Sonya S. Nikolova, **Jordan Boyd-Graber**, Christiane Fellbaum, and Perry Cook. **Better Vocabularies for Assistive Communication Aids: Connecting Terms using Semantic Networks and Untrained Annotators**. *ACM Conference on Computers and Accessibility*, 2009, 8 pages.

Links:

- Ratings [http://wordnet.cs.princeton.edu/downloads.html]
- Slides [http://docs.google.com/present/view?id=d94mvcj_346c5z3p4hq]

Contact Jordan Boyd-Graber (jbg@boydgraber.org) for questions about this paper.

Better Vocabularies for Assistive Communication Aids: Connecting Terms using Semantic Networks and Untrained Annotators

Sonya Nikolova, Jordan Boyd-Graber, Christiane Fellbaum, and Perry Cook
Princeton University
35 Olden Street
Princeton, NJ 08540, USA
{nikolova, jbg, fellbaum, prc}@princeton.edu

ABSTRACT

The difficulties of navigating vocabulary in an assistive communication device are exacerbated for individuals with lexical access disorders like those due to aphasia. We present the design and implementation of a vocabulary network based on WORDNET, a resource that attempts to model human semantic memory, that enables users to find words easily. To correct for the sparsity of links among words, we augment WORDNET with additional connections derived from human judgments of semantic similarity collected in an online experiment. We evaluate the resulting system, the visual vocabulary for aphasia (ViVA), and describe its potential to adapt to a user's profile and enable faster search and improved navigation.

Categories and Subject Descriptors

K.4.2 [**Computers and Society**]: Social Issues—*Assistive technologies for persons with disabilities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Human Factors, Design, Experimentaton

Keywords

Assistive Communication, Aphasia, Adaptive Tools, Semantic Networks, Visual Vocabulary

1. INTRODUCTION

In this section, we briefly review a debilitating language disorder known as anomic aphasia. We then introduce the idea of an adaptive and adaptable vocabulary which we designed to help people with language impairments find words easily. Our work is guided by the idea that technological tools can be effective aids in helping people with language impairments find the words they wish to express. Furthermore, we believe that the effectiveness of such tools can be enhanced by our knowledge of human semantic memory.

1.1 Aphasia

Aphasia, an acquired disorder that impacts an individual's language abilities, affects close to one million people in the United States alone [26]. It is often acquired as a result of stroke, brain tumor, or other brain injuries. The resulting impairments to the ability to understand and produce language vary and affect an individual in any combination. Even though rehabilitation can alleviate the level of impairment, a significant number of people with aphasia are left with life-long chronic disability that impacts a wide range of activities and prevents full re-engagement in life.

Research and commercial efforts have shown that technology has the potential to help individuals with aphasia communicate and thus regain some of their independence and social life. Designing tools for this population, however, is particularly challenging due to the variability of impairments. Thus, some researchers have advocated addressing the heterogeneity of the user population by providing flexible and customizable solutions [7, 17, 28]. Despite efforts to design adaptive assistive tools for elderly and cognitively impaired people, none has been adopted by the majority of aphasic individuals. In fact, most assistive tools for people with aphasia focus on essential therapeutic efforts and the recovery of basic language function. Thus, they do little to leverage the skills of individuals with some residual communicative ability [5]. We attempt to partly fill this void by designing a tool to assists people with anomic aphasia; people who have some remaining communication abilities but experience problems with lexical access.

Assistive communication tools often fail to meet the needs of their users because they do not provide an intuitive and quick way for selecting words when composing phrases for communication [7]. Users, particularly individuals with anomic aphasia, are confused by arbitrary organization of vocabulary terms and the absence of specific words from the given vocabulary. The real difficulty is in providing a flexible system in terms of adding new vocabulary items, adapting to users, and minimizing the complexity of navigating the vocabulary. While we are interested in addressing all of these issues, the work described in this paper focuses on vocabulary navigation.

1.2 ViVA: Visual Vocabulary for Aphasia

Vocabulary expressiveness, organization, and retrieval often hurt the effectiveness of assistive communication devices. Although initial vocabulary sets can be formed from words frequently used by the target population, no packaged system has the depth or breadth to meet the needs of every individual. In addition, most of the existing visual vocabularies have a lexical organization scheme based on a simple list of words. Some word collections are organized in hierarchies, which often leads to deep and non-intuitive searches; others are simply a list of arbitrary categories which cause excessive scrolling and a sense of disorganization. To address these issues, it is important to build a vocabulary that can be easily maintained and enhanced, and that offers improved navigation and search capabilities.

We have developed a multi-modal visual vocabulary that enables the user to compose sentences and phrases efficiently by providing flexibility through both adaptivity and adaptability. The visual vocabulary for aphasia (ViVA) organizes the words in the vocabulary in a context-sensitive network tailored to a user profile that makes finding words faster. ViVA is designed to reorganize and update the vocabulary structure according to user preferences and system usage statistics.

