A Communication Aid with Context-Aware Vocabulary Prediction

by

Ewa Dominowska

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Computer Science and Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2002

Author . . . .
Department of Electrical Engineering and Computer Science
May 24, 2002

Certified by . . . .
Deb K. Roy
Assistant Professor
Thesis Supervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

A Communication Aid with Context-Aware Vocabulary Prediction

by

Ewa Dominowska

Submitted to the Department of Electrical Engineering and Computer Science
on May 24, 2002, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Computer Science and Electrical Engineering

Abstract

Communication disorders prevent over two million Americans from leading normal, productive lives. A speech-impaired individual has to rely on communication devices to interact with other people, but current communication aids are often ineffective. This thesis presents the design and implementation of a communication aid with a novel method of lexical prediction. Vocabulary pre-selection is done based on environmental context cues. The aid's interface is an electronic, dynamically changing picture array with touch-based access. The aid arranges lexical items within semantically coherent categories and uses semantic frames to facilitate grammatical communication.

Word prediction is achieved through cluster generation. We introduce a novel word clustering algorithm, the *Decision Tree Cluster Convergence* algorithm. This algorithm generates contextual categories that are selected for the user when appropriate contexts are detected.

The contributions of this project include the introduction of semantic frames into an AAC system, a scheme for improvement of lexical prediction though the use of contextual information and the design and implementation of a communication aid that includes both features.

Thesis Supervisor: Deb K. Roy
Title: Assistant Professor

Acknowledgments

I would like to thank my advisor Professor Deb Roy, Professor Rupal Patel, Robert Ragno, Professor Patrick Winston, Professor Robert Berwick and Sheel Dhande for all the help with ideas used in this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Motivation and Background

Many people with disabilities are unable to verbally communicate with those around them. This setback might be caused by motor-control difficulties, such as not being able to control the muscles responsible for producing speech; or it might be caused by a neuro-physiological problem. The problem could either be congenital or a result of trauma, ensuing in an inability of the brain to create speech in the first place. In the US alone, 500,000 people have *cerebral palsy* [15], another 2.1 million are speech-impaired [56], and a further 9 million are illiterate.

Many people have no verbal communication skills at all and as such need some form of prosthetic aid known as an Augmentative and Alternative Communication (AAC) device. AAC denotes the use of techniques to supplement or replace speech as communication. Almost everyone uses augmentative modes of communication to some extent, whether in the form of gestures and facial expressions or long-range technology such a fax machines and computers. However, individuals with severe communicative difficulties must rely on AAC to meet their most basic everyday needs. These people view communication not as a social interaction, but as the sole way to fulfil their most basic needs, such as medical emergencies, requests for food, or other help.

There are many forms of AAC devices ranging from very basic pointing boards to complex computer systems. A device needs to be chosen to suit the needs and abilities of an individual; however, people with communication disorders often have other physical problems that prevent them from using complex aids. For instance, using a keyboard is often not a viable option. The two most common methods of operating an AAC device are low-resolution *direct selection* and *single switch* access. In the direct selection method a user tries to communicate by directly pointing to meaningful symbols or objects. However, the user's movements are not exact. The user cannot apply sufficient pressure to activate a switch and cannot sustain constant directional positioning. In the single switch access method users select letters or images one-by-one as a system scans though the available choices. Both these methods are prone to errors and result in extremely slow communication rates. As a consequence, even with a help of an AAC device individuals cannot join in a normal conversation.

This thesis tries to address some design deficiencies in the existing systems. We are trying to build an aid that will accelerate communication rate by reducing the number of selections. The two acceleration methods we introduce in this paper are *semantic frames* and *contextually informed word prediction*.

1.1 A Need for Devices to Aid in Communication

Speech-based communication is one of the unique traits that makes us human. We use communication to interact with others, to request help, and to share knowledge. The ability to communicate is important for healthy social interaction. Appropriate communication and interaction speeds appear critical for competent communication performance, and are linked to academic, social, and employment success. Thus, it is necessary to provide all people with tools to aid communication. Further, it is necessary to make these tools adequate for the tasks in which the users might be involved. But what happens when a person is no longer able to communicate effectively? The field of augmentative communication addresses the loss or lack of communication a person may experience due to neurological and motor deficits.

Speech is the most common communication medium. When we want to share information with other people we usually talk with them. Normal human speech rates are around 180 to 220 words per minute, while handwriting averages about 20 words per minute, and typing ranges from 80 to 100 words per minute for well-trained typists [37]. Establishing communication rate is a composite process. We need to consider several different variables:

1. The language representation method employed for accessing core vocabulary;

2. The selection rate;

3. The error recovery.

The task becomes even more complex when trying to establish communication rates for individuals with disabilities. Communication deficiencies are often linked with other physical or mental problems. Disabled individuals often require multiple modes of communication to accommodate for emergency situations and technical problems. For example when a person uses a hi-tech communication device he must also maintain a low-tech backup that he can use in case of battery failure. Thus, when establishing communication rates for AAC users we need to consider additional factors:

1. Average communication rate;

2. Peak communication rate;

3. Communication rate by language representation method;

4. Language representation method usage for spontaneous communication;

5. Selection rate;

6. Word selection errors per word selected;

7. Spelling errors per word spelled [38].

These additional factors make it hard to compare rates between users of AAC devices and people with no communicative impairment. However, extensive investigation of AAC users has estimated an average message production rate of 10 words per minute. This rate is about 20 times slower than the rate for non-disabled individuals. This lack of speed has been characterized as the *"rate problem"* [9] and is a major barrier to normal functioning for people requiring assistive technology. To provide adequate communication, it is necessary to bridge the rate gap.

Some proposed acceleration methods include optimized access, reduced access to production ratio, and efficient categorization of information retrieval. This project tries to improve upon existing acceleration methods by providing semantic sentence templates, by using whole word images for fast selection, by providing efficient, personalized categorization, and by providing the user with lexical prediction based on his/her environment.

1.2 Augmentative and Alternative Communication Users

The general characteristics of individuals who are potential users of assistive communication technology include:

1. Inability to communicate effectively with speech (nonverbal);

2. Only partial ability to communicate with speech (has speech but is not understood by most listeners or speech is not functional);

3. Adequate speech but requires an augmentative device for purposes of writing or carrying on long conversations.

Our goal is to design a general AAC system; however, our initial target users are individuals with severe speech impairment, and some degree of motor impairment which includes group 1 and group 2 individuals. Initially we targeted people affected by cerebral palsy. Most individuals who suffer from cerebral have no significant cognitive deficits, which makes their inability to communicate even more of a burden. Cerebral palsy is congenital, and it is a stable condition. The prevalence of *dysarthria* in cerebral palsied people ranges from 31% all the way up to 88%. Other groups of people who could benefit from the proposed aid design would be individuals affected by amyotrophic lateral sclerosis (75% dysarthria and severe motor neuron degeneration), Parkinson's disease (50% dysarthria and impaired movements), and brain-stem stroke (severe dysarthria to anarthria for individuals with "locked-in syndrome") [63].

AAC interventions, including the use of electronic devices, have long been recognized as appropriate for individuals with cerebral palsy. In a survey of 66 non-speaking persons with cerebral palsy, the majority used augmentative communication systems (57.5%) and accessed them through direct selection in 62% of the cases [48]. In addition, there are many case examples of individuals with cerebral palsy and severe dysarthria who use AAC devices effectively in their daily lives. There are many men and women ranging in ages from 20 to 65 who lecture and have written extensively about the impact of AAC devices on their lives, such as Bob Williams, Peg Johnson, and Jennifer Lowe. All these people are productive and highly respected members of their communities. They represent just a small sample of the growing number of individuals with cerebral palsy in the United States who use AAC devices effectively to do things they would otherwise not be able to accomplish. Now, it is necessary that the available AAC technology become adequate to meet their needs.

1.3 Necessary and Desired Design Criteria

To build an effective communication aid it is necessary to establish a set of guiding *design principles*. The designer needs to be aware of how exactly the aid will be used. Goodenough-Trepagnier has tried to compile a list of design goals that should be considered in the development phase. These goals include: learnability, consistency, immediate utility, spontaneity, minimum motor demand, minimum attention shifting, and goals that fit people [35].

In this project we established the following set of essential features:

1. *Learnability* - the system is predictable and relatively stable;

2. *Fast message construction and propagation* - number and variety of steps and decisions required to complete a task is kept to a minimum;

3. *Error correction* - errors are easy to detect and easy to fix;

4. *Optimal core and auxiliary vocabularies* - the vocabulary is sufficient to address user needs;

5. *Adaptation* - the system is appropriate for the skill level and personal preferences of each user.

Now, we are going to look at each of these principles in more depth.

1.3.1 Learnability

The first step in effectively communicating through an AAC device is acquiring the skills necessary to use the device. Communication systems should make the steps for message construction self evident (or *"operationally obvious"* as termed by Light and Lindsay, [54]). This allows for easy and fast skill acquisition. An early learner requires much of the knowledge needed to operate the system to be externally available (e.g., prompts or help files displayed in text, graphic, or auditory form). It is necessary to provide easy access to help functions and extensive support for error diagnosis and correction. Learners at all stages of skill acquisition benefit from a system that is consistent, predictable and coherent (i.e., a system having a high degree of internal organization) [18].

In the system described in this paper we offer a high degree of immediate help and user feedback. For example, each symbol can be annotated with a text label. As each new message is constructed the user is provided with immediate sentence construction feedback. Finally, even though we are proposing a dynamic system, we try to maintain a level of consistency. For example, the top-level categories and special symbols are always available at one touch even as the user moves through a hierarchy of categories.

1.3.2 Fast message construction and propagation

As mentioned earlier, speed of communication has been identified by AAC users as having the greatest impact on communication success [78]. Thus, when constructing a message, the number and variety of steps required to complete the task should be kept to a minimum. Moreover, the user should not be required to make too many decisions; or the decisions that should be made must be routine, repeated decisions. Since we are concentrating on an *image-based direct selection* system it is necessary to ensure that the selection process is fast and smooth. The selection technique should:

- Facilitate effective control through inaccurate or imprecise pointing
- Support learners at all stages of skill acquisition
- Not require timely responses
- Be controllable through a variety of controlling actions

We have based our system on direct selection of icons. The number, size, and complexity of the images can be easily controlled, thus making for fast selection. Furthermore, the number of necessary actions is greatly decreased as a result of hierarchical ordering, template semantic frames, and contextual vocabulary prediction.

1.3.3 Optimal core and auxiliary vocabularies

Since we are working with a word-based system, it is necessary to decide what vocabulary should be easily accessible and how it should be organized. Choosing a vocabulary size requires a tradeoff between versatility and speed. It is desirable to have access to a very large vocabulary, at least on the order of tens of thousands. However, access to 20,000 words would require the user to navigate through 666 boards with 30 images each. Even with a *hierarchical* system and the assumption that we can create a fully balanced categorical structure, this would require a minimum of three choices per word. However, investigation into categorization of words shows that categories tend to vary extensively in size [58], [30] making balanced structures unachievable. So how many words should we make available?

