Sequence Package Analysis and Soft Computing: Introducing a New Hybrid Method to Adjust to the Fluid and Dynamic Nature of Human Speech

Amy Neustein

Abstract. At Linguistic Technology Systems, we are using Sequence Package Analysis (SPA) to architect a new, pragmatically-based part of speech tagging program to better conform to the fluidity and dynamism of human speech. This would allow natural language-driven voice user interfaces and audio mining programs – for use in both commercial and government applications – to adapt to the *in situ* construction of dialog, marked by the imprecision, ambiguity and vagueness extant in real-world communications. While conventional part of speech (POS) tagging programs consist of parsing structures derived from syntactic (and semantic) analysis, speech system developers (and users) are also very much aware of the fact that speech recognition difficulties still plague such conventional spoken dialog systems. This is because the inherent inexactitude, vagueness, and uncertainty that are inextricable to the dynamic and fluid nature of human dialog in the real world (e.g., a sudden accretion of anger/frustration may transform a simple question into a rhetorical one; or transform an otherwise simple and straightforward assessment into a gratuitous/sardonic remark) cannot be adequately addressed by conventional POS tagging programs based on syntactic and/or semantic analysis. If we consider for a moment that the biological organism of the human mind does not appear (for the most part) to have much difficulty following the vagarious ebb and flow of dialog with remarkable accuracy and comprehension, so that business transactions and social acts are consummated with a fair amount of regularity and predictability in our quotidian lives, why can't we design spoken dialog systems to emulate the human mind? To do this, we must first uncover the special *formulae* that humans regularly invoke to understand human-to-human dialog which by virtue of its fluid and dynamic constitution is often punctuated by ambiguities, obscurities, repetitions, ellipses, and deixes (indirect

Amy Neustein
Founder and CEO of Linguistic Technology Systems,
800 Palisade Avenue, Suite 1809, Fort Lee, NJ 07024, USA
e-mail: amy.neustein@verizon.net

E. Corchado et al. (Eds.): SOCO 2011, AISC 87, pp. 1–10, 2011.
springerlink.com ©   Springer-Verlag Berlin Heidelberg 2011

referents) – the same stubborn and ineluctable features of natural language which individually and collectively impede the performance of speech systems. Using a unique set of parsing structures – consisting of context-free grammatical units, with notations for related prosodic features – to capture the fluid/dynamic nature of human speech, SPA meets the goal of soft computing to exploit the tolerance for imprecision, uncertainty, obscurity, and approximation in order to achieve tractability, robustness and low solution cost. And as a hybrid method – uniquely combining conversation analysis with computational linguistics – SPA is complementary to artificial neural networks and fuzzy logic because in building a flexible and adaptable natural language speech interface, neural networks, or connectionist models, may be viewed as the natural choice for investigating the patterns underlying the orderliness of talk, as they are equipped to handle the ambiguities of natural language due to their capacity, when confronted with incomplete or somewhat conflicting information, to produce a fuzzy set.

Keywords: Sequence Package Analysis, Part-of-Speech Tagging, Artificial Neural Networks, Fuzzy Logic, Conversation Analysis, Natural Language Understanding, Soft Computing, Voice-User Interface, Audio Mining.

1 Introduction

A decade ago I began to openly address the polemics of deriving programming rules from conversation analysis [1], a rigorous and empirically based method of breaking down spoken communication into its elemental form of conversational sequences and speaking turns (and parts of turns) within those sequences to learn how speakers demonstrate through the design of their speaking turns, their understanding and interpretation of each other's social actions including the wide spectrum of emotions embedded within those actions [2].

On one side of the aisle were those who fervently believed that the fluid, dynamic, and in situ production of human speech made it intractable to the design of simulacra [3,4] because "the prospect of constructing a simulation of ordinary conversation is going to be lacking in procedures for achieving [the] essential feature of projecting turn completion, and thus the management of turn transition will not be arranged in the way that it is in conversation" [5]. Those who held this belief reasoned quite persuasively, in some quarters at least, that because "possible [turn] completion is something projected continuously (and potentially shifting) by the developing course and structure of talk" [6] human dialog was found to be too unpredictable and changeable, moment to moment, to be reduced to a set of programming rules [5].

