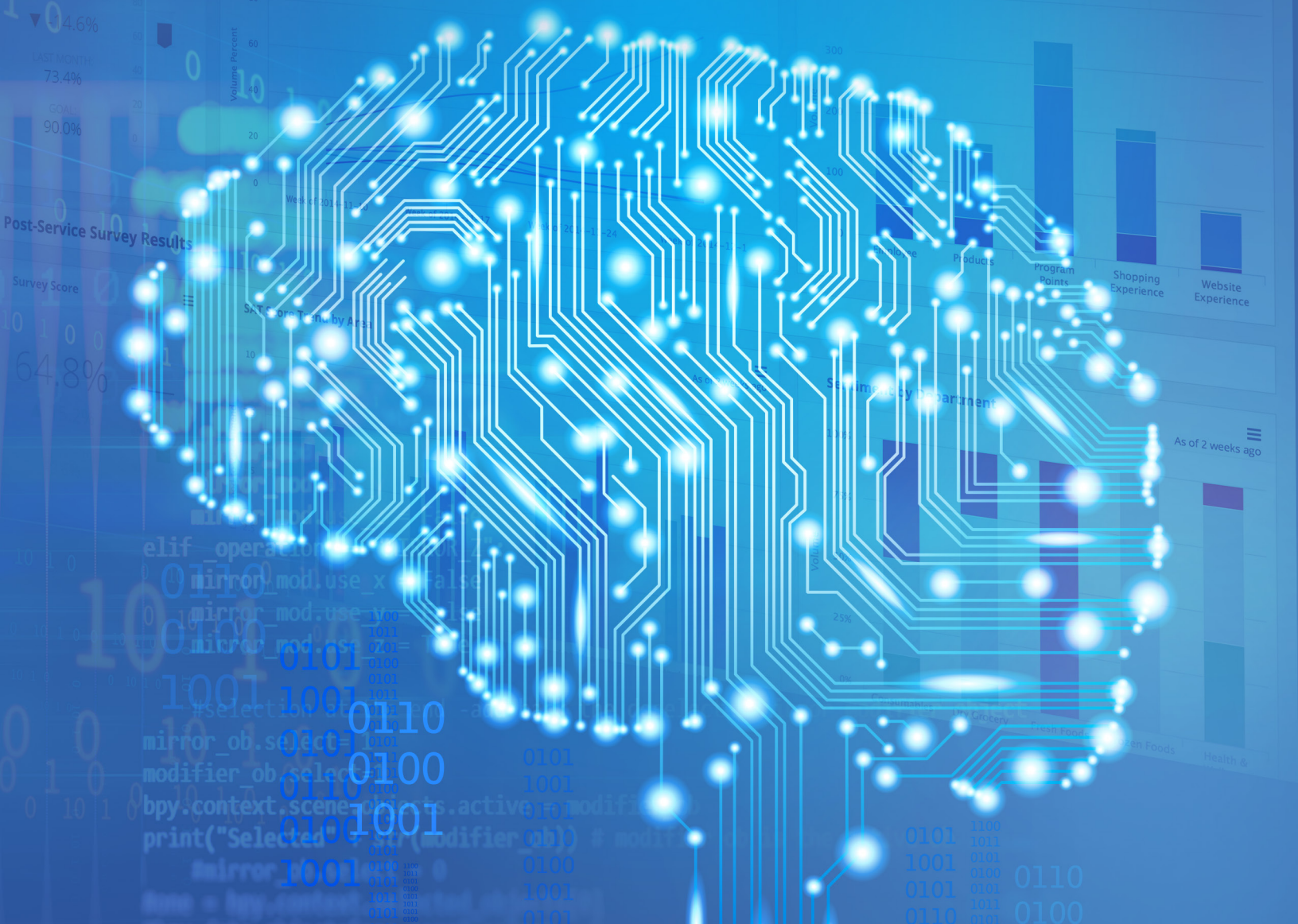


# DEBUNKING NATURAL LANGUAGE PROCESSING

AN INSIDE LOOK AT THE ENGINE THAT POWERS TEXT ANALYTICS AT CLARABRIDGE

A CLARABRIDGE EBOOK

by Ellen Falci Loeshelle,  
Director of Product Management, Clarabridge



# Table of Contents

**Introduction .....1**

**Part 1: Understanding the Meaning of Words ..... 3**

Recognizing Named Entities.....3

Detecting Products, Brands, Companies and Industries.....6

Detecting People, Phone Numbers and Email Addresses .....9

Detecting Events.....12

Emojis and Emoticons .....14

**Part 2: Going Beyond Words—Understanding the  
Meaning of Sentences and Documents.....17**

Detecting Intent and Actionable Sentences.....17

Analyzing Emotion and Effort.....20

Detecting Spam.....23

**Part 3: Important Linguistic Considerations .....25**

Translation .....25

Sarcasm .....28

**Conclusion .....30**

About the Author.....31

About Clarabridge.....32

# Introduction

When I started working at Clarabridge six years ago, natural language processing (NLP) and text analytics were very technical-sounding terms. My family and peers were baffled when I tried to explain the field that I worked in. Since then, though, the term NLP has experienced a rebirth as marketers have tried to capitalize on the perceived sexiness and sophistication of related technologies.

While it still preserves a very specific meaning in academia and research organizations, NLP has been watered down in the customer experience (CX) space and often just refers to lists of words and topics. This change in semantics is unsurprising. As NLP toolkits like Stanford NLP, NLTK, spaCy, IBM's Watson and Google's SyntaxNet

become more consumable, they are quickly divorced from their technical origins. One no longer needs to be a software developer or a computational linguist to use NLP technologies. These toolkits “blackbox” a lot of their complexity so that users can focus on the outputs, not on the algorithms. However, as a result, many users leverage only a small fraction of these systems' capabilities by focusing on outputs that are easy to consume (e.g., word clouds). In an attempt to recapture the complexities of NLP, some individuals have embraced a related academic term, natural language understanding (NLU), to refer to computational approaches to understanding meaning behind language, instead of just abstract patterns.

Let's be honest: Just grouping data into topics or categories isn't enough anymore. For a while, it was exciting and novel to be able to categorize your data into topics with a few keywords and phrases.

It streamlined old, subjective processes, such as manually reading through and categorizing customer feedback data, and added consistency and efficiencies into CX workflows. Some tools (including Clarabridge) offer features that help you uncover new topics that can then be added into your reporting, but over time there are fewer and fewer truly new categories to add. We've reached the long tail. It's useful to monitor these categories, but there is a limit to what they can truly accomplish for you. Simply analyzing volumes of topics and their associated sentiment is insufficient to find actionable and meaningful insights.

