natural language processing to evaluate the vast universe of online content. Google wants to give you the most helpful resource first when you search for something, so the findings of its natural language processing system influence where a web page will rank in the search results that Google returns.

Specifically, Google's natural language processing systems analyze the syntax of an article and look at its structure, including the structure of its sentences and use of nouns, verbs, and other parts of speech. It evaluates whether the content is grammatical and whether the author used language properly. The system also assesses what audience the content is appropriate for (e.g, scientists or elementary school students). Using entity recognition, natural language processing enables Google's systems to understand what is inside images and videos.

These techniques analyze the emotions in content, such as whether reviews are positive or negative and how intensely those emotions are felt. They look at content to determine whether it has an emotional impact and how people react to it. Natural language processing helps Google analyze why a blog post about a puppy that saves a child is so powerful for readers.

Google's natural language processing systems are built on e syntax and semantic algorithms. Syntactic algorithms perform tasks like part-of-speech tagging, morphological segmentation, and parsing. Semantic algorithms perform named entity recognition and coreference resolution, the task of finding every expression that refers to the same entity in a text. The company says it "focus[es] on efficient algorithms that leverage large amounts of unlabeled data" and that it is "interested in algorithms that scale well and can be run efficiently in a highly distributed environment."

The Future of Natural Language Processing

Natural language processing has a long way to go, but it is an area of intense research because it is critical to making AI easier for people to use. Natural language processing can help improve what we might think of as *user experience* — how easy and intuitive it is to communicate with AI systems. Natural language processing has a major role here. If AI systems understand you like a human would — and that includes not being stumped by

technical imperfections in speech — they become much easier to deal with. This ability, in turn, will make us more willing to use and interact with AI and will speed the proliferation of AI technologies.

As we make advances in natural language processing, here's what the future likely holds:

- **Better Chatbots for Customer Service**: Faster bots with greater functionality will handle more tasks and interact with customers more seamlessly. Chatbots will become more valuable to businesses as they get better at understanding customers' emotional states and needs.

- **More Natural User Interfaces**: Technologists aim for us to be able to use machines with an invisible user interface. This means that it will feel like we are interacting directly with these systems rather than issuing commands or pushing buttons. These systems would leverage natural language processing to understand what we say or write no matter how we express it or what we are doing.

- **Greater Emphasis on Natural Language Generation**: So far, systems have focused mostly on understanding our everyday language. But as natural language processing gets more sophisticated, the focus will shift to making systems that can communicate with us more naturally.

- **Deeper Understanding**: Despite the advances in natural language processing, it still falls painfully short much of the time. Users of Google's Translate app are often stumped or amused by the contorted and unnatural translations that it provides. This is because natural language processing does not encompass the deep understanding and subtleties that humans bring to language, largely because a person's perceptions are shaped by three-dimensional interactions with the real world, not algorithms. This is one of the great frontiers for natural language processing researchers.

- **Efficiency Aids**: While this horizon is long, natural language processing will continue to increase our efficiency with things like medical diagnostic assistance, virtual health assistants, and interactions with devices connected by the Internet of Things (IoT).

Exciting Startups in the Natural Language Processing Space

Startup companies are doing a lot of exciting work in natural language technologies. Here are some of the interesting players currently in this space:

Klevu: It provides a natural language processing-powered e-commerce site search tool that is capable of learning. Klevu is geared for small and medium-sized web stores, and is used by thousands of online stores around the world.

EnglishCentral: This is an online English language learning platform that provides what it calls English conversation solutions, teaching users how to use English words

in context.

Yummly: This is an app and website with a semantic recipe search function that allows users to search by ingredient, diet, nutrition values, price, cuisine, time, taste, and allergy. It also learns about users' likes and dislikes.

Vurb: This is a social activity search engine, recommendation generator, and social network that Snapchat acquired in 2016 for $200 million.

Insight Engines: This company creates smart search assistants with natural language processing capabilities that make data widely accessible across an organization.

MindMeld: This is a platform for building intelligent conversational interfaces that work across applications and devices. It is used to power voice and chat assistants. The company is a recognized leader in natural language processing.

Desti: This is a smart travel planner that allows users to plan trips without complicated search interfaces. Because Desti is powered by natural language processing, its users can search using multiple criteria besides dates and prices.

MarketMuse: A platform for digital marketers, it helps create high-quality content by examining topic relevance, identifying subpar content, and identifying topics that haven't been covered.

Kngine: This is a simple question-answering engine designed to provide direct answers to questions.

Agolo: This is a content summarizer that is faster and more wide ranging than a human being. It's meant to facilitate strategic decision making.

AddStructure: It provides a semantic product search technology that is designed to understand search criteria in a conversational form.

- **NetBase**: This is a social media analysis tool that provides real-time insights. NetBase gauges customer sentiment and monitors brand reputation.

- **Inbenta**: It provides a number of natural language processing-powered customer service solutions, including chatbots and an intelligent search function.

Free and Helpful Resources on Natural Language Processing

If you want to learn more about natural language processing, there are a lot of resources available, some of them free or low cost. These include the following courses, lectures, platforms, slides, and books:

- There are many online courses, such as <u>this one</u> by the Higher School of Economics and <u>this course</u> from Stanford University on machine learning by noted researcher Andrew Ng.

- The University of Michigan's <u>video course</u> on natural language processing is good for relative beginners and has short segments on key topics.

- A technical <u>series of 18 videos</u> by Stanford University lectures on natural language processing with deep learning.

- The <u>Natural Language Toolkit</u> is a platform for Python programs that work with language data. It offers tools and a discussion forum as well as an opportunity to try your hand at natural language processing.

- These are the slides for <u>Oxford University's deep natural language processing course</u>.

- <u>Jurafsky, Daniel and Martin, James H. *Speech and Language Processing (2nd Edition)*</u>

- <u>Manning, Christopher D. and Schutze, Hinrich. *Foundations of Statistical Natural Language Processing*</u>

- <u>Bird, Steven. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*</u>

- <u>Ingersoll, Grant S., Morton, Thomas S., Farris, Drew. *Taming Text: How to Find, Organize, and Manipulate It*</u>

Discover the Power of Natural Language Processing with Smartsheet

Whether you're ready to implement natural language processing techniques or simply want to increase efficiency in any area of your business, it's evident that streamlining repetitive, manual tasks can lead to increased time savings and productivity. One tool that can help automate these tasks is Smartsheet, an enterprise work management platform that fundamentally changes the way teams, leaders, and businesses get work done. Over 74,000 brands and millions of information workers trust Smartsheet as the best way to plan, track, automate, and report on work.

You can build automated business processes without a single line of code, complex formulas, or help from IT. Achieve faster progress by creating automated approval requests and automated update requests that are triggered based on preset rules. Use Smartsheet to automate and streamline the following processes: time card tracking, sales discounts, procurement, HR hiring, content approvals, and more. Plus, Smartsheet integrates with the tools you already use, to seamlessly connect your efforts across applications.

https://youtu.be/4EqkIFD4n98

Get from idea to impact, faster, by building powerful, automated business processes in Smartsheet.

Try Smartsheet for Automating Business Processes