Beyond Linear Syntax: An Image-Oriented Communication Aid

Rupal Patel, Sam Pilato, and Deb Roy
Northeastern University

Abstract: This article presents a novel AAC communication aid based on semantic rather than syntactic schema, leading to more natural message construction. Users interact with a two-dimensional spatially organized image schema, which depicts the semantic structure and contents of the message. An overview of the interface design is presented followed by discussion of its implications and limitations. Potential benefits of the new design include more fluid, expressive and efficient face-to-face communication for individuals with severe speech and motor impairments across a broad range of ages and linguistic abilities.

Keywords: Human-machine interaction, Communication aid, Semantics, Syntax, Image-oriented composition

Nearly two million Americans who have severe speech and motor impairments must rely on alternative and augmentative communication (AAC) systems to express their needs and desires. AAC aids include physical objects, picture symbols, sign language, alphabet boards, adapted keyboards, electronic interfaces with words and phrases, and a myriad of other cues or devices that facilitate expressive language (Beukelman & Mirenda, 1992). AAC users are a diverse group varying in age, motor and sensory abilities, cognitive abilities and linguistic abilities. The work described in this paper is focused on preliterate AAC users who require image-based communication devices yet whose cognitive and linguistic abilities show promise for significant future gains in expressive communication.

Image-based AAC devices provide users with a set of iconic symbols that can be combined to construct messages. With the introduction of affordable, portable computing technologies, numerous touch screen based devices have been developed that allow users to interactively select multiple symbols to construct messages. Virtually all image-based AAC devices of this kind use a similar strategy of message construction, which is based on the linear word ordering of English. For example, to generate "I want a large ice cream", the user must select symbols corresponding to 'I', 'want', 'large', and 'ice cream' in precisely this linear sequence.

Many AAC users have difficulties with this process of message construction. Their utterances are often limited to simple two-three word sequences (Udwin & Yule, 1990; van Balkom & Welle Donker-Gimbrere, 1996). In addition, the grammatical completeness and accuracy of messages is often impaired. Van Balkom and Welle Donker-Gimbrere (1996) documented that many AAC users employ unusual syntax in their constructions. For example, they may use girl + house + go (subject, object, verb) or house + go + girl (object, verb, subject) when trying to formulate "the girl is going home" (i.e., girl + go + home). We believe that part of the problem is in the message construction process imposed on users of current AAC systems. We do not believe that message construction is most naturally achieved through the linear concatenation of syntactic units. In this paper we describe a significantly different interaction process based on semantic rather than syntactic frames, and which takes advantage of the two-dimensional spatial configuration of icons, enabling a new form of message construction.

Our approach is inspired by the ideas underlying case grammars (cf. Fillmore, 1968). Case grammars focus on the functional relations between the verb of a sentence and other sentence elements. For example, in the sentence "I want a large ice cream" the main verb is 'want' which takes an agent ('I') and an object ('ice cream'), which in turn can take a modifier ('large'). Empirical (Griffin, 1998; Griffin & Bock, 2000) evidence also supports the notion of the verb as the central focus during sentence planning and execution. Structured by case grammar rather than linear syntax, our interface allows the user to construct messages by first selecting the verb, and then specifying the agent, object, and various other verb-dependent message components. The interface is designed for flexibility in the ordering of symbol selection. The case based approach provides a general framework for interaction.

Our second main innovation is in design of the display used during message construction. Again, our goal was to break out of the linear sequencing paradigm. Rather than displaying symbols corresponding to each word in linear order, we have developed a visual language in which thematic roles are translated into two-dimensional spatial relations between symbols (see also Ingen Housz, 1996). For example, the icon symbolizing the agent always appears above that of the verb, and the object appears to the right of the verb. Users can directly manipulate this two-dimensional display to edit and construct messages. The resulting message is a visual depiction of how the various message components interact.

In this paper, we describe the design of an image-based AAC communication aid that enables users to efficiently construct and deliver messages within a semantic schema framework that facilitates communicative expressiveness. Our goals were threefold: to (a) improve communication efficiency, (b) improve communication naturalness, and (c) facilitate improved expressive language skills.

We begin with an overview of the interface design, discuss the individual components, and elaborate on the rationale behind various interface decisions. We discuss the implications of this work on vocabulary selection, communication efficiency and seamless modifications to communication aids through the lifespan. We then discuss some of the obstacles encountered, and some of the planned future directions of this work.