Recent research in linguistics indicates that a core vocabulary of fewer than a thousand words might be sufficient to provide semantic coverage for nearly every idea expressible in a human language. The most-well know example of trying to express a complete language with a restrained vocabulary is called *"Basic English"*. Basic English, however, is just one among many. Using 800 words and appropriate morphology for those words, a person can express nearly everything that can be conveyed using the English language [62].

Spoken language studies on elicited and spontaneously generated speech have shown that the 100 most frequently occurring words of a linguistic sample typically account for more than 60 percent of the total words communicated. The top 100 to 200 words usually account for 80 percent of the total words communicated. The phenomenon of a core of words with high frequency accounting for a majority of words communicated is not limited to English [6].

Based on these findings we decided to provide a *core vocabulary* of about 500 words. Additionally, the aid provides access to *auxiliary vocabularies* through the use of an iconic keyboard. Further addition of new icons and words is almost trivial. New words can be acquired from the WWW, from existing clipart, or by scanning printed images.

1.3.4 Error correction

Another design consideration that is critical at all stages of communication aid use is error strategies. The possibility of making errors should be minimized as much as possible. Error correction should be simple and direct. There should be no opportunity to make additional errors in the process of correcting an error. It is estimated that even skilled users spend a quarter of their time making and correcting errors. Errors are "disorienting," frequently disrupting automatic performance and therefore requiring conscious intervention before the "rhythm" of automatic operation can be regained. Our system provides continuous feedback during message construction, which allows for easy error detection. Once an error is detected, a one step undo process allows for fast correction.

1.3.5 Adaptation

One of most important questions to consider when designing a communication system is what properties of that system make it most conducive to skill acquisition. Human factors literature contains abundant and frequently conflicting design guidelines [46]. However, it is clear in the literature that a system that is optimally configured for the novice user is usually not optimally configured for the skilled user [49]. The access system therefore needs to be sufficiently flexible to accommodate the learner at all stages of skill acquisition. Therefore, we decided to make our system adaptive. A user can start with a small vocabulary and only a few basic word-categories. They can use an image set that they are already familiar with. As their usage skills progress, the lexicon can be expanded to provide the means for spontaneous social interaction.

1.4 Existing Communication Aids

Existing communication aids range from charts, bracelets, and language boards all the way to advance computer systems. With these aids, objects may be represented by pictures, drawings, letters, words, sentences, special symbols, or any combination thereof.

Electronic devices are available that can speak in response to entries on a keyboard or other methods of input. Input can come from any number of different switches that are controlled with motions as simple as the push of a button, a puff of air, or the wrinkle of an eyebrow. The possibilities increase virtually every day. Augmentative communication users don't stop using speech. When speech is used with standard and special augmentative communication, not only does communication increase, but so do social interactions, school performance, feelings of self-worth, and job opportunities.

Existing communication technology can be divided into 7 categories (commeracially avaliable examples in parenthesis):

1. Low-Tech Communication Board (Pick N Stick)
2. Low-Tech Communication Book (Porta book)
3. Simple Voice Output Device (Step by Step, Cheap Talk)
4. More Extensive Voice Output Device (MessageMate, Macaw)
5. Voice Output Device with Dynamic Display (Dyna Vox, Vanguard)
6. Voice Output Device with Icon Sequencing (Liberator, Alpha Talker)
7. Voice Output Devices that Rely on Spelling (Link, LightWRITER)

Many aids combine two or more of these categories. Also, most users employ multiple aids within their everyday lives. Nevertheless, we will try and describe aids representative of each category.

1.4.1 Simple communication boards and books

This category includes communication boards consisting of graphic symbols, pictures, or objects. The boards may consist of one or multiple pictures and various sizes depending on user ability. Various display set-ups may be used. Aided Language Stimulation techniques can be used to model use of communication boards and increase a child's understanding of messages. This type of communication board may be used as the only AAC option for a young child or a child with a very limited vocabulary or as a back up to a more complex voice output device. Often, communication displays are made for specific activities by selecting items from an appropriate category. The displays can also be selected for social situations and taken into the community when a voice output device may not work as easily or as well. Communication boards are also often used to encourage the use of visual language strategies such as calendars, schedules, and step-by-step directions.

Graphical communication boards have no or low technology requirements. They are referenced by pointing, eye gazing, touching, or scanning that is activated through switches or by indicating to an adult. The symbols on a graphical system may be represented visually, aurally, and/or tactilely. These systems tend to be relatively inexpensive and are often homemade.

A disadvantage of these types of graphical systems, however, is the fact that the partner has to pay attention to the communication board and not to the individual using the board. Similarly, the user has to look at the board too, instead of at the partner's face. In addition, another method is required to obtain the attention of someone who is not close by or not attending. Finally, the aids can be used to converse only with people who are familiar with the symbol system being used.

1.4.2 Simple and layered voice output devices

Voice Output Communication Aids (VOCAs) are also graphical systems, but, unlike communication boards, are high technology devices that output speech. VOCA typically refers to a dedicated electronic speech apparatus. Computers can also be placed under this category since they can also be used as speech output devices. (Even cassette tape recorders can, in some instances, accomplish the same goals as a VOCA.)

Simple and low-cost VOCAs provide voice output with one set of messages available to the user at a time. Pressing a key or cell produces one message. These devices may have any number of messages. The overlay within such a device must be physically changed and the device reprogrammed to change the messages.

One way to expand the capability of such a device is leveling or layering. A layered device is capable of storing several levels of messages. Each layer can be programmed with different messages. Changing from one level to another requires pushing a button (or sliding a switch) and physically changing the overlay.

Vocabulary must be programmed into a VOCA, which may be done at least partially by the manufacturer or entirely by the purchaser.

1.4.3 Voice output devices with advanced features

Dynamic displays

A *"dynamic"* or *"transient"* display is a screen that changes in response to user input. Laptop computers with high-resolution displays have made dynamic display technology portable and useful for AAC devices. Persons using AAC devices with static displays rarely have access to more than 128 symbols/messages unless their overlays are changed or they learn codes [83]. Dynamic display augmentative communication devices present options for arranging vocabulary in a series of simpler displays. The user has access to multiple screens by navigating through a hierarchical structure. This is particulary important for beginning communicators who face severe and multiple challenges. Especially young AAC users do well with activity-based displays that are organized consistently from one page to the next [14]. Adding the component of active control that the dynamic display provides allows users to be more successful in communicating.

AAC devices with static displays place cognitive, motor, perceptual and learning requirements on those using the devices to communicate. Many feel transient displays substantially reduce these demands, as well as enhance rate. The dynamic display technology allows individuals to change screens quickly and to configure the size, color, and arrangement of symbols, words, and phrases on their screen. They can select (or construct) messages without remembering codes or physically switching overlays. Also, sufficient computer memory allows storage of libraries of pictures, symbols, animation, text, sounds, and speech.

Features inherent in dynamic displays also carry cognitive, motor, visual-perceptual, and learning loads that challenge some users. For example, moving between and among screens to construct messages requires visual attention and decision-making. Although memory demands are reduced, automaticity in generating messages (or parts of messages) may be difficult to achieve. Off-the-shelf computers are not designed to be mounted on wheelchairs and used as communication aids; however, concerns raised about their ruggedness remain unsubstantiated.

Voice output devices with icon sequencing or Minispeak

Another way to allow for increased vocabulary is *semantic compaction*. Semantic compaction, often referred to as *icon sequencing*, is a way of organizing language that uses an ordered array of pictures to code vocabulary. The user presses one, two, or three keys in sequence to produce one message. The most commonly used paradigm for icon sequencing is called Minispeak. Minispeak is a pictorial system that allows for fast and accurate access to language through a process of using multiple meaning icons.

Voice output devices that rely on spelling

This type of device allows the user to type either through direct selection or through scanning. Once a word has been typed the device either speaks or prints out the message. This category of device requires the user to have good spelling skills. These devices often feature abbreviation expansion features to make possible the storage of longer messages with a few keystrokes for activation.

1.4.4 Problems with existing AAC devices

Most problems with AAC technology are related to specific devices. Unfortunately, there is a list of issues that are prevalent among many of the aids.

1. Small vocabulary

   All static and most dynamic devices provide the user with only limited vocabulary. Often when extensive vocabulary is available it covers only very specific topics and still lacks many core words. For example, a computer-based system may be able to store thousands of messages, or a picture board can be made with many pages. On the other hand, a system based on tangible objects will be extremely limited due to the size of the objects.

2. Often not comprehensible by conversation partners

   Some aids output regular speech, but sign language or other symbol systems can be like a foreign language and must be learned everyone in the user's social community.

3. Slow communication

   The speed of communication plays a large role in conversational quality. Interactions between AAC users and non-users tend to be imbalanced, with non-users dominating conversations and users primarily in the role of respondent. Studies have shown that one of the main stumbling blocks to equalizing their standing is the speed with which the AAC user can converse.

4. Requires expert configuration (not universal across uses, people, situations)

   Selecting and configuring an aid for a user requires the help of many specialists. A *speech-language pathologist* needs to establish the user's ability to understand language (oral and written). They need to establish how the user interacts with different communication partners. They measure the muscle control for speech and the pronunciation of speech sounds in order to establish the appropriate vocabulary for use with the augmentative communication system. Then an *occupational therapist* establishes the muscle control of different body parts with and without special equipment for different body positions, the ability to tell differences in size, color, and shape, and requirements for mobility and seating. A *physical therapist* measures muscle strength, range of movement, flexibility, balance, and coordination. Physicians of various specialties, such as pediatricians, neurologists, otolaryngologists, orthopedists, psychiatrists might also be needed. A *rehabilitation engineer* decides on the mounting of devices and switches and the design and development of customized parts. A *social worker* judges the individual's total living situation (family structure, finances, etc.) and the need for additional community resources. A *psychologist* estimates the individual's learning potential and the need for individual and family counseling. A *vocational counselor* determines the individual's potential to hold a job and helps with the identification of career goals. An *audiologist* evaluates and treats hearing loss. The *manufacturer/distributor* of communication devices determines possible applications and modifications of a device, provides repair information, and establishes possible sources of funding.

5. Expensive

   Expense must be considered in terms of both money and time. High technology devices can cost several thousand US dollars. Low technology communication systems can often be homemade, but the time required for constructing all of the overlays or vocabulary pages necessary for many different activities in different environments may be very large. In addition, it is important to consider how difficult it would be to replace a device if it were lost, stolen, or irreparably broken.

6. Static (topic changes very hard)

   High technology devices must be programmed. Many graphical devices, both high and low technology, require overlays or pages to be constructed. Given the number of activities – and, therefore, messages – in which a user may be involved at home, at school, and in the community, updating can require a substantial amount of work.

7. Requires power

   If a device uses batteries, how long will the batteries last before needing to be recharged? A dead battery without a replacement could leave a child without his or her primary means of communication.

8. Reliability

   With any hi-tech devices users are likely to encounter breakdowns. It is important to consider how easy it is to repair the device. Sometimes high tech devices like voice output communication aids (VOCAs) and computers take weeks to be fixed.

9. Low portability

   Many devices are large and cumbersome. They make it impossible for a user to communicate while walking or otherwise moving from place to place, playing, riding in a car, etc. Especially in the case of children who are small, it may be difficult for them to handle a large or heavy device whether they have physical disabilities or not.

