Next Generation Utterance-Based Systems: What Do Pragmatic Studies Tell Us About System Design?

Kathleen F. McCoy

Department of Computer and Information Sciences, University of Delaware
Newark, DE

Linda Hoag

School of Family Studies and Human Services, Kansas State University
Manhattan, KS

Jan Bedrosian

Department of Speech Pathology and Audiology, Western Michigan University
Kalamazoo, MI

Abstract

Utterance-based AAC systems hold the promise of faster/easier communication. Over the past 10 years, we have conducted a number of investigations into the kinds of pragmatic choices that a user of such a system will have to make and the effects of these choices on public attitudes and conversational behaviors. In this paper, we discuss some of the technological implications of the findings from these experiments.

Introduction

We know that, no matter how good an utterance-based system is, there will at times be a pragmatic mismatch between the most appropriate available prestored utterance and the particular discourse context in which the user finds him/herself. It is impossible to anticipate all text needs. This paper focuses on the underlying technology of utterance-based systems, as opposed to the devices that run these systems, and what is going to allow these systems to work best to promote the user's communicative success. By "technology" we are referring to what is behind the utterance-based system that allows it to deliver those prestored messages when wanted. The technology choices include an organizational scheme for messages that allows desired messages to be navigated to and selected, as well as system support for full-utterance use. An organizational scheme involves how the messages are stored and in what sequence they are presented to the user. As pointed out by Todman, Alm, Higginbotham, and File (2008), different kinds of organizations/technologies of prestored utterances are possible, and we must carefully match appropriate technologies to the kinds of prestored utterances and corresponding communicative contexts of interest. What is appropriate for chatting with friends, for example, is different from what is needed to tell stories or to interact with service providers in public settings.

Our research applies to independent adults who require the use of augmentative and alternative communication (AAC) and are literate and cognitively high functioning. In addition, we focus on face-to-face goal-oriented interactions in public settings with service providers as partners. These situations are arguably ones in which fast communication is most important. For these situations, we have developed and tested a model of the effects of pragmatic trade-off choices when prestored text is used as opposed to slowly constructing messages from scratch. The ultimate goal of this research is to inform the design and development of utterance-based system technology to put the user of such systems in the best public light with respect to accomplishing his/her goals. The purpose of this paper is to discuss the technological implications of our experimental findings.

Motivation for Utterance-Based Technology

There has been a long tradition of research in utterance-based technologies (e.g., Alm & Arnott, 1998; Alm, Neumann, & van Balkom, 2000; Alm, Arnott & Newell, 1992; Alm, Todman, Elder, & Newell, 1993; Dye, Alm, Arnott, Harper, & Morrison, 1998; Higginbotham, Lesher, Moulton, Adams, & Wilkins, 2005; Higginbotham, Wilkins, Lesher, & Cunningham, 2006; Higginbotham, Wilkins, Lesher, & Moulton, 1999; Lunn, Todman, File, & Coles, 2004; Todman, 2000; Todman, 2008; Todman & Alm, 1997; Todman, Alm, Higginbotham, & File, 2008; Vanderheyden, 1995; Vanderheyden, Demasco, McCoy, & Pennington, 1996). There are a number of products on the market that incorporate utterance-based technologies (e.g., TalkBoards, InterAACT, systems with Visual Scene Display capabilities). The promise of this technology is to increase speed of communication. It may also help reduce cognitive load. On the other hand, if care is not taken, such systems may add to both the amount of time it takes to select a message and the associated cognitive load.

To appreciate the problem, first consider speed of communication. It takes time for the user to execute each keystroke. A literate user will be able to communicate in unlimited variety, but at the cost of speed; each word of each utterance takes an average of 4-5 keystrokes. This reality has led to advances such as abbreviation expansion, semantic encoding, word prediction, or whole word selection with single keystrokes. If words, for example, are associated with keystrokes, then, as the number of available words gets too large, an organization is required and a user must navigate between pages of available words. This leads to costs: time and keystrokes spent navigating to the desired word and cognitive load spent remembering what words are stored and how to navigate to them.