Juxtaposed to the naysayers was a small, but progressive, group of sociolinguists who drew analogies between the human mind – showing how speakers engaged in dialog are routinely found to "work actively to find meaning for the term that makes [most] sense within [the] context" so that they can effectively overcome the vagueness and ambiguity of human communications caused by the inexorable context-dependent meaning of utterances which gives several possible, and sometimes conflicting, interpretations of the same utterance – and "the grammar a chart parser operates on will [have] alternative patterns against which the [speech] input can be matched" [7].

Those among this progressive group of socio-linguists, looking at the socially competent human organism as a model for the design of natural language-driven speech based interfaces, opined "it is clear conversation analysis must have a role in Natural Language Understanding because there is a sense in which [conversation analysis] is just a small sub field of artificial intelligence" [8]. Their sympathizers, in fact, candidly pointed out that "in order to design computer systems which either simulate, or more ambitiously reproduce the nature of human communication, it is necessary to know about the ways in which everyday (conversational) interaction is organized" [9]. So far, with all the pronouncements of these progressivist socio-linguists, they have yet to introduce a detailed method that shows how best to use conversation analysts' empirical findings on the orderly sequences that emerge as indigenous to the talk to successfully build simulacra that model human dialog.

2 A New Hybrid Method

I couldn't rest easily knowing that conversation analysis which offered a rigorous empirically-based method of recording and transcribing verbal interactions – using highly refined transcription symbols to identify linguistic and paralinguistic features, including some of the most critical prosodic data needed by speech system developers, such as stress, pitch, elongations, overlaps, cut offs, accelerations and decelerations and marked fluctuations in intra-utterance and inter-utterance spacing – had merely in its most elementary form (namely, incorporating some of the basic features of the turn-taking model, such as "barge-in" capabilities) been hearkened by computational linguists, who work closely with speech system designers and speech engineers in building spoken dialog systems. Something had to change, and that meant that the barriers that were keeping computational linguists on one side of the room and conversation analysts on the other had to be stripped down, and for good. No longer could one hide behind the asseveration that the "inferential possibilities of a sentence" were refractory to programming rules [3]. Nor could one be expected to accept with complete credulity that the "rules operating in conversation" are not "codifiable or reducible to an algorithm" either, for that matter [4].

Using a pragmatically-based part of speech tagging program to capture the fluid/dynamic/changeable nature of human speech, Sequence Package Analysis (SPA) meets the goal of soft computing to exploit the tolerance for imprecision, uncertainty, obscurity, and approximation in order to achieve tractability, robustness and low solution cost. As a hybrid method – uniquely combining conversation analysis with computational linguistics (something which has never been done before) – SPA is complementary to artificial neural networks and fuzzy logic because in building a flexible and adaptable natural language speech interface, neural networks, or connectionist models, may be viewed as the natural choice for investigating the patterns underlying the orderliness of talk, as they are equipped to handle the ambiguities of natural language due to their capacity, when confronted with incomplete or somewhat conflicting information, to produce a fuzzy set.

For SPA, the primary unit of analysis is the sequence package in its entirety, rather than an utterance, a sentence or an isolated syntactic part, such as a subject, verb, object [2]. By parsing dialog for its relevant sequence packages, the SPA designed natural language interface extracts important data, including emotional content, by looking at the timing, frequency and arrangement of the totality of the context-free grammatical components that make up each sequence package. As a soft computing method attuned to human-like vagueness and real-life uncertainty, SPA recognizes that natural speech consists more of a blend of sequences folding into one another than a string of isolated keywords or phrases. In keeping with this posture, a sequence package analysis, by virtue of its capacity to map out the orderly sequences that emerge as indigenous to the talk, can therefore be viewed as one way of providing a spoken dialog system with a clear, unambiguous schematic design that makes up the context (in situ construction) of the talk. As such, the goal of soft computing to exploit the tolerance for imprecision, obscurity, vagueness – which in speech may often take the form of repetitions, ellipses, deixes (indirect referents), metaphoric and idiomatic expressions – in order to achieve tractability, robustness and low solution cost, may be better met by spoken dialog programs which employ SPA.