10. Fragile

    Technology devices often include many delicate parts, such as liquid crystal displays (LCD) or memory modules. Yet, these devices need to be usable in a variaty of situations and environments. Children are not careful when using electronic equipment, including AAC devices.

11. Lack of independance

    Many systems require the assistance of a partner. The partner is usually in control of scanning and awaits a signal from the child indicating message choice. Examples of this are if the partner recites the available choices (auditory scanning), or points from one picture to another on a picture board (visual scanning).

12. No long distance use

    A picture board or sign language necessitates the partner being close enough to see the picture or sign that is being indicated. An eye-gaze system is even more demanding and requires that the partner be positioned so that he or she is able to tell at what the disabled person is looking.

13. Prone to errors

    When a user is pointing at a picture on a picture board, he or she may not know whether the correct picture is being indicated until the partner responds.

14. Hard to learn

    Often there is a trade-off between ease of learning and system flexibility for both devices and symbol systems. For example, a VOCA with only three buttons may be very easy to comprehend, but it offers limited opportunity for growth. Among symbols, pointing to tangible objects may be easy to learn but is less flexible and convenient than graphic symbols.

15. Problems with outdoor use

    Rain and darkness can be issues, but computer screens and LCD screens can be difficult to read in bright sunlight too.

16. Lack of intimacy

    Many people feel that interacting though a device in a social interaction reduces the feeling of closeness and intimacy.

In Chapter 3 of this Thesis we will describe in detail the interface design process and the final system. Then we will concentrate on message rate improvement through *word prediction* in Chapter 4 and *contextual categories* in Chapter 5. We will try to prove the benefits of our method using collected usability model data (described in Chapter 6) by evaluating generation of contextual categories in Chapter 7. Chapter 5 gives detailed descriptions of all issues related to generation of contextual categories. Chapter 8 contains a short summary of the project contributions, and Chapter 9 discusses future extensions.

Chapter 2

Background on Image-Based Languages

Once the decision to use images was made, it was necessary to select a specific pictorial language. When choosing a universal set of symbols, it is essential to consider how easy they are to learn and how effective they are in conveying information. Symbols can range from photographic images to complex line drawings to simple signs. Moreover, numerous precompiled symbol sets are available commercially.

In order to select a suitable set, it is first necessary to choose the desired criteria. We preferred to employ symbols that are easy to use and accurately depict the word meaning. In addition, we wated to minimize ambiguity yet convey enough detail to communicate complex concepts. Symbols have been described in terms of their "iconicity" (ease of recognition) [11], ranging from transparent (highly resemble their referent) to translucent (some relation to the referent), to opaque (relationship to referent is not obvious). One must be careful, however, not to assume that comprehension of symbols follows an obvious hierarchy [12] from concrete objects to printed words and letters.

There is a wide array of symbol sets to choose from. We will concentrate on "aided" symbols: those that require something external to the body to represent meaning, such as a book, a board, or a device. Then, we will briefly mention "unaided" symbols, which only use movements or sounds from the human body to represent meaning (e.g. sighing, speech, etc.).

The initial distinction that we need to make is between photo-realistic images and simple line drawings. For example, PictureThis uses a database of 2,900 high-quality photos.

The aid also offers the option of 1,000 additional pictures for Functional Living Skills and Behavioral Rules and a further 2,000 photos for School Routines and Rules. Just looking at these numbers makes it obvious that there are problems with the use of photo-realistic images. Such images often contain too much detail. They are also very poor in conveying abstract and metaphorical concepts.

The most widely used systems in AAC technology are drawing based. There are many styles and many complete commercial sets available. The list included here is intended to be representative, consists only of the major systems, and is by no means comprehensive.

2.1 Simple Line-Drawing Systems

2.1.1 Picture Communication Symbols

Picture Communication Symbols (PCS) 2-1 is the most commonly used symbol set in the field of AAC technology. Though PCS were created to depict apparent meanings, they may depict abstract vocabulary concepts as well [43]. The PCS set was originally designed to create professional looking communication aids both quickly and inexpensively. The images are now also used extensively in unlimited types of learning activities and lessons. The over 3,000 symbols create one of the largest picture symbol sets available today. In addition, existing familiarity with a symbol set may increase the potential to improve an AAC user's ability to develop language.

The PCS symbols are:

- Considered to be simple, clear drawings that are easily recognized
- Appropriate for all age levels
- Divided into six main word categories of people, verbs, nouns, descriptive, social, and miscellaneous
- Easily combined with other symbol systems, pictures, and photographs for individualized communication aids

Figure 2-1: Example: PCS images take from the PCS DonJonston web page: `http://www.donjohnston.com/catalog/piccomd.htm`

2.1.2 Picsyms and DynaSyms

There are about 800 Picsyms 2-2, and approximately 2,000 available as Dynasyms in the Dynavox. The symbols can be pictographic or as abstract as needed. The shape or manner of representation of a symbol provides a clue to the semantic category that the item belongs to. For example action is represented by an arrow indicating the direction of the action. There are guidelines for creating new vocabulary. The drawback is that many of the symbols look very visually "busy."

Figure 2-2: Example: Picsym images.

2.1.3 Makaton vocabulary

Makaton 2-3 is a unique language offering a structured, multi-modal approach for the teaching of communication, language and, literacy skills. It was devised for children and adults with a variety of communication and learning disabilities. Makaton is used extensively throughout the UK and has been adapted for use in over 40 other countries. It is loosely based on line drawings of manual (British Sign Language) signs. Makaton was first developed in the early 1970s as part of a research study into the functional communication needs of a group of disabled people. It was the first British project to use signs and, later, symbols. The Makaton Core Vocabulary of essential concepts for everyday needs was updated in 1996 to reflect the communication needs of the UK's modern multi-cultural society. The additional over 7,000 concepts in the Makaton Resource Vocabulary were identified in the 1980s with signs and symbols to match, covering a huge range of the social, educational, and personal needs of Makaton users [81].

Figure 2-3: An example of a Makaton symbol taken from the makaton web page: `http://www.makaton.org/`

2.1.4 Signalong

Signalong is another example of a sign-only system. It was established in the 1980s and first developed a basic vocabulary of concepts to cover daily needs. The icons resembled the Makaton Core Vocabulary in size and content. Since then it has developed further selections of signs related to topics similar to the Makaton Resource Vocabulary. Signalong also uses signs from British Sign Language. The symbols are meaning based and some of them are sign linked. There are guidelines for creating new vocabulary. For example, if it is not possible to create a pictographic symbol, a sign linked one can be used instead. Color is used to highlight the focus of the symbol.

2.1.5 COMPIC

COMPIC pictographs are a communication resource developed in Australia by COMPIC. The image set consists of a library of 1,670 pictographs (computer-drawn line drawings), which represent words, objects, activities, and concepts. The pictographs are accompanied by the relevant word or phrase. The vocabulary includes adjectives, verbs, nouns, questions, phrases, colors, numbers, and quantities. The pictographs are designed according to international standards. They are clear, very iconic, and suitable for children and adult applications and thus have been widely accepted.

2.1.6 Paget Gorman

PGSS (en example image shown in Figue 2-4) is a specifically generated language with topic-based signs. The system is based on 21 standard hand postures and 37 basic signs used in different combinations. Very complex and exact finger and hand positions are used. Meaning can be changed by adding identifying features to the basic signs. PGSS mirrors spoken language exactly; it has a one-to-one sign-to-word correspondence.

Figure 2-4: An example of a Paget Gorman symbol taken from the PGSS web page: `http://www.pgss.org`

2.1.7 Signed English

Signed English uses signs from British Sign Language with finger-spelling as well as specifically generated signs and grammatical markers. Like PGSS, it mirrors spoken language exactly. Additional signs for vocabulary development are available from BSL.

2.1.8 Rebus

A Rebus, shown in Figure 2-5, is a picture that visually or nominally represents a word or a syllable. For example, a rebus of a knot could be used to symbolize either “knot” or “not.” Many of the icons are easy to recognize (“transparent”) and only a few are more abstract and must be learned. Currently, there are about 4,000 Rebus Symbols available, using a combination of symbols and letters. Rebus symbols represent grammatical markers in text form and were originally devised to help develop reading skills. New rebuses can be created quickly and easily by following the guidelines given in the revised Rebus Glossary.

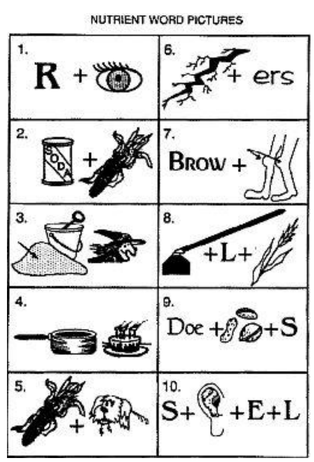

Figure 2-5: An example of messages written using rebus symbolics.

2.1.9 Pictograms

Pictograms are white symbols on a black background as shown in Figure 2-6. They are useful for some individuals with visual impairments.

Figure 2-6: An example of a Pictogram symbol taken from:`http://www.zygo-usa.com/books.htm`

2.2 Complex Line-Drawings

The symbol systems described below are much more complex than the ones just discussed. This does not mean that the symbols themselves are complicated. The *systems* are complex because the symbols can be combined according to particular rules to create novel utterances.

2.2.1 Blissymbolics

Blissymbolics is a very large symbol set that permits individuals to create novel utterances and thus can be characterized as generative. It was originally developed by Charles K. Bliss (1897–1985) for the purpose of international communication and was first used in communication with children with physical disabilities in 1971 by an interdisciplinary team at the Ontario Crippled Children's Centre (now the Bloorview MacMillan Centre).

Blissymbolics is a language currently composed of over 2,000 graphic symbols. Figure 2-7 shows examples of blissymbols. Bliss–characters can be combined and recombined in endless ways to create new symbols. Bliss-words can be sequenced to form many types of sentences and express many grammatical capabilities. Simple shapes are used to keep the symbols easy and fast to draw. Because both abstract and concrete levels of concepts can be represented, Blissymbolics can be used by both children and adults and is appropriate for users with a range of intellectual abilities.

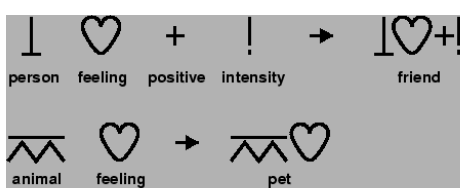

Figure 2-7: An example of Blissymbolics icons taken from the Blissworld web page:
`http://www.symbols.net/blissworld`

The system of Blissymbolics has several features that could make it the preferred means of communication for non-speaking individuals. In addition, it has no requirements for literacy or language background.

Blissymbolics is a language with a wide vocabulary, a grammar that allows for sentences in the past, future, and present tenses, and markers for possession, plurality, questions and commands. There are many strategies within the system of Blissymbolics that enable the user to create new symbols. Blissymbolics is a totally generative system, with each new symbol interpretable by the receiver through analysis of the components. In the same way that letters represent sounds when used to create words in print, meaning-based Bliss units are sequenced to define the meaning of each compound symbol. Since there are a limited number of elements (called key symbols), the learner needs only to master the meaning of approximately 100 visual symbols.

