

SpeechTek 2001 Expo and Conference

White Paper

Why Linguistics is Important for the Design of a Non Fictional “Hal”
By Amy Neustein, Ph.D.

1. Introduction

An artificially intelligent agent equipped with natural speech capabilities such as “Hal” – the computer character from 2001: A Space Odyssey – does not seem far fetched when we consider how the field of linguistics, with its wide spectrum of methods to study interactive speech, provides the building blocks for spoken language systems that simulate human dialog. But how do we get from enticing visions of talking robots to the realistic production of such simulacra?

At the past spring conference and expo of the AVIOS (Applied Voice Input/Output Society), keynote speaker Dr. David Nagel, head of AT&T Labs, emphasized the necessity of marshalling the resources of a variety of different disciplines in order to achieve fully interactive speech interfaces that can simulate human discourse. The panel here today (**Technical Track: Why Linguistics is Important**) accordingly reflects a broad range of disciplines, including computational linguistics, human factor design and conversation analysis.

The need for sound linguistic methods has become ever greater as voice-enabled technology is used more and more in mobile devices, automotive applications, interactive toys and smart-home controls, among other uses. The increasingly sophisticated applications of voice-enabled products and services that allow users to speak naturally, rather than to follow a scripted text of menu choices, make it imperative for software

engineers to draw from the rich field of linguistics to make these systems live up to their potential.

II. The Disciplinary Divide

In my article “Using Sequence Package Analysis to Improve Natural Language Understanding,” which appeared in March 2001 in the *International Journal of Speech Technology* (Vol. 4, Issue 1, pp. 31-44), I discussed two linguistic methods that are often at variance with one another in their approach to the design of spoken language systems: computational linguistics and conversation analysis. Whereas computational linguistics focuses on grammatical discourse structure, conversation analysis focuses on social action and its associated interactional features.

In *Conversation Analysis: Principles, Practices and Applications* (1998) Hutchby and Wooffitt argue for the necessity of using conversation analytic principles in the design of spoken language systems after pointing out the “tendency for computational linguists to discuss aspects of conversational organization...in the abstract, removed from empirical methods” (p. 244-245). Hutchby and Wooffitt, conclude 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” (p. 241).

However, applying conversation analytic research findings in the design of speech interfaces is not without its difficulties. In *Computers and Conversation* (1990), Button contends that the rules operating in conversation are not givens, nor are they finite: they are not “codifiable” or “reducible to an algorithm” but are, instead, “resources” (p. 84) for speakers to discover as their talk becomes, in situ, that of an achieved orderly form of

social activity. This is why there are conversation analysts who are fundamentally opposed to deriving programming rules from conversation analytic findings. In *Computers, Minds and Conduct* (1995), Button, Coulter, Lee, and Sharrock argue, “whilst having a finite set of rules might seem to be a nice convenience for computational linguistics, the fact (is) that in practice the inferential possibilities of a sentence...(are) not constrained in that way...” (p. 176).

Even so, computational linguists are already using bits and pieces of conversation analytic findings, such as the turn-taking model for the allocation of speakership rights. This illustrates the need for the use for conversation analysis in building dialog systems in spite of its inherent difficulties. However, without the assistance of trained conversation analysts, this piecemeal use of conversation analytic research findings by computational linguists may not be adequate for the design of an artificially intelligent, speech-enabled computer with human-like speech recognition accuracy and communication abilities. Thus, it may be the disciplinary conflicts within the broad field of linguistics, rather than the inadequacy of the methods available to us, that impede progress in the design of truly intelligent devices that use natural language understanding.

Against this current of disciplinary conflicts, there have nevertheless been some clarion calls over the years for collaboration between computational linguists and conversation analysts. In a paper presented at the *Thirteenth Scandinavian Conference on Linguistics* in 1992, McIlvenny and Raudaskoski argued “that interactional and linguistic concerns will have to be mutually addressed in computational linguistics. Interactional demands simply cannot be ignored in spoken language artifact design...If we understand computational linguistics in the broad sense of modeling language use and

structure using computers as a tool and with language technology as a product, then it should be clear that interactional concerns are crucial” (*Proceedings of the Thirteenth Scandinavian Conference on Linguistics*, 1992, p. 274).

And surprisingly, even among the skeptics – those conversation analysts who are wedded to the belief that natural language systems cannot possibly simulate human dialog – there are those who strongly encourage system designers in the building of speech interfaces to incorporate critical features of the turn-taking model (e.g., turn transition relevance), which they refer to as “the development of functional equivalents to the organizational activities...engaged in by speakers and hearers” (Button and Sharrock, 1995, p. 122). This is a small step toward an open collaborative relationship between computational linguists and conversation analysts in the interest of designing fully interactive conversational interfaces. But we need much more.

III. The New Frontier

As speech recognition software and speech synthesis become perfected over time, designing systems that simulate human dialog – that might even one day be “human” enough to pass the Turing Test – becomes much less science fiction and more reality. By integrating the many disciplines and sub disciplines in linguistics we can provide a rich corpora of knowledge to serve as the foundation for artificial agents that can perform human tasks.

Allow me to propose three new ways speech interfaces can more closely resemble human dialog via the application of conversation analytic research findings to the existing speech recognition systems built on computational linguistics.

- 1) **Conversational Dialog**: Use of natural speech rather than menu driven voice applications that follow a scripted text. To take this technology to its next level, a system should someday be able to understand a speaker who does NOT use the appropriate key words, such as when a speaker attempts to make an airline reservation, becomes frustrated, but does NOT request transfer to a human “agent” or “operator.” At present, in the absence of the use of those key words (and when there is no manual zero out option) the system would not transfer a frustrated user to an agent for assistance. A system that would indeed understand the interactional signs of user frustration not based on *key words*, but rather based on discrete patterns of sequentially organized conversational features that are consistent with frustration, would bring us closer to a real life “Hal.”
- 2) **Idiomatic Expressions**: Non literal words or word phrases that are used for their symbolic meaning. Everyday language is punctuated by idioms. However, it is far too costly for a system to be equipped with enough intelligence about the world at large to grant accurate meaning to every idiom in the English language. Given that in everyday language speech idioms are rarely taken literally but are rather granted their symbolic meaning, how do we get a speech interface to do likewise? The answer may lie, first and foremost, in the study of how and when idioms are used in interactive dialog. Then, algorithms may be formulated that depict common patterns of usage of idioms so that a system may be able to spot an idiom by virtue of “how” and “where” it appears in interactive dialog.

3) **Empathy**: A display of understanding of what the speaker is trying to convey. True “Hal”-like features someday would include a display of proper empathy by an intelligent agent. When callers seek assistance from human operated help-lines, they often show signs of needing empathy from the human agent. It is at such junctures in a help-line call that a human agent can be most useful by placating an irate caller through showing support for the legitimacy of the caller’s grievance. A natural dialog system that can algorithmically map out the common sequentially organized conversational features of a caller’s attempts to elicit empathy, can more responsively handle a caller’s complaints. A system that is more responsive to the caller’s emotions is a system that can better simulate human dialog.