In the rest of this paper, we first discuss some relevant background work. We then describe the design of ViVA and how it incorporates semantic measures to improve vocabulary navigation. We conclude with a discussion of results from an initial evaluation of our adaptive approach to vocabulary organization.

2. RELATED WORK

In this section, we briefly discuss some existing assistive tools for people with cognitive and language impairments. We then introduce WORDNET, a semantic network that serves as the framework for our vocabulary, and a semantic association measure which augments WORDNET and can be used to improve vocabulary organization and navigation.

2.1 Assisting People with Language Impairments

Existing augmentative and assistive communication (AAC) tools for people with language impairments share a number of attributes. They make use of picture-based representation of concepts and often provide a multi-modal interface that combines images, text, and speech audio. Most assistive devices also enable phrase or sentence composition by assembling words in a linear fashion. For example, if a user wants to communicate the phrase "I am hungry," she needs to find the icons for the pronoun "I," the verb "am," and the adjective "hungry" and arrange them in the correct order.

TouchSpeak [27], an AAC software which evolved from the PCAD [28] project, incorporates all of the above-mentioned characteristics, including the emerging trend of giving more control over the system to the user by providing customization capabilities. TouchSpeak offers a hierarchical vocabulary that can be enhanced with images from the user's personal collection or ones that have been taken with the device's camera [27]. Lingraphica [15] is another commercial AAC device which is designed specifically for people with aphasia and is based on the C-VIC and VIC systems [22]. In addition to images, sound and text, Lingraphica's multi-modal vocabulary includes animations for verbs and users can add to the collection personal images, and video and audio clips.

There has also been a concerted effort in harnessing computer technology not only to provide means for rehabilitation in the form of medical training, but also to assists users with language impairments with everyday task and thus enable them to regain some of their independence and improve their quality of life. Research efforts have resulted in systems that support storytelling, email, appointment scheduling, photo management, and cooking [9, 23, 7, 2, 25].

Available AAC tools share one disadvantage: their vocabularies consisting of thousands of words are challenging to navigate [5]. Similar to TouchSpeak, Lingraphica's vocabulary is hierarchical, but it attempts to mimic real-life situations by grouping words according to shared context. If you need to find "milk," for example, you enter the "kitchen," then the "fridge" category, then you find the "dairy," and finally you see the icon for "milk". This organization is not necessarily intuitive for all users since, for example, some people associate the word "milk" with the "drinks" category while others may prefer a "location" association and keep "milk" in the "fridge." This problem is somewhat alleviated by allowing the user to customize the vocabulary categories, but deep hierarchies have other disadvantages as well. The user can, for example, easily get lost browsing which can cause frustration and discourage her from exploring the vocabulary in the future. During the evaluation of ESIPlannerII [7], a daily planner designed for people with aphasia, speech-language pathologists suggested that flat categories are less confusing and easier to navigate, but populating them with sufficient number of words creates the problem of excessive scrolling. Patel et al. [19] designed a communication aid that improves on the traditional linear syntax by representing the semantic content of a phrase in a special schema. Flat categories were still used for organizing the vocabulary and the authors have not reported results from an evaluation of the tool so it is unclear whether their semantic approach to phrase composition affected the aid's effectiveness. No results have also been reported yet on the effectiveness of VocaSpace featured in a new software product Proloquo2Go [4] which targets a wide user group of people who experience difficulties speaking. The VocaSpace vocabulary is rich in words which are organized in functional categories such as greetings and questions, and common word categories such as colors, places, and clothes. The vocabulary can be extended and personalized, and provides phrase starters such as "I want to" and "I need" [4].

2.2 Speaker's Mental Lexicon

To address the problem of cumbersome vocabulary navigation, we base ViVA's vocabulary structure on theories that explain how the human mind organizes words. We fist appeal to the psychological literature on speakers' "mental lexicon," where words are stored and organized in ways that allow efficient access and retrieval. Every speaker has experienced the inconvenience of temporarily impaired semantic connections (the so-called tip-of-the-tongue (TOT) phenomenon). This inability to retrieve a specific word needed to express a given concept can be due to a variety of causes such as fatigue or interference from a word that is morphologically or phonologically similar to the target word. People with anomic aphasia can be thought of as experiencing a chronic and severe case of TOT, as they have persistent diffi-

culties accessing and retrieving words that express intended concepts.

