Sequence Package Analysis

A New Global Standard for Processing
Natural Language Input?

Amy Neustein, Founder and CEO,
Linguistic Technology Systems

Amy Neustein

As speech technology continues to evolve, so do the heated debates among speech system developers over the best way to design natural language dialog systems. On one hand, there are speech recognition designers who champion strictly grammar-based speech recognition systems that find the best possible word match for the string of phonemes presented to the recognizer. On the other hand, there are those who reject context-free grammars altogether, claiming such methods of processing natural language input fail to reflect how humans comprehend speech.

Among those who reject the context-free grammar (CFG) approach to designing speech recognizers, some have actively championed a conceptually-based, Artificial Intelligence (AI)-driven speech recognizer that relies heavily on relevant knowledge domains when processing speech input. In such a system, recognition of the user's speech input is more dependent on conceptual content derived from relevant knowledge domains (e.g., flight arrival and departure times) than on matching each of the user's phonemes – regardless of the context – against the speech system's application vocabulary.

Against this backdrop of contrasting approaches to Natural Language Processing, a new method that is neither entirely based on context-free-grammar, on the one hand, nor domain-driven, on the other hand, has emerged as an alternative approach to understanding natural language dialog. This new method is called Sequence Package Analysis, or SPA. [1]

Three computer scientists at Oklahoma State University's Computer Science Department, in a recent review of the literature dealing with different methodological approaches to mining call center data, point to SPA as a method that can "caption the text to which data mining is applied" so as to enable the "capturing of early warning signs of caller frustration." [2]

At last December's LISA Forum – *Accelerating Global Understanding* – held in Washington, D.C. (December 8-12, 2003), this author demonstrated how the SPA approach to building dialog

systems can meet the speech industry's expanding globalization and localization needs in the following workshop sessions: 1) *Filling the Global Communications Gap*; and 2) *Building Standards for Global Speech Applications*. (LISA members may click **here** to access the presentations.)

An SPA-driven speech system would be better able to make sense of vague requests for assistance.

SPA looks at dialog as a socially organized activity in which participants demonstrate - through the design of their speaking turns - their understanding and interpretation of each other's social actions. It transcends the polarized approaches to the design of speech systems (which juxtapose word spotting parsing techniques to those that are AI-driven and concept-based). It is particularly suited to meet globalization and localization requirements. Here's why: SPA looks at language for its social architecture, rather than for its grammatical discourse features (e.g., syntactical rules). This means that it can be readily applied to a myriad of other languages and dialects. This is because *all* forms of interactive dialog, regardless of their underlying grammatical discourse structures, are ultimately defined by their *social* architecture.

SPA is an effective tool that can speed up the development of grammars that better match how callers truly express themselves across many different languages and dialects.

How Does Sequence Package Analysis Work?

SPA maps out the social architecture of language in the form of "sequence packages": a series of related turns and turn construction units (lexically bounded parts of turns at which point the turn may conceivably, but not necessarily, be yielded to the next speaker) that are discretely packaged as a sequence of conversational interaction. These sequence packages may in fact contain the kinds of data that are usually discarded by most speech recognizers, such as pauses, changes in rate of speech, elliptical phrases and repetitions. These data, that appear useless to most speech systems, are actually crucial for performing sequence package analysis on natural language dialog because they point to the discourse interactive work being done by speakers as part of their situated achievement of socially organized talk.

SPA technology provides one possible solution to "voicemail hell."

The kinds of social activity found in such sequence packages consist of requests addressed to help-line desk agents, reporting on problems and the giving of information, and other socially related tasks. A request for assistance addressed to a help-line desk agent might consist of the following sequence package: a vague and fuzzy assessment made by the caller about his/her difficulty with a particular product or service; then, a series of pronouns in place of nouns, which appear more as a circumlocution than a clear identification of the problem; and finally, a repetition of the preliminary assessment that was nebulous to begin with.

SPA offers the field of speech recognition design true multilingual functionality.

Notably absent from such a request, given in such a roundabout way, are the keywords in the speech application vocabulary. This poses a big problem for a conventional word-spotting speech recognizer, which will find such a data entry to be useless in its processing of speaker input. Even if the recognizer were to incorporate conceptual content, we would still have no guarantee that such vague dialog could even tap into the relevant knowledge domains in the first place.

Even when confronted with a language such as Arabic or Farsi, the SPA-driven speech system can extract those sequence patterns that are indigenous to that language.