Interface Design: Structure and Function

The communication aid runs on a touch activated tablet computer. It consists of two main areas: a sentence construction workspace, and a set of vocabulary panels (see Figure 1). The user composes a sentence by selecting lexical elements from the vocabulary panels, which the system inserts into the semantic schema in the sentence construction workspace. A second diagram in that workspace depicts the sentence-in-progress in a corresponding linear form, ready for output as text or speech. Figure 1 illustrates a fully constructed sentence: "I want another red cap."

The vocabulary panels are organized into three sections. The leftmost vocabulary panel contains verbs, the middle panel contains lexical categories, and the rightmost panel contains lexical items within a chosen category. The user first selects a verb. The system then displays a semantic template for that verb which is filled by selecting the appropriate vocabulary items from the lexical category and/or lexical item panels.

The interface also displays a set of message parameters, which the user controls to directly affect the contents and expression of each

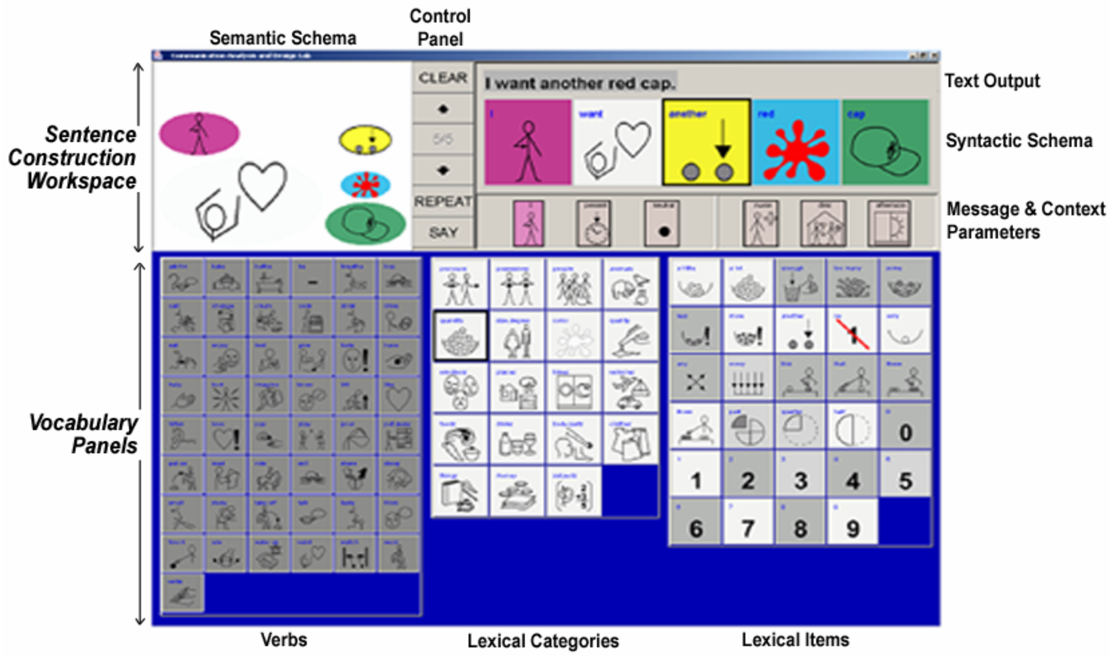

Figure 1. Image-oriented messaging interface.

sentence; a set of context parameters, which track sensed aspects of the communication environment to continually optimize context-specific vocabulary; and a set of messaging controls for working with and delivering constructed sentences.

All system components share a set of design elements. First, all representations of words, phrases, and parameter values are pictorial line drawings, with optional text labels. Second, all interface components and their individual elements have fixed, predictable spatial positions. The visual presentation of the interface can be dynamically adjusted on the basis of predictive algorithms that analyze usage patterns and context. Vocabulary items are differentially shaded along a discrete set of levels that range from white to dark gray, according to each item's predicted likelihood for inclusion in the current sentence frame. Likelihood measures are based on both linguistic and user-specific usage data. In the following sections we elaborate on each interface component.

Sentence Construction Workspace

A semantic schema with fillable slots is the primary focus of attention within the sentence construction workspace (see Figure 2). The user begins message construction by first selecting a verb. The system then generates a unique semantic schema associated with that verb. The pictorial representation of the schema includes the verb as the core meaning of the sentence as well as satellite slots than can be filled by lexical items that fulfill each argument role.