2.2.2 Minspeak

Baker [4] designed an iconic encoding technique referred to as semantic compaction, or Minspeak. Minspeak uses sequences of icons that are combined to generate new words, phrases, and sentences in voice-output AAC devices. Semantic compaction or iconic encoding is typically considered to be a rate enhancement technique for retrieval of vocabulary from AAC systems [10]. Iconic encoding can also serve to expand the number of accessible vocabulary items stored in an individual's AAC device if he or she is using a static display. In particular, it can enable an individual's vocabulary set to expand from approximately 120 items to a vocabulary reaching approximately 2,000 words. An example of a board using Minspeak symbols is shown in figure 2-8.

The advantage of static displays is that they promote automatic access in the same way as touch-typing. Motor planning supports selection because the keys and vocabulary remain in the same physical position [1].

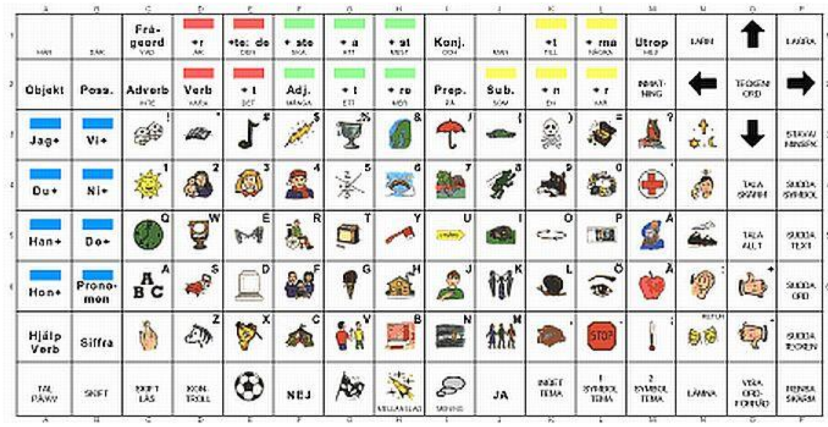

Figure 2-8: An example of a Minspeak board taken from the Minspeak web page: `http://www.prentrom.com/speech/minspeak.html`

2.2.3 Elephant's Memory

The Elephant's Memory (EM) is a pictorial language consisting of about a hundred and fifty combinable graphic elements. Figure 2-9 shows an example sentence written using EM symbols and sentence construction rules. Its aim is to be culture and background independent. EM builds a transitional space between natural languages and is oriented primarily towards children. The system is an exploratory tool promoting new approaches to the concept of linguistics. The pictograms develop a visual link between the members of a community and provide original material for families and educators to encourage dialogue and creativity.

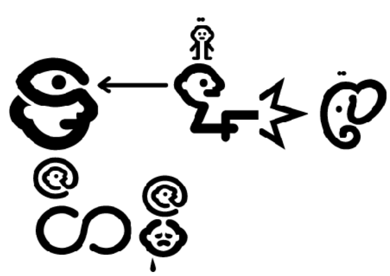

Figure 2-9: An example of an Elephant's Memory sentence taken from the EM web page: `http://www.khm.de/˜timot/ Sentences.html` .
The sentence says: "*Seeing elephants shot by people makes me cry.*"

2.2.4 Orthographic symbols

Orthographic symbols are generally used in the form of spelling boards or word boards. Text is immediately understood by most people. Access to individual letters provides the user with unlimited vocabulary. In public settings, text is perceived as more socially acceptable than symbols or pictures. Although a user may recognize a large number of words, they may not be able to spell additional words to expand their vocabulary. Words tend to be viewed as having only one meaning; therefore the listener may not think as divergently with a text-based user as with a symbol user. Paper and pencil are generally needed to write down what the user is saying in order to provide reinforcement for them and to act as a memory jogger for the listener.

2.2.5 Phonetic picture language

Phonetic picture-writing is image based writing that can be read phonetically. Its ideograms (picture symbols) are composed of special letters. When reading the ideograms, the result is not English language but an artificial language, which might serve as an international auxiliary language, as a kind of Esperanto with associated picture-writing. The original purpose of phonetic picture-writing was to create an artificial language that could be used to express everything both optically and acoustically. The ideograms can be combined into scenes, and these scenes can be spoken as sentences. Such an artificial language has many advantages: quick optical surveying, quick writing of the simply-shaped letters, very clear phonetics (good for speech recognition by computer).

2.3 Evaluation of Image Based Languages

There is no consensus about which symbol set is superior. However, there are some consistent comparison results between individual systems. Rebus and PCS are considered to be the most translucent overall; however, a number of Blissymbolics, Picsyms, and PIC symbols have also rated as highly translucent [16]. The PCS symbols were found to be more translucent than the DynaSyms, which in turn are more translucent than the Blissymbols for

European/American, Mexican, Chinese, and African-American individuals [39]. The same study also shows that some symbols were perceived differently by persons from different backgrounds. A comprehensive study was conducted by Mirenda et al [59]. The symbol sets included non-identical objects, miniature objects, identical colored photographs, non-identical colored photographs, black and white photographs, PCS, Picsyms, Rebus, Self-Talk, Blissymbols, and written words. Real objects were found to be the most readily recognized and Blissymbols and written words the most difficult. PCS symbols seem to be consistently rated as the most transparent 2D images. Symbol sets were also studied to determine if normally developing 3-year-old children showed any significant differences in terms of transparency and learning rates. PCS and Picsyms were found to be more transparent and easier to learn than Blissymbols, regardless of the symbol category [60].

Chapter 3

Design and Implementation of the Communication Aid Interface

A well-designed AAC device should be able to address the personal needs of each individual, yet it should also be general enough for a large population of users. The interface ought to facilitate easy skill acquisition and at the same time be sufficiently versatile not to limit social interaction. The interface design process represents a constant evaluation of trade-offs, since a wrong decision at this level of construction could make the difference between a useful and a useless device. In this Chapter we explain our development process and the choices that we had to make during the early design stages.

3.1 Language Representation Method

The language representation method is the interface between the means of selection and the generated communication. The three most common language representation methods are alphabet-based, single meaning pictures, and semantic compaction [71]. These methods provide access to the aid's lexicon.

Alphabet-based methods include spelling, word prediction, and letter coding. Spelling is attractive because it requires only a small symbol set. In many languages, 25-30 letters (depending on the language) and the "space" character allow the user to spell any word. The adverse aspect of alphabetical methods is the number of selections that must be made to convey a meaningful message. If a person has any motor difficulties, the selection process may be slow and result in an inadequately low communication rate. One effort to expedite the process was the development of word prediction systems in which a computer guesses the word that is being spelled. Letter codes, or abbreviations, can also be used with some success. However, conflicts arise even with a small vocabulary, reducing the possibility of this method developing as a viable option.

The semantic compaction language representation method uses short sequences of symbols from a small symbol set to define words and commonly used phrases [5]. The multi-meaning icons can be fit in a single selection area, so multiple pages are not necessary. As with other methods, training is needed for proficient use of semantic compaction. However, the static user interface of the semantic compaction method encourages the selection process to become automatic. Unlike word prediction, the semantic compaction keystroke reduction, relative to spelling, results in enhanced communication rate. The adverse feature of this method is that only those words previously stored in the system can be selected and spoken.

The third method, single meaning picture systems, uses arrays of unique meaning images. In contrast to semantic compaction, in this system each word in the vocabulary is represented by a different image. Hundreds of pictures are needed for a modest vocabulary size. Many meanings must be taught since pictures do not naturally represent most words. To construct a message an individual selects a number of icons forming a useful concept or even a complete sentence. Since only a limited number of images can be displayed concurrently these systems often have multiple display boards. On low-tech systems a caregiver needs to change the aid's overlay to provide the user with access to new vocabulary. In high tech systems, the displays are usually dynamic allowing vocabulary access through hierarchically arranged multiple displays.

To fulfill the majority of the communication needs of a disabled individual it is often necessary to use more than one language representation method. For example, core vocabulary may be accessed using semantic compaction or whole words, while extended vocabulary may be accessed using either alphabet-based methods or single meaning pictures, depending on the literacy skills of the user.

When deciding whether to use textual or visual representation of words in this project, the following considerations were taken into account. Textual representations can be more compact, more flexible, and easier to learn. However, they require literate users, make quick selection harder, and impose a single language. In addition, most existing communication aids are image based. As we are targeting people affected by cerebral palsy who are often preliterate, we decided to use an image-based system for access to core vocabulary. The core vocabulary included in the aid is relatively large; thus, the system we chose for access to extended vocabulary is an image-based alphabet (3-1). The alphabet can be used to spell any word whenever necessary. Image-based boards are practical because they are easier and faster to scan, and they can be used by preliterate children or alliterate individuals, as well as non-English speaking users. Based on an evaluation of all these factors, we resolved to use a *single-image* system. Furthermore, the icons can be annotated by text to ease the learning curve where appropriate.

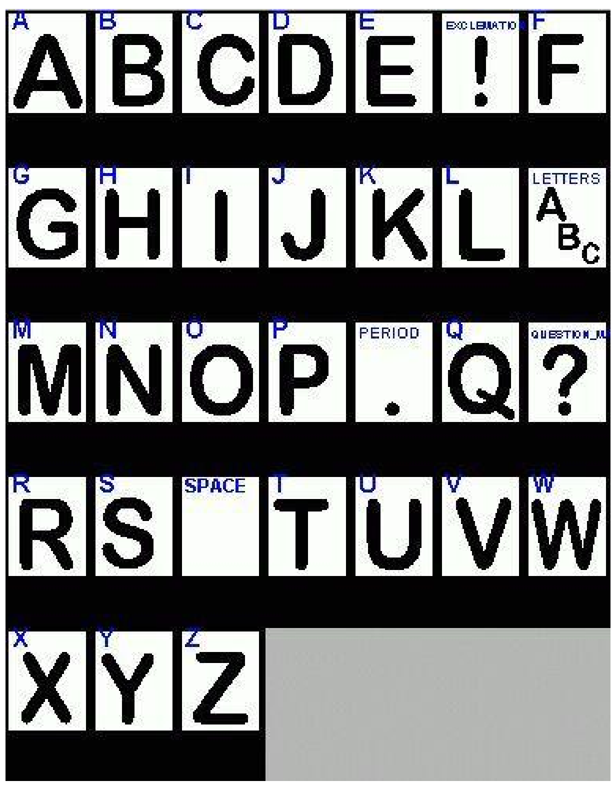

Figure 3-1: An image-based alphabet array can be used to access any word through spelling.

After extensive research we decided to use drawing-quality images. This level of detail facilitates the communication of targeted information. Using photo-realistic imagery might obscure the message with unnecessary information. Even though they are not optimal, we found PCS symbols to be the best available both because of their transparency and because of their prevalence in the field. The set includes images not only for physical objects but also for more abstract concepts such as feelings and quantities, allowing for a diverse vocabulary.

Therefore, our initial implementation used the PCS symbols. However, still not fully satisfied with the transparency and consistency of the images, we started work on the development of a new picture language. As the basis for this language, we use the Elephant's Memory image set. We chose this system because of its flexibility. We used some of the preexisting images and developed some other images for words in our lexicon that were not adequately represented. EM explores relative positioning for meaning purposes. This decreases the number of required symbols and gives a more consistent look to each phrase, speeding up communication.