### Going Beyond NLP

In order to deliver on NLU, a system must have mature NLP capabilities. At Clarabridge, we've spent over ten years developing our proprietary NLP engine, which has afforded us an extremely strong linguistic foundation upon which we can build value-added features. You may have noticed a shift in our attention over the past few years toward development of features that help you understand your data in new ways, with less effort. We are bullish about the value in leveraging semantic analysis, emotion analysis, effort analysis and more on top of topic analysis—but not in place of it. For example, imagine if you could automatically find all suggestions from your customers, isolate every time a customer praised an employee or even find all instances

of high customer effort. We advocate for layering new types of language analysis to expose new kinds of insights. The more layers, the more potential insights.

In this ebook, I will be highlighting the features, functionality and services that help Clarabridge users go beyond simple topic and sentiment analysis. We will explore each area in-depth to show how these new methods and types of analysis aren't so far-fetched after all.

**We are bullish about the value in leveraging semantic analysis, emotion analysis, effort analysis and more on top of topic analysis—but not in place of it.**

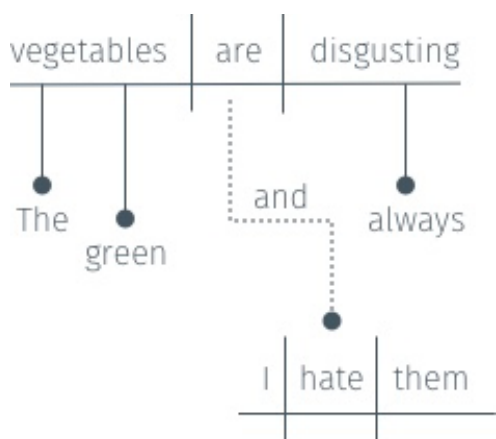


# PART 1

## UNDERSTANDING THE MEANING OF WORDS

### Recognizing Named Entities

When we think of grammar, some of us are still haunted by nightmarish flashbacks of diagramming sentences in middle school. A mess of lines and dashes, the diagrams were intended to help us understand how words related to each other.



Source: WikiHow - How to Diagram Sentences

As a self-proclaimed language nerd even as a teen, I loved diagramming sentences (thanks, Mr. Dopko!), but in reality doing so carries very little value in communicating effectively with my peers. Beyond parts

of speech, noun clauses, subjects and predicates lies a world of complex meaning embedded in words and phrases. Putting words together in the proper order is a skill we master as toddlers, but truly understanding what words mean requires a lifetime of focus. The same pattern is true of an NLP engine. The rules of English (or any language) are relatively straightforward to teach a machine. Understanding the subtleties of meaning, on the other hand, is fraught with challenges. Take, for example, this sentence:

"There's a red and green striped jaguar in my garage."

How would you interpret it? This sentence may refer to a uniquely colored feline or a one-of-a-kind automobile. Both of those interpretations are valid, but each would result in a tremendously different reaction from an audience. Let's look at a slightly different version of this sentence:

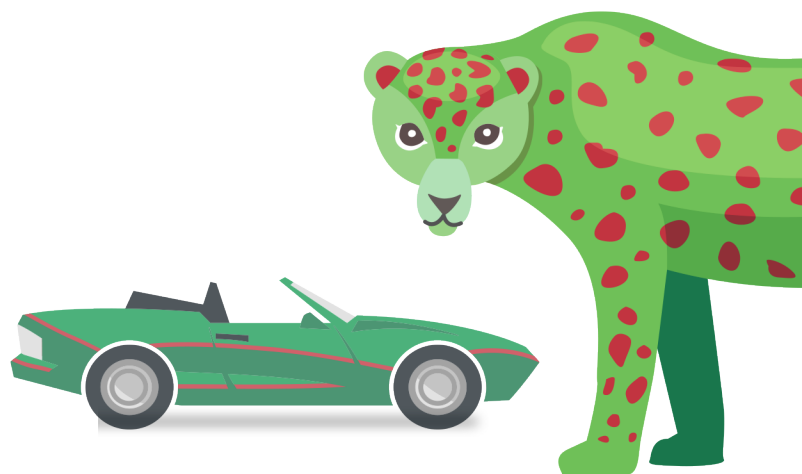
“There’s a red and green striped Jaguar in my garage.”

This small orthographic change (capitalizing the J in “jaguar”) pushes the interpretation toward recognizing it as a brand name instead of a mammal.

Named Entity Recognition (NER), or the ability to deduce whether a word, or sequence of words, is a proper noun, is a well-studied, challenging subject in the language community. There are both linguistic and statistical approaches to tackling the problem. For example, we can teach the NLP engine to look for small clues in capitalization, the presence of suffixes (e.g., LLC, Inc., etc.), changes in punctuation and part of speech sequences. Machine learning systems can be trained to spot certain patterns as well. These strategies perform decently well on well-formatted text; however, when these essential clues are missing (which is often the case in social media data), it becomes nearly impossible to disambiguate meaning accurately.

Unfortunately, spotting these terms within a sentence is only half the battle. There are a multitude of kinds of named entities. Within your dataset, you might see products, brands, companies, people, events and locations. As you can imagine, being able to analyze any of these groups of terms in customer experience

“There’s a red and green striped Jaguar in my garage.”



feedback would provide added value on top of a list of the topics in your data. You probably already know the top ten topics in your survey data by heart. But could you tell your stakeholders the top ten products mentioned? Top ten employees? Celebrities? Cities? Brands? Looking at your data through the dimension of named entities opens up a whole new world of analysis in CX data.

I particularly love that using named entities to drive analysis starts to expose unexpected truths. So many of the entities mentioned have nothing to do with your brand and would not have been noticed if you were only looking for your direct

brands or competitors. For example, when analyzing data about the vintage soda Surge, we found a huge surge (pun intended) in mentions of eBay as Surge enthusiasts were trying to buy and sell bottles in online auctions. We worked with a major US airline that received significant backlash for switching the brand of coffee served on board. Another online retailer faced negativity when its online payment system was suddenly incompatible with PayPal. Our use case list is chock-full of these kinds of stories in which one company's issues were actually based on biases (or preferences) for other brands or individuals that were completely unrelated to their own. If these companies only looked at their known topics and their own list of products, they would have been completely blind to these insights from tangentially related organizations, products or brands.

While topic analysis is commonplace—and I'd argue required—to be competitive in the customer experience analytics marketplace, the ability to analyze named entities is actually rare. In order to offer this kind of semantic analysis, a tool must have a very mature NLP foundation that serves up enough information at the part of speech and grammar levels to give a Named Entity Recognition module a fighting chance.