The number and type of argument roles vary across verbs, but each role has a predictable location within the two-dimensional semantic schema as well as a distinct color code. For example, the AGENT role is found to the upper left of the verb image, as a pink-shaded oval.

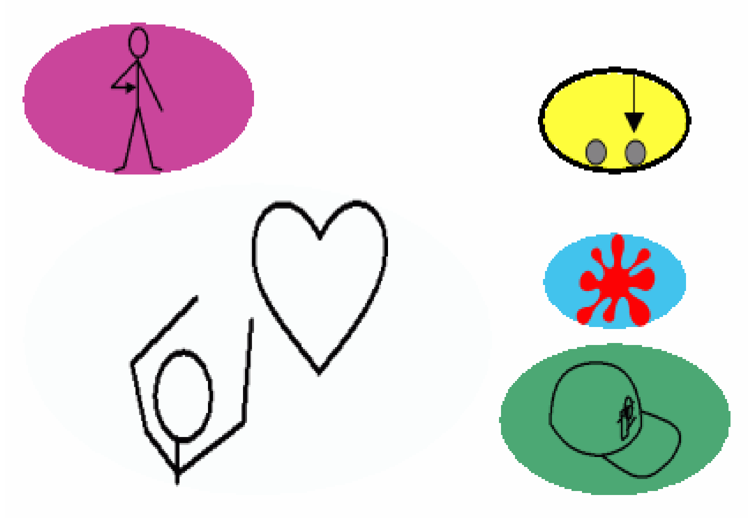

Figure 2. Semantic Schema for the verb 'want'.

To allow more expressive constructions, the semantic schema also includes sub-roles (as smaller ovals) associated with the main role arguments to the verb. For example, an OBJECT role may be filled with a noun, while its QUALITY sub-role might be filled with an adjective that modifies that noun. A black border around an oval slot signifies the current focus. For example, in Figure 2 the yellow COUNT sub-role has the focus. The user may select any slot to change the focus and override the default sequence of content specification.

Once the verb has been selected, the user continues constructing a message by using the vocabulary panels to select a desired category and then a desired lexical item for each role. Each selection fills the role with the chosen lexical item, and advances the focus to a vacant role.

Through the differential shading of the vocabulary items, the system encodes which lexical items within each category are most appropriate for each slot. Particular items are thereby highlighted or darkened — recommended or discouraged — but the user is ultimately allowed to put any lexical item into any slot. The user may opt to fill only some of the slots, and may even actively exclude a slot, whether it is filled or still empty. An excluded slot is depicted as superimposed by a translucent white veil.

A second, synchronized message construction representation parallel to the semantic schema is depicted as a linear sequence referred to as

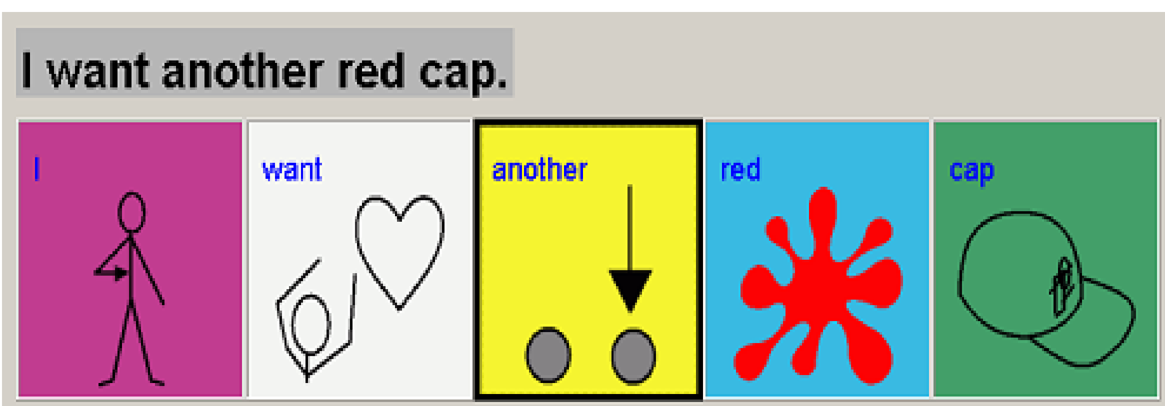

Figure 3. Linear diagram and text corresponding to semantic schema.

