W²ANE: When Words Are Not Enough

Online Multimedia Language Assistant for People with Aphasia

Xiaojuan Ma, Sonya Nikolova and Perry R. Cook
Department of Computer Science, Princeton University
35 Olden Street, Princeton, NJ 08540 U.S.A.
{xm,nikolova,prc}@cs.princeton.edu

ABSTRACT

In this paper, we introduce W²ANE, an Online Multimedia Language Assistant for individuals with aphasia, a language disorder that affects millions of people. W²ANE offers a rich online multimedia library (OMLA) supported by an adaptable and adaptive vocabulary scaffold (ViVA). The system, accessible over the Internet, provides a platform for applications such as looking up unknown words, constructing phrases for communication, practicing pronunciations, and accessing content. W²ANE also enables resource sharing and remote collaboration.

Categories and Subject Descriptors

H5.4. Information interfaces and presentation (e.g., HCI): Hypertext/Hypermedia.

General Terms: Design, Human Factors

Keywords: Aphasia, multimedia, adaptive vocabulary.

1. INTRODUCTION

Language is essential to communication and information exchange. Over one million people in North America have aphasia, a language disorder caused by a stroke, brain tumor, or brain injuries [14]. The abilities of people with aphasia to produce or comprehend written or spoken English are impaired in different degrees and in various combinations. Thus they are often unable to retrieve information from common channels of communication and information such as daily conversations, newspapers, books, and the Internet.

Current systems for people with aphasia rely on physical devices like laptops and mobile devices that have a number of shortcomings such as portability and synchronization constraints. To address some of these concerns, Boyd-Graber et al. [4] combined a desktop computer and a personal digital assistant to offer flexibility of input, a large display and mobility for communication on the go. We take this idea further by putting the assistive communication content online which provides additional flexibility in finding, using and exchanging content.

In this paper, we introduce W²ANE, an integrated online multimedia language assistant for people with aphasia. It provides a smart multimedia (visual and auditory) vocabulary that aims to enhance the comprehension of information as well as self-expression for people with aphasia. W²ANE consists of an adaptive and adaptable lexical structure, a multimodal vocabulary with image/icon/animation/video/audio-to-concept associations, and web interfaces to access the vocabulary. The W²ANE system also aims to provide a platform for assistive communication applications which enable people with aphasia to find and share information. Aphasic individuals, with assistance from caregivers and speech language pathologists when needed, can work on rebuilding their vocabulary, comprehending phrases, and expressing themselves. The system will also help users to reengage into social settings through information sharing. For example, a group of people with aphasia we worked with are fans of baseball. They created a small library of words enhanced with images representing baseball related concepts in order to practice spelling and pronunciations. In another example, an aphasic individual took many pictures when on a trip to D.C. and he loved sharing them when talking about his experience. With the help of W²ANE, he would be able to tag the photos and upload them online to share his experience as well as the new terms he learned.

MM'09, Octber 19–24, 2009, Beijing, China.

2. RELATED WORK

Thornburn et al. [16] showed that even with the language center partially damaged, aphasic individuals tend to retain their ability to perceive pictorial representations. Thus, although text alone is not sufficient to support communication for people with aphasia, other visual and audio clues can be added for assistance. A number of existing assistive systems utilize pictorial representations for communication and language rehabilitation. For example, Danielsson and Jonsson [6] explored the use of digital photographs as a language for people who have limited or non-existent speech and/or written language. Audio clips and music have also been used for therapy in aphasia rehabilitation [11]. Other examples include [9] and [17]. While all existing assistive communication tools are beneficial in different ways, they share the common drawback of relying on a physical device that constraints the user. Here we mention a few examples.

Lingraphica [10], a communication aid specifically designed for people with aphasia, runs on a laptop computer sold as a dedicated device. It has a rich vocabulary that users can browse to find words and to compose sentences. Because Lingraphica is too bulky to carry around, Boyd-Graber et al. [5] designed a Desktop-PDA system. The desktop component is based on Lingraphica and the mobile component is an extension that allows for communications to be available when people need them outside of their home. The Portable Communication Aid for Dysphasic people (PCAD) [17] also provides mobility, but both systems face other limitations (limited storage, small screen, synchronization).

While building on these lessons, W²ANE differs in that it covers more modalities of multimedia data, not only icons, images, animations, and speech audio, but also videos and non-speech audio. In addition, the vocabulary is organized in a dynamic network based on how the human brain organizes concepts. Finally, it provides an additional mode of flexibility by making content accessible through the Internet, not bound to a particular device, eliminating physical

constraints, and also allowing people with aphasia to share resources and update information online.