**In order to offer this kind of semantic analysis, a tool must have a very mature NLP foundation that serves up enough information at the part of speech and grammar levels to give a Named Entity Recognition module a fighting chance.**

# Detecting Products, Brands, Companies and Industries

In the last section, we talked about the value of named entities in customer feedback data and the challenges that NLP engines face in identifying them in text. Finding named entities in your customer feedback data opens the door to new kinds of business questions and answers.

However, finding them in text is only the first step. The next challenge is determining what type of named entity you have at hand—whether it’s a product, brand, company, industry or another type of proper noun.

Take the word “Georgia,” for example. Georgia could be a state, a country, a city, a person, a brand of coffee, a university, a movie, a song or even a battleship. It is often difficult to differentiate which one is intended, even for humans. “I love Georgia” could legitimately be referring to any of those items. For some NLP systems, the distinction of one named entity type to another is irrelevant. But for a customer experience NLP engine like Clarabridge’s, the distinction carries importance as it could completely change how a business understands its customers’ needs and preferences.

In order to differentiate these types, Clarabridge combines linguistic approaches and knowledge-based approaches. We look for clues such as the presence of a suffix (e.g., LLC, Inc., PLC, etc.) that would indicate if an entity is an organization or the existence of a version number that would indicate a product. We combine those clues





with dictionaries of known products, brands and organizations and tie it all together with an out-of-the-box hierarchy. Therefore, if a consumer mentions trying to integrate his or her iPhone with a car, we will also reflect it as a reference to the Apple brand, Apple Inc. and the Computer Software industry.

In academia, there are four standard types of named entities: persons, locations, organizations and products. However, when Clarabridge approached the challenge of interpreting named entities, we found that this four-type classification was too narrow to meet the needs of those analyzing customer feedback.

So, we started by expanding the dual classification of product and organization entities to a four-part hierarchy that linked products to brands to companies to industries. Why? Many of our customers and their competitors own multiple products in multiple brands that roll up to a single organization. The understanding of brand was an essential tier for analysis and product improvement.

Take General Mills, for example, which owns dozens of brands under the General Mills label. The General Mills team would want to track Yoplait, Pillsbury, Larabar and Old El Paso, to name a few. Most consumers would not even mention “General Mills” when talking about their moldy Yoplait yogurt or their delicious Larabar. We had to bridge

the gap between what consumers say and how businesses think about their offerings.

Here at Clarabridge, we feel strongly that we should put our NLP to work for you, which is why we have four attributes that automatically populate with business named entities every time you load data: Product, Brand, Company and Industry. These attributes provide immediate value as you can determine which are the top products or brands mentioned in your data within seconds of the data loading—no category model or keyword search required!

How are these semantic attributes best leveraged? Here are a few key use cases for these features:

### 1. Competitor Analysis

Search your data for mentions of your competitors. Look at corresponding volume and sentiment to assess market share and health. Consider these mentions in conjunction with a customer journey topic model to determine areas in which your

**Here at Clarabridge,  
we feel strongly that  
we should put our  
NLP to work for you.**

competition shines and areas in which it fails. Or, search for sentences that mention both you and your competition comparatively. Armed with this information, you can design your business strategy to capitalize on your advantages.

### 2. Brand or Product Analysis

Identify references to your brands and your products in your data. Consider these mentions using sentiment, emotion and effort analysis to determine how your customers relate to your offerings. Look at these mentions with filters for “suggestions” or “requests” to identify areas of opportunity to improve your offerings.

### 3. External Brand Analysis

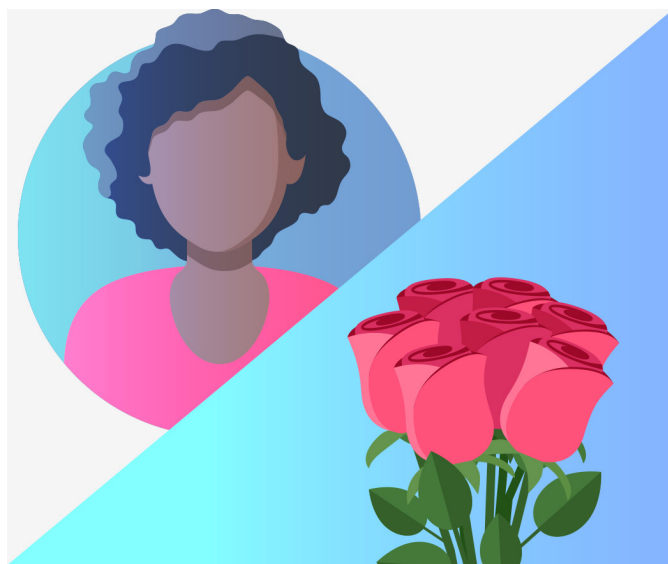
Your customers’ worlds are bigger than just you (sorry to burst your bubble!). Identify mentions of other non-competitor brands in your data that may be affecting your customers’ experience with your offerings. Cross-analyze these with mentions of effort or top topics to find intersections between offerings. Seek to capitalize on your customers’ preferences or biases for other brands that will help them adopt your products more seamlessly.

We’ve found great value in the above-mentioned use cases, but we are continually impressed by the new ways in which our customers use them. The list above only scratches the surface.

**Your customers’ worlds are bigger than just you (sorry to burst your bubble!). Identify mentions of other non-competitor brands in your data that may be affecting your customers’ experience with your offerings.**

# Detecting People, Phone Numbers and Email Addresses

“A rose/Rose by any other name would smell as sweet”—or would she? Was Juliet simply referring to the flower or was it a subtle reference to a woman? Another lover, perhaps? A literary scholar or even a high school English student would surely conclude the former. However, if you give this task to a computer, the answer may not be so obvious.



In the last section, we discussed the challenges that machines face in trying to detect the subtleties of named entities in text and the importance of being able to

interpret them. The difference between generic and specific versions of brand or product names like “apple” and “Apple” is significant, and the same is true for generic nouns and human names like “rose” and “Rose.” In the customer experience industry, it would be misleading to misinterpret a name like “Bill” as a statement of money owed and catastrophic to misinterpret a name like “Sue” as a legal action. I don’t think it’s an overstatement to say that the ability to both identify and classify these words correctly is mission-critical in customer experience management.

The four-part classification of named entities common in academia (persons, locations, organizations and products) doesn’t quite align to the needs of a customer experience program. In order to help our users answer questions about their businesses, Clarabridge offers four hierarchical business entities as described in the previous section: Product > Brand > Company > Industry. Similarly, we aligned an expanded set of functionality around the named entity classification for person. Why? Two reasons. First, the success and failure of every organization is tied to its

employees and second, the livelihood of every organization is tied to its customers. Without a doubt, **people** are the essential cogs in the business workflow; we designed our Person detection functionality to cater directly to this keystone.