3.2 The Lexicon Selection Process

In augmentative communication, various language corpora assembled throughout the past 20 years have shown that children, adolescents, and adults use the same core vocabulary. Further, these studies show that the same core vocabulary is used across environments.

Everyday speech is made up of core vocabulary and grammatical morphemes, yet these are not the focus of vocabulary development in augmentative communication. Instead most systems focus on the *"power words"* in each environment. Fringe vocabulary is considered powerful because simply by mentioning a word, it is possible for a conversational partner to fill in the blanks. However, when we allow the conversational partner to fill in the blanks, what we are doing is allowing him or her to guide the conversation, direct its contents, as well as many other things typically developed speakers would never allow [6].

Core vocabulary words are those few hundred words that constitute the vast majority

(85-95%) of what is said. Extended vocabulary words are the remaining words, generally numbering in the thousands, that make up the remaining 5-15% of communication. Kucera and Francis' 1967 study [47], which analyzed written language samples for word frequency, revealed that the top 10 words account for 24% of written text, the top 50 words account for 41.2% of written text, and the top 100 words account for 48.1% of written text.

We have chosen to make the vocabulary for our aid very flexible. We made it easy for the user to add and remove words. Also, the aid will remove or re-categorize words based on usage. However, it was still necessary to decide on some initial vocabulary that could be used to "bootstrap" the communication aid. This vocabulary was compiled based on existing core vocabulary usage frequency lists, the vocabulary list for Odgen's Basic English, and statistics run on collected experimental data.

3.2.1 Frequency lists

Since high frequency words are representative of the most commonly used words it seems reasonable to assume that they are the most useful for communication. Unfortunately, frequency lists are compiled based on specific data corpora, which are not always are representative of the needs of each individual. To adjust for this, we used five different generic frequency lists. One of the lists we compiled using the Brown corpus [19]. The other four are pre-existing lists compiled for various uses:

- Core Vocabulary List: High Frequency Words Vocabulary for Young Adults
- Unabridged Vocabulary Lists with Use Statistics for Non-Disabled 20-30 Year-Old Adult Vocabulary
- AAC User Vocabulary List
- Universal English Frequency List

All except the last one are taken from Barkeley Memorial Augmentative and Alternative Communication Center's web site. A general ranking was established using all lists and assigning weights based on their order within each of the five lists. The lists were limited to 1000 elements. If a word was not present in a list, a penalty was assigned.

3.2.2 Basic English

The vocabulary list for Ogden's Basic English is based on the notion that many words are redundant and that combinations of simpler words can replace many more complex ones. The language was built based on the 25,000-word *Oxford Pocket English Dictionary*. The result was 850 words that can be used to describe 90% of the concepts in that dictionary. In addition to reducing the necessary vocabulary, the shortened list simplifies the effort to learn spelling irregularities. The rules of usage are identical to English so that the practitioner communicates in perfectly good, yet simple, English. Basic English was developed by Charles K. Ogden in 1930 [62]. The words in Ogden's list were used by us, as a filter to reinforce the concept of a simple core vocabulary.

3.2.3 Contextual frequency lists

Finally, a more representative list has been generated based on data collected from everyday conversations (described in detail in Chapter 6). Non-staged everyday conversations have been recorded in such settings as kitchens, bathrooms, bedrooms, bus rides, and stores. All conversations were collected for the same person in order to establish a consistent user profile. Only one side of the conversations was recorded due to privacy issues. The data was transcribed and associated with corresponding general positioning information. A frequency list was extracted from this data and was used in combination with the other frequency lists when establishing the complete word ranking. This contextually infromed frequency list was given havier weight in the process.

3.3 Picture Book Format

The interface of the device is modeled on the Picture Book framework, which is a very commonly used setup for AAC devices. Choosing a familiar interface avoids the problem of requiring the user to learn a new interface. The device also makes it possible to use a set of images that the user is already familiar with, when it is necessary to ease the transition. The communication aid interface consists of four major parts (as shown in figure 3-2).

The uppermost horizontal part of the screen is dedicated to functional components and immediate visual feedback of message construction. The remainder of the interface is dedicated to the aid's vocabulary. The lexical part is further divided into three sections: sentence template array (left column), category array (middle columns), and lexical array (right columns). Area specific layout considerations are described in the following section.

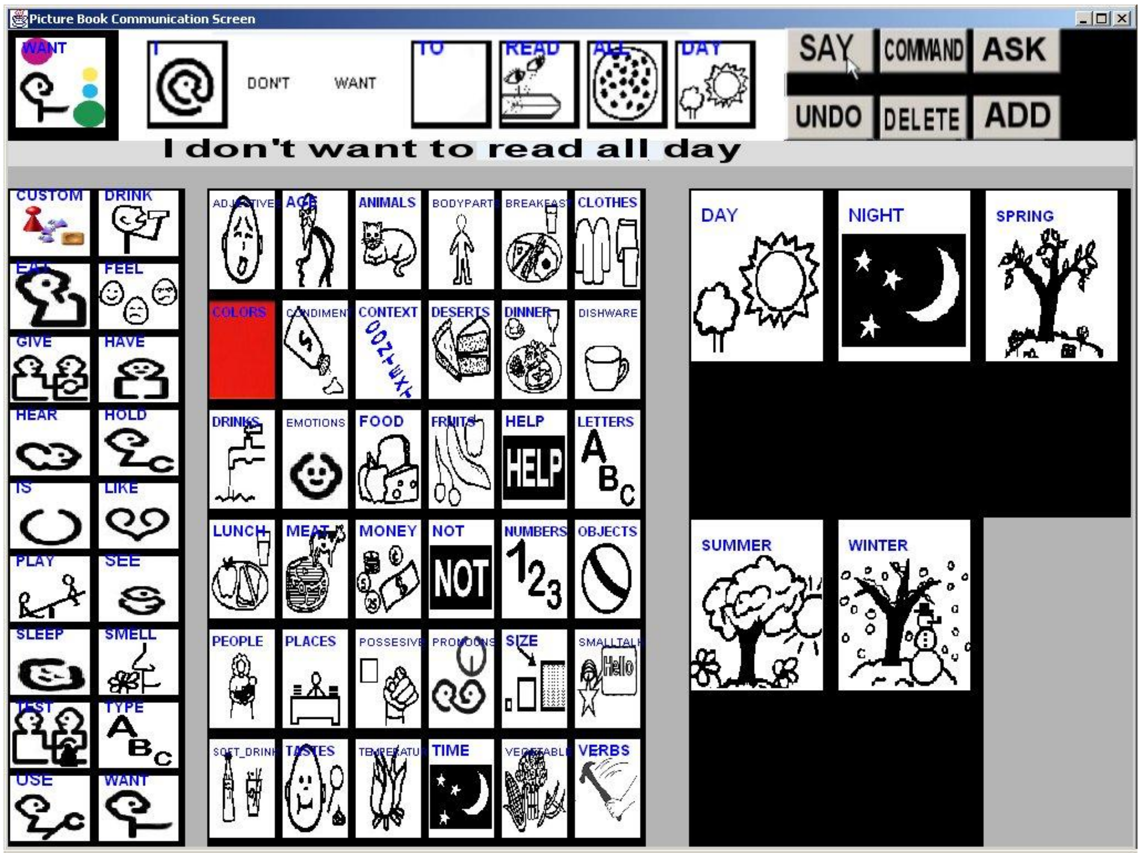

Figure 3-2: The interface has four major parts. The top of the interface shows all functional components. The bottom gives access to the lexicon. The left part shows a list of avaliable semantic frames. The center gives access to all categories, and the right part shows all vocabulary available for the currently selected category.

3.4 Layout Organization

Blackstone [13] shares an organizational arrangement of vocabulary as outlined by Janice Light. It includes schematic, taxonomic, semantic-syntactic, and alphabetic groupings. Schematic organization of vocabulary (based on biographical information and life experiences) tends to be particularly helpful for young children and people with memory/cognitive deficits or dementia.

The goal of communication boards/displays "is to arrange language in space so individuals can, by selecting from the available options, say what they wish to say as quickly as possible, and can do so with a minimal amount of effort."[13] This purpose might seem straightforward, but the process of creating functional layouts is very complex.

3.4.1 Symbol layout

When deciding where and how to display images there are a number of issues to consider:

- Display of symbols vs. display of functional elements

  In our design we draw a clear distinction between the placement of lexical symbols and functional elements. We are designing a dynamic display AAC, yet we keep all functional elements in place. Functional elements pertaining to sentence construction are all located in the top area of the screen as shown in Figure 3-3.

Figure 3-3: All functional elements of the aid are located on the top part of the screen.

  The functional part of the display has three subparts:

  1. *Current sentace frame* (the left most element in Figure 3-3)

     This section shows an image of the currently selected semantic frame with the avaliable slots.

  2. *Current sentence state* (the center part of Figure 3-3)

     This shows the sentence as it is being constructed. Each selected word fills in the next empty slot.

  3. *Controls* (the right part of Figure 3-3)
     There are six control buttons:

– SAY - outputs the currently constructed message
– COMMAND - outputs the current phrase as a command
– ASK - allows the user to construct a *wh-question* using the current message
– UNDO - undoes the last slot filled in the current frame
– DELETE - deletes the next avalible slot in the current frame
– ADD - adds a new slot before the first unfilled slot in the the current frame

- Number of symbols

  The number of symbols depends on the number and size of categories. The user can directly manipulate the quantity of images displayed simultaneously. The number of symbols in this aid ranges from 3 icons in the age category to 43 in the desserts category.

- Modularity vs. large communication fragments

  The aid offers three levels of communication. A user can choose from a precompiled list of small talk phrases such as those shown in Figure 3-4, iconic word representations, and single alphanumeric characters shown earlier in Figure 3-1.

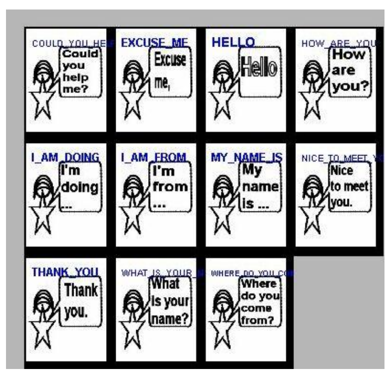

Figure 3-4: The user has access to a category of commonly used small talk phrases.

  To establish some level of consistency words are not reordered within a category. Instead all most appropriate words are available through the context category (Figure 3-5).

- Frequency or use ordering

  A special CONTEXT category allows the user to acess high frequency vocabulary appropriate for the current situation.

Figure 3-5: The contextual vocabulary changes based on the current environment. This figure shows four different vocabularies with their corresponding contextual environments. Starting from the top: bathroom, bedroom, generic, and kitchen.

3.4.2 Hierarchical vocabulary structure

Due to display size limitations and in order to facilitate easier access, the system will not simultaneously display the complete lexicon. Since different images will be available on different boards it is necessary to establish board ordering and access. We have decided to semantically group the lexicon through hierarchical categorization. This means that categorization is done on the basis of meaning rather than alphabetical order. For example, dog will be classified as a pet, which in turn will be classified as an animal, which in turn will be classified as a living thing. The hierarchy is intended to be very shallow. Most words will belong to only one category but some will span across several categories; for example fruit could be dessert or a plant. In order to maximize search efficiency the structure is intended to form a well-balanced tree. This will provide access to all words through only one to three selections.