These two issues are exacerbated when whole utterances are stored. Indeed, as most recently reiterated by Todman et al. (2008), “What has proved more difficult is finding ways to structure the storage of phrases so they can be quickly recalled when needed” (p. 236). Crucial to the success of utterance-based systems, then, is an organizational strategy that (a) allows users to navigate to the place in the system where the desired prestored utterance can be found (keeping in mind that this will cost keystrokes); (b) once that navigation is done, supports users in communicating a number of turns with few keystrokes (otherwise, the benefit of prestored utterances is lost, due to the cost of navigating between turns); and (3) helps users remember what perstored messages (so that time is spent selecting messages rather than searching for an appropriate message).

While existing technologies attempt to provide solutions to these issues, it is important to note that no solution is going to be perfect; it is impossible to come up with a solution that optimizes each of these issues for every situation for all possible prestored utterances. The question becomes, then, one of optimizing: Can we characterize situations where a series of prestored messages is likely to be successfully used quickly with a minimal amount of cognitive load?

To answer this question, we turned to a research paradigm for investigating the outcomes of using prestored text when the optimum message is not available versus slowly constructing a “better” message from scratch. Can we find evidence through research that will guide us in understanding what combination of technological choices we should make? We believe our work on building and testing a model of pragmatic choices, described in a companion article previously published in *Perspectives* (Hoag, Bedrosian, & McCoy, 2009), is very informative with respect to this task.

Findings

Two of the findings from our experiments (Bedrosian, Hoag, & McCoy, 2003; Hoag, Bedrosian, McCoy, & Johnson, 2004, 2008; McCoy, Bedrosian, Hoag, & Johnson, 2007) are most critical for our purposes here. First, in all the public settings examined, utterance-based system users were rated significantly worse and their communication goals were met with the least success when they quickly delivered a prestored message with partly-relevant information, no matter the degree, as compared to all of the other types of prestored message violations (i.e., messages with not enough information, too much information, or repetitive information). Second, with respect to the types of prestored message violations, ratings of utterance-based system users were either higher, no better, or lower than the messages delivered with delays in order to construct them from scratch, depending on the public setting. Likewise, with respect to the types of prestored message violations, the communication goals of the user were met either with more success, no better success, or less success than the slowly delivered messages, again depending on the public setting.

These experiments have taught us several lessons:

- A system that solely uses prestored text will not work.
- The storage scheme should not make it easy to provide partly relevant messages.
- It must be easy/seamless to edit existing messages or create new messages at the time they are delivered.

Phases of System Use

The following table outlines utterance-based system use and indicates how this use is affected by the System Storage scheme that mediates between the two phases of preparation and use.

Preparation	System Storage	Use
User anticipates text needs and enters a set of prestored messages for storage in the system	**System** must provide a storage scheme to: • Organize messages • Help user in the preparation phase to anticipate text need • Save time in the preparation phase in entering prestored messages • Allow user to access desired message(s) quickly • Help user remember relevant messages during a conversation	• Easy and fast access to **relevant** prestored messages when needed • Ability to quickly edit/augment existing message to fit current needs • Ability to quickly compose new messages when prestored messages are not appropriate

In the preparation phase, the system should make the storage of prestored text as easy as possible. This may require helping the user anticipate situations where prestored utterances are appropriate and guiding him/her in determining what utterances to include and where to put them. This latter issue depends crucially on the organization scheme that is responsible for providing the organization that dictates what prestored utterances are available on a given page and whether and how the user must navigate to get to additional prestored utterances once an utterance is selected.

The use phase of the system should allow the user to quickly navigate to desired messages, but should also help the user remember what messages are available and allow quick editing of existing messages or composing new messages without losing the place in the system. A few systems on the market (e.g., InterAACT) do allow this seamless switching. Our research provides evidence that this is a beneficial feature that could be expanded in future systems.