Clarabridge offers three attributes that populate automatically when you load data that assist in identifying individuals: Person, Phone Number and Email Address. These attributes provide immediate value as you can determine which of your employees are mentioned in your data as well as identify which customers are seeking engagement by providing their contact information.

Prior to the release of these features, identifying these kinds of attributes was extremely difficult in Clarabridge. Users

would add the top male and female names from the most recent census to a category node. But, as you can imagine, names like Mark, Sue, Will, Bill, April, May, June, Ray, Angel, Guy, Bob, Virginia, Rose, Ruby, Grace, Dawn, Amber, Joy, Terry, Penny, Kay, Violet, Daisy and Barb caused a real headache! Similarly, there was no simple way to query for all of the different variations in the structure of phone numbers or to get a list of all email addresses mentioned. The power of a mature NLP engine is that it can use techniques that go far beyond simple keyword matches to find these kinds of entities. Our approach, which combines both lists of known names and language-specific linguistic rules, yields higher precision and higher recall of detected names and contact information.

How are these semantic attributes best leveraged? Here are popular use cases for these features:

### 1. Employee Praise and Critique

Discover which of your employees are mentioned in your data using the Person attribute. The praise or criticism that customers provide about an associate can be used to reward or modify performance internally. The Person attribute is highly valuable for operational use cases where “on the ground” employees frequently interact with customers.

**Without a doubt, people are the essential cogs in the business workflow; we designed our Person detection functionality to cater directly to this keystone.**

### 2. Customer Engagement

Quickly identify customers who are seeking engagement with your brand by finding all records that contain contact information using our Phone Number and Email Address attributes. Customers may or may not be explicit when soliciting a response. Picking up on subtle cues like the presence of a phone number or email address in text and then proactively contacting the customer is sure to make a positive impression.

### 3. Influencer Identification

Not only will Clarabridge pick up on the names of your employees, it will also find names of other individuals in the data. For example, it's not uncommon to see names of performers, politicians, business executives and other celebrities bubble up in reports for Person. It can be incredibly insightful to discover these external influences on your customers' perceptions. When analyzing Las Vegas casino reviews, we found lots of negative mentions of Elton John. Turns out that he had been canceling lots of shows and customers were not pleased!

The above-mentioned use cases have led to valuable insights for our customers, and we're excited to see how they continue to discover more and expand our use case library.

**The power of a mature NLP engine is that it can use techniques that go far beyond simple keyword matches to find these kinds of entities. Our approach, which combines both lists of known names and language-specific linguistic rules, yields higher precision and higher recall of detected names and contact information.**



# Detecting Events

For many people, September 19 is a *remarrrrrkable* occasion for which pirate memes and puns abound. National Talk Like a Pirate Day, invented in 1995 as an inside joke between two friends after a contentious racquetball game, has become an internationally acclaimed parodic sensation. (September 19 also happens to be my birthday.)

Events bookmark our perception of time and give us a shared cultural experience. Specific events also help keep certain businesses afloat. Many retailers make more money between Black Friday and Christmas than they do the entire rest of the year.

In the previous two sections, we discussed some of the ways in which Clarabridge uses Named Entity Recognition to find meaning within customer feedback. Many customer experiences are dictated not just by specific products, brands, companies, etc. but by the events for which those items were purchased. Events, of course, make up a broad category that includes national observances, religious holidays, cultural events, seasons and life milestones, to name a few. Some

events take place on specific days, some span weeks or months, and each comes with its own traditions and customs.

For those in the customer experience space, understanding consumer behaviors in relation to specific events can be extremely valuable information as it can power development, marketing and commercial decisions. In addition to detecting the named entities described in previous sections, Clarabridge automatically tags events of all kinds. Our dictionary of events is geographically diverse and includes entries for all of the event sub-types mentioned in the previous paragraph, including Diwali, Halloween, Chinese New Year, Easter, Shabbat, The Olympics, The Oscars, Black History Month, back-to-school, weddings,



baby showers, graduations, funerals and more. It even includes emojis! For events that have a dedicated emoji such as 🎄 or 🍁, Clarabridge detects this emoji as an event and maps it to its standard text form for a more complete understanding of customer feedback.

How are these event attributes best leveraged? Here are popular use cases for this feature:

### 1. Seasonal Trends

Track mentions of events in conjunction with discussions of sales and promotions (Independence Day Sale, Black Friday, Cyber Monday) to determine which ones are generating buzz. Filter down these mentions to those discussing availability to determine if demand is outpacing supply at critical times of the year.

### 2. Gift-Giving Opportunities

Determine for which occasions products are being purchased by individuals for themselves or as gifts for others. Many products and brands are favored for certain milestones. Mentions of weddings, engagements, baby showers, graduations and so on may help highlight how best to market and price items to target specific buyers.

### 3. Product Uses

Discover how products are being used by analyzing them in conjunction with

mentions of cultural events or occasions. Understanding that a refrigerator is used for Shabbat or an oven for a Christmas dinner can shape the features that the product should offer and the marketing strategies that are used to get it to the most ideal consumers.

These examples have led to valuable insights for our customers, and we're excited to see how they continue to discover more uses for this feature.

**For those in the customer experience space, understanding consumer behaviors in relation to specific events can be extremely valuable information as it can power development, marketing and commercial decisions.**

# Emojis and Emoticons

In 2015, the Oxford English Dictionary named 🤩, aka the “Face with Tears of Joy” emoji, as its word of the year. This news may have prompted you to feel like 😂, 🤔, 🙄, 😞, 😠 or 😊. Emoticons<sup>1</sup> and emojis<sup>2</sup> are widely attributed as emblems of Internet communication, but, contrary to popular belief, emoticons are actually not a new invention.

At the end of the nineteenth century, Ambrose Bierce, an American editorialist, suggested different ways to manipulate punctuation to better represent tone. He proposed using an open parenthesis flipped on its side to express wry smiles and remarks.

Throughout the next century, many writers hinted at the benefit of having a special set of symbols to indicate emotion, tone or intention, but it remained a fringe movement for over a century. Then, with the advent of the Internet, the emoticon went viral.

Although emoticons and emojis found their first home on social networking platforms, they weren’t homebodies for long. As speakers embraced their utility, these new punctuation marks quickly infiltrated other communicative spaces on both digital and analog platforms. You’ll spot emojis in Amazon reviews, survey responses, private chat dialogues and even handwritten notes. At first, they were substitutes for other concepts; they now have taken on meanings of their own.