Experimental evidence — including evidence from TOT states induced in the laboratory — suggests that words are organized in a speaker's mental lexicon by various similarity relations, in particular phonological and semantic similarity. For example, subjects in word association experiments overwhelmingly respond with *husband* to the stimulus *wife* [18]. Semantic priming [24], a robust and powerful tool for the experimental investigation of cognitive processes, relies on the semantic relatedness of the prime and an experimental target: responses to the target are faster when it is related to the prime as in the classic case *doctor-nurse*. Spreading network activation models [8] assume that presenting a prime stimulus word activates the corresponding representation in lexical memory and that this activation spreads to other related nodes, thus facilitating the processing of related target words. The semantic network WORDNET [16, 11] is a large-scale lexical database inspired by network theories of semantic memory that accommodate the spreading activation paradigm among related words and concepts. Taking advantage of the knowledge encoded in WORDNET, we attempt to build a system that can compensate for some of the missing semantic connections in a user's mental lexicon.

2.3 WordNet and Evocation

WORDNET has a rich structure connecting synonymous words, called synonym sets or synsets, to one another. Depending on the part of speech, synsets are interlinked according to specific meaning relations. For example, verb synsets are connected by a variety of lexical entailment pointers that express manner elaborations `[walk]`-`[limp]`, temporal relations `[compete]`-`[win]`, and causation `[show]`-`[see]` [11].[1] The links among the synsets structure the noun and verb lexicons into hierarchies, with noun hierarchies being considerably deeper than those for verbs.

Despite these connections, WORDNET's internal density is insufficient– there are too few cross-part-of-speech links and connections among the synsets. Boyd-Graber et al. [6] enhanced WORDNET with thousands of new links based on the measure of "evocation", i.e. how much one concept brings to mind another. Evocation aims to add cross-part-of-speech links that allow for connections among entities (expressed by nouns) and their attributes (encoded by adjectives); similarly, events (referred to by verbs) can be linked to the entities with which they are characteristically associated. For example, the intuitive connections among `[traffic]`, `[congested]`, and `[stop]` can be clearly conveyed using evocation. We intend to exploit the structure of WORDNET enriched with evocation to improve navigation and speed up word finding.

3. THE DESIGN OF VIVA

All users tend to rely on consistency and stability within an interface. This dependency is even more pronounced among users with cognitive impairments. In order to address this concern and still achieve our goal for ViVA to be flexible, we explore a mixed-initiative approach to customization, an

[1] Throughout this article we will follow the convention of using a single word enclosed in square brackets to denote a synset. Thus, `[dog]` refers not just to the word dog but to the set – when rendered in its entirety – consisting of `{dog, domestic dog, canis familaris}`.

effective blend of automation and direct manipulation [12]. ViVA is both **adaptable**, able to be customized by the user, and **adaptive**, able to dynamically change to better suit the user's past actions and future needs. We believe that this approach will enable 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 two main components of ViVA, the user preference module and the active learning module, are shown in Figure 1. The user preference module allows the user to add and remove vocabulary items, group them in personalized categories (for example a "Favorites" folder or ideas related to "Family"), enhance words with images and sounds, and associate existing phrases and sentences with a concept. In addition to practical concerns of having sufficient vocabulary terms to express the needed concepts, the ability to customize a system invests in the user a sense of ownership and empowerment. This attachment to the system, brought about by a sense of accomplishment, is an important aspect of the rehabilitation process [2]. We do not elaborate on the design of the adaptable component in this paper, and its evaluation is part of our future work. Instead, we concentrate on describing the adaptive side of ViVA and we present results from an initial evaluation of this component.

The adaptive component, the active learning module, updates the vocabulary organization based on the usage of the system, user preferences, and a number of semantic associations. For example, if the user wishes to compose the phrase "I need an appointment with my doctor" and she searches for `[doctor]` first, the vocabulary network centered on `[doctor]` may look like the one shown in Figure 1. The links between the words may exist because the user has previously composed sentences using `[doctor]` and `[medication]` or using `[doctor]` and `[appointment]`. The concepts `[hospital]` and `[doctor]`, for example, may be linked because of a prediction based on known synset association measures and usage. In addition, the user may be able to find the phrase "Need appointment with my doctor" right away if she had already composed it in the past. In order to have words related to the context of the communication surface faster, we account for the links that the user has created in the past by assembling words into phrases. In addition, we introduce links suggested by three word association measures— evocation, word dependency, and word proximity. Based on these features, ViVA also predicts a set of words that could assist a user in phrase composition. Thus, we integrate a level of intelligence into the vocabulary organization informed by user-specific information and general knowledge of human semantic memory.