Implications for Storage Scheme

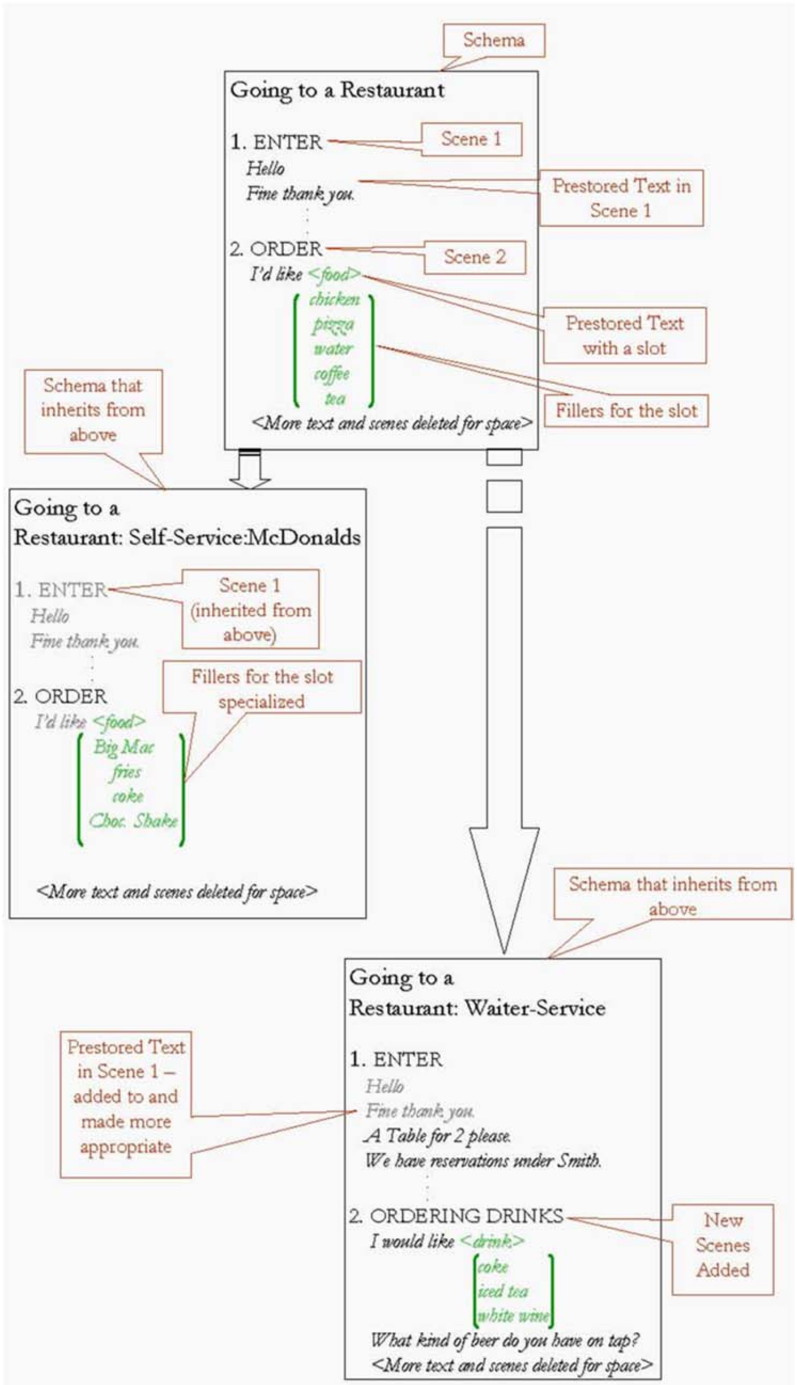
As our experiments have shown, the relevance of a prestored message is the most important aspect with respect to conversational rules. Thus, we must acknowledge that effective prestored message use is going to be limited to situations where a series of relevant prestored messages is available, so that little or no navigation is required between messages. In goal-oriented interactions in public settings, an obvious choice is to organize and store messages appropriate for a prototypical (scripted) sequence of events, such as purchasing an item at a store or ordering a meal at a restaurant. Such settings demand fast communication and constitute domains sufficiently restricted that it is conceivable to anticipate enough text needs to make utterance-based technology necessary. This was the basis for a prototype system called SchemaTalk (Vanderheyden, 1995; Vanderheyden et al., 1996), as well as other systems based on the notion of Scripts (e.g., Dye et al., 1998).