Emojis go far beyond faces, though. In 2016, the avocado emoji gave us a new way to communicate our brunch plans. In 2018, the badger, lobster and llama emojis gave us new ways to describe our trips to the zoo or the aquarium. But, there’s more than meets the 👁. Many of these emojis have multiple meanings that add unique dimensions to analysis. In the summer of 2018, we sought to analyze social



<sup>1</sup> Emoticons, a portmanteau of “emotional icons,” consist of punctuation marks, letters or numbers that are used to express an emotional state, such as :) or :(.

<sup>2</sup> Emojis are pictographs of faces, characters, objects, actions and symbols, such as 🤩 or ❤️.

comments about the FIFA World Cup when suddenly we found ourselves confronted by a fierce herd of goats: 🐐. Our brief confusion subsided as we realized the use of the emoji revealed which player each country believed was deserving of the Greatest Of All Time (G.O.A.T.) moniker! Similarly, you might look for 🔥 to determine which products socially hip customers find attractive or exemplary, as in “lit,” or 100 where they are in full agreement with a new idea or campaign. Keep your eyes peeled for innovative uses of emojis; you never know what you’ll find!

In Clarabridge’s NLP engine, emojis and emoticons are treated the same as words. Variations on common emojis or emoticons are mapped to standard tokens. For example, :) and :D both map to :). These tokens can be tuned to contain sentiment. The meaning of most emoticons and emojis are context-sensitive. However, there is a small subset

of unambiguous emojis and emoticons to which Clarabridge assigns out-of-the-box positive or negative sentiment values.

Analysts can choose to analyze emojis in much the same way that they would approach words. Here are the top use cases for emoticons/emojis:

### 1. Emoji Cloud

Create a word cloud of just emojis to explore how customers are reacting to your brand. Emotional faces and other symbols can quickly give you a sense of brand perception and may even clue you in to product uses.

### 2. Correlation with Products and Services

Drill down on specific products, services and experiences to see which emojis are used in conjunction with each one. Or, drill down on a specific emoji to see which products, services and experiences are being discussed. Understanding reactions to these offerings can assist in designing, marketing and promoting your products and services.

### 3. Emoji Trends

Follow specific emojis over time to see how emotions and impressions are changing. Take advantage of a viral trend or get ahead of a major customer issue. Emojis can expose feelings that sometimes are lost in words.

**The Clarabridge NLP engine treats emojis and emoticons the same as words.**

Emojis have become ubiquitous in our digital world. With increased adoption and pervasiveness of emojis in all channels, it's paramount to consider them as full-fledged components of customer feedback. They reveal unique messages and definitely invite a certain degree of playfulness in interpretation of feedback!

This section wraps up our discussion of named entities. I hope you found it eye-opening to explore the value behind these special elements. We will turn our attention next to a slightly bigger space by moving from a discussion of the meaning of words to one on the meaning of complete sentences and documents.

**Emojis have become ubiquitous in our digital world. With increased adoption and pervasiveness of emojis in all channels, it's paramount to consider them as full-fledged components of customer feedback.**

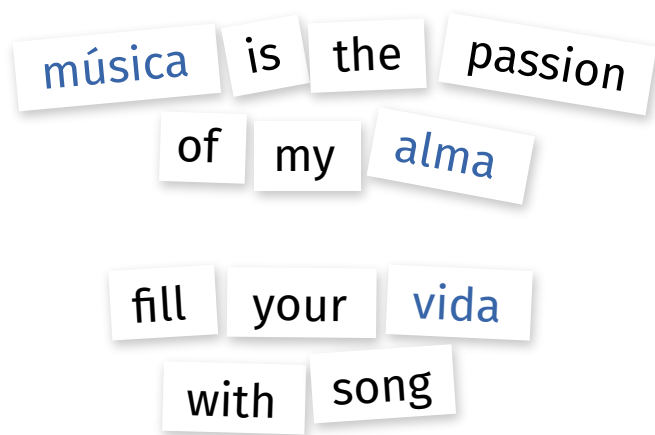


# PART 2

## GOING BEYOND WORDS—UNDERSTANDING THE MEANING OF SENTENCES AND DOCUMENTS

### Detecting Intent and Actionable Sentences

Throughout college, our apartment refrigerator was littered with Magnetic Poetry magnets from the various languages that my roommates and I spoke or studied.



English, Spanish and Yiddish nouns, verbs and adjectives plus corresponding function words such as prepositions, conjunctions and pronouns, as well

as a variety of profanity, allowed us to imagine novel phrases that admittedly often had NSFW intent. Being able to reorganize words tactilely propels us to reimagine their meaning and how they could be reorganized to achieve new intent. While we are fixated briefly on each individual magnet, it is the complete clauses or sentences that yield paradoxical, nonsensical or absurdly amusing meanings that cause us to guffaw.

In part one of this ebook, we focused our discussion on the value of extracting named entities, or proper nouns such as brands, locations and persons, from customer feedback data. These words often carry more weight than other nouns as they refer to a specific entity that reveals preferences and biases that could affect a customer's perception or experience. However, like the Magnetic Poetry example, individual words are so much more meaningful when in the context of a phrase or sentence.

We normally don't spend a lot of time thinking about the intent of a sentence that we are uttering. Each sentence is just a piece of a larger dialogue used to convey our message. However, when analyzing customer feedback, being able to isolate certain kinds of intents actually lends substantial power. Compare "Your website sucks!" and "Your website would be so much easier to use if the chat box didn't cover up the login area!" While we might be drawn to the obvious negativity in the first sentence, it is the second one that we would deem as actionable.

**When analyzing customer feedback, being able to isolate certain kinds of sentences actually lends substantial power. For example, compare "Your website sucks!" and "Your website would be so much easier to use if the chat box didn't cover up the login area!"**

It offers an explicit suggestion that unlocks valuable information we can use to identify specific pain points and to design customer-centric solutions.

Clarabridge uses semantic analysis strategies to identify 20 different kinds of intents specifically relevant for customer experience analytics. Together, there are four actionable types (Suggestions, Requests, Cries for Help, Churn), five sentiment-bearing types (Generic Apathy, Generic Criticism, Generic Praise, Recommend, Not Recommend), six question answer types (No Comment, Don't Know, Everything, Yes, List, Cross-Reference), three social remarks types (Hello/Goodbye, Thanks, Laughter), and two legal disclosure types (Disclosure, Mini-Miranda). These sentence types give Clarabridge users the power to segment their data in ways that go far beyond traditional topic analysis.

When some users struggled to identify this kind of feedback using keyword searching, we realized that the variation in language was too great to solely rely on this archaic strategy. We went back to the drawing board and returned with a machine learning-based solution that leverages linguistic metadata produced by the Clarabridge NLP to identify sentence types rather than simply keywords and phrases.