3.1 Core Vocabulary

Customization of the vocabulary can be a powerful feature, but we still need an initial organization to allow the user to successfully use ViVA from day one. We selected ViVA's initial vocabulary set such that it is a collection of commonly used words as well as ones relevant to our target population, people who have aphasia. The vocabulary comes from two sources: the "core" WORDNET consisting of frequent and salient words and the visual vocabulary of an assistive device for people with aphasia created by Lingraphicare [15]. We used all synsets from the core 1000 synsets used in the experiment by Boyd-Graber et al. [6],
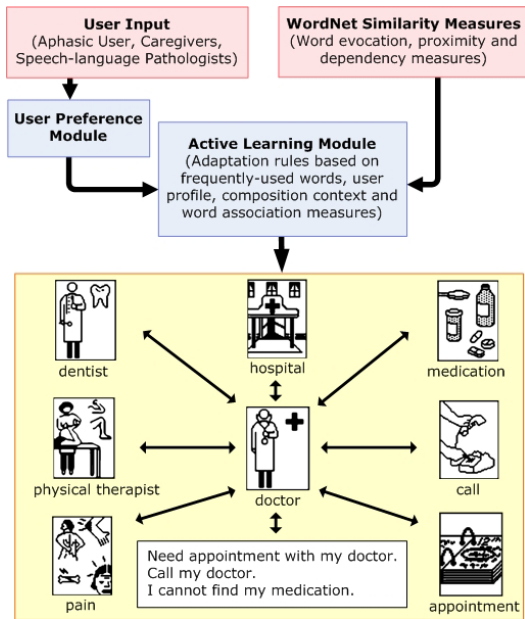
Figure 1: Schematic of components of a system to assist individuals with aphasia.

Figure 2: Example stimulus for collecting evocation ratings. A user rates 50 pairs in a single sitting.

all verbs in Lingraphicare's vocabulary, and all nouns and adjectives in both Lingraphicare's vocabulary and the core 5000 synsets.

Lingraphica's vocabulary represents each concept with a triplet of an image, a sound clip, and the corresponding text. This allowed us to apply a form of coarse disambiguation. For each concept in Lingraphicare's vocabulary, we selected the corresponding concept from WORDNET to create a single, unified representation of the vocabulary.

3.2 Augmenting the Vocabulary with Higher Evocation Ratings

While the evocation set created by Boyd-Graber et al. [6] provided us with an initial collection of human semantic association ratings, many of those ratings were zeros since the synsets were selected and paired randomly. We used this initial set to generate a list of word pairs that were likely to result in higher evocation ratings. To collect ratings for this new set, we adjusted the original experimental setup so that it is less laborious and expensive.

Many natural language processing tasks require human annotation that is expensive and time-consuming to collect on a large scale. Recently, a new and more efficient method for collecting sizeable sets of inexpensive annotations from a broad pool of human contributors has emerged. Snow et al. [21] demonstrated that labels acquired through Amazon Mechanical Turk [3] from non-expert annotators are in high agreement with gold standard annotations from experts. The positive results of their work motivated us to collect the evocation ratings that we needed using the same online tool.

Amazon Mechanical Turk (AMT) is an online community that provides access to a vast pool of individuals, called workers, who complete tasks for a monetary reward. The people who post work on AMT, called requesters, design a task and rules for completing it, determine the reward, and release it in the pool of available work. Workers browse the
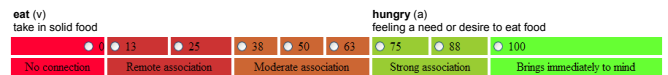
work pool and select the tasks they wish to do. Once a task has been completed, its requester can approve or reject the results. The approval rate, which provide a minimal level of quality control, is part of a worker's profile visible to all requesters. It is represents how often the individual's work has been found satisfactory.

3.2.1 Method

A machine learning algorithm selected the synset pairs to be rated by the AMT annotators. We used many of the features found to be predictive of evocation including those based on WORDNET connectivity [13], pointwise mutual information based on words appearing in the same sentence, and context similarity. We duplicated high evocation pairs (having a median rating of greater than 15) to create a high-recall training set, trained a classifier using AdaBoost [20], and then took the subset of all pairs of synsets in our vocabulary labeled as having a high predicted evocation by our learning algorithm. These pairs were the ones selected to be rated via AMT.

We created two thousand human intelligence tasks (HITs in AMT jargon) consisting of 50 pairs each. We asked that 15 people complete each task. The design of the template that we posted on AMT provided anchor points on a scale from 0 to 100 to be used for rating evocation (see Figure 2). Raters were first presented with the following set of instructions:

1. Rate how much the first word brings to mind the second word using the provided scale.
2. The relationship between the two words is not necessarily symmetrical. For example, "dollar" may evoke "green" more than the reverse.
3. Pay attention to the definition of the words given on the second line; words can have more than one meaning. For example "dog" (the animal) would not bring to mind "bun" (the piece of bread you serve with a hot dog).
4. The letter in parenthesis signifies whether the word is a: an adjective, n: a noun or v: a verb.
5. Don't use information from your personal life. For example, if you had a dog named "bog" you personally would associate "bog" and "dog," but the average person wouldn't.
6. Don't use the spelling of words to make your decisions. For example, even though "bog" and "dog" rhyme, they are not associated.
7. We cannot offer you a generous reward for your time, but we greatly appreciate your sincere effort. There are a few pairs with known average ratings embedded in the HIT. If your ratings for those pairs do not fall within generously set acceptance bounds, we will have to reject your responses. On the other hand, you will receive a BONUS of $0.02 for each response set that falls within our bounds of known correct answers.

