Symposium: Software Agents with Natural Language Capabilities – Where are we?

Jerry T. Ball
Human Effectiveness Directorate
Air Force Research Laboratory
Mesa, AZ 85212-6061
(480) 988-6561 ext. 457
Jerry.Ball@mesa.afmc.af.mil

At BRIMS 2003, the importance of developing software agents with Natural Language Processing (NLP) and spoken language capabilities was emphasized. Our software agents will not be fully formed until they can communicate with other software agents and human participants in human terms. In this symposium, we will look at the current state of the art and review some projects that are developing software agents with NLP and spoken language capabilities.

Order of magnitude improvements in computing power and memory size, advances in Linguistic and Cognitive theory, the development of large structured knowledge bases and annotated language corpora, and speech recognition and synthesis systems have made the prospects for the successful development of software agents with NLP and spoken language capabilities far more likely than in the past. Unlike earlier efforts which often underestimated the difficulty of building such systems, many of these difficulties have been identified and tools and technologies developed to overcome them.

Listed below are some of the theoretical and technical advances that enhance the prospects for the development of software agents with NLP and spoken language capabilities:

- The advent of Cognitive Linguistics, with its focus on meaning, as an alternative to Generative Syntax, with its focus on linguistic form, independent of meaning
- Extensions to Generative Syntax like Head-Driven Phrase Structure Grammar (HSPG) which explicitly supports the representation of meaning and addresses processing issues
- The introduction of hybrid symbolic/subsymbolic systems capable of modeling a wider range of human behaviors, including verbal behaviors, than purely symbolic or subsymbolic systems
- The successful development of speech recognition and synthesis technologies
- The development and availability of large language and knowledge systems that provide the extensive knowledge needed to support NLP (e.g. WordNet, CYC, FrameNet)
- The development of statistical techniques to support lower level modeling of language (e.g. Latent Semantic Analysis) and the coalescing of statistical and connectionist techniques
- The development of cognitive architectures and modeling environments that facilitate the development of cognitive models with NLP/speech capabilities (SOAR, COGNET, ACT-R)
- The development of virtual environments and interfaces that provide platforms for the creation of interacting software agents

Although much progress has been made, there are still major theoretical and technical barriers that must be overcome before software agents with NLP/speech capabilities become commonplace:

- Successful integration of NLP with speech recognition/synthesis technologies
- Managing the structural and lexical ambiguity of language
- Modeling the context of use of language
- Dealing with non-literal uses of language
- Handling ungrammatical and degraded input and recovering from errors
- Dealing with the massive amounts of encyclopedic and linguistic knowledge involved in comprehension of unrestricted language
- Integrating the theoretical and technological achievements from multiple disciplines into a coherent system

The tools, technologies and theoretical advances listed above are already being used in the development of software agents with NLP/speech capabilities. Several specific projects are reviewed. The presenters will address the following questions as part of their presentation:

- What recent theoretical/technological advances have most influenced your research?
- What major theoretical/technical barriers still remain?
- What major political barriers (e.g. lack of sustained funding) exist?

The AWO Tutor

Bruce Roberts
BBN Technologies
BRoberts@bbn.com

The Air Weapons Officer (AWO) tutoring system developed by Aptima and BBN Technologies under contract to AFRL/AFOSR effectively demonstrates the possibility of developing software agents with NLP/speech capabilities. Flying in an AWACS aircraft with its extended sensor range, an AWO is responsible for advising, monitoring, and directing the actions of pilots during a mission. In effect, the AWO extends and maintains the situation awareness of the pilots, communicating predominately via speech. In the tutoring system, the user assumes the role of an AWO working in concert with one or more groups of fighter aircraft. The user both initiates exchanges and responds to pilot's requests. The pilots themselves are synthetic agents, which are able to communicate among themselves as well as with the AWO; they also exhibit reasonable combat flying skills. A tutor software agent observes the interactions between the AWO and the pilots, along with events unfolding in the simulated airspace, in order to assess the quality and timeliness of the communications and, in some cases, to suggest actions to the AWO in real-time.

The AWO tutoring system combines speech recognition and synthesis technology with language parsing technology in its pilot and tutor software agents that communicate with an AWO trainee. The AWO Tutor is impressive for its breadth of coverage and points the way to the development of systems with more extensive NLP capabilities. The current level of success of the AWO Tutor hinges on the use of a naturally restricted vocabulary and the intentional brevity of communication between pilots and AWOs. Nonetheless, the AWO Tutor is one of the few systems of software agents that attempts to integrate NLP capabilities with speech recognition and synthesis.

Mission Rehearsal Exercise

Jonathan Gratch
Institute for Creative Technologies
Gratch@ict.usc.edu

The Mission Rehearsal Exercise (MRE) project of the University of Southern California's Institute for Creative Technologies is focused on the development of virtual humans—autonomous agents that support fact-to-face interaction in virtual worlds. The project is currently funded by the Department of the Army and was previously funded by the Office of Naval Research. These virtual humans are graphically depicted in a virtual world, communicate with other virtual humans and real humans in natural language using speech recognition and synthesis and language comprehension and generation technologies, perform virtual actions, and even display emotions.