The premise behind the Schema Talk system is derived from Schank and Abelson (1977), who suggested that people develop mental scripts (or schema) from past common experiences that they call up in similar situations. These scripts capture typical sequences of events (each called a scene) that can be expected in a given situation. People use these scripts to determine how to act and how to interpret language in these situations. Schema Talk uses the notion of these schemata to organize prestored utterances. The system represents the scenes in a typical sequence of events (i.e., the first scene, the second scene, etc.) and places in each scene the prestored utterances that might be used in carrying out the action associated with the scene.

Users of the system, when finding themselves in a situation for which they have a schema with prestored text (e.g., going to a restaurant) navigate to the schema (i.e., go to the page that has the first scene and the prestored utterances associated with that scene). They then select any number of prestored utterances from this scene. The utterances are spoken by the device. Depending on how the specific interface is set up and how much text is stored, users might see on the screen utterances from several scenes at once. If they select text from a scene “further along” in the script, the system automatically scrolls so that the prestored messages from the earlier scenes are no longer visible, but prestored messages from scenes beyond the “now” point are. Thus, users would step through the available utterances as they are actually doing the actions, with prestored messages available for quick selection.

The following figure, adapted from Vanderheyden (1995), illustrates the concept. Notice that the figure shows a hierarchical representation. In the preparation phase, previously described, the highest level of abstraction (Going to a Restaurant) might be used to provide the most general utterances useful in any restaurant. In the more specific schema (Going to a Restaurant: Self-Service-McDonalds), the general scenes and the utterances within them will be inherited. By this we mean that the utterances in the most general script would automatically be made available in the more specific instance. The inherited information is shown in gray in the figure to indicate that it would need to be entered only once (at the higher level in the hierarchy). But, notice that, within the McDonalds schema, the user can provide more specific information (in the figure, the food slot has been modified to include items found specifically at McDonalds). The Going to a Restaurant: Waiter-Service schema shows that not only can new utterances be added to an existing scene (see the ordering scene), but new scenes can be added as well (e.g., an ordering-drinks scene appears in this more specific schema).

In the use phase, the users navigate through the pages to the desired schema (e.g., to the schema lowest in the hierarchy that is appropriate for the setting.) Once they navigate to this schema, they are shown the scenes and the available utterances from which they can select.



Consider how this organizational scheme approach differs from, for example, a Visual Scene Display that shows a picture of a waiter-service restaurant and includes prestored utterances located in various hot spots of the picture. The Visual Scene Display is a static representation that shows objects. While messages associated with actions can be stored with the objects (e.g., a request for a table for two can be associated with the doorway to the restaurant), such a representation has limitations. First, it does not help the user remember which messages are stored and how to get to them. This is in contrast to the premise behind a script, which presumably mirrors the way/order that a person thinks about the activities in a given, predictable situation (Schank & Abelson, 1977). Second, complications may arise when an object can be associated with different actions. For example, one might prestore a message for ordering an entrée “behind” the menu. However, the menu might actually be required for ordering several different things—drinks, appetizers, entrée, and dessert—at different times during the meal. Which message should be stored with the menu? How does the user remember where the message is for ordering these various items? In addition, some of the actions in a restaurant are associated with objects that are present earlier in the dining experience (e.g., a menu for ordering); others, with objects that appear only at the end of the meal (e.g., a check for paying). It could be confusing for a user to see both a menu and a check on the table at the same time. Finally, in the figure there are several prestored utterances associated with the ENTER scene in the Restaurant: Waiter-Service schema. Several of these utterances might be used in succession—upon entering a restaurant and then moving to the next scene. It is not clear how to achieve the same effects in an object-oriented representation. An experimental comparison between these two organizational schemes is warranted.

Conclusions

For a number of years, we have been studying what happens when utterance-based technologies are used in the inevitable situation where there is a pragmatic mismatch between the situation in which the user finds him/herself and the available prestored messages. Our goal has been to develop a pragmatic model of communication that will inform the development of future technologies. We have illustrated here one of the lessons on future technology development that our studies have identified.

Acknowledgments

This research was supported by grants (5 R01 DC03670-06) from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders. The authors wish to thank their many consultants and graduate research assistants.

References

Alm, N., & Arnott, J. L. (1998). Computer-assisted conversation for nonvocal people using prestored texts. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 28*(3), 319-328.