The last instruction was included to forewarn annotators that sloppy contributions such as clicking all zeros will not be rewarded and to encourage them to invest some additional effort for a small bonus. We embedded five checks, unknown to the annotators, in each task which were later used to

Filtering Method	Correlation with Mean	Correlation with Median	Number Pairs
All Checks	0.54	0.54	46900
Most Checks	0.45	0.43	55400
Some Checks	0.37	0.34	56850

Table 1: Correlation of the mean and median against evocation annotations collected by trained undergraduate annotators.

determine the validity of the gathered results. Annotators were paid $0.05 to complete a task.

3.2.2 Results

We collected ratings for 107,550 synset pairs over a period of three months. The average time to complete the task of rating 50 pairs was 3.6 minutes, resulting in an average pay of $0.74 per hour (an order of magnitude less than painstakingly trained undergraduate annotators [6]). To ensure the quality of the ratings and a consistency with previous results, we used embedded checks to decide which submitted tasks were valid. The ratings for four of those checks were collected from the dataset available from Boyd-Graber et al. [6]. The fifth check required annotators to rank a pair consisting of the same synset, for example [help] and [help]. We ran three different reliability tests depending on the number of checks we wanted satisfied. If the annotator's rating for the fifth check was 100 and a number of the remaining checks were met within certain acceptance bounds, the annotations were considered valid. The acceptance bounds were defined as follows.

As in the task, the scale of 0 to 100 was split into 5 intervals, $\{[0 - 10], [10 - 30], [30 - 70], [70 - 90], [90 - 100]\}$. If an annotator's rating fell within the same interval as the corresponding check or in the upper half of the immediately lower interval or the lower half of the immediately higher interval, the rating was considered valid. The first reliability test required **all** checks to be met. For this set, 43.4% of the pairs were rated as having no association and 2.7% fell in the category immediately brings to mind. The second reliability test required **most**, three or more, checks to be met in addition to satisfying the complete-evocation check. The final and most relaxed reliability test required **some**, two or more, checks to be met in addition to the complete-evocation check. Table 1 shows the number of synset pairs for each of the reliability levels, and Table 2 has explicit examples of mean evocation ratings for the three levels. Finally, Table 1 shows mean and median correlation of the three reliability sets against the ratings provided by undergraduate students in [6]. As expected, the results are noisier when fewer checks are applied. The set of synsets where all checks were met results in the highest correlation to the original evocation data. This correlation on a very difficult task is sufficient to show that with good quality control, gathering ratings through AMT was a valid approach. While AMT annotators seemed to rate on average evocation lower than the trained annotators(Figure 3), ratings from the untrained online annotators correlated well (0.54) with those collected by trained undergraduate annotators.

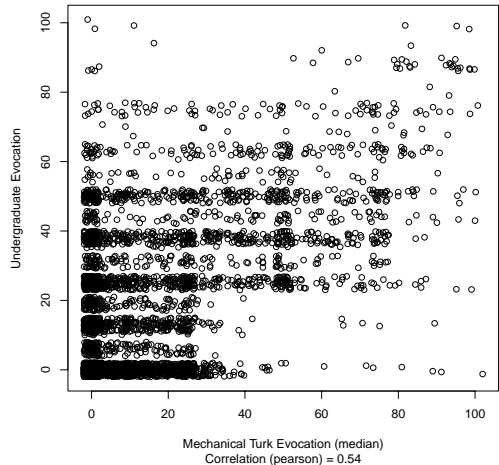## 4. EVALUATING VIVA WITH SIMULATED USAGE DATA

Figure 3: Ratings from untrained annotators on the web correlated well (0.54) with those collected by trained undergraduate annotators. Points are jittered.

It would be challenging to evaluate an adaptive tool—one that is supposed to assist people with everyday communication—in short laboratory studies. On the other hand, asking users to incorporate an early prototype in their everyday life can cause them much frustration, especially when they experience cognitive difficulties. Thus, as a first step, we concentrated on evaluating the backend adaptive functionality of ViVA by using simulated usage data in the form of sentences gathered from blogs of elderly people [1]. After training the vocabulary with the simulated usage data, we examined how the system performs given new sentences composed from the same blogger. Our hypothesis was that it will become easier to link words in the process of composing a phrase because meaningful user-specific connections are discovered from usage statistics.