2.1 BNF (Backus-Naur Form)

Using SPA, I have designed a BNF (Backus-Naur Form) table consisting of 70 Sequence Packages – a typology of parsing structures representing the pragmatic (inferential, interpretative, context-dependent, connotative) aspects of communication – that capture the affective data found in natural speech, blogs and emails [2]. Particular attention was paid to the fact that it was precisely these dynamic, fluid, and changeable features of dialog that stirred such strong incredulity among conversation analysts over the construction of algorithms to enable a computer to understand (and replicate) human dialog, that I was at pains to construct a BNF table that would allow flexible pattern recognition and co-existing probabilities so that the fluidity of natural language can be effectively managed, measured, manipulated by the spoken dialog system, rather than hinder its performance.

To accomplish this task, I set out to build a BNF table that while consisting of a set of non-terminals – context-free grammatical units and their related prosodic features for which there is a corresponding list of interchangeable terminals (words, phrases, or a whole utterance) – it also provided for the intricate incremental design of complex grammatical structures from their more elemental units. As such, much of the complexities, subtleties, convolutions, reflexivities, circumlocutions, and intricacies fundamental to human dialog can be more accurately represented by a BNF that has built in multi-tiered grammatical structures – so that natural language dialog systems equipped with such capabilities may exploit the tolerance for the imprecision and vagaries extant in interactive dialog. A "very angry complaint," for example, could be illustrated on this BNF table as the natural accretion of more elemental parsing features – assertions, exaggerations and declarations – so as to effectively notate such pragmatic aspects of communication [2].

Among these pragmatic aspects of human dialog – demarcated by this specially designed part-of-speech (POS) tagging program – are speakers' in situ achievement of one's social status, power, and hierarchical relationship vis-à-vis the other conversational interactant, as demonstrated by the livid customer who reprimands the call center agent for failing to answer his/her service request. In truth, though the angry caller may never use such keywords as requesting a "transfer" to a "supervisor" in his/her interactions with the customer care and contact center agent, the caller's anger/frustration would not elude a pragmatically-based POS tagging program which is built on a typology of parsing structures (whose timing, frequency and arrangement make up distinct sequence packages) which is aimed at exploiting tolerance for obscurity and ambiguity regnant in interactive dialog.

2.2 Domain-Independent

Sequence packages are frequently transferable from one contextual domain to another. What this means is that many of the same sequence package parsing structures (whether they are single or multi-tiered) found in call center dialog may be found, for example, in conversations between terror suspects, doctors and patients, or teachers and students. This is not to say that subject domain would not influence the frequency of the occurrence of certain pragmatically-based parsing structures in spoken communications. For example, in doctor-patient dialog, inasmuch as it is the doctor, and not patient, who directs the dialog in the form of directed questioning, one would find a higher rate of question-answer sequences in medical encounters than in a casual conversation between two friends [10]. Nevertheless, the same BNF table of parsing structures can be used to analyze conversations across many different domains because grammatical units, and their more elaborate arrangements as complex grammatical structures, are generic to human communication.

2.3 Language-Independent

In addition to being domain-independent, SPA is also language-independent. By focusing on the social organization of talk, rather than on a sentence or an isolated syntactic part, this new hybrid method for designing pragmatically-based POS tagging programs may be applied to a wide range of other languages because "all forms of interactive dialog, regardless of their underlying grammatical discourse structures, are ultimately defined by their social architecture" [11].