3. W²ANE

In this section, we introduce the architecture of W²ANE and briefly discuss its different components. W²ANE consists of three main parts (Figure 1): a library of rich multimedia-word associations called Online Multimedia Language Assistant (OMLA), an adaptable and adaptive vocabulary (ViVA: Visual Vocabulary for Aphasia) which enables efficient vocabulary navigation and word retrieval, and web interfaces for users to navigate and search for library items. Currently, different components are constructed and tested separately and the system is at its early stage of integration.

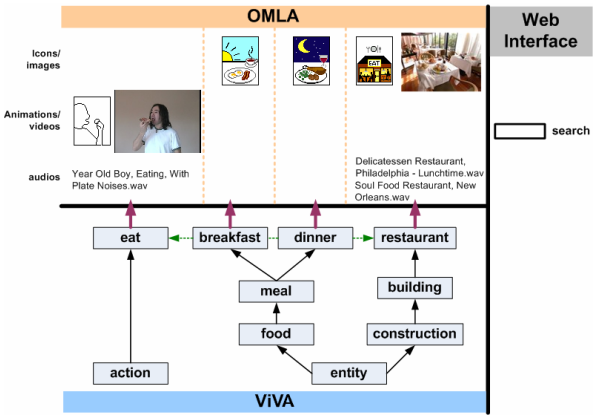

Figure 1. W²ANE architecture.

3.1 Visual Vocabulary for Aphasia

People with aphasia, especially those impaired by anomic aphasia, experience persistent difficulties accessing and retrieving words. To help these users with word finding, we appeal to the psychological literature on speakers' "mental lexicon", where words are stored and organized in ways that allow efficient access and retrieval. We also exploit the structure of WordNet [7], a large-scale lexical database inspired by network theories of semantic memory, as well as a set of word association measures.

Most existing assistive communication tools organize their vocabulary either in hierarchies which tend to be deep and un-natural or in a long list of arbitrary categories. Such vocabulary organization often results in fruitless scrolling, backtracking, and ultimately frustration. To address these issues, we have designed a visual vocabulary for aphasia (ViVA) that is both *adaptable*, customizable by the user, and *adaptive*, able to dynamically change to better suit the user's past actions and future needs. This mixed-initiative approach enables the user to feel in control by making changes and anticipating ones that have been initiated by the tool while still allowing adaptive methods to help determine where and when changes are required. The vocabulary's adaptable component allows the user to add and remove words, group them in personalized categories, and associate existing phrases with a concept. The adaptive component updates the vocabulary organization based on the usage of the system, user preferences and a number of semantic association measures. For example, if the user wishes to compose the phrase "I need an appointment with my doctor" and she searches for *doctor* first, the words *medication* and *appointment* may surface (see Figure 2), because they have been linked to *doctor* due to past usage, while *hospital* and *doctor* could be linked due to prediction based on word association measures. In addition, the user may be able to find the phrase "Need appointment with my doctor" right away if it has been composed in the past. Thus, the vocabulary tailors the word organization according to both user-specific information and general knowledge of human semantic memory.

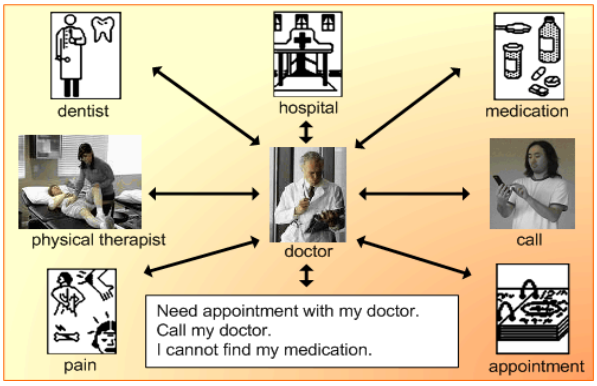

Figure 2. ViVA network overlayed with OMLA.

ViVA currently compensates for some of the missing semantic connections in a user's mental lexicon by incorporating in the vocabulary network evocation, a word association measure that indicates how much one concept brings to mind another. Using machine learning techniques, the structure of WordNet [6], and an initial collection of evocation ratings [2], we generated a list of word pairs that would provide high evocation ratings. Over three months, we collected ratings for 107,550 word pairs through an online experiment published on Amazon Mechanical Turk [2]. We collected ratings from untrained online annotators that correlated well (0.60) with those collected by Boyd-Graber et al. [2] from trained annotators for 90% of the word pairs [15].