4.1 Method

We propose to build a vocabulary network that enables efficient navigation and retrieval of words due to links between words based on word association measures, usage data, and adaptive reorganization of those links informed by system usage. To test this idea, we augmented the vocabulary hierarchy of Lingraphica, a commercial device designed for people with aphasia, and observed how it performs compared to the original. We chose Lingraphica's vocabulary organization for a baseline because, to our knowledge, Lingraphica is the only commercial device, successfully sold in United States of America, specifically designed for people with aphasia. Its design has evolved through a number of years based on input of speech language pathologists and users with aphasia; thus, it presents a realistic and practical standard to evaluate against.

We started off with two basic lexical inventories. One was Lingraphica's current hierarchy of words and the other one was this same hierarchy augmented with links between words based on the evocation data that we collected. We will refer to the augmented hierarchy as the ViVA vocabulary. If the evocation between two words was ranked to be higher than 30 (moderate or higher evocation), a link was

Number of Checks			Trained	Synset 1	Synset 2
All	Most	Some	Undergraduates		
50	10	61	88	trust.v.01	responsible.a.01
39	44	41	44	surgeon.n.01	responsible.a.01
25	18	22	42	deservingness.n.01	exceed.v.02
29	30	30	20	television_receiver.n.01	performance.n.02
46	57	62	19	log.n.01	leaf.n.01
34	33	31	16	diligence.n.02	craft.n.04
25	20	27	16	abundant.a.01	harmony.n.02
23	19	18	0	eyelid.n.01	wrist.n.01
25	28	26	0	reason.n.02	reference_point.n.01
4	5	9	0	spread.n.05	pill.n.02

Table 2: Examples of mean evocation ratings given three different methods to ensure rater reliability. For comparison, evocation ratings from trained undergraduates are also shown.

introduced. The evocation subset used to create these links was the intersection of the whole data set and the words contained in the Lingraphica vocabulary. We constrained the data to the Lingraphica vocabulary in order to be able to present a fair comparison of the two approaches to vocabulary organization.

We built five simulated usage data sets from paragraphs collected from five elderly individuals whose blogs were readily available online [1]. They covered topics such as cooking, gardening, health and family. The text from each blogger was broken into one thousand sentences. We then extracted the nouns, verbs, and adjectives from each sentence and created word pairs connecting every two neighbouring words. For example, processing the sentence "I checked my credit balance and called the dentist.", resulted in the following set of pairs: {checked, credit}, {credit, balance}, {balance, called}, and {called, dentist}.[2] The usage data was also filtered so that it contains only words that are part of the Lingraphica vocabulary. The five collections of usage word pairs were used to test the performance of the ViVA vocabulary in comparison to the Lingraphica vocabulary.

We selected randomly 80% of the word pairs from the simulated usage data to create a usage set. This set was used to further enhance the ViVA vocabulary with direct links between words simulating past usage of the system. We used these links and an equal number of links based on the evocation data to train the vocabulary. The remaining 20% of the simulated usage data were used as a testing set. To predict new links between synsets in the vocabulary, we ran a logistic regression using as input the training set. The features of the input vectors for the logistic regression were a usage score (0 or 1), an evocation score (0 or the corresponding mean from the collected ratings), a score based on semantic distance introduced by [13], and one last score based on semantic relatedness as computed by [14]. For the purposes of training, a link between two synsets was assigned if they were connected due to usage or due to evocation rating greater than 30 (moderate of higher evocation). To avoid having the evocation and usage scores completely guide the outcome, if both of these scores were zero, the higher of the semantic distance and semantic relatedness scores determined whether the two synsets should be linked or not. The links suggested from the outcome of the logistic regression

were incorporated in the vocabulary and we studied whether the paths between the words in the testing set were shorter compared to the ones in the LG vocabulary. The results are presented in Table 3 and we discuss them in the following section.

4.2 Results

The experiment of creating a vocabulary network based on links created due to the collected evocation data, simulated usage data, and predicted connections showed improvement over the original Lingraphica hierarchy. Adding evocation and simulated usage data links alone resulted in shortening the distances between approximately 44% of the words that appeared next to each other in a sentence from the usage sets. Predicted links due to logistic regression improved the results by 8% on average. As seen in Table 3, there was little variation across the different usage profiles. On average, 22% of the paths became shorter by two or more steps. Table 4 has some specific examples of shorter ViVA paths.

We constrained our working vocabulary only to words available in Lingraphica to be able to draw a clear comparison with a practical baseline. As seen in Table 3, this eliminated approximately 45% of the data for all of the simulated usage data sets. We also used only part of the synset pairs from the evocation data set; 43% were excluded for this experiment, because one or both of the words in a pair were not part of the Lingraphica vocabulary. Constraining the data eliminated a number of links that could have shortened paths between related concepts even further.

We performed an additional naïve baseline test to show that our improvement in the distances between usage related words due to link prediction cannot be achieved simply with a random increase in the density of the vocabulary network. We contrasted adding links predicted using logistic regression by adding the same number of links chosen randomly to the initial Lingraphica vocabulary. As shown in Table 3, there was still an improvement on some of the path distances, but it was minimal.

