Natural language understanding

Learning to speak customer-ese.

NUANCE

Table of contents

In recent years speech recognition systems have made impressive advances in their ability to provide rich and natural interactions with callers. The term Natural Language (NL)—and its family of synonymous abbreviations: Natural Language Understanding (NLU), Natural Language Processing (NLP), and Natural Language Technology (NLT)—have been held up as the panacea for improving the accuracy of a speech system. Natural language suggests that, rather than constrain a caller to a list of choices, the caller can say anything in any context and be understood by the system. Speech vendors trumpet their NL capabilities, linguists at research labs talk about advancements in NLP technology, contact centers evaluating vendor proposals ask for explanations of available NLU features and industry press proclaim NLT as the future of speech recognition.

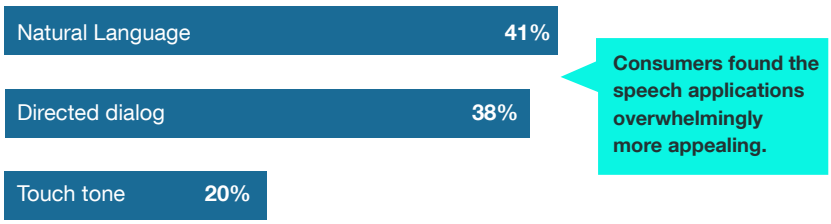
The problem is that the term "natural language" has become meaningless. This standard speech industry terminology is anything but standard. It implies that you could call up a travel service and say, "I'd like a flight to London for the 25th, for two adults, no wait make that two adults and an infant on my lap, and can I have two aisle seats next to each other?" The term natural language is vague and overused, full of promise, yet promising nothing. For it to mean something, one needs specifics.

In this paper, the specific benefits of natural language in speech recognition are explained and the approaches examined in detail. By the time you've finished reading, you'll understand natural language and how it can be effective in improving the overall accuracy of your speech-based IVR system while providing callers a more engaging experience.

Speech-enabled automation
Interactive Voice Response (IVR) systems are at the heart of today's contact center self-service solutions. Without IVR systems to automate calls, most contact centers would be overwhelmed by operating costs and by angry customers facing long queue times. But by automating common customer service requests, customers can resolve many issues quickly and agents can focus on helping the customers who truly need them.

Speech-enabled IVRs improve the overall automation rate of self-service systems. Unlike an IVR with only touchtone options, a speech-based self-service system can handle more complex tasks, like updating an address, and can simplifying existing tasks to further increase automation. The resulting increase in automation rates mean more satisfied customers and lower costs for the contact center.

Automation Preferences
"Which of the example systems do you find most appealing?"



Source: Harris Interactive

Nevertheless, the automation rate can only be as good as the accuracy during the caller's entire interaction with a speech recognition system. Several recognition factors can limit the automation rate of a speech system.

Noisy backgrounds, confused callers, very heavy accents, and mumbled responses can make it hard for a system to complete a call. Of course, those factors can affect the ability of a live agent to complete those calls as well. And despite much advancement in technology, it is still possible for a speech recognition engine to simply make a mistake.

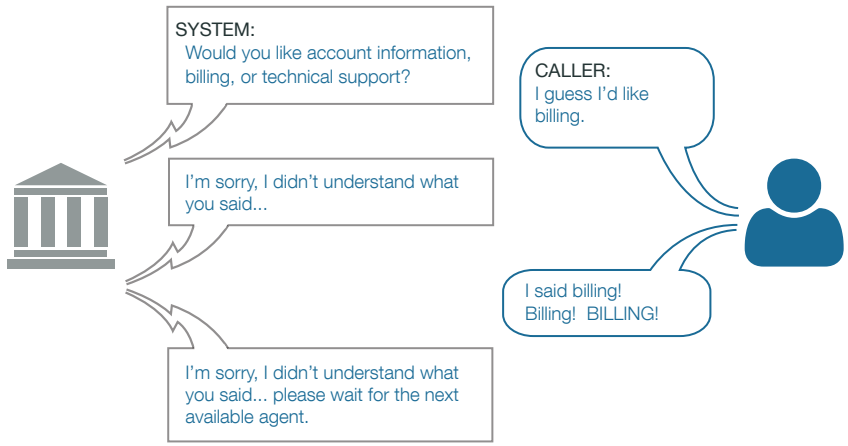
Learning to speak customer-ese

As all contact center agents know, callers truly believe they are speaking in a clear and obvious manner when they are explaining their concerns. In fact, each caller is really speaking in 'customer-ese', their own personal dialect that agents and IVR systems need to decode. This interpretation process is very similar to learning a new foreign language. To address the challenge many speech recognition systems have a list of acceptable phrases, known as the grammar, that the system is programmed to recognize at any given time. Grammars ensure that a speech recognition engine knows that the phrase "new to speech" is not a "nudist beach", and that "agent," "operator," and "I want to speak to someone" all direct the caller to a live person.