Alm, N., Arnott, J. L., & Newell, A. F. (1992). Prediction and conversational momentum in an augmentative communication system. *Communications of the ACM (Association for Computing Machinery), 35,* 46–57.

Alm, N., Neumann, H., & van Balkom, H. (2000). Scripts on an AAC system. In *Proceedings of CSUN 2000,* Northridge, CA.

Alm, N., Todman, J., Elder, L., & Newell, A. F. (1993). Computer aided conversation for severely physically impaired non-speaking people. *Proceedings of the INTERACT ’93 and CHI ’93 Conference on Human Factors in Computing Systems, ACM,* 236-241.

Bedrosian, J., Hoag, L., & McCoy, K. (2003). Relevance and speed of message delivery trade-offs in augmentative and alternative communication. *Journal of Speech, Language, and Hearing Research, 46,* 800-817.

Dye, R., Alm, N., Arnott, J. L., Harper, G., & Morrison, A. (1998). A script-based AAC system for transactional interaction. *Natural Language Engineering, 4,* 57–71.

Higginbotham, D.J., Lesher, G.W., Moulton, B.J., Adams, K., & Wilkins, D. P. (2005). *The Frametalker Project: Building an utterance-based communication device.* Proceedings of CSUN.

Higginbotham, D. J., Wilkins, D. P., Lesher, G.W., & Cunningham, B. (2006, August). *The Frametalker Project: Building an utterance-based communication device.* Paper presented at the meeting of the International Society for Augmentative and Alternative Communication, Düsseldorf, Germany.

Higginbotham, D. J., Wilkins, D. P., Lesher, G. W., & Moulton, B. J. (1999). Frametalker: A communication frame and utterance-based augmentative communication device. *Proceedings of the 22nd RESNA Conference*, Long Beach, CA.

Hoag, L., Bedrosian, J., & McCoy, K. (2009). Theory-driven AAC practices with adults who use utterance-based systems: The case of conversational rule violations. *Perspectives on Augmentative and Alternative Communication, 18,* 103-108.

Hoag, L., Bedrosian, J., McCoy, K., & Johnson, D. (2004). Informativeness and speed of message delivery trade-offs in augmentative and alternative communication. *Journal of Speech, Language, and Hearing Research, 47,* 1270-1285.

Hoag, L., Bedrosian, J., McCoy, K., & Johnson, D. (2008). Hierarchy of conversational rule violations involving utterance-based augmentative and alternative communication systems. *Augmentative and Alternative Communication, 24*, 149-161.

Lunn, J., Todman, J., File, P., & Coles, E. (2004). Making Contact in the workplace. *Communication Matters, 18*, 25-28.

McCoy, K. F., Bedrosian, J. L., Hoag, L. A., & Johnson, D. E. (2007). Brevity and speed of message delivery trade-offs in augmentative and alternative communication. *Augmentative and Alternative Communication, 23,* 76-88.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum.

Todman, J. (2000). Rate and quality of conversations using a text-storage AAC system: A training study. *Augmentative and Alternative Communication, 16*, 164-179.

Todman, J. (2008). *A series of short videos that illustrate the TALK communication aid and the science underlying its development.* Retrieved February 4, 2009, from http://www.dundee.ac.uk/psychology/jtodman/video2.htm

Todman, J., & Alm, N. (1997). TALK Boards for social conversation. *Communication Matters ,11*, 13-15.

Todman, J., Alm, N., Higginbotham, J., & File, P. (2008). Whole utterance approaches in AAC. *Augmentative and Alternative Communication, 24,* 235-254.

Vanderheyden, P. B. (1995). *Organization of pre-stored text in alternative and augmentative communication systems: An interactive schema-based approach* [Technical Report #AAC9501]. Wilmington, DE: Applied Science and Engineering Laboratories. Retrieved February 15, 2009, from http://www.asel.udel.edu/nli/pubs/1995/Vanderheyden95b.ps

Vanderheyden, P. B., Demasco, P. W., McCoy, K. F., & Pennington, C. A. (1996, June). A preliminary study into Schema-based access and organization of reusable text in AAC. In *Proceedings of RESNA '96 19th Annual Conference.*