

**Amy Neustein, Ph.D.  
Founder and CEO,  
Linguistic Technology Systems  
1055 River Road  
Suite: 1013  
Edgewater, NJ 07020  
201-224-5096**

**Presentation at AVIOS 2002 (May 10<sup>th</sup>) (paper published in AVIOS proceedings)**

## **Title: Sequence Package Analysis: A Data Mining Tool to Speed Up Wiretap Analysis**

In the wake of the September 11<sup>th</sup> tragedy, the need for newer and more advanced methods for analyzing a jumble of recorded dialog is more urgent than ever. The problem is especially pronounced when suspects engaged in a dialog consciously avoid the use of “key words” that can identify names, places and actions.

Sequence Package Analysis (SPA), a novel approach to natural language understanding first introduced by the author in last year’s *International Journal of Speech Technology*, “Using Sequence Package Analysis to Improve Natural Language Understanding” (Vol. 4, No. 1 pp. 31-44), provides a data mining tool which offers one solution to this problem.

The purpose of SPA is to better understand natural language dialog, particularly when speakers avoid using standard key words (including names, locations and dates) or deviate from the use of standard grammatical forms to express themselves. In cases of terrorism, in particular, one often finds speakers deliberately avoiding the mention of names and places that may arouse suspicion; they often speak in a peculiar *code*, characterized by the use of cryptic terminology.

The Cutter Consortium wrote in *NextInterface This Week* (March 22, 2002) that “one of the benefits that emerging audio mining technologies provide is faster and more efficient monitoring of potential threats in an increasingly security-conscious world.” For example, “security personnel can obtain critical information from hours of recorded phone calls or radio transmissions much more rapidly than before . . . [by combining] speech recognition, language processing, and intelligent indexing and search algorithms to transcribe the content of video or audio broadcasts into computerized text information.”

SPA contributes to these emerging audio mining technologies by offering a natural language processing tool that goes beyond conventional discourse grammars that search for words and word strings. SPA works by examining a series of related turns and

construction units, discretely packaged as a sequence. Such sequences can make up anything from a single ‘episode’ of talk to an entire dialog. Because sequence packages emerge as a dialog unfolds, such sequences can only be discerned via the application of discourse grammars that are flexible enough to arrange the unfolding dialog in the form of sequence packages. In contrast, the application of standard grammars that use preset word strings cannot recognize the individual, idiosyncratic features of dialog, arranged as sequence packages. And certainly in conversations where sensitive information is deliberately masked, these idiosyncratic features are likely to be most prominent.

## 1.1 Advantages of Sequence Package Analysis

One critical advantage of a data-mining tool that identifies the structural organization of sequentially based interactive dialog, as opposed to spotting key words, is its application to many different languages and dialects. Because SPA is able to decipher a speaker’s *conversational sequence patterns* as opposed to simply spotting a preset glossary of lexical items, a wide array of languages and dialects can be analyzed using this method. The unit of analysis is therefore the sequence package rather than a glossary of words that are indigenous to an individual language. Sequence patterns can vary from language to language, but the SPA approach allows for the discernment of the sequentially arranged dialog patterns particular to the language under study. This makes SPA applicable to a multitude of languages. And it is far more scalable to build a discourse grammar that can identify sequence packages in a variety of tongues than to build into a speech recognition engine an infinite lexicon to serve multi-lingual purposes.

Another advantage of Sequence Package Analysis is its capacity to perform audio text mining in real time, rather than merely from prior recordings of conversations. Because the operational capability of a SPA data-mining tool allows for automated data mining while a conversation is occurring, a human analyst can be brought in immediately when high alarm content is produced in the dialog. Given the recent upsurge in wiretapping activities, a new tool that can wade through audio data in real time (as well as after the fact) and determine how best to make use of a human analyst’s time and resources is undoubtedly a valuable asset to any intelligence community. And it is already becoming evident that new and sophisticated tools are greatly needed to manage this appreciably large volume of wiretap data.

## 1.2 Demonstration

The following example shows how applying a SPA approach to wiretapped dialog can flag important security information. The dialog below consists of two speakers planning a covert operation. One speaker is “educating” the other about current plans and operations by referring to a new location at which to meet. (The name of this location is outside of the lexical entries and therefore could not be located by standard grammars.) The speaker who introduces this new location does not want to explicitly highlight the location as a new site by making an introductory comment about its newness or prior unfamiliarity. However, the speaker demonstrates in his well-orchestrated dialog that he clearly recognizes the importance of making sure the other speaker comprehends the

significance of this new site rather than allow it to just pass over his head. What this example demonstrates is that standard grammars with preset key words would not be able to spot this important security information – but that a SPA driven data mining tool would.

Speaker “A”: “Come to the intersection near Juniors?” 0.2-0.5 (the question mark indicates an upward intonation followed by a brief pause, between 0.2 and 0.5 seconds)

Speaker “B”: 1.2 (over a one second pause)

Speaker “A”: “You know the thoroughfare with the big traffic light?”

Speaker “B”: “Juniors, yeah.”

Here is the sequence package found in the dialog above:

- a. noun referent (“Juniors”) marked by an upward intonation that implies the other speaker might not be familiar with this place that is called “Juniors.”
- b. A brief pause (0.2-0.5 seconds), which gives the listener the opportunity to chime in and show recognition or make a clarification request.
- c. Silence by the listener (a pause of one second or longer indicates difficulty in a conversation, usually caused by lack of comprehension or confusion).
- d. Clarification of the noun referent (“You know the thoroughfare with the. . .”) by the first speaker.
- e. Recognition evidenced by second speaker who repeats the noun referent (“Juniors”) that was initially the source of the recognition trouble; followed by a recognitional marker (“Yeah”).

In this example, “Juniors” is packaged within a series of turns whose structural organizational format can be mapped into a discourse grammar. The components of this sequence package (“a-e”) are also generic enough to allow a discourse grammar supplemented by SPA to scan any conversation to find a speaker’s introduction of new names and places. In this fashion, a sequence package approach to data mining permits the uncovering of vital information that exists outside of the lexicon of names and places. In so doing, SPA can actually build a lexicon of critical names and places by using sequence package formulae for analyzing audio text.

## 2.1 Private Industry Applications

Government-sponsored research into audio mining of wiretap communications will inevitably lead to private industry applications. Dr. Chin-Hui Lee (formerly Director of Dialogue Systems Research, Bell Labs, and recipient of the Distinguished Lecturers for the year 2000 by the IEEE Signal Processing Society) sees the value of audio mining in our current industry, pointing to “the increasing bandwidth and the explosive growth in internet traffic [and] more and more multimedia data, including audio, video, image,

speech and text, becoming available on the web and wireless phones, internet appliances and PDAs.” Lee concludes, “since information rather than raw data is what users are looking for, intelligent organization and presentation of such information is critical for the next-generation network services.”

Curt Hall, senior consultant for Cutter Consortium, lists in *NextInterface* (March 22, 2002) five commercial areas for audio mining: 1) technical support centers and help desks; 2) call centers; 3) broadcast media; 4) corporate communications and public relations departments; and 5) conference managers.

A new tool that can uncover speakers’ conversational sequence patterns, such as SPA, is useful both to the government and private industry. Now that we have significantly improved means of organizing large volumes of audio data, new linguistic tools for analyzing these data are sure to find a welcome place in the speech technology industry.