If implementing a speech-enabled self-service system is like teaching a foreign language, then a grammar is like a translation book, mapping common phrases to their meaning. It works quite well… as long as people stick to those phrases. So what happens if a phrase is not in the book?

When a caller believes that a speech system has misrecognized his or her response, it is more likely that the caller spoke something which was out-of-grammar—the system simply wasn't expecting the caller to respond quite like that.

The Need for Natural Language
A Typical Example



An analysis of a number of representative speech applications confirms that out-of-grammar errors outnumber misrecognition errors by a factor of as high as 5-to-1. Put simply, the problem is not recognizing what the caller said… it's knowing what the caller meant. Therefore, to increase the accuracy of any system—improving its automation rate, making customers happier with the use of the system, and lowering overall costs for the contact center—a critical factor is to reduce these out-of-grammar errors.

So how can we teach a speech recognition system to speak this language of the customer? Different natural language capabilities provide different approaches to tackling that problem.

Make the translation book bigger
You can make the grammar contain a larger list of entries to appear like natural language is at work. This works best when trying to pick between a limited number of responses – for instance, a confirmation dialog may recognize yes, yup, you betcha, sure, heck yeah, as well as many other variations of a positive response.

This straightforward, brute-force technique can provide a semblance of natural language understanding. However, the approach has significant drawbacks as the list of possible responses gets longer and longer. It becomes harder to tune and maintain a system while keeping track of all these options. Even worse, it is possible for the overall performance of the system to decrease as the recognition engine struggles to compare the caller's response to such a potentially confusing list of phrases.

The reality is that it is impossible to foresee and capture every conceivable response variant even in a very large grammar. To learn to speak customer-ese, the self-service system needs more natural language capabilities.

Learn to recognize variations of phrases
Since every caller expresses themselves in a unique manner, often adding extraneous information to their response, focusing on recognizing certain keywords is another approach for avoiding out-of-grammar errors and achieving higher automation rates.

Keyword spotting, sometimes referred to as robust parsing, is used when a caller says what the system is expecting but surrounds their response with filler phrases. Instead of simply saying "Yes" they may actually say "Um, I think so, yes". The keyword spotting approach is most effective with short menus and yes/no dialogs which typically account for more than 70% of speech-enabled application interactions.

This statistical technique is based on comparing vocabularies of filler phrases created from transcriptions of previous calls or from classes of common phrases refined over time. Applying these vocabularies can provide an impressive 30% or more increase in recognition accuracy.

As callers have more flexibility in how they can exactly respond, the caller's perceived accuracy is typically very high when using keyword spotting. As with any increase in accuracy, callers also appreciate interactions that avoid retry prompts and repeated confirmations.

Although the self-service system has already taken a big step towards learning to speak customer-ese, the caller's responses are still limited to words that have been chosen for them rather than allowing them to use their own words. There are still more natural language capabilities to explore.

Become fluent
Rather than focusing only on matches within and around a grammar, another approach for determining meaning is to ask callers to describe a problem in their own words.

This approach often begins by offering callers an open-ended prompt such as "how may I help you?" Callers respond by describing in their own words why they are calling, perhaps by saying "there's a strange charge on my credit card statement". The system then uses sophisticated statistical modeling to identify key phrases and look for a match in a list of example phrases that have known meanings. In this case, the likely meaning would be associated with billing even though the word 'billing' was never spoken.

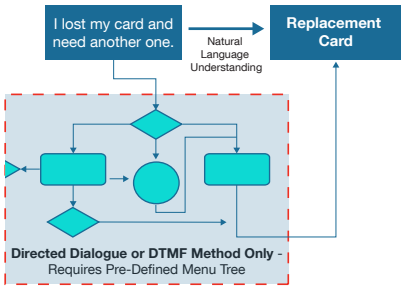
SmartListener™ technology increases automation by interpreting the meaning of caller responses that include filler phrases that do not exactly match expected dialogs.

Although developing the detailed statistical models involves manual efforts to define and assign the agreed upon meanings to each transcription, the result is very high automation rates and high customer satisfaction. The key to extracting the correct meaning from this open-ended approach is the proper tagging of previously transcribed phrases. As with grammars, if the caller gives a completely unexpected response, like talking about 'pizza toppings' to an electricity provider's system, there will not be a match.