An SPA-driven speech system, on the other hand, would be better able to make sense out of such vague requests for assistance. Other speech recognition methods can only identify a string of phonemes or knowledge domain concepts. SPA, on the other hand, is able to identify the common kinds of conversational sequence patterns found in natural language dialog, such as those that habitually entail circumlocutions, ellipses and other vague features.

SPA-driven systems have the potential for a higher rate of speech recognition accuracy than conventional recognizers that are limited to the identification of keywords.

But the value of SPA does not end there. What SPA truly offers the field of speech recognition design is its multilingual function: its ability to detect conversational sequence patterns cutting across many different languages. This means that one need not build an entirely new application to accommodate each language and dialect, as one might be required to do when building a voice application based on word spotting or conceptual speech. While some critics may object that this rosy picture does not tell the whole story – some sequence patterns may indeed be *sui generis* (indigenous) to a particular language or dialect – most sequence patterns do traverse many different languages. This is because human language communication throughout the world is such a highly organized form of social activity that many languages inevitably share features in common. SPA meets both globalization and localization requirements by mapping out sequence patterns in many languages and dialects at the same time.

In designing speech applications that can be used with a variety of languages and dialects, SPA serves another function as well. This method of processing natural language input can expand a speech system's application vocabulary to include the additional words and phrases that are essential to the voice application. SPA expands the application vocabulary as it uncovers new words and phrases which correspond to the sequence package templates contained in the SPA-designed dialog system. Thus, within the configuration of the sequence package itself, a new word or phrase, which until then was unknown to the speech system, can be identified and consequently incorporated into the application vocabulary. This is a much quicker approach to learning new words than conventional parsing methods that try to find new semantic relations among sets of words or their morphological equivalents. In short, SPA provides a heuristic approach to speech system design that clearly has a double advantage: it not only permits a speech recognizer to understand natural language input even when the words used by the speaker were not initially included within the speech system's application vocabulary, but it dramatically

shortens the time it takes to develop the grammars that drive these voice systems. And by falling back on sequence packages to help spot OOV (out of vocabulary) words, SPA-driven systems have the potential for a higher rate of speech recognition accuracy than conventional recognizers that are limited to the identification of keywords.

Where Can Sequence Package Analysis Be Used?

SPA technology can be used in a number of different environments with an eye toward its multilingual applications. Here we discuss two of those environments: 1) call centers; and 2) government security.

I. Call Centers

The customer call center, which made its appearance in Western European countries well over two decades ago, has now become an indispensable asset to American businesses. Most of these call centers use some form of an IVR (Interactive Voice Response) system for either routing customers' calls to a human agent or for entirely processing these calls through automation. But with the changeover from menu-driven systems to fully interactive natural language dialog systems, especially those that encourage mixed initiative dialog (dialog that asks for more than one thing at a time), callers may experience difficulty with the system understanding less than perfect natural language input. This is because callers often grope for words, missing the keywords found in the voice application's vocabulary. Callers' speech may also contain circumlocutions, ellipses, hedges, pauses and disfluencies. The end result is that when the system fails to understand the caller, asking the caller to repeatedly state his request, he/she can become frustrated and either "zero out" or hang up.

SPA technology can be used to analyze practically any language.

SPA technology provides one possible solution to "voicemail hell" (the slang termed used to describe getting stuck in an IVR system that fails to properly route or handle the call). In an SPA-driven IVR system, a user's dialog entries would be parsed by grammars that locate specific conversational sequence patterns where important information can be found. The principle operating here is that even if users obscure their requests with the use of vague descriptors, making it hard to locate the crucial information couched in their natural language dialog entries, their conversational sequence patterns can themselves be readily detected. This enables SPA grammars to pinpoint the location in the user's talk (e.g., the middle segment of the utterance) where critical data appear. The dialog system can then go ahead and read back that portion of the talk to the user in an effort to gain a better understanding of what the user is trying to say.

The litmus test for the effectiveness of natural language processing methods is
their capacity to adapt to other languages and dialects.

SPA grammars may be readily applied to other languages, because all languages adhere to a highly organized social architecture discernible in the form of conversational sequence patterns. For example, disfluencies, though they may differ in lexical structure (the Americans say "uh," "um," and the Norwegians say "eh," "m" and "hm"), nevertheless fit the same kinds of

conversational sequence patterns. They are used in the dialog at around the same location points to accomplish the same kind of interactive work. This should not come as a surprise, considering that natural language dialog has been shown to be a well-organized social activity in which disfluencies can be seen as an intrinsic feature of linguistic interaction across many different cultures. [3]