How is the Intent Detection (Sentence Types) attribute best leveraged? Here are popular use cases for this feature:

### 1. Isolation of Suggestions and/or Requests

Let your customers guide you to success by isolating their suggestions/requests in your dataset. Analyze the top topics of suggestions to understand where your business could improve; track KPIs for this customer group to determine how any product or service deficiencies are affecting their loyalty. Look at topics of requests (questions) to determine where you may need to amp up your messaging or clarify content on your website.

### 2. Noise Filtering

Hearing the customer's voice can be challenging when it's swallowed up in a sea of non-actionable messages. Use our non-actionable sentence types to filter out the content that you don't want to see. This step helps to part the seas, so to speak, allowing you to hear the feedback that you were missing.

### 3. Route Cries for Help or Threats of Churn

Don't miss out on an opportunity to help your customer! When customers ask for a return contact, leave their phone number or email address in their response, or threaten to leave your business, they are pleading for assistance. Use the Cries for Help or Churn sentence types to automatically trigger an alert and a case within our Case Management system so that it gets routed to the agent who can best help.

**Let your customers guide you to success by isolating their suggestions/requests in your dataset. Analyze the top topics of suggestions to understand where your business could improve.**

# Analyzing Emotion and Effort

“In Whoville they say that the Grinch’s small heart grew three sizes that day.” —Dr. Seuss

In late 2015 and early 2016, the digital world got sentimental. A place that had been predicated on instant gratification and click rates suddenly had a change of heart—quite literally. Our favorite digital platforms started embracing their feelings. Twitter replaced the star with a heart to indicate a favorite tweet. Facebook supplemented its iconic thumbs up with a heart and four other faces to indicate a wider array of feelings. Slack followed suit as well by giving users the ability to react to any message with hundreds of different symbols and shapes. These new symbols afforded us the ability to “respond” in a way that matches much more closely how we feel and why we feel that way.



In the customer experience world, sentiment is our thumbs up or our heart. It provides an arbitrary quantification of the polarity of the experience. At first, it was our best available mechanism to quantifying how positive or negative a customer’s experience was, but we continue to hear today that sentiment is not intuitive and many organizations struggle to communicate its meaning to their stakeholders and decision makers. Over time, sentiment has become a crutch for understanding the customer. It’s ubiquitous in CX dashboards; red and green word clouds or bar charts haunt us like the Ghost of Christmas Past. However, like the old social media icons, sentiment, while a good barometer for the quality of an experience, may be insufficient to express the subtleties of a customer’s true feelings.

Simon Sinek, in his famous [TED talk](#), points out that the inner two sections of our brains (the limbic brain) are responsible for generating our feelings, behaviors and decisions. Interestingly, it is this area of the brain that has no capacity for language. So, it’s no wonder that we often struggle to articulate how we feel about our experiences and why we feel that way. It’s certainly unfair, then, to attempt to quantify a customer’s experience simply by looking at positive and negative words in a survey

response or a chat dialogue. In order to really digest these experiences, we need tools that help to categorize and quantify our emotions and the level of effort that we exert when interacting with brands and services. In the past, sentiment has been the proxy for this kind of analysis. We can take a giant step forward by using emotions and effort as the clues that help us truly understand the impact that we have on our customers' limbic systems. (Read: how likely they are to be a brand advocate, to purchase or to churn!)

**In order to really digest these experiences, we need tools that help to categorize and quantify our emotions and the level of effort that we exert when interacting with brands and services.**

Our toolbox for analyzing customer experiences continues to grow here at Clarabridge. In 2016, we introduced an Emotions model specifically designed for customer experience analysis. It's important to realize that not all

emotions are created equal. In customer experience, "grief" is less common and less actionable than "confusion" or "surprise." Our Expanded Emotions model offers nearly 50 distinct emotions that help CX professionals align customer feedback to specific and actionable feelings. In 2018, we introduced the Clarabridge Effort Score, a new metric that gives analysts the power to quantify how easy or how difficult different experiences were. This new metric is powered by a machine learning algorithm that can pick up on the nuances of language. Not only will words like "easy" or "hard" affect the score, but so will implied effort as reflected in comments such as "I had to call five times" or "It truly was a breeze."

### The Clarabridge Effort Scale



Both emotion and effort are leading indicators of loyalty. By paying attention to shifts in these dimensions, customer experience analysts can get ahead of burgeoning issues. Our customers use emotion and effort analysis tools to simplify complicated and confusing parts of their products, to mitigate anxiety regarding service offerings, and to capitalize on excitement and ease in product marketing. We love this new movement toward



emotion and effort. Unlike sentiment analysis, which we've had to teach to our employees and customers year after year, the reaction that we get to emotion and effort analysis is that it is intuitive. Period. We humans have spent our whole lives learning how to interpret and respond to each other's feelings. We don't need to rely on mechanisms that we invented simply for the purpose of putting numbers on a page.

We believe that this is just the beginning of our journey into analyzing and designing customer experiences with an empathetic bent. Our lab is busy prototyping other tools and techniques for analyzing both emotion and effort that will give Clarabridge users more humanism in the way they interact with data. Make no mistake: Focusing solely on sentiment is focusing on the effect and missing the cause. Language tells a much richer story when we can analyze it in multiple dimensions, and we are so excited to see what stories our customers find by leveraging the new tools for emotion and effort analysis in our toolbox.

**Let your customers  
guide you to success  
by isolating their  
suggestions/requests  
in your dataset.  
Analyze the top topics  
of suggestions to  
understand where  
your business could  
improve.**

# Detecting Spam

Waitress:

Well, there's egg and bacon

Egg, sausage and bacon

Egg and spam

Egg, bacon and spam

Egg, bacon, sausage and spam

Spam, bacon, sausage and spam

Spam, egg, spam, spam,

bacon and spam

Spam, sausage, spam, spam, spam,

bacon, spam, tomato and spam

Spam, spam, spam, egg and spam

Spam, spam, spam, spam, spam,

spam, baked beans, spam,

spam, spam and spam

Ah, Monty Python at its finest. And, recently, an apt description of my email inbox and my voicemail. Spam, Hawaii's favorite processed meat and Western culture's favorite protein to mock, has an impressive résumé. Hormel offers 15 flavors of Spam and has sold over 8 billion cans across 44 countries since its introduction in 1937. But, Spam's influence isn't just culinary; it has also had an understated influence on our technological world. Inspired by the Monty Python sketch, certain abusive users of Bulletin Board Systems and Multi User Dungeons would repeat "spam" a massive number of times to scroll other previous messages off the screen. Soon, the term became a moniker for the unwanted junk we find on the Internet that obfuscates the content that we actually want to see.