Thus, in assisting a multitude of other languages to exploit the tolerance for the imprecision, uncertainty and obscurity found in their own regional dialects respectively, this new, pragmatically-based part of speech tagging program helps meet the soft computing goals to achieve tractability, robustness and low solution cost, by employing neural networks which are equipped to handle the ambiguities of natural language because of their capacity, when confronted with incomplete or somewhat conflicting information , to produce a fuzzy set – a group of candidate patterns, each with a known likelihood of being the actual pattern for the representation of the data so far given to it.

2.4 Granularity

It is the characteristic extemporaneity of interactive dialog, and its multiple possibilities for sequence development (e.g. a conversation closing sequence may contain "topical expansion features to totally reopen the talk" rather than close it down [12]; or a help-oriented service delivery sequence may suddenly metamorphose into an inflammatory argument sequence in which the recipient of the proffered help challenges and/or rejects assistance [13, 14]), which makes it imperative for NLU (natural language understanding) algorithms to be guided by probabilities –keeping all of them simultaneously active at all times – rather certainties. Granular computing which works best with soft information, by performing data abstraction and deriving knowledge from that information, is a most natural feature of this new hybrid NLU method for handling the vagaries of talk, because it can effectively comb through a morass of spoken language data marked by characteristic ambiguity, imprecision, and obscurity, to isolate the "granules" of linguistic data that are of critical importance to a business enterprise that is trying to boost customer retention or to a government agency working on increasing homeland security.

Below are a couple of illustrations of how SPA, in its attempt to recognize and exploit the knowledge present in linguistic data at various levels of resolution or scales (making it part of a large class of methods that provide flexibility and adaptability in the resolution at which knowledge or information is extracted and represented) assigns a numeric value to interactive dialog in a customer care and contact center to show the level of agitation of the customer who avails himself/herself of a help-line to fulfill service requests. In these examples, presented below, the calls were answered exclusively by human agents as opposed to an automated call center in which consumer requests are handled by Interactive Voice Response (IVR) systems.

3 Illustrations

The following two examples show how call center dialog achieves its score on the customer anger/frustration index, by adding up the relevant parsing structures that comprise the sequence package of anger/frustration found in the talk. Given the empirical basis of the SPA hybrid method of analyzing natural language dialog, all illustrations are drawn from actual conversations that have taken place in the call center [15]. The examples below are drawn from recordings of a software helpline for some of the earlier versions of the Microsoft Windows program [16]. The punctuation symbols below are purely acoustic and not grammatical: question marks appear mid-sentence to indicate an upward query at that location point in the dialog; no punctuation appears at terminal sentence position unless the inflection has dropped; and if inflection has risen an exclamatory marker is used.

3.1 High Anger Level

Caller: Absolutely unbelievable! What is your? name
Agent: Mr. Smith
Caller: Well! I intend to take this much further…This is just absolutely ridiculous!

In this illustration, above, though the descriptors used ("absolutely unbelievable" "absolutely ridiculous") inhere what is known in as a "high salience value" since they frequently co-occur with the emotion class "anger" (as opposed to a low salience value that is ascribed to more neutral words, such as "continue" or "yes," which do not co-occur with a strong emotion class) [17, 18], there are still no findings of any standard "catch" phrases or keywords in this caller's dialog with the call center agent to signify an irate caller. The caller's exasperation with the customer service agent can nonetheless be detected by tallying up the scores given to individual parsing units. The dialog sample is broken down into parsing structures representing the pragmatic aspects of communication. (For the purposes of this illustration I will not be addressing the smaller grammatical units that make up the larger parsing structures that I indicate below, since it is a given that a spoken language system would naturally identify the smaller units that make up these larger parsing structures.) Each terminal is given its corresponding pragmatically-based POS tagging structure with an associated numeric value, the total of which constitutes the anger/frustration index:

Absolutely Unbelievable! <Exaggerative Qualifier> (8)
What is your? name <Identification Request> (non sequitur; accusatory tone as indicated by displaced (mid- sentence) inflection) (9)
Well! <Exclamation with Prosody> (7)
I intend to take this much further…<Declarative Assertion> (9)
This is absolutely ridiculous! <Exaggerative Qualifier> (8)

Total Score for Customer Anger/Frustration Index: 41

By mapping out the pragmatically-based parsing structures in this dialog, SPA as a hybrid method can take what ordinarily might appear as ambiguous, imprecise, tortuous dialog and find the critical indicia of the caller's emotions (anger/frustration) that portend serious consequences for the enterprise, such as an increased risk to customer retention.

3.2 Moderate Anger Level

Caller: I'm just asking a question..I am just wondering whether or not I should install MS Word

In this second example, as in the first, the caller does not use any catch phrases or keywords to signify an angry/frustrated caller. In fact, in this example, above, he uses the *positive* indicative voice twice (I'm; I am), rather than the negative

("I am not"). A speech system designed to spot negative grammatical constructions for signs of anger or frustration might have overlooked the caller's emotions in this dialog sample, whereas a speech analytic program that performs pragmatic POS tagging would not have been misled so easily by positive grammatical constructions in so far as pragmatically-based speech systems go beyond the elemental grammatical units to explore the incremental arrangement of increasingly complex grammatical structures that are built upon their more elemental parts.

I'm just asking a question <Formulation> (5)
I'm just wondering <Repeat Formulation> (7)
Whether or not I should install MS Word <question> (6)

Total Score for Customer Anger/Frustration Index: 18

While the anger/frustration index in this instant case is less than half the score of the prior dialog example, the speaker's use of two formulations – grammatical devices that permit a speaker to use some part of the dialog to "formulate" or "sum up" the activity he is presently engaged in [19] (in this case example the activity that is summed up is the caller's asking of questions of a help-line desk agent) – in tandem order to one another clearly indicate anger/frustration. The reason for this is that a caller would not ordinarily preface his/her inquiry with "I'm just asking a question, I'm just wondering whether or not" –prefaces that appear more like a declaration than a simple request for help – unless the caller feels that his/her inquiry has not been properly addressed by the call center agent in the first place. For this reason, the formulations present a red flag; the second formulation is given a somewhat higher anger/frustration index than the first, as it indicates escalation in the speaker's emotional state. Moreover, the question that follows the two prefatory formulations is assigned a moderate (to high) level of anger by virtue of its sequential placement following the two formulations, whereas had it appeared in the dialog as a straightforward questions sans a preamble ("Can you tell me whether or not I should install MS Word?"), it would have been assigned the value of "1" – the lowest level on the anger/frustration index.

4 Coda

In the past three decades that I've worked as a socio-linguist, I have witnessed an impressive sea change in the acceptance of artificial neural networks, fuzzy logic, evolutionary algorithms and other major components of soft computing, inasmuch as computer scientists, along with speech system designers, computational linguists and engineers, have begun to acknowledge that real-world problems present with human-like vagueness and real-life uncertainty that demand the flexibility and adaptability uniquely offered by hybrid methods of soft computing whose goal is to exploit the given tolerance of imprecision, partial truth, and uncertainty of any given problem so as to achieve tractability, robustness and low solution cost. Given the fact that soft computing techniques complement (rather than compete) with one another, those in the field of soft computing have set a stellar example that has resonated loudly among member of the hard computing community who have observed how partial truth, imprecision, obscurity, and approximation can be rendered, using the best hybrid soft computing methods, tractable and robust.

Perhaps this is the reason for the sea change in attitude toward acceptance of soft computing methods? If so, I look forward with much alacrity to the next three decades, as I am sure the other members of the soft computing community do as well.