AutoTutor

Natalie Person & Art Graesser
Institute for Intelligent Systems
person@rhodes.edu

AutoTutor is a web-based intelligent tutoring system developed by an interdisciplinary research team at the University of Memphis' Institute for Intelligent Systems. This team is currently funded by the Office of Naval Research and the National Science Foundation. The Tutoring Research Group (TRG) has conducted extensive analyses of human-to-human tutoring, pedagogical strategies, and conversational discourse. This research has provided the empirical and theoretical foundations for developing intelligent tutoring systems that help the students learn by engaging them in a natural language conversation about a particular subject matter. Currently AutoTutor comes in two forms: The computer literacy version is designed to help students learn basic computer literacy topics covered in an introductory course (e.g., hardware, operating systems, and the Internet). The conceptual physics version is designed to help students learn Newtonian physics. AutoTutor can also be adapted for use in a variety of other content domains.

AutoTutor delivers dialog moves with an animated conversational agent and students type in their answers via the keyboard. Dialog moves are selected from a curriculum script that contains a variety of dialog moves for particular topics. An example question in the conceptual physics AutoTutor is “Suppose a boy is in a free-falling elevator and he holds his keys motionless right in front of his face and then lets go. What will happen to the keys? Explain your answer.” Such questions are designed to require about a paragraph of information (3-7 sentences) to answer. However, initial answers to these questions are typically only 1 or 2 sentences in length, even though students have more knowledge that is relevant to an answer. This is where tutorial dialog is particularly helpful. AutoTutor pumps the student for more information, prompts the student to fill in missing words, gives hints, fills in missing information with assertions, identifies and corrects bad answers and misconceptions, answers students’ questions, summarizes answers, and provides feedback (positive, neutral, or negative) that reflects the quality of the student’s input. In essence, constructivist learning is implemented in a mixed-initiative dialog.

Previous analyses of AutoTutor’s dialog moves revealed that AutoTutor adapts to the user’s performance level in terms of dialog move selection. TRG researchers have documented the correlations between the distribution of AutoTutor’s dialog moves and students’ pre-test performance scores. Optimally, students should supply the majority of the information, and the incidence of dialog moves should be Pumps > Hints > Prompts > Assertions. Specifically, high-performing students should receive a higher proportion of pumps and hints (dialog moves that require more from the student), whereas low-performing students should receive a higher proportion of Prompts and Assertions (dialog moves in which the tutor supplies more information). Results indicate that the correlations between the dialog move proportions and the students’ pretest performance scores are indicative of an adaptive tutor (i.e., Pumps, r = .49; Hints, r = .24; Prompts, r = -.19; and Assertions, r = -.40). Thus, AutoTutor generates more pumps and hints for high performing students and more prompts and assertions to low performing students.

SCOTT

Wayne Zachary
CHI Systems
wzachary@chisystems.com

Synthetic Cognition for Operational Team Training (SCOTT) is a team training architecture that integrates intelligent guided practice pedagogy [1] with synthetic environments populated with intelligent interactive synthetic players. An example is the SCOTT E-2 system [2], developed under ONR support, which applied the SCOTT architecture to the problem of cross-platform training for the tactical aircrew on-board a Navy E-2C aircraft. In aviation, as in so many other defense/aerospace environments, verbal transactions are the primary medium of establishing and maintaining coordination both within platform crews and between and among other platforms. Moreover, in such teamwork situations, 'listening in' behaviors, in which players (human or synthetic) gain key contextual information by listening in on multiple verbal transactions among other parties over the voice networks, are just as important as direct verbal behaviors. The synthetic players in SCOTT E-2C were therefore required to interact not only with human players verbally, but also to support 'listening in' behaviors. Listening in, as well as obeying the general protocol used by people for voice-net management, added additional temporal and interactional complexities to the synthetic players. For example, it required them to be aware of what they were (separately) hearing in each ear (which in the E-2C domain may be listening to a separate net), to be aware of when they were speaking (being able to stop in mid-sentence when a higher authority demanded 'silence on the net') and to be aware of how the act of speaking would interfere with the listening process. The fact the multiple synthetic players were active at the same time in SCOTT E-2 also meant that the speech generation had to support the creation of multiple easily recognizable synthetic voices (particularly for the listening in behaviors). Finally, the full understanding of the *verbal* transactions required each synthetic player to develop a full understanding of the *tactical situation* as well, which in turn required a sophisticated reasoning structure, which was developed using the COGNET/iGEN cognitive agent software. The SCOTT E-2 technology has subsequently been reapplied in several other training systems, including the STRATA system being developed under DARPA's DARWARs program.

References

[1] Zachary, W., and Cannon-Bowers, J. (1997) Guided Practice – a New Vision for Intelligent Embedded Training. *Proceedings of Human Factors Society 41st Annual Meeting*. Santa Monica, CA: Human Factors Society, pp. 1111-1112.

[2] Zachary, W., Weiland, W., Stokes, J., Scolaro, J. and Santarelli, T. (2004) Using Synthetic Naturalistic Worlds to Train Teamwork and Cooperation. In S. Schliffit, L. Elliott, E. Salas, & M. Coovert (eds.) *Scaled Worlds*. Surrey, UK: Ashgate Publishing