Spam content can be a real burden to those trying to find meaning from their customer feedback data. It can increase the effort needed for analysis and can cause misinterpretation of customer needs. In most cases, customer experience analysts want to analyze the customer's organic voice, not the inorganic voice of automated bots or fake customers. Looking at volumes and sentiment of mentions of specific products or topics without regard to whether the content is spam can be very misleading.



Not all spam, though, is created equal. By classifying the types of inorganic messages that appear frequently in customer feedback data, we can gain a better picture of how the market views a brand or product. The types of spam present in each dataset may vary. On social media, auto-generated messages in content such as news headlines, reviews and articles abound. In email sources, job requests or solicitations for corporate sponsorship may get in the way. Analyzing the language used in the inorganic messages bares its own utility. When these spam documents mention a brand or product, they may reveal how customers and potential customers use, advertise and perceive that brand or product.

A tool that blindly looks at words and phrases is limiting itself to parsing of discrete words; a tool focused on understanding will tease out the organic versus inorganic messages and classify them into their associated types. By exposing these types to end users, such a tool empowers analysts to isolate the true feedback and determine meaningful insights.

The Clarabridge proprietary Content Type Detection feature uses a machine learning algorithm to identify and tag spam content and then classify it into subtypes: advertisement, coupon, link or undefined. Users can choose either to purge any spam

documents upon ingesting or to retain them for analysis. Analysts can also customize this feature by training the algorithm on their own data and with custom sub-types. With the Clarabridge Content Type Detection, users can go beyond basic topic and sentiment analysis. Considering the integrity of a message provides a different dimension and a unique analytical lens for many stakeholders that would be missed or misinterpreted if all documents were viewed the same way or were viewed purely by their independent words. Just the same as how “egg and spam” is not the same as “spam, spam, spam, egg and spam.” Bon appetit.

**A tool that blindly looks at words and phrases is limiting itself to parsing of discrete words; a tool focused on understanding will tease out the organic versus inorganic messages and classify them into their associated types.**

# PART 3

## IMPORTANT LINGUISTIC CONSIDERATIONS

### Translation

Bad translations are ubiquitous in our globalized world. There are seemingly infinite listicles showcasing very comical mistranslations. My favorite service, though, might be TranslationParty, which takes your sentence and translates it back and forth through Google Translate between English and Japanese until it reaches equilibrium. The result is inevitably a bastardized but hysterical interpretation of your original sentence.

Translation services used to be human-centric. A bilingual speaker would read the original document and then translate it into the target language. With the advent of the Internet and improvements in computing power, translation services have become machine-centric. They have matured greatly over the past few years and have arrived to the consumer market in many different forms. We see them in free web tools like

Google Translate or Bing Translate that support text-to-text translation. In the past few years, we've also seen the growth of speech-to-speech and image-to-text translation tools. The futuristic image of having an earpiece translate audio in near real time is no longer science fiction. And, for common language pairs like English to Spanish or French to German, the quality is actually pretty decent.

YOU SAID:

When words fail, music speaks.

INTO JAPANESE

文字がダメなら音楽で。

BACK INTO ENGLISH

If the character is not good, with music.

INTO JAPANESE

性格が良くない場合は音楽付き。

BACK INTO ENGLISH

Music is included if the personality is not good.

INTO JAPANESE

性格が良くない場合は音楽が含まれます。

BACK INTO ENGLISH

Music is included when personality is not good.

INTO JAPANESE

性格が良くない場合は音楽が含まれます。

BACK INTO ENGLISH

Music is included when personality is not good.

**Equilibrium found!**

However, neither human nor machine translation is perfect. Both are vulnerable to misinterpretation. Machine translation excels at translating certain words, especially nouns. Verbs and adjectives, on the other hand, are more difficult to translate automatically; they often have subtleties that are lost in translation. While it's not so hard to translate a pronoun, a preposition or a conjunction in isolation, machine translation can end up completely reorganizing them throughout the sentence. Add idioms, hyperbole, sarcasm and named entities to a sentence and we have significantly complicated the task for both a human and a machine. While analyzing a translated text, a human can make inferences about the meaning even when the grammar or vocabulary are flawed. However, when we ask an NLP engine to interpret an imperfect translation, we are asking for trouble.

Analyzing text through computational means is also a very difficult task. It's taken the industry many decades to produce suitable, scalable systems that work on well-formatted, natively written text. As we've established, translating text often warps keywords and grammatical elements, which makes natural language processing even more challenging. Attempts at feeding poorly translated text through an NLP engine may result in failures to identify negation, modification, sentiment, associated words and topics.

The subtlety of meaning, emotion and intent are inevitably lost, jeopardizing your ability to relate to your customer. While NLP engines (including Clarabridge's) can be used to analyze both native and translated customer feedback, the output from text in its original language will always be superior in quality and accuracy.

There certainly are cases where translating text is an essential component of business. Marketers must translate web content to target new markets. News agencies must translate headlines in international environments. This pattern does not apply to customer experience management or customer experience analytics. In the CX ecosystem, native-language text reigns supreme. You should make every effort to preserve your customers' intended messages in their truest form. By doing so, you improve your ability to understand the messages fully and to offer customers the most empathetic service possible.

A comment I often hear is "But Ellen, I don't have anyone on my team that speaks [insert your exotic language here]!"

That is a fair point. It's better to understand your customers partially than to ignore them completely for not speaking your lingua franca. In this situation, I encourage you to translate the keywords in your topic categories rather than translating the text itself. This procedure is most



successful when there is a human in the loop. Translating topics is not the same as translating prose. A translator must consider the domain and data source when translating keywords. (It was a comical moment when we were auditing our Automotive category template and discovered that “starter” had inadvertently been translated as “appetizer.” The translator had not properly considered the domain of the data!) By translating the keywords that you’re querying for instead of translating the original data, you don’t lose the meaning from your customer’s voice and you can still present your data and insights through whatever lens you prefer.

Machine translation technologies are rapidly evolving and may reach a point where meaning and intent are not warped through translation. But in the meantime, I implore you not to sacrifice meaning for expediency. Your customers will thank you for not playing the telephone game with their feedback!

**You should make every effort to preserve your customers’ intended messages in their truest form. By doing so, you improve your ability to understand the messages fully and to offer customers the most empathetic service possible.**

# Sarcasm

When I tell people that I work on text analytics products, their first question is often “But, how do you handle sarcasm?” My typical response—“About as well as a human, which is to say not very well at all”—is equally sincere and sarcastic, a poetic homage to the difficulty of the linguistic problem.