Taxonomic groupings (category based) are also useful for clarifying or for conveying simple messages such as requests. However, because they tend to have one word class per page (such as nouns to depict animals, clothing items, or other categories), it is very hard to use taxonomic displays for full conversations. To mediate this we have augmented taxonomic grouping with a cross-category contextual display.

3.5 Use of Semantic Frames in Place of Syntactic Frames

Yet another approach to word prediction is introduced through sentence frame systems. A *fame system* assigns a type label to each lexical item. Then as the user selects words, each new word is associated with a consecutive *slot* within the initial frame. Each new word reduces the number of frames satisfying the lexical history. Reciprocal reduction of possible frames limits the number of lexical items that fit the current pattern. In this way we can greatly limit the number of items that the user is allowed to select.

The literature distinguishes three types of frame systems:

1. Surface Syntactic Frames

   – These are mainly verb and noun structures. They use grammatical constraints and parts of speech information for frame design.

2. Surface Semantic Frames

   – These utilize an action-centered meanings of words. They include qualifiers and relations concerning participants, instruments, trajectories and strategies, goals, consequences and side-effects.

3. Thematic Frames

   – These use scenarios concerned with topics, activities, portraits, and setting. They encompass outstanding problems and strategies commonly connected with a topic.

3.5.1 Syntactic frames

The most common, although the least effective, are syntactic frame systems. Syntax pertains to the rules for the formation of grammatical sentences in a given language. The arrangement of words in sentences, clauses, and phrases, and the study of the formation of sentences and the relationship of their component parts can be used to build a set of valid frames for a language. In a language such as English, the main device for showing the relationship among words is word order. For example, in the sentance “The girl fights the boy.” the subject is in the initial position, and the object follows the verb. This means that Syntactic Frames build on the position in which a word occurs relative to other word classes (where a word class denotes part of speech designation) in the same phrase (e.g., "Det N" is a syntactic frame for nouns indicating that nouns are preceded by determiners).

The members of syntactic frames (lexical labels) are composed of one or more words that form a phrase. One type of phrase is usually interchangeable with another phrase of the same category. The most common categories are NP (noun phrase), VP (verb phrase), PP (prepositional phrase), ADJP (adjective phrase), S (sentence), and word level parts of speech. This allows for recursion of constructs.

The most common lexical specification of syntactic frames is one of subcategorizational structure of verb description. Abstracting away from theory-and system-specific coding protocols, we may characterize a frame as a generalization over different syntactic contexts required by a verb and that are associated with the same syntactic behavior.

Syntactic frames can usually be automatically constructed with no or little ambiguity. The number of phrase types recognized by syntactic frames is finite and minimal. The phrases are constructed from syntactic constituents. These constituents occur either as arguments or as lexicographically relevant modifiers [57].

The types are the following:

1. Noun phrase types
   (a) Noun phrase (the witness)
   (b) Non-maximal nominal (personal chat)
   (c) Possessive NP (the child's decision)
   (d) Expletive there (there was a fight)
   (e) Expletive it (it's nice that you came)
2. Prepositional phrase types PP
   (a) Prepositional phrase (look at me)
   (b) PP with gerundive object (keep from laughing)
   (c) Particle (look it up)
3. Verb phrase types
   (a) Finite verb phrase (we ate fish)
   (b) Bare stem VP (let us eat fish)

(c) To–marked infinitive VP (we want to eat fish)
(d) WH–VP (we know how to win)
(e) Gerundive VP (we like winning)

4. Complement clause types

(a) Finite clause (it's nice that you came)
(b) WH–clause (ask who won)
(c) If/whether clause (ask if we won)
(d) Gerundive clause (we saw them running)
(e) To–marked clause (we want them to win)
(f) For–to-marked clause (we would like for them to win)

Frame semantics characterizes the semantic and syntactic properties of predicating words by relating them to semantic frames. These are schematic representations of situations involving various participants, props, and other conceptual roles, each of which is a frame element.

Syntactic frames view language as a linear string. Frames can be organized in a structure that can be modeled using an inheritance lattice. They range from being very general, like case frames [31] or other simple event schemas underlying thematic roles, to being very large, lexically specific ones. The most interesting frames are those at an intermediate level of specificity that encapsulate generalizations about the semantic and syntactic properties of word classes that are overlooked by thematic role analyses.

3.5.2 Semantic frames

Many sentences can be strongly constrained by the verb or verb-noun combinations that they use. For example, the sentence that describes the action of giving will most likely have a giver, a receiver, and an object that is being transferred. Peshkovskij claims that "…verb usage is the basic form of our linguistic celebration. The predicate–the verb–is the most important member of the sentence and of our speech in general" [68]. Thus, it should be possible to tightly constrain the lexicon given a central verb.

A semantic category covers a class of words similar in respect to meaning, for instance the class of ANIMATE, HUMAN, ADULT nouns. This is unlike the syntactic category, which covers words, which occur in similar syntactic contexts (traditionally these cover NOUN, ADJECTIVE, ARTICLE, PRONOUN, VERB, ADVERB, PREPOSITION, CONJUNCTION, INTERJECTION).

Frames can be defined for each verb type by providing:

- A list of slots (complements) having phrasal realizations
- Categorical and morpho-syntactic information for the lexical unit
- Control information for equal and raising verbs
- Surface order information

A frame represents synthetically a set of possible syntactic structures associated with the verb. In a frame, the combinatorial realizations of slots are supposed to yield possible surface instantiations of the frame. In other words, for some syntactic structure associated with a verb, each slot of the verb frame will either have some realization, or it will have none if the slot can be optionally left unrealized. Constraints on realization give criteria to split frames in some cases. Further, slots are not independent of each other from the point of view of their realization. They are linked through semantic constraints.

Slots provide information about subcategorized elements. Slots can be accessed with an index, imposing canonical order in the frame and semantic compatibility. A canonical order can be given over slots based, for instance, on a canonical order given over the functions. This criterion for ordering is not formally constrained by the frame, but each slot necessarily has an index that gives its range in the frame. This canonical order is not necessarily linear. A Boolean attribute optionally specifies if a slot has to be realized.

Slot realizations are characterized syntactically and semantically by:

- Parts of speech category label
- Function
- Semantic class restrictions
- Thematic role

Semantic class is one of the semantic characteristics of slot realization. The semantic class expresses a restriction on the semantic interpretation of the slot realization as belonging to a certain semantic class and semantically constrains the set of possible realizations. Semantic class is characterized by:

- *Name* (e.g. animate, human);

- *Class* – hierarchical link to another semantic class; simple or multiple hierarchies can be described through this link.

To constrain each message through a verb selection is necessary to classify verbs by type. One such verb classifications was introduced by Levin [52], and another is prominent in *WordNet* [58]. Since sets of semantic components can be associated with lexical items, in particular primary senses of verbs, which will account for most of their syntactic behavior, it is possible to implement verbs as sets of features. This provides a more flexible representation than the rule-based Lexical Conceptual Structures, allowing for more robust processing and best partial matches [66]. In addition, this approach is more amenable to empirical methods, since a distributional analysis of syntactic frames should provide critical information regarding a verb's semantic classification, not just for English but for all languages [65]. These semantic classifications, although potentially quite diverse, should share key cross-linguistic semantic components as suggested by Talmy [76] and Jackendoff.

For the present purposes, by *lexical selection* we mean the specific lexical requirements that a verb imposes on its subcategorized context, e.g. the selection of a particular preposition introducing a given complement. All lexicons considered in this project express, in some way or another, this type of information with varying degrees of granularity. Typical instances of lexical selection are bound prepositions, particles, complementizers, impersonal subjects, and clitics. Bound prepositions and particles are specified in all lexicons through listing and/or by indication (when possible) of a relevant class of prepositions or particles (as in the case of locative prepositions).

3.5.3 Thematic frames

Finally, *Thematic Frames* are a somewhat differently constrained version of *Semantic Frames*, even though they are distinguished from them in the literature. Thematic role is one of the non-mandatory semantic characteristics of slot realization. The thematic role expresses the role of the slot realization in the set of semantic participants of the situation or event associated with the frame. The thematic role is characterized by:

- *Name*;

- *Role features* – the set of elementary characteristics of the thematic role, e.g. cause, change, etc.

The lexical roles hierarchy includes the following roles in descending order: agent, beneficiary, recipient and participant, instrument, patient, theme, location, motive. The action is constant within a frame; thus, it is not a viable slot option.

An example of a basic verbal communication frame would contain slots such as:

- Speaker (A person who performs the act of verbal communication)
- Addressee (An actual or intended recipient of a verbal message)
- Message (A communicated proposition)
- Topic (The subject matter of a message)
- Medium (A physical channel of communication)
- Code (The language or other code used to communicate)

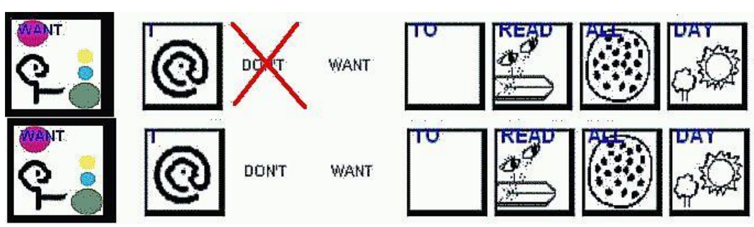

Figure 3-6: An example of a one–step undo of default negation.

These slots all derive their names from the concepts explicit in a basic communicative event. A true frame representation cannot just consist of a list of role names, but it must characterize the actual events that it describes.

In our design we have tried to combine the related concepts of semantic and thematic frames. Our frames are constrained by the action. We have developed a system in which semantic themes are first selected, and then completion is achieved by filling thematically appropriate lexical items into the set of semantic slots. All slots need not be filled to create a complete sentence. Often a number of optional slots is provided to facilitate richer communication. The optional slots are usually for count nouns, temporal descriptors, and other modifiers. However, some slots have to be realized in order to communicate a message. For example, in a “give” frame, if a sentence does not realize the item that is being transferred it is not permissible. Each semantic frame might also have a set of default words that can be deselected. One example would be negative voice as shown in Figure 3-6. The user must deselect the negative markers to construct a message with a positive voice.