SPA can benefit call center technology in yet another way. Enterprises record thousands of hours of calls between customer service agents and callers, thus creating a gold mine of data for gathering business intelligence about the customer or client. Yet, little of this fertile resource of customer data is ever sufficiently mined to learn useful information about customers' needs and preferences. The reason for this is that when customers fail to use the expected keywords in articulating their complaints and requesting assistance, standard data mining programs experience great difficulty uncovering information about the nature and frequency of customer complaints. Data mining programs that use SPA could surmount this common obstacle. By looking for conversational sequence patterns, an SPA-driven program could parse the audio recordings of human-to-human dialog in much the same way that natural language input would be parsed by SPA-driven IVR systems. The only difference is that, instead of parsing the dialog while the call is in progress, the calls would be mined afterwards to glean important business intelligence data: what are customers really complaining about; are they being properly serviced; and what kinds of products do they prefer?

Here is an example of how this would work. A customer calls his telephone provider to find out about an attractive package of amenities but fails to use certain standard industry words found in the application vocabulary. Rather than inquiring about "call waiting" and "three-way calling," he asks the customer service representative if he can receive "another call" while still on the "present one," and whether he can "get another person on the call?" While the human agent probably would have understood what this customer was asking, a conventional automated mining program would be stumped because the caller failed to use the keywords in the voice application vocabulary. In most cases, the enterprise would therefore be unable to gather this important CRM (customer relationship management) data about the kinds of telephone service products their customers prefer.

The SPA-driven audio data mining program would process this conversation by first identifying the conversational sequence patterns in which this caller made these important requests. Then, the program would add to the speech application's vocabulary all of the words and phrases that the caller used to make the request which differed from the standard industry terms for those same call features. By adding these so-called "alternate" words and phrases, the program would gain the ability to mine future customer calls more effectively, so as to extract important business intelligence data that would have eluded systems limited to spotting standard industry terms referring to products and services.

SPA is an effective tool that can speed up the development time for the construction of grammars that better match how callers truly express themselves across many different languages and dialects. This is so because sequence packages, which are so easy to identify and also generic enough to have multilingual applications, can readily capture callers' rich vocabularies consisting of alternate forms of expression.

II. Government Security

In the wake of 9/11, the NLECTC (National Law Enforcement & Corrections Technology) News Summary profiled SPA as a tool that can "help law enforcement better weed through wire-tapped conversations to learn of possible terrorist plots [by looking] for certain dialog sequences." [4] NLECTC recognized that SPA technology might improve the detection of threats to public safety. The reason for this is that while suspects may try to alter their speech (refraining from the use of certain words that raise red flags or by masking their speech in a special "code"), it is much more difficult for them to alter their conversational sequence patterns. Take the example of two suspects concerned about a possible wiretap speaking with one another about a well-rehearsed plan of attack. While they might refrain from identifying names and locations, they would still show a marked increase in their use of pronouns – a sequence pattern that is often found when speakers go over "familiar" material.

As mentioned earlier, SPA technology can be used to analyze practically any language because of the generic nature of conversational sequence patterns. Even when confronted with a language whose sequence patterns are distinctive rather than generic, such as Arabic or Farsi, the SPA-driven speech system can extract those sequence patterns that are indigenous to that language. In this way, it can locate in the dialog important information that might otherwise have escaped the notice of government agents.

In the last analysis, the litmus test for the effectiveness of natural language processing methods is their capacity to adapt to other languages and dialects. Based on that criterion, it seems clear that SPA will be part of the story of improved natural language processing, helping to make speech systems meet globalization and localization requirements.

References

[1] Neustein, A. (2001). Using sequence package analysis to improve natural language understanding. *International Journal of Speech Technology* 4(1):31-44.

[2] Paprzycki, M., A. Abraham, and R. Guo. (2004). Data mining approach for analyzing call center performance. *The 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*. Ottawa, Canada. In *Lecture Notes in Computer Science*. Germany: Springer Verlag.

[3] Erard, Michael. "THINK TANK; Just Like, Er, Words, Not, Um, Throwaways." *New York Times*, January 3, 2004, Section B, p. 7.

[4] "Linguistics Expert Predicts Voice Technology Will Play Pivotal Role in Spotting Terrorists." NLECTC (National Law Enforcement & Corrections Technology) News Summary, Thursday, October 18, 2001.

Amy Neustein, Ph.D. is the Founder and CEO of Linguistic Technology Systems, a New York-area based think tank for natural language voice applications. Dr. Neustein serves on the Editorial Advisory Board of *Speech Technology Magazine* and on the peer review panel of the *International Journal of Speech Technology*. She can be reached at lingtec@banet.net.