5. DISCUSSION AND FUTURE WORK

We considered two different sources of data to evaluate our proposed vocabulary system, ViVA. Since we envision a tool that will help compose phrases and sentences for communication, we thought about using data from repositories of switchboard exchanges. However, there would not have been an easy way to filter this data such that they are ap-

---

[2]Because Lingraphica and WORDNET both have different senses for homographs, we assumed that all words represented the most frequent sense for that word.

Data set	Total Usage Pairs	Intersection with Lingraphica	Shorter Paths after Evocation & Usage	Shorter Paths after Prediction (additional increase)	Paths Shorter by two or more Steps	Naïve Baseline Performance
Usage 1	5844	2539	43.9%	8.1%	27.9%	1.6%
Usage 2	6164	2914	42.4%	5.8%	25.3%	0.6%
Usage 3	3497	1537	44.6%	6.9%	15.6%	0.3%
Usage 4	4500	2077	46.5%	9.7%	21.8%	1.2%
Usage 5	4910	1865	42.3%	7.6%	19.2%	0.7%

Table 3: The vocabulary network augmented with links between words based on usage, word association measures and predicted associations decreased the browsing distance between related words.

Word 1	Word 2	Lingraphica Path	ViVA Paths
rice	cheese	rice-home-dictionary-things-house-kitchen-refrigerator-dairy products-cheese	rice-cheese
get	ticket	get-home-dictionary-things-leisure-outings-movies-ticket	get-buy-ticket
baby	brother	baby-more people-people-family-family relations-brother	baby-brother
hard	try	hard-home-dictionary-actions-communicating-thinking-try	hard-teach-try
table	drink	table-home-dictionary-things-food-drink	table-glass-drink

Table 4: Examples of browsing paths between related words shortened by ViVA. In Lingraphica, often the only way to reach the second word, having found the first, is to default to the home category which is the root of the vocabulary hierarchy.

propriately representative of our target user population. Instead, we decided to use informal text written by elderly people, describing their daily issues and routines. We collected sentences from the postings of elderly bloggers which served as simulated usage data. All bloggers were above the age of 60 which places them in and age range where there is a higher risk of suffering from stroke and thus acquiring aphasia. Even though they probably lead more active lives than people with aphasia, they have similar social interaction and needs, which makes their communications relevant to the type of data our tool will be handling.

We constrained our vocabulary to the intersection of the data we collected and the Lingraphica vocabulary, but one could imagine a more powerful solution with ViVA's organization overlaid with a rich image library such as ImageNet [10], an inventory that populates WORDNET synsets with hundreds of clean, high resolution images. To enrich our vocabulary network with links based on evocation, we used data collected from Amazon Mechanical Turk workers. The carefully filtered results were comparable to ratings provided by trained annotators, but even the inherent noise in the evocation data should be viewed as a positive outcome. This variation reflects idiosyncrasies in world knowledge and only by accepting this reality and incorporating it into assistive technologies can we hope to build devices that can truly help a heterogeneous target population. The AMT approach to collecting semantic data also introduces an exciting avenue for interface design by allowing hundreds or thousands of individuals with diverse backgrounds to help inform the design of assistive technology. It ensures that no single person has unreasonably shaped the interface and that the design reflects a broad spectrum of society.

The positive results from the evaluation of ViVA's adaptive component are encouraging. However, this evaluation is only preliminary and aimed to show the potential of this alternative vocabulary organization approach. It is too early to claim that such a vocabulary organization will in practice be of assistance to a user. We plan to continue this work and investigate how people with aphasia respond to a changing vocabulary organization and whether they will take advantage of customization features that we will provide. We have developed a working prototype of a phrase composition tool that relies on the described framework and we will be testing it with our target population. We believe that taking an adaptive and adaptable approach to designing a multimodal visual vocabulary will ensure that ViVA addresses the communication needs of our heterogeneous user population. One of a few inherent challenges in our work is finding a balance between managing the vocabulary automatically and providing a stable and dependable interface.

6. CONCLUSION

We presented the design of a mixed-initiative visual vocabulary for aphasia (ViVA) that aims to address vocabulary organization and navigation problems that are prevalent in existing assistive communication tools. ViVA implements adaptable techniques in order to allow the user to customize the tool and adaptive techniques to be able to tailor itself to better fit usage patterns and user needs. To assist people with aphasia in finding words faster and in a more personalized way, we attempt to exploit theories of human semantic organization. We use as a framework WORDNET, an electronic lexical database, whose design mimics the mental lexicon. In addition, we augment ViVA with links between words based on a large data set of evocation ratings that we collected from participants in an online experiment.