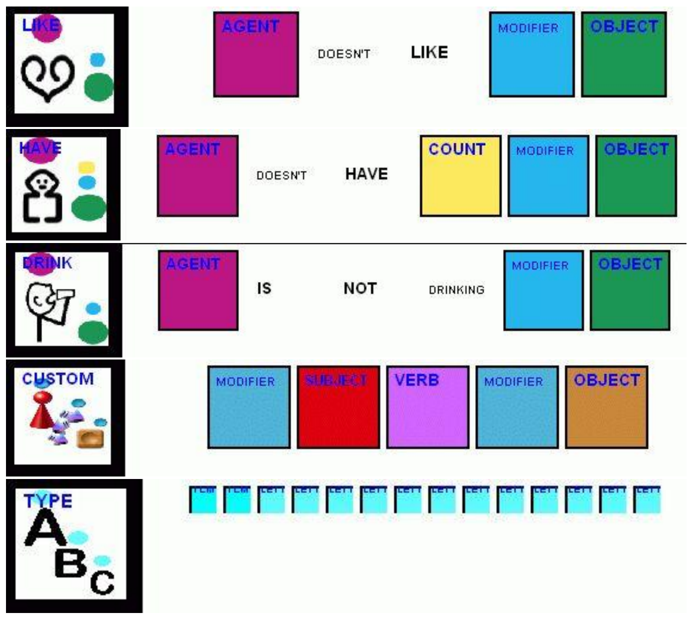

Figure 3-7: Examples of semantic frames. The *like frame* and *have frame* are rather generic templates. A *drink frame* shows a better example of constrainment. The object in this frame has to be a liquid. The user has the option of bypassing the frames by using a *custom frame* or a *type frame*.

Figure 3-7 shows an example of several semantic frames. To realize a frame the user selects a word for each slot from the lexicon through the image display.

3.6 Message Construction

3.6.1 Input mode

Currently the input mode for the device is a touch screen. The aid does not require exact touch resolution, since the average number of images per display is 15-20. An even coarser icon placement may be enforced if necessary.

The interface is designed in a modular fashion (a client-server setup), which makes it easy to substitute other input modes in place of touch. Some input modes that are currently under consideration include gaze tracking, head switch scanning, and haptic feedback controllers for more customized input.

3.6.2 Building a message

Existing image boards require the user to construct a sentence though valid sequential symbol selections. The user needs to select the subject first then the verb and finally the object. This type of message construction is not dependant upon the specific information that the user is trying to convey. Further, it increases the cognitive load by requiring the user to remember partial sentences and be aware of grammatical constraints.

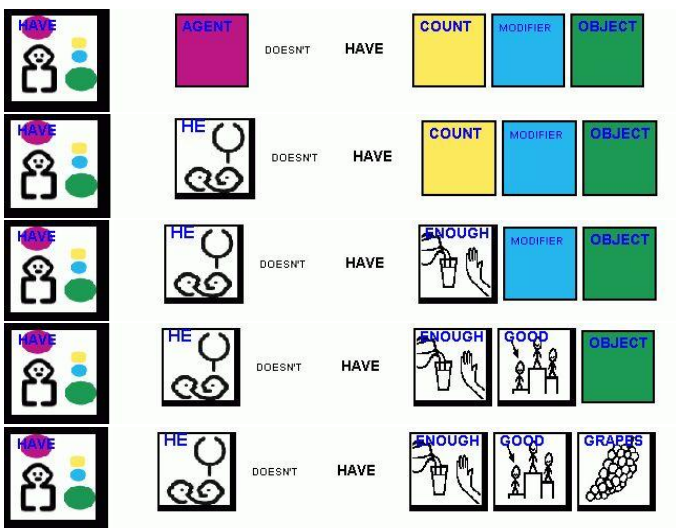

Figure 3-8: Example of the use of a frame. We start with an empty *have frame*, and fill each slot by selecting the corresponding slot realizations from the available vocabulary.

In contrast, our system lets the user select a relevant sentence frame. This makes the message construction much simpler. Now the user only needs to choose words that will fill in the meaning-carrying bits. Only minimal vocabulary is required for generation of complete messages as shown in Figure 3-8. The user does not need to be aware of the grammatical constraints of the language. More importantly, the user does not need to fill in irrelevant information in order to produce complete grammatical sentences.

Given the right sentence frame, the realization of the message is accomplished by the selection of lexical items. Since lexical items are organized into semantic categories, the user might need to first choose the appropriate category. When a category icon is selected, the aid's lexical array is updated accordingly. For example, selecting the "animal" category would update the lexical array to include icons for "dog," "cat," "monkey," etc. The lexical array can also be updated by the user's situational context. For example, if the user is in a restaurant, food items would be called up into part or all of the lexical array. The notion of contextual category is described in detail in Chapter 5.

Chapter 4

Background on Word Prediction

As mentioned earlier, one technique often used to improve communication rates in AAC devices is *word prediction*. Word prediction aids in sentence construction by suggesting and predicting words. Word predictors can learn as the correct text is being typed. This dramatically aids the development of literacy and increases communication accuracy.

In AAC, word prediction technology is usually used to assist with text entry. Software packages predict the word you are typing and the next word based on word frequency and context. They may also include features such as spell-checking as you type, speech synthesis, and hotkeys for frequently used words. Word prediction is particularly useful for slow typists, probe or pen users, and people with motor impairments. Word prediction is often used to help people with some types of spelling or word retrieval difficulties or to reduce the number of keystrokes required.

It is important to distinguish between the two major types of word prediction:

- *Word completion*, often referred to as word prediction, describes the notion of word prediction based on letters typed for the given word;
- *Word prediction*, sometimes referred to as phrase completion usually denotes sequential prediction of words based on words already chosen for a particular sentence.

Average rates for machine word prediction are considered to be 48% for simple prediction (pure statistical prediction) and 54% for advance prediction that includes grammatical, lexical, and/or contextual information.

4.1 Necessity for Word Prediction in AAC

When using single meaning pictures or alphabet-based boards the user often has to use many selections in order to convey even a simple message. With single meaning pictures a simple concept requires a number of pictures equal to the number of words in the vocabulary. A normally developing three-year old has a vocabulary of about 1000 words [64]. A set of 1000 pictures becomes very difficult to access in a way that would allow fluent communication to occur. For example, this relatively small vocabulary would require 20 screens of 50 pictures each. So in order to access a specific word a user might need to navigate through a series of screens.

Similarly, with text based aids the user needs to construct each word using letters. In basic English average sentence length is 15 to 20 words and the average length of an English word is around twelve letters. This means that a user might need over 100 selections to construct a sentence in simple English. A method is required that will improve the rate at which users can participate in a conversation using all types of AAC devices.

Prediction systems are amongst the methods used to facilitate rate improvements. They attempt to predict the user's intended next word using various methods of computation. Many methodologies employ statistical methods within the user's language, ignoring any grammatical information. Thus many of the predictions produced are grammatically incorrect and as such wasteful to the user.

Many word prediction programs were first developed for students whose keyboarding skills were severely limited by their motor impairments. For students who cannot use the standard keyboard, alternate methods for selecting letters–switches, trackballs, head and mouthsticks–can be slow, increasing the gap between generating ideas and capturing them in writing. Word prediction reduces this gap simply by reducing the number of selections necessary for encoding words.

4.2 Types of Word Prediction

Word prediction systems have provided more and more advantages in the last few years, including features such as grammatical prediction, frequency, recency, and the ability to easily add words.

Intelligent word prediction often uses both knowledge of syntax and knowledge of word frequencies to predict the next word in a sentence as the sentence is being entered, and it updates this prediction as the word is typed.

Using grammar modeling, a system can first narrow the range of words that could follow the current one to those that are grammatically correct. When the more normal statistical methods are overlaid onto this method then the hit rate (the rate at which the system predicts a word correctly) will be higher. When using a grammatical model it is assumed that the words in the sentence are entered in correct grammatical order. The accuracy of the system's predictions will depend on the accuracy of the syntactic rule probabilities and word frequencies.

4.2.1 Word prediction in typing

In a letter-based system the first word of the sentence is predicted on the basis of the frequency of words in the sentence-initial position. Each word is incorporated into an augmented transition network (ATN), with a branch for every acceptable syntactic interpretation of the sentence segment so far.

The grammar contains probabilities for each ATN structure, and the lexicon contains the frequencies with which words occur in each of their possible subcategorizations. The combination of these statistics results in a list of possible next words, and a number of most probable words is offered to the user. The user may reject the predicted word list and begin entering the word by hand. As each letter is entered, words that are not consistent with the word segment entered so far are filtered out, and the prediction list is updated.

EZ Keys by Words+ has time-saving features, including dual word prediction and abbreviation expansion. When the user begins to type a word, EZ Keys displays a table of the six most frequently used words that begin with those letters. The user selects the appropriate word from the display, and EZ Keys instantly types the remainder of the word. In addition, EZ Keys features next word prediction, where the program actually learns your word patterns and displays a list of the last six words you used in conjunction with the previous word. Simply select the correct word and EZ Keys types it for you. World-renowned astrophysicist Stephen Hawking uses Words+ software to deliver lectures around the world.

4.2.2 Frequency-based word prediction

The basic prediction techniques rely on the use of frequencies of individual words independent of context. Statistical word prediction language statistics are a useful information source for most of the existing commercial and academic word-prediction systems. The statistical information used in prediction systems consists mostly of n-gram language models.

Most of the early word-prediction programs, especially those that were developed as writing aids for people with disabilities in the early 1980s, used word frequency information (i.e. a word unigram model) to predict words for current or next position. These predictors ignore all previous context, and use only word frequency information in their prediction process. In order to improve the accuracy of predictions, some predictors consider a larger context and use word sequence statistics, such as word bigram or trigram models.

A major drawback of both unigram and bigram predictors is that they do not consider the syntactic and semantic structure of the sentence, and therefore there is a possibility of predicting words that are syntactically and/or semantically inappropriate for the desired sentence position. Another disadvantage of such predictors is that they consider only a small context.

WordQ software is a writing tool that was developed at the Bloorview MacMillan Centre, Toronto, Canada, with support from the Ontario Rehabilitation Technology Consortium [61]. WordQ can be used along with standard Windows word-processing software. It has a word prediction module that suggests words to the user and a speech module that provides spoken (text-to-speech) feedback.

The WordQ prediction module uses word unigram and word bigram statistics along with optional customization to each user's vocabulary in order to suggest appropriate words. Wherever the bigram probability exists, the WordQ prediction algorithm suggests words according to that probability. In this case, the predictor suggests words whose bigram probabilities are higher according to the given previous word. Otherwise, it uses the unigram language model to predict the next word, i.e., it ignores the previous word and predicts only according to the words' frequency of use. WordQ is adaptive, and thus the vocabulary used for prediction is customizable depending on the user's level and the specific topic selected by the user.

4.2.3 Syntactic word prediction

Prediction systems that attempt to guess what the user will say next based purely on statistics can result in the prediction of unusable words. This imposes additional cognitive load on the user, decreasing the text composition rate. Various studies confirm that removing the syntactically inappropriate words from the list of suggestions improves the comfort of users, even if the actual physical effort saved does not change much [2]. Therefore, a system should only produce syntactically correct predictions.

Therefore, various prediction systems have been developed that attempt to include syntactic information in their prediction process. The goal of syntactic prediction is to ensure that the system does not suggest grammatically inappropriate words to the user. Some of these systems consider part-of-speech tag information of words as syntactic information, while others use a parser to build the syntactic structure of the whole sentence.

An important issue, however, is that the prediction system is required to be as fast as possible. But the more information sources the predictor uses to suggest words, the slower the system is. Also, it has been found in some experiments that by including syntactic information in prediction, appropriate words are removed from the prediction list almost as often as inappropriate words are prohibited from appearing on the list [2].