3.2 Online Multimedia Language Assistant

The Online Multimedia Language Assistant (OMLA) incorporates a variety of multimedia data that we have collected and evaluated experimentally. These multimodal representations (Figure 3), including images, icons, animations, videos and non-speech audio enhance the words in the vocabulary in order to assist understanding and communication for people with aphasia. The image-concept associations came from the ESP Game dataset [14] which ensures that each image is capable of eliciting the associated word.

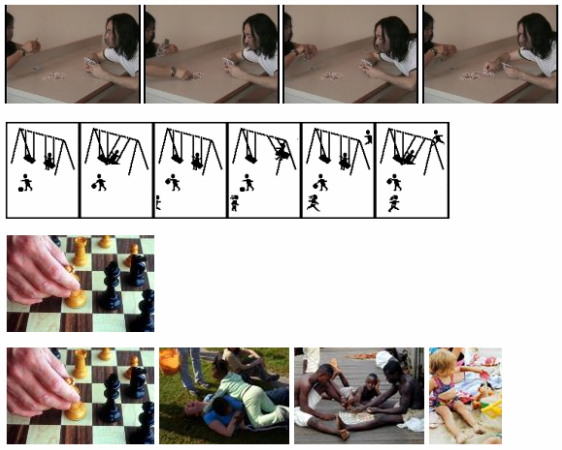

Figure 3. Video, animation, and images for "play".

Image labels from agreement of human players are more accurate than the ones provided by common image search engines (e.g. [9]). For each intended concept, we subjectively selected the image that

best expresses it from a set of ESP images tagged with its synonym set (synset). Although such a technique still requires human intervention, it is less cumbersome than drafting an icon, and could even be automated.

The icon/animation-concept associations were created in a similar scheme. About 2000 icons were derived from Lingraphica. However, since the Lingraphica vocabulary has not been disambiguated, the words linked to the icons were first assigned to the corresponding synsets. New creations and modification were made keeping the Lingraphica style when necessary. For example, some commonly used verbs (such as pay) do not have an animation associated or share the same icon with another concept.
The video-concept (more specifically video-verb) associations were designed by our group [13]. Current online video resources such as YouTube [19] are too noisy, while computer vision databases have cleaner but limited video data. We selected 48 most frequently used verbs and filmed short video clips following the rules: four reviewers picking the best script out of five, single actor with extra hands/feet if necessary, pure background, no sounds or texts, and no lip language. Each video clip is of a length of about three seconds, and applied special effects like fast forwarding if necessary.

Non-speech audio is a new stimulus brought into language support systems for people with aphasia. Little work has been done exploring incorporating it into communication. OMLA intends to provide a set of clean unified short (five seconds) environmental sound clips that are proved to be able to evoke corresponding concepts. The audio-concept associations were constructed based on the BBC Sound Effects Library [3]. We parsed all the words in file names (detailed labels of the sound), disambiguated them, and assigned the audio clips to related synsets. A large scale online study has been completed, providing reliable assessment for the proposed audio-word associations.

3.3 Web Interface and Applications

All of the OMLA data is stored on a server, which users can access through web interfaces that W^{2}ANE provides. One of the basic provided functions is searching (Figure 4). When users type in a word, all the associated multimedia representations will be displayed. Users can browse the images and icons, listen to the sounds, and watch the video. People with aphasia can also practice the pronunciation following the speech sound. If the word has multiple senses in ViVA, the users will be asked to pick the target sense. ViVA also suggests words that are frequently used together with the target word to assist phrase composition.

Another interface to W^{2}ANE is a tree display of the location of a specified word (together with the associated multimedia data) in ViVA (Figure 5). Since each word can have more than one visual/audio representation attached, users can pick a particular mode for display. The tree structure lists hyponyms, hypernyms, and other words that are connected to the target word based on semantic measures such as evocation. Currently, the search and tree structure interfaces are mostly accessible to caregivers, SLPs, and aphasic individuals who still have high verbal functions. We are working on building corresponding aphasia-friendly versions.

We have also built a popup pictorial dictionary with which people can instantly lookup visual representations of unknown concepts on any web pages. A usability test of the dictionary has been conducted with 10 normal able people. The design will be further modified to fit the needs of our target population.