References

1. Neustein, A.: Using Sequence Package Analysis to Improve Natural Language Understanding. International Journal of Speech Technology 4(1), 31–44 (2001)
2. Neustein, A.: Sequence Package Analysis: A New Natural Language Understanding Method for Improving Human Response in Critical Systems. International Journal of Speech Technology 9(3-4), 109–120 (2008)
3. Button, G., Coulter, J., Lee, J.R.E., Sharrock, W.: Computers, Minds and Conduct. Polity Press, Cambridge (1995)
4. Button, G.: Going Up a Blind Alley: Conflating Conversation Analysis and Computational Modeling. In: Luff, P., Gilbert, N., Frolich, D.M. (eds.) Computers and Conversation, pp. 67–90. Academic Press, London (1990)
5. Button, G., Sharrock, W.: On Simulacrums of Conversation: Toward a Clarification of the Relevance of Conversation Analysis for Human-Computer Interaction. In: Thomas, P.J. (ed.) The Social and Interactional Dimensions of Human-Computer Interfaces, pp. 107–125. Cambridge University Press, Cambridge (1995)
6. Schegloff, E.A.: To Searle on Conversation: A Note in Return. In: Verschueven, J. (ed.) Searle on Conversation. Pragmatics and Beyond New Series, vol. 21, pp. 113–128. John Benjamins Publishing Co., Amersterdam (1992)
7. Gilbert, G.N., Wooffitt, R.C., Frazer, N.: Organizing Computer Talk. In: Luff, P., Gilbert, N., Frohlich, D.M. (eds.) Computers and Conversation, pp. 235–257. Academic Press, London (1990)
8. Hirst, G.: Does Conversation Analysis Have A Role in Computational Linguistics? Computational Linguistics 17(2), 211–227 (1991)
9. Hutchby, I., Wooffitt, R.: Conversation Analysis: Principles, Practices and Applications. Polity Press, Cambridge (1998)
10. Frankel, R.: Talking in Interviews: A Dispreference for Patient-Initiated Questions in Physician-Patient Encounters. In: Psathas, G. (ed.) Interaction Competence, pp. 231–262. University Press of America, Washington, D.C (1990)
11. Neustein, A.: Sequence Package Analysis: A New Global Standard for Processing Natural Language Input? Globalization Insider, XIII(1,2) (2004)
12. Button, G.: Moving out of Closings. In: Button, G., Lee, J.R.E. (eds.) Talk and Social Organization, pp. 101–151. Multilingual Matters, Clevedon (1987)
13. Sacks, H.: posthumous publication of Harvey Sack's lecture notes. In: Jefferson, G. (ed.) Lectures on Conversation, vol. 11, p. ix-580. Blackwell, Oxford (1992)
14. Jefferson, G., Lee, J.R.E.: The Rejection of Advice: Managing the Problematic Convergence of Troubles-Telling and a Service Encounter. Journal of Pragmatics 5, 399–422 (1981)

15. Neustein, A.: Sequence Package Analysis: A New Method for Intelligent Mining of Patient Dialog, Blogs and Help-line Calls. Journal of Computers 2(10), 45–51 (2007)
16. Emmison, M.: Calling for Help, Charging for Support: Some Features of the Introduction of Payment as a Topic in Calls to a Software Help-Line. In: Symposium on Help-Lines, Aalborg, Denmark, September 8-10 (2000)
17. Lee, C.M., Narayanan, S.S.: Toward Detecting Emotions in Spoken Dialogs. IEEE Transactions on Speech and Audio Processing 13(2), 293–303 (2005)
18. Schmitt, A., Pieraccini, R., Polzehl, T.: For Heaven's Sake, Gimme a Live Person! Designing Emotion-Detection Customer Care Voice Applications in Automated Call Centers. In: Neustein, A. (ed.) Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics, pp. 191–219. Springer, Heidelberg (2010)
19. Heritage, J.C., Watson, D.R.: Formulating as Conversational Objects. In: Psathas, G. (ed.) Everyday Language: Studies in Ethnomethodology, pp. 123–162. Irvington Publishers, New York (1979)