An initial evaluation of the adaptive capabilities of ViVA showed that the tool shortened the paths among words that

were associated by usage. The results demonstrate a potential for a realistic improvement of the vocabulary organization which we plan to investigate further involving users with aphasia.

In addition to the direct application of our results to improving assistive communication tools for people who have aphasia, we expect that the lessons learned during the design and evaluation of ViVA will have some broader applications and contributions. We will work towards narrowing the existing gap in understanding how mixed-initiative tools can assist users with other language and communication impairments. Finally, our mixed-initiative approach to addressing customization and flexibility in a communication system will very likely be applicable to other domains such as in the design of tools for the elderly and of educational tools for children and for foreign language learners.

7. ACKNOWLEDGMENTS

8. REFERENCES

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

[2] M. Allen, J. McGrenere, and B. Purves. The design and field evaluation of PhotoTalk: a Digital Image Communication Application for People with Aphasia. In *Proc. ACM SIGACCESS 2007*, pages 187–194. ACM, 2007.

[3] Amazon.com, Inc. Amazon mechanical turk. https://www.mturk.com.

[4] AssistiveWare. Proloquo2go. http://www.assistiveware.com/.

[5] D. R. Beukelman and P. Mirenda. *Augmentative and alternative communication: Management of severe communication disorders in children and adults.* Brooks Publishing Company, 1998.

[6] J. Boyd-Graber, C. Fellbaum, D. Osherson, and R. Schapire. Adding dense, weighted, connections to WordNet. In P. Sojka, K.-S. Choi, C. Fellbaum, and P. Vossen, editors, *Proc. Global WordNet Conference 2006*, pages 29–35, Jeju, South Korea, January 2006. Masaryk University Press.

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

[8] A. M. Collins and E. F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, November 1975.

[9] E. Daeman, P. Dadlani, J. Du, Y. Li, P. Erik-Paker, J. Martens, and B. D. Ruyter. Designing a free style, indirect, and interactive storytelling application for people with aphasia. In *INTERACT*, pages 221–234, 2007.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[11] C. Fellbaum. *WordNet : An Electronic Lexical Database*, chapter A semantic network of English verbs. MIT Press, Cambridge, MA, 1998.

[12] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proc. CHI 1999*, pages 159–166. ACM Press, 1999.

[13] J. Jiang and D. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc. on International Conference on Research in Computational Linguistics*, Taiwan, 1997.

[14] D. Lin. Automatic retrieval and clustering of similar words. In *Proc. of COLING-ACL*, pages 768–774, Montreal, CA, 1998.

[15] Lingraphicare Inc. Lingraphica. http://www.aphasia.com/.

[16] G. A. Miller. Nouns in WordNet: A Lexical Inheritance System. *International Journal of Lexicography*, 3(4):245–264, 1990.

[17] K. Moffatt, J. McGrenere, B. Purves, and M. Klawe. The participatory design of a sound and image enhanced daily planner for people with aphasia. In *Proc. CHI 2004*, pages 407–414. ACM Press, 2004.

[18] H. Moss and L. Older. *Birkbeck Word Association Norms*. Psychology Press, 1996.

[19] R.Patel, S. Pilato, and D. Roy. Beyond linear syntax: An image-orientation communication aid. *Assistive Technology Outcomes and Benefits*, 1(1):57–67, Fall 2004.

[20] R. E. Schapire. The boosting approach to machine learning: An overview. In D. D. Denison, M. H. Hansen, C. Holmes, B. Mallick, and B. Yu, editors, *Nonlinear Estimation and Classification*. Springer, 2003.

[21] R. Snow, O'Connor, D. Jurafsky, and A. Ng. Cheap and fast—but is it good? evaluating non-expert annotations for natural language tasks. *Proc. of EMNLP-08*, Jan 2008.

[22] R. D. Steele, M. Weinrich, R. T. Wertz, M. K. Kleczewska, and G. S. Carlson. Computer-based visual communication in aphasia. *Neuropsychologia*, 27(4):409–426, 1989.

[23] A. Sutcliffe, S. Fickas, and M. Sohlberg. Investigating the usability of assistive user interfaces. *Interacting with Computers*, 15:577–602, 2003.

[24] D. Swinney. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, (18):645–659, 1979.

[25] K. Tee, K. Moffatt, L. Findlater, E. MacGregor, J. McGrenere, B. Purves, and S. S. Fels. A visual recipe book for persons with language impairments. In *Proc. CHI 2005*, pages 501–510. ACM Press, 2005.

[26] The National Aphasia Association. Aphasia: The facts. http://www.aphasia.org.

[27] TouchSpeak & TypeSpeak. http://www.touchspeak.co.uk/.

[28] M. van de Sandt-Koenderman, J. Wiegers, and P. Hardy. A computerised communication aid for people with aphasia. *Disability and Rehabilitation*, 27(9):529–533, 2005.