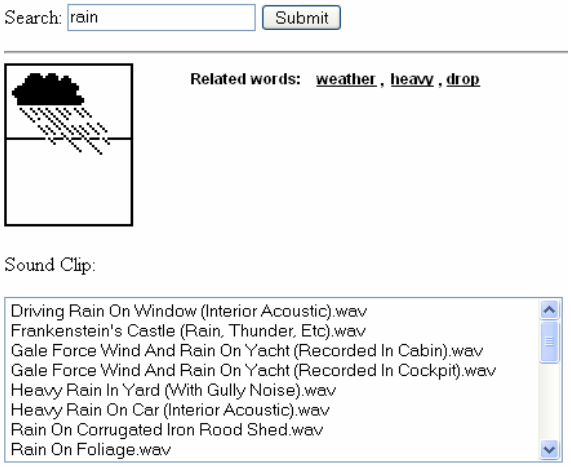

Figure 4. Online search interface.

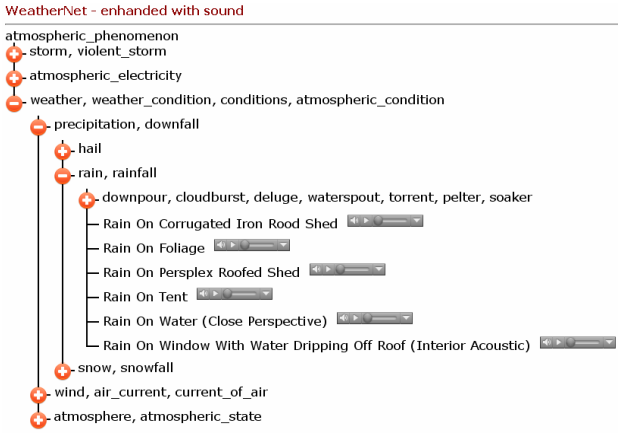

Figure 5. Online tree structure interface.

The dictionary allows users to select what visual stimulus they prefer displayed based on their ability and personal preferences. They can also expand and customize the vocabulary by uploading new representations such as personal photos.

4. EVALUATION

Currently, the two main components of W^{2}ANE, ViVA and OMLA are implemented and tested separately. At this early stage of design and evaluation, we chose to first test the components with normal able people for the following reasons: 1) to better modify and enhance the system design and function, we need to collect user feedbacks early on, which is hard to accomplish with a population with communication difficulties; 2) it is inappropriate to lay the burden of early design decisions on our target population with big variances in their disabilities, since it may be hard to justify if the task failure comes from system flaws or the disability of the users. However, we did try to represent our target users' best interests in many aspects in our experiments. For example, our testing materials and scenarios come from topics that people with aphasia encounter daily, as well as suggestions from SLPs, who have valuable domain knowledge.

As a first step to evaluating ViVA, we assessed its backend adaptive functionality by using simulated usage data in the form of sentences gathered from blogs of elderly people [1]. We compared ViVA's performance in connecting words related due to usage against the vocabulary hierarchy of Lingraphica [10], a popular assistive device

for people with aphasia. First, we trained ViVA with usage data from one blogger's profile, and then examined how it performs given new sentences from the same profile. Connecting words due to evocation and usage resulted in shortening the browsing distances between approximately 40% of the words that appeared next to each other in a sentence from the usage sets. Using logistic regression, we predicted further links between words which improved the results additionally by 8% on average. In addition, 39% of the paths became shorter by two or more steps due to ViVA's vocabulary organization. A naïve baseline test showed that our improvement in shortening the distances between related words could not be achieved simply with a random increase in the density of the vocabulary network [15]. The preliminary evaluation of ViVA's prototype shows the potential of our alternative vocabulary organization to adapt and suggest useful words based on semantic measures and usage statistics. Next, we plan to investigate how users who have aphasia respond to the proposed adaptive vocabulary. To inform the design of ViVA's adaptable component, we are studying a large set of usage data collected from Lingraphica's customer support and will also interview people with aphasia who use assistive communication tools.

To start evaluating OMLA, we examined how well the multimedia data can illustrate and evoke concepts. We first evaluated the efficacy of images and icons in conveying all parts of speech with 24 senior citizens, and conveying nouns with 50 people with aphasia. It was concluded that the ESP images from the web worked as well as stylized icons [12]. Later, we compared four visual modes (a single static image, a panel of four images, an animation, and a video clip) in their effectiveness of illustrating common verbs for 82 young and old adults [13]. Results showed that videos outperformed the other modes, especially for contact and motion verbs. Feedback from the studies also helps us refine our guidelines for creating/collecting multimedia data for OMLA.