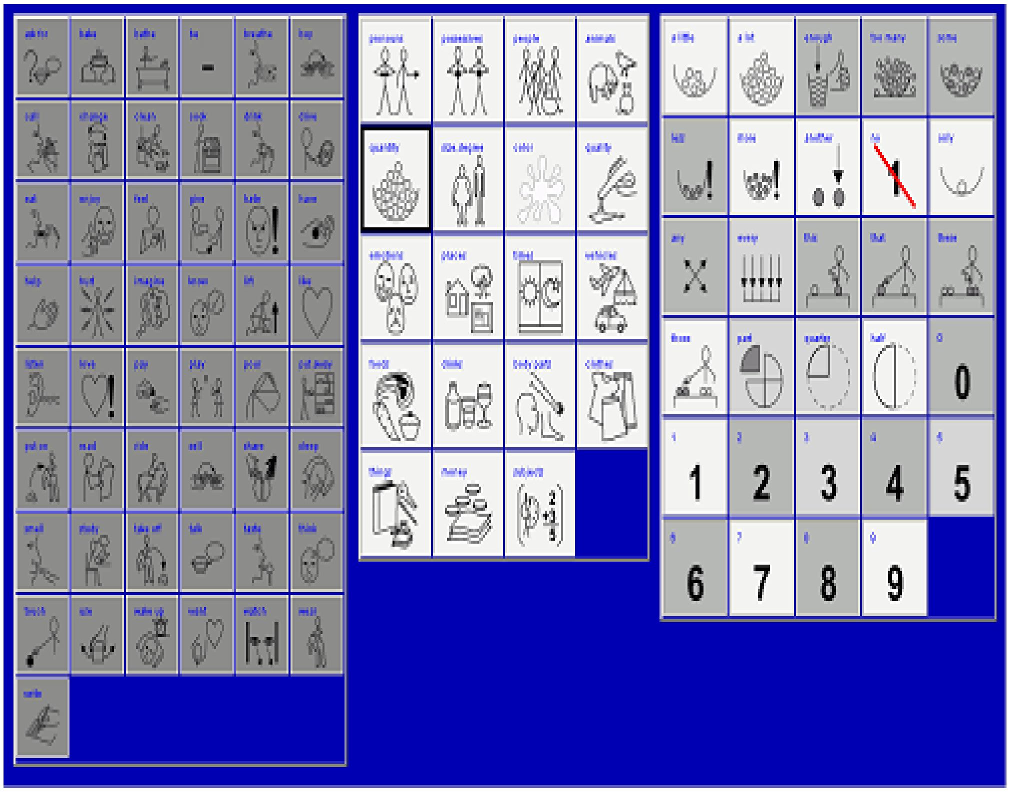

Figure 4. Vocabulary panels for verbs, categories, and the 'quantity' category.

the syntactic schema (see Figure 3). This serves as an intermediate representation between the semantically motivated message construction workspace and the syntax-governed form required for generating text and spoken sentences. The text of the sentence-in-progress is displayed above the syntactic schema.

While the semantic schema does not impose any particular sequence on slot filling, the syntactically-organized linear schema form requires a strict sequence. The user may manipulate either the semantic or syntactic schema interchangeably.

Vocabulary Panels

The verb and category panels have a fixed set of items (see Figure 4). The contents of these panels, however, can be customized to meet the needs of individual users. Once the user selects a category, it is marked by a black border (e.g. the 'quantity' category is selected in Figure 4) and its contents are displayed in a third panel. Items in the lexical panel are the only vocabulary items that come and go over time, as the user changes categories or as the system senses different contexts. The user selects a lexical item to insert into the current role slot. If the slot is already filled, its content is replaced. While the vocabulary panels currently have only two levels, we are exploring novel methods to visualize and navigate through multiply layered vocabulary.

Message and Context Parameters

The user can modify a fixed set of message parameters (i.e. reference, tense, utterance type) that directly affect the contents and expression of each sentence (see Figure 5). For example, when the reference parameter is set to 'I', sentences created with any semantic schema will by default adopt 'I' as the agent. All message parameters are *sticky* in that their values are carried on to subsequent sentences unless explicitly altered by the user (see

Todman, 2000) for time and effort savings benefits of sticky parameters). In addition, a set of context parameters (i.e., location, communication partner, time of day) can be set by the user or sensed by the system. These parameters are an additional means to enhance communication rate and relevance. We have previously developed methods for automatic sensing of situational context (Dominowska, Roy, & Patel, 2002). In the future we plan to integrate these two lines of work.