Sarcasm is mired in a complex web of deep cultural knowledge, emotional sensitivity and individual awareness. Some cultures (looking at you, my British friends) find comfort in the dryness of the humor, whereas others find it distasteful or even disrespectful. Sarcasm can be subtle or obtuse; flattering or insulting; funny or offensive. In sarcastic expressions, the words used are only a small piece of the puzzle. Tone matters, body language matters, context matters. Given the high degree of tangible and intangible awareness necessary for sarcasm to succeed, it is perhaps unsurprising that non-native speakers struggle to master both delivery and understanding of sarcasm in learned languages.

All of these components make sarcasm an extremely difficult linguistic problem

to study. Linguists have classified sarcasm into specific sub-types including irony, satire, passive aggression and flattery. They’ve determined that your brain works differently when processing sarcastic comments in comparison to sincere ones. Others have identified the facial tics that betray an otherwise earnest face. Most of this research, though, is conducted through individual face-to-face analysis. Conducting broad analyses of sarcasm through text analytics is nearly impossible. Text-only communication lacks tonal and visual cues, making it highly susceptible to misunderstanding and misinterpretation. Other clues for correct interpretation of a comment may be baked directly into the medium and shared among some of its participants, but these clues may be imperceptible or opaque to outsiders.

An NLP engine is not a native speaker of any human language. It understands the rules or the patterns that its human overlords have programmed into it, but it lacks any linguistic intuition. It can understand words used, relationships between those words, and maybe even emotion and intent, but it fails when meaning transcends content. It may understand its context or its purpose but will undoubtedly fail when other niche cultural or societal knowledge is injected

into a witty retort. As humans, we could all listen to or read the same passage and walk away with different understandings of its intent. An NLP engine fares about the same. In some situations, it will interpret an ironic comment correctly; in other cases, it will completely miss the mark and produce the exact opposite sentiment as a native speaker would otherwise expect.

The Clarabridge NLP engine errs on the side of sincerity, but, given the flexibility in our sentiment engine, users have the power to customize rules to support common sarcastic phrases that appear in their dataset. I've found success in customizing rules for two specific types of sarcasm.

**As humans, we could all listen to or read the same passage and walk away with different understandings of its intent. An NLP engine fares about the same.**

1. Speakers, in an attempt to underscore their emotions, may associate positive actions with negative aspects (or

vice versa) such as in the sentences "I love sitting in traffic" or "going to the dentist is the best." Clarabridge understands word associations, parts of speech and sentiment and allows users to leverage this word-level metadata in the construction of sentiment rules. Users could construct a rule that negated every positive verb associated with "dentist" or "traffic" or "[insert emotionally charged word from your industry here]."

2. Social media posts are now often suffixed with #sarcasm, #sarcastic or #not to aid a reader in interpretation of an otherwise ambiguous post. Users can tune sentiment based off of specific hashtags and positions of these hashtags within posts.

Sarcasm is, without a doubt, an important part of our communicative tools as social beings. However, sarcastic expressions are relatively rare in most types of text. The ability to detect sarcasm through computational means should not be a make-or-break point when deciding which NLP tool to use. We at Clarabridge will continue to investigate sarcasm in customer feedback and how we may be able to improve sentiment accuracy for these expressions, but for now I'll quote the Comic Book Guy from "The Simpsons": "Sarcasm detector? Now that's a really useful invention."

# Conclusion

With increasing pressure on customer experience teams to deliver value and customer-centric solutions, there's no time to meander around looking for insights.

Topic reports lack the oomph needed to truly change a business for the better. By analyzing the intents, emotions and effort expressed in feedback data, analysts can better understand their customers' needs and wishes and cut through all of the noise with a few swift clicks. It's no longer sufficient just to know **what** your customers are talking about; you need to also understand **how** and **why** they're talking about it.

**It's no longer sufficient just to know what your customers are talking about; you need to also understand how and why they're talking about it.**

If you intend to make your customers the centerpiece of your customer experience strategy, you must also make them the key part of your customer experience analytics. Understanding your customers, just like deeply understanding your family, your friends, your boss, your co-workers, or your partner, is really about going beneath the surface by discovering what makes them tick. Each person or group of people is much more than a list of topics or a set of sentiment scores. They are nuanced individuals who are constantly evolving with new ideas, opinions and connections. It is only after we truly digest the factors that influence their experiences and their perspectives that we can empathetically and effectively design solutions that cater to their needs and preferences. A customer experience strategy that matches solutions to true needs will be much more intuitive and successful than one that takes surface-level assumptions and offers hasty answers. You owe it to yourself and your organization to spend more time getting to know your customers. After all, you're a customer too, aren't you?

As you embark on this mission, I encourage you to think carefully about the tools that are in your toolbox. If you endeavor to truly understand customer feedback, you need to leverage tools that go beyond

## DEBUNKING NATURAL LANGUAGE PROCESSING | PART 3

topics and sentiment and that have a strong linguistic foundation to unlock the potential of NLU features. Shortcuts don't pay off in this game. Clarabridge is built on the hypothesis that unstructured data holds a treasure trove of value for customer experience analytics. We started by building an NLP engine customized for customer experience and have built the supplemental assets around it. This mature foundation has afforded us opportunities to offer more than word clouds and other elementary façades for faux NLP engines. Our engineering efforts go far beyond improvements to our grammatical parsing. We strive to innovate and develop tools that give you the power to unlock the inherent value of text—a tenet that is core to our identity as a company.



### About the Author

**Ellen Falci Loeshelle** is a Director of Product Management at Clarabridge where she is responsible for developing and executing on the product strategy for the data integration, NLP, enrichment and analytics parts of the Clarabridge product suite.

Throughout her six years with Clarabridge, she has worked with customers to develop innovative solutions to customer experience challenges within the context of linguistics and technology theory. Prior to her current role, Ellen also worked as a Business Consultant and as the Product Manager of the Data Acquisition and NLP teams at Clarabridge. Ellen is a proud alumna of the University of Virginia and Georgetown University.

# About Clarabridge

Clarabridge helps the world's leading brands understand the true voice of their customers. Companies leverage Clarabridge's omni-channel capabilities to listen to all customer interactions and conversations, analyze this data via a best-in-class AI-powered text analytics engine, and get actionable insights to make mission-critical decisions.

### ADDITIONAL STRENGTHS INCLUDE:

- A world-class analytics platform with beautiful, interactive and easily customizable dashboards that can be tailored for every role in the organization, advanced predictive algorithms and sophisticated case management workflows
- A patented, best-in-class Natural Language Processing (NLP) platform specifically designed for Customer Experience Analytics that combines the latest AI and Machine learning technologies and offers accurate and nuanced topic, emotions, effort and intent detection
- Clarabridge was named a Leader in Customer Feedback Management Platforms and Text Analytics, as well as a Strong Performer in Speech Analytics by Forrester Research