As on-going work, we are verifying the non-speech audio-concept associations via Amazon Mechanical Turk [2]. We posted over 300 short sound clips for 184 words and asked participants to answer what is the source of the sound, where they are likely to hear the sound, and how the sound is made. We hope to examine if the sound actually evoke the target concept based on large number of labels collected this way. A follow-up study is being designed to further assess the ability of environmental sounds to persuade meanings in context of daily topics.

5. CONCLUSIONS AND FUTURE WORK

We described the architecture and design of W^2ANE, a smart online multimedia language assistant for people with aphasia. Concepts associated with a variety of multimedia representations are organized in a customizable adaptive vocabulary hierarchy and made available over the Internet. Individuals with aphasia can look up words for information comprehension, communication, and language rehabilitation.

We are currently working on improving and evaluating the different components of the system which will be integrated into a complete functional system. We are considering enhancing W^2ANE with ImageNet [5], an image data base that maps tens of millions of clean web images to WordNet [7]. Now that we have set the foundations of the aphasia-friendly online multimedia library, we will use it to build applications that will allow people with aphasia to find and share information, and collaborate more efficiently with their speech therapists and among themselves.

6. ACKNOWLEDGMENTS

We thank the Adler Aphasia Center, Princeton Senior Resource Center, the Princeton Sound Lab and the UBC Aphasia Group for their help. We are grateful to the Kimberly and Frank H. Moss '71 Research Innovation Fund of Princeton Engineering, and to Microsoft Research for their support.

7. REFERENCES

[1] Ageless Project. http://jenett.org/ageless/. January, 2009.

[2] Amazon Mechanical Turk. https://www.mturk.com. 2008.

[3] BBC Sound Effects Library. www.sound-ideas.com/bbc.html. 2007.

[4] Boyd-Graber, J., Fellbaum, C., Osherson, D., and Schapire, R. Adding Dense, Weighted Connections to WordNet. *Proc. Thirds International WordNet Conference*. Masaryk University Brno, 2006.

[5] Boyd-Graber, J., Nikolova, S., Moffatt, K., Kin, K., Lee, J., Mackey, L., Tremaine, M., and Klawe. M. Participatory design with proxies: Developing a desktop-PDA system to support people with aphasia. *Proc. CHI 2006*, 151–160.

[6] Danielsson, H. and Jonsson, B. Pictures as language. *Proc. International Conference on Language and Visualization 2001*.

[7] Deng, J., Dong, W., Socher, R., Li, L-J., Li, K. and Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. *Proc. CVPR 2009.*

[8] Fellbaum, C. WordNet: Electronic Lexical Database, A semantic network of English verbs. MIT Press, 1998.

[9] Google Image Search. http://images.google.com/.

[10] Lingraphica. http://www.lingraphicare.com/. January, 2009

[11] Lucia, C. Toward Developing a Model of Music Therapy Intervention in the Rehabilitation of Head Trauma Patients. *Music Therapy Perspectives*. 4, 34-39.

[12] Ma, X., Boyd-Graber, J., Nikolova, S., and Cook, P. Speaking Through Pictures: Images vs. Icons. *Proc. ASSETS09 (to appear).*

[13] Ma, X. and Cook, P. How Well do Visual Verbs Work in Daily Communication for Young and Old Adults? *Proc. CHI2009,* 361-364.

[14] National Aphasia Association. http://www.aphasia.org.

[15] Nikolova, S., Boyd-Graber, J., Fellbaum, C. & Cook, P. Better Vocabularies for Assistive Communication Aids: Connecting Terms using Semantic Networks and Untrained Annotators. *Proc. ASSETS'09 (to appear).*

[16] Thorburn, L., Newhoff, M., and Rubin, S. Ability of Subjects with Aphasia to Visually Analyze Written Language, Pantomime, and Iconographic Symbols. American Journal of Speech Language Pathology, 4(4): 174-179, 1995

[17] Van de Sandt-Koenderman, M., Wiegers, M., and Hardy, P. A Computerized Communication Aid for People with Aphasia. *Disability Rehabilitation*, 27(9): 529-533, 2005

[18] Von Ahn, L. and Dabbish, L. Labeling Images with a Computer Game. In *Proc. CHI 2004*, 319-326. ACM Press, 2004

[19] YouTube. http://youtube.com.