When there is no match, the system will offer a fall-back alternative that uses traditional grammars to determine the caller's intent such as "I'm sorry I did not get that, would you like mortgages, credit cards, …" Experience has shown that applying the keyword spotting approach to these fall-back grammars often leads to a successful interaction.

Appropriate for determining the meaning of any open-ended response, this powerful approach is highly effective for call routing applications. When callers use their own words, there tends to be drastically fewer zero-outs and misroutes as complicated menus are bypassed and confusing or arcane jargon is avoided.

Natural Language Understanding technology increases automation by correctly interpreting the meaning behind open-ended caller responses allowing callers to use their own words rather than words chosen for them.



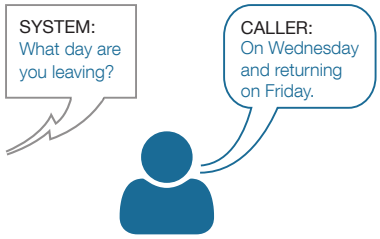
Provide a more flexible script

Accurately recognizing the meaning of what a caller says is only part of how natural language can improve automation. The next step in learning to speak customer-ese is for the self-service system to have a conversation with the caller.

The conversational approach relies on dialogs that continuously adapt to the information provided. This way the caller has a personalized experience while they control the conversation. For example, callers often provide more information than prompted for. A system that can respond to varying amounts of information will have more productive and shorter calls. Notice in the dialog at the right the caller provides the 'extra' information of the return date allowing the system to avoid having to ask for it.

More than collecting information, a natural language conversation needs to flow smoothly. A system that constantly asks for confirmations creates a disjointed conversation that callers tend to reject. However systems that can handle corrections and verifications by dynamically embedding the confirmations in the next prompt are more engaging, leading to better automation rates. Notice in the dialog below how the system confirms the change to Boston when the caller requests a window seat.

Conversational dialogs can handle more than one piece of information



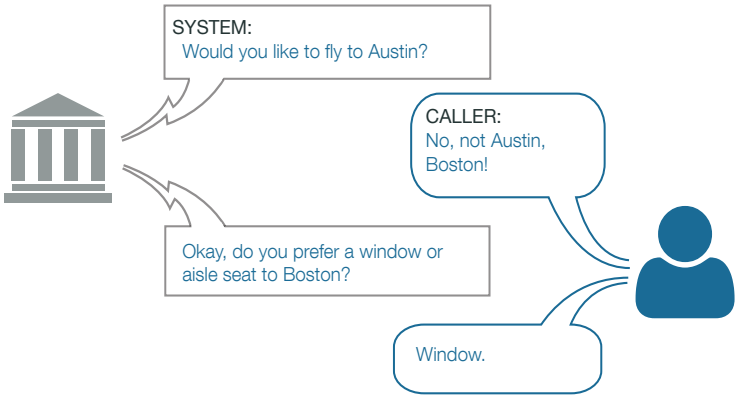
Natural Language Understanding
Caller provides responses the way they want to and the system accepts and interprets the information.

AIRLINE SYSTEM with Natural Language Understanding:

Information to Collect

- ❑ Departure City
- ☑ **Departure Date**
- ❑ Number of Adults
- ❑ Number of Children
- ❑ Return City
- ☑ **Return Date**
- ❑ Seating Assignments

Conversational dialogs can build in confirmations without having to explicitly ask for them



This dialog flexibility is controlled by rules designed into prompts themselves so the system can respond and guide callers rather than dictate to them. Perhaps the defining characteristic of a natural language experience is engaging callers with intuitive and direct conversations.

Conclusion

In challenging economic conditions, the importance of cost-savings can not be underestimated. By applying effective natural language techniques to a speech-enabled self-service IVR system, automation rates will increase leading directly to lower costs. This paper has shown that different approaches to natural language can be applied throughout the caller's interaction to achieve these improved automation rates. The three complementary natural language approaches of increasing the recognition accuracy within and around grammars, allowing callers to give open-ended responses, and interacting with adaptive dialogs all lead to more efficient and appealing calls.

Nuance Adaptive Dialog Modules are specialized building blocks used to efficiently create intuitive speech-enabled dialogs that drive automation.

About Nuance Communications, Inc.

Nuance Communications is reinventing the relationship between people and technology. Through its voice and language offerings, the company is creating a more human conversation with the many systems, devices, electronics, apps and services around us. Every day, millions of people and thousands of businesses experience Nuance through intelligent systems that can listen, understand, learn and adapt to your life and your work. For more information, please visit nuance.com.

NUAN–CS–385–01–WP, Feb 17 2015