Another possible drawback of a syntactic predictor is that it cannot adapt itself to the user's vocabulary easily since all the words in the lexicon need to be tagged with their syntactic category and some other grammatical features. The Predictive Program is a program used as a writing aid for people with disabilities. It was developed at the University of Washington [79]. When the user of the system enters the first letter of the desired word, the system presents 6 to 10 possible completions for the existing word prefix. The predictions are presented according to the grammatical information of the preceding word(s) and the frequency of English words. Unfortunately, it is not clear how the prediction algorithm uses the grammatical information.

Windmill is a MS-Windows system that uses syntactically correct prediction. It uses an augmented phrase structure grammar to parse semi-complete sentences. The system is designed as an expert system with a rule base providing possible sentence-completion sequences from which parts of speech for the next word can be deduced. A type-tagged lexicon is then consulted to obtain words that fit these types. A statistical method, based on frequency and recency of usage, is applied to this output to produce an ordered prediction list of the top five words that are offered to the user [84].

4.2.4 Semantic word prediction

Semantic word prediction uses the notion of word association within a prediction system. Word association can increase significantly the efficiency (keysaving) a word predictor can achieve. Word association can be incorporated into a predictive algorithm without using syntactic or higher linguistic information, and so the algorithm is not language specific. The technique may therefore be applied to many languages with little modification [7].

Another way to perform semantic word prediction is through the use of semantic classification. Semantic classification can be used to aid retrieval of pre-stored messages [50] or when creating customized vocabulary for particular users in special situations [75]. Also, some situations are sufficiently important that a topic-specific set of words and phrases would be useful, even if those situations only account for a small proportion of the AAC user's time. Interacting with a doctor might be one such case.

4.2.5 Complex word prediction

Prophet is a frequency-based adaptive lexical prediction program that has been being developed since 1982. It is used mainly by persons with motor handicaps or linguistic disabilities such as mild aphasia and dyslexia. Prophet is currently undergoing major development to improve the prediction function. The goal is to enhance the current prediction capability by extension of scope, addition of grammatical, phrasal, and semantic information, and abandonment of a strictly frequency-based system for a probability-based one, allowing information from multiple sources to be weighted appropriately for each prediction. For this to be possible, a database of grammatically tagged corpora needs to be constructed [40].

The eventual goal of a language model is to accurately predict the value of a missing word given its context. One approach to word prediction is based on learning a representation for each word as a function of words and linguistic predicates in its context. This approach raises a few new questions. First, in order to learn good word representations it is necessary to use an expressive representation of the context. Second, since the number of words "competing" for each prediction is large, there is a need to "focus the attention" on a smaller subset of these.

4.3 Contextualy Grounded Dialogs

Most existing word classification systems depend only on statistical information. However, observation of real life conversations shows that in the context of a communication there is additional information available. Human conversations are coherent in that they follow a subject. Many conversations are dependent on simple contextual information such as current or past locations and times of the conversation occurences. Further, most conversations are reactive, so information can be acquired from all participants. Finally, conversations tend to be speaker specific; it is possible to learn not only the precise vocabulary used, but also vocabulary patterns reoccurring within the scale of days and in relation to physical locations.

A number of efforts have been made to apply models of conversation in order to introduce prediction into augmentative communication systems. At the word level, word predictors, which employ both frequency and recency information, are applying a model of word usage that is based on the observation that an individual's daily discourse usually consists of a relatively small number of words drawn from a large store of possibilities. At the level of an entire utterance, systems have been developed that apply predictive techniques using some sort of model of the conversation to refer to. These attempts have been based on knowledge derived from discourse and conversation analysis about openers, closers, back channeling, and stepwise topic change. The mechanisms underlying such prototypes have included transition networks, semantic networks, hypertext, fuzzy information retrieval, and scripts [3].

Communication between non-AAC users is usually cooperative, bi-directional, and multimodal. In the context of AAC, a conversational partner often becomes actively involved in the construction of the augmented speaker's message. The partner may ask questions, repeat part of the augmented speaker's utterance, or simply nod and smile in agreement. This feedback may in turn affect the message being produced by the augmented speaker.

Computer-based AAC systems currently "see" only the words the user selects. The less information the user provides as input, the less the likelihood of accurate output. Studies on manual AAC systems suggest, however, that other modes of communication may be preferred. For example, children with cerebral palsy chose to use vocalizations, gestures, or eye gaze as modes of communication far more often than their manual symbol boards in interactions with their mothers or their speech therapists. These multimodal modes of communication may be critical for fully understanding augmented interactions [80].

4.4 Clustering and Classification for Category Generation

Direct use of statistical language models for word prediction is difficult since pure statistical approaches encounter problems in reliably estimating probabilities from corpora given that many events are rarely seen. Word clustering and other similarity-based approaches have been proposed in order to improve generalization and surmount the sparseness problem [19].

The basic approach is to enhance the performance of word-level predictors by combining evidence from events that occur in the neighborhood of the candidate words with evidence accumulated from events that occur around words that are similar to the candidate words.

This requires some definition of word similarity or the closely related notion of word clusters (classes). Thus, a crucial ingredient of this approach is the ability to estimate word clusters from data before the word-level predictors can be used.

Assume, for example, that the task is to predict which of the verbs $v_1; v_2; \ldots v_k$ is more likely to occur in a given context. A plausible approach [23] is to consider a bigram language model and gather statistics on the occurrences of pairs (v,o), where v is one of the candidate verbs and o is the object of that verb. At decision time, the statistics collected on the current object o is used to determine the most likely verb. This approach, however, is prone to the sparsity problems; in the likely event that the current object o has never been observed with a candidate verb, its likelihood to occur is estimated by interpolating from other objects o that are similar to the current object o.

This has motivated studies of word similarity and clustering [67], [23], [53] so as to improve performance on these decision tasks.

In the next chapter we introduce the concept of similarity and clustering. We survey the existing methods and present a novel approach to learning word representations as a simple function of the context in which they occur.

Chapter 5

Generation of Contextual Categories

To improve the communication rate, we are trying to improve the word prediction accuracy. We need a way of generating categorical sets of probable words. To properly generate these word clusters we need to be able to classify a situation as either one that has already been encountered or as a novel one. In order to both form new categories and to properly categorize words it is necessary to establish a goodness criterion. The quality of local selection can be estimated by comparison with other possible selection. In order for such a comparison to work it is necessary to establish a distance metric.

In this chapter we will first motivate the use of clustering. Then we will introduce concepts relevant for clustering. The concept of distance and the different distance metrics will be discussed in detail. Then we will give a general overview of existing clustering methods. Finally, we will introduce a novel clustering method loosely based on Decision Trees classification algorithms.

5.1 Motivation for Use of Clustering

Clustering methods allow us to find structure within our data. One example of structure information is similarity of data, where similar information is collected within a cluster and dissimilar information is split between clusters. For our purpose we would like to find out what words are used within similar contexts. Therefore if we run a clustering algorithm on contextually tagged speech information we should be able to establish coherent word categories that can be used for predication whenever the user reenters a similar contextual situation.

Now we will cover some basic terminology and summarize existing clustering methods.

- Cluster

  A cluster is an ordered list of objects, which have some common characteristics.

We will describe in more detail the notions of similarity and distance measures.

5.2 Similarity Measures

A similarity measure $SIMILAR(C_i, C_j)$ can be used to represent the similarity between two clusters. Typical similarity generates values of 0 for clusters exhibiting no agreement among items, and 1 when perfect agreement is detected. Intermediate values are obtained for cases of partial agreement.

There are some heuristics which should be followed when using the notion of similarity. Similarities should be normalized by size of the cluster. The term weighting may not be of particular value. (For a cluster centroid, weights might even be re-computed based on ranks. For example, consider a cluster centroid with t terms, so the one with largest weight has rank 1 and the term with lowest weight has rank t. Assume we want a linear decrease in weights, from value 1 for the rank 1 term down to value $(1 - c)$ for the term with rank t, where c might be say 0.9. Then, we might use the following value for the weight of the term with rank i:

$$1 - c \cdot \frac{(i-1)}{(t-1)})$$

For all measure the numerator is based on the inner product, and the denominator is a normalization factor. For example for the Dice method you would multiply numerator by 2, and normalize by the sum of both sums of squares and for the Jaccard method you would normalize by the sum of both sums of squares, less the inner product cosine

5.3 Distance Metric

The distance between two clusters involves some or all elements of the two clusters. The clustering method determines how the distance should be computed.

5.3.1 Semantic distance

Semantic distance is a psychological construct that has been used to locate concepts along various dimensions of meaning [73]. A semantic distance effect is an inverse correlation between semantic distance and relevance assessment. This means that near descriptors are assessed as relevant, far descriptors are assessed as non-relevant. Further, relevance assessments are expected to decline systematically with semantic distance. One implication of a semantic distance effect is that the indexer-generated, topically relevant term would be assessed as most relevant.

5.3.2 WordNet

WordNet is a hand compiled lexical reference system whose design is inspired by current psycholinguistics theories of human lexical memory [30]. WordNet organizes English nouns, verbs, adjectives and adverbs into synonym sets (synsets), each representing one underlying lexical concept. The synsets are related through such relations as hypernyms, meronyms, synonyms, antonyms and many others.

WordNet was developed by the Cognitive Science Laboratory at Princeton University under the direction of Professor George A. Miller (Principal Investigator).

The ontology of WordNet is organized into hierarchies of semantic types: For example:

- Entities – concrete
  - Animate
    - Animal → Mammal → Human
    - Plant
  - Inanimate substances

        * Solids
        * Liquids
        * Gasses

- Entities – abstract
    - Event
        * Military
        * Athletic
    - Emotions

Semantic relations in WordNet include Inheritance, IS-A relations, Supertype/subtype, Hypernym/Hyponym, Part-Whole, meronym, and Synonyms.

WordNet is a lexical reference that is organized into synonym sets – related concepts.

The 5 out of 25 top most nodes in WordNet include:

- - act, action, activity
- - animal, fauna
- - artifact
- - attribute, property
- - body, corpus

5.3.3 WordNet distance measure

WordNet Distance measure reference to a familly of metrics that realy on word on WordNet's hierarchical ordering of words for simillarity calculations.

Our system computes semantic relatedness using number of links WordNet distance. Where the number of links is defined as the distance between the two words extracted from the unweighted graph representing hypernym relations [20]. The measurements are further scaled by a priori word usage and the semantic depth [51]. The categorization system uses entropy to establish actual information content [70]. Word frequencies have been established based on the Brown Corpus [19]; however, they are also augmented by frequencies from the everyday usage data described in Chapter 6.

5.3.4 Temporal distance measure

The temporal distance is a simple arithmetic norm. Times points are signified by a time measurement the time of day in each data point. The relative distance between two points can be calculated by subtracting the smaller time point from the larger one. If the points are more then 12 hours appart, the temporal distance is the difference between 24 hours and the two point time distance.

$$temporalDist = \min{(t_2 - t_1), (24 * 60 * 60 - (t_2 - t_1))} \tag{5.1}$$

5.3.5 Spatial distance measure

The distance metric used is a simple square distance:

$$physDist = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \tag{5.2}$$

where x and y are the latitude and longitude coordinates. If the difference between the two latitude or the two longitude distances is less than a small threashold, the distance is assumed to be zero. The usage of general positioning (GPS) data allows us to establish physical location