Session History

The user has access to a complete session history of both delivered and not-yet-delivered sentence workspaces, for browsing, editing, and re-delivery. This access is tightly integrated with the messaging controls for the current sentence workspace. The user may leave a workspace containing a not-yet-delivered sentence, to browse or create other workspaces in the history, and return to it at a later time. If a message has not been delivered, the system auto-copies it before editing to keep a complete work history.

Messaging Controls

The messaging controls allow the user to SAY (deliver via text and speech) the current sentence, or to REPEAT the most recently delivered sentence (see Figure 1). The user may also CLEAR the current workspace. In addition, the user can navigate UP (earlier) and DOWN (later) the session history to reuse previously constructed text or to repeat previous sentences. In future usability testing we plan to assess the added value of recycling sentence fragments and repeating previous text for maintaining dialog and improving communication efficiency and effectiveness.

Design Issues

Many design issues arise when developing a new interface, which pertain to the overall functionality as well as the characteristics and roles of individual components. Addressing these concerns will require extensive, well-designed and executed laboratory and field testing of device learnability and usability. Such testing is of course a long-term and ongoing process of discovery, interleaved with iterative design and development. Nevertheless, at this point we would like to clarify some initial design issues, and some choices we have made that we think will lead us in an informative and fruitful direction.

Semantic Schema

The use of a semantic schema is intended to reduce the linguistic demands of message construction that are imposed by syntactically ordered message construction systems. The aim is to move away from the linear ordering and into the realm of meaningfully structured visual images. Semantic frames provide scaffolding for users to compose complete sentences (cf. Fillmore, 1968; Levin, 1977, 1993; Van Valin, 2004; Kingsbury, Palmer, & Marcus, 2002). We believe that this kind of representation is more accessible to non-literate and pre-literate communicators, yet can also effectively serve linguistically skilled users.

A two-dimensional spatially-organized image can express semantic relationships between words and concepts that are often lost in the

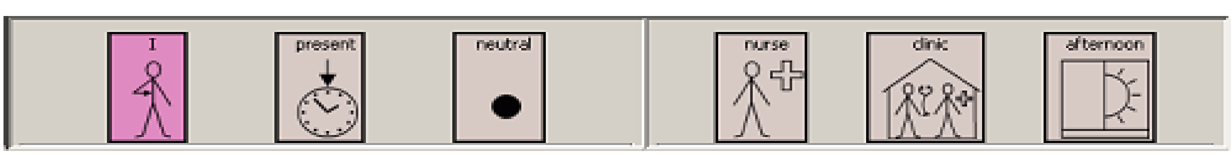

Figure 5. Message parameters (left side) and context parameters (right side).

linear organization of written text. The semantic schema is directly manipulable to give it a real-world "tangibility" which may provide an additional modality of communication.

We are initially working with roughly 50 verb frames, each with up to three main argument roles and up to four sub-roles that modify the main roles. These verb frames were chosen based on projected user needs for face-to-face interaction across a range of social contexts. While we expect the complexity and completeness of message construction to improve over time, our main goals are to promote learnability, expressivity, and communication effectiveness.

Symbol Set

The major lexical elements in our interface are visual symbols accompanied by text. This was an explicit decision in order to serve the needs of non-literate/pre-literate users. Several factors influenced our choice of a particular symbol collection. Within sentence constructions we use different color backgrounds to code roles, and within vocabulary panels we use different grayscale backgrounds. This led to a strong preference for line drawings, and minimal use of color. To reduce the learning curve for the symbol system itself, we decided that the standard symbol set should be pictorial, rather than abstract, and have no strong prior schema for composing elements that may conflict with our own semantic schema design.

We chose to use the Widgit Rebus Symbol Collection (Detheridge, Whittle, & Detheridge, 2002) as our base symbol set. Widgit's line drawings are relatively transparent and systematic in their representation of words and concepts. The collection has substantial field experience behind it, and also includes images for "parts-of-speech" beyond nouns, verbs, and adjectives. We work with a subset of the Widgit Rebus vocabulary, organized into our own categories.

Vocabulary Size and Organization

We currently have a small and simple vocabulary organization, designed to meet our immediate research and development needs. Besides verbs, we provide access to roughly 400 lexical items in roughly 20 categories. As we extend the vocabulary, our intent is to stay in the realm of face-to-face interaction. To this end, we are exploring vocabulary access techniques that minimize extensive navigation or re-arrangement of the visible layout given the increased cognitive burden they impose.

Session History

The session history is an essential feature of the interface given the immense cost of message construction for users of AAC devices. Rather than having to generate novel messages from the ground up, the user may access previous messages that fit their needs and use them as is, or make minor changes before use. Either way, many costly selection actions are saved by the use of an integrated message history buffer. Allowing immediate editing of any image in an on-line session history is a time-saving convenience whose usability and natural feel must be tested.

Input Modality

To adequately support pointing gestures on a touch tablet, we constrained the size of buttons and selectable regions. Furthermore, the geometric layout of elements is informed by common usage patterns. All selection operations are upon discrete elements, to allow the system to accept a variety of input methods. For example, a fully able communication partner might prefer to make selections by point-and-click operations using a standard mouse and screen configuration.

On the other hand, the system can be adapted for users with severe motor control disabilities, who cannot use a touch screen and thus require input from switch-controlled tabbing or scanning interfaces.

Discussion

In this section, we discuss several potential limitations to our approach, ways in which we plan to address these concerns, and future directions of this work. We conclude with a case example of a potential user and a set of testable claims as to the benefits of our interface on the end user.

Limitations and Future Directions

The interface requires some basic level of linguistic and cognitive functioning. While we believe it is less than that required in linear syntactic ordering, the user must nonetheless have symbolic reference and categorization abilities. To ensure that the interface has continued relevance over the user's lifespan, we plan to extend the interface complexity toward simpler and more immediate representations.

The visible vocabulary size of any image-based system is limited by the physical *real-estate* of the display. While layering images would enable access to larger vocabularies, there is an inherent trade-off between size and cognitive demands due to search, navigation, categorization, memory and attention load. While some symbol systems such as Blissymbols (Bliss, 1965) and semantic compaction (Baker, 1982, 1986) facilitate symbol combination, they are dwarfed by the generative power of orthography. We use Widgit Rebus symbols with our schematic layout to provide flexibility of meaning and message complexity from simple sentences through to highly modified and embedded clauses.

Though we try to minimize changes in vocabulary layout, some layering is unavoidable and may be visually distracting to some users. As we further tailor vocabulary subsets to track the changing context, we may make the visibility and placement of items even less predictable. To balance vocabulary and *real estate* trade-offs, we differentially shade items based on likelihood measures where others might spatially reorganize them. This changing matrix of shades, however, imposes its own cognitive load. As a start, we can disable shading or reduce the number of levels, for those users who see it as a distraction rather than a benefit.

In the long run, we envision an AAC device that is highly tuned and responsive to the patterns of activity and situational context of the user. As a step towards this vision, we are developing a set of situational context sensors that will allow the system to respond to real-time changes in the user's communication preferences as a function of sensed context. In this way we hope to emulate how human communication partners use their knowledge of the world and of given situations to facilitate conversation with an AAC user. Access to context-dependent vocabulary will enable users to construct messages about the here-and-now in an efficient manner, thereby increasing opportunities for more natural and satisfying communicative interactions.

Outcomes and Benefits

The AAC interface we have presented is designed with several major benefits in mind for the user. Many of these benefits hinge on our ability to provide a single interface that is accessible across a range of ages, accommodates to changing needs, and promotes and supports developing linguistic and cognitive abilities. Such an interface must be highly scalable to afford a seamless increase in sentence and/or image complexity, vocabulary size, and communicative functions. We provide a case of an example user and a set of testable claims that illustrate the potential benefits of our interface.

Paul is a 10-year-old child with spastic cerebral palsy. Although he cannot read yet, Paul demonstrates only mildly delayed cognitive abilities when compared to age matched peers. His mobility is seriously compromised requiring the use of a powered wheelchair. For the past two years, Paul has relied on a picture-based communication aid in which sentences are constructed by linear ordering of symbols as his primary means of communication. His rate of message construction is slow and labored and he often experiences physical fatigue after prolonged use.

The ease and rate of face-to-face dialog will be improved using ready-made templates, in the form of semantic schemas and Paul's own past constructions. The ability to reuse and recycle fragments and wholesale messages will have a significant impact on the appropriateness and timeliness of his responses. As a result of improved communication rate and appropriateness, family members, teachers, peers and other communication partners may perceive Paul to have greater communicative competence. The consequences of these perceptions are perhaps as real as his abilities.

Message construction will be more natural and easier to learn compared to Paul's current linear composition system. We believe the semantic schema framework emulates the process of message construction during natural message formulation, whether speaking or writing. Manipulating pictorial symbols in a spatially organized schema may provide a more direct link between the message Paul wishes to convey and how he goes about constructing it. For example, when Paul constructs the message, "I want another red cap", he can begin to see the visual correspondence between argument roles and the type of lexical items that can fulfill those roles. To fill in the satellite slots for the 'want' semantic schema, Paul must consider the following questions: Who wants the cap? What kind of cap? Whose cap is it? Does he have a cap like that already? etc. The spatial and color-coded organization of the semantic schema guide Paul in constructing a complete sentence.

The interface also suggests without enforcing, syntactically proper choices through highlighting the most likely lexical items. While the syntactic schema and the text output are useful for message delivery, they also promote Paul's expressive language and literacy skills. Over time he may internalize common patterns across semantic schemas such as the relationships between roles and the lexical items that can fulfill those roles.

Long-term experience with a single interface that grows with Paul's changing needs rather than having to migrate from image-sequencing devices to text-composing devices will have numerous financial, social, and educational benefits. Rather than expending time and energy into learning novel system rules and organization, he can spend his time learning to read and engaging in more fulfilling communicative interactions.

While the above scenario may seem idealistic, we believe it is possible. Usability testing of the interface with AAC users such as Paul is currently underway in our laboratory. Ultimately, generative and creative use of language within the semantic schema framework may better support Paul in achieving socially satisfying communicative interactions.

Acknowledgments

The interface presented in this paper builds on previous software written by Ewa

Dominowska and Simon Shamoun. The material presented is based upon work supported by the National Science Foundation under Grant No. 0083032.

References

Baker, B. (1982). Minspeak: A semantic compaction system that makes self-expression easier for communicatively disabled individuals. *Byte*, *7*, 186–202.

Baker, B. (1986). Using images to generate speech. *Byte*, *11*,160–168.

Beukelman, D., & Mirenda, P. (1992). *Augmentative and alternative communication: Management of severe communication disorders in children and adults*. Baltimore, MD: Brookes.

Bliss, C. K. (1965). *Semantography*. Sidney, Australia: Semantography Press.

Detheridge, T., Whittle, H., & Detheridge, C. (2002). *Widgit Rebus Symbol Collection*. Cambridge, UK: Widgit Software Ltd.

Dominowska, E., Roy, D., & Patel, R. (2002). An adaptive context-sensitive communication aid. In *Proceedings of the 17th Annual International Conference Technology and Persons with Disabilities,* Northridge, CA. Retrieved June 15, 2004, from http://www.csun.edu/cod/conf/2002/proceedings/109.htm

Fillmore, C. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 1-90). New York: Holt, Rinehart, and Winston.

Goodenough-Trepagnier, C. (1994). Design goals for augmentative communication. *Assistive Technology*, *6*(1), 3–9.

Griffin, Z. M. (1998). *What the eye says about sentence planning*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, IL.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274-279.

Ingen Housz, T. (1996). *Elephant's memory: An interactive visual language*. Retrieved February 26, 2004, at http://www.khm.de/~timot/PageElephant.html.

Kingsbury, P., Palmer, M., & Marcus, M. (2002). Adding semantic annotation to the Penn TreeBank. In *Proceedings of the Human Language Technology Conference*. San Diego, CA: Publisher. Retrieved June 8, 2004, from http://www.cis.upenn.edu/~ace/HLT2002-propbank.pdf.

Levin, B. (1977). *Mapping Sentences to Case Frames* (Working Paper 143). Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago, IL: University of Chicago Press.

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

Udwin, O., & Yule, W. (1990). Augmentative communication systems taught to cerebral palsied children: A longitudinal study: I. The acquisition of signs and symbols, and syntactic aspects of their use over time. *British Journal of Disorders of Communication, 25*, 295-309.

Van Balkom, H., & Welle Donker-Gimbrere, M. (1996). A psycholinguistic approach to graphic language use. In S. von Tetzchner & M. Jensen (Eds.), *Augmentative and alternative communication: European Perspectives* (pp. 153-170). London: Whurr Publishers.

Van Valin, R. D. Jr. (2004). *Lexical representation, co-composition, and linking syntax and semantics.* Retrieved February 24, 2004, from http://wings.buffalo.edu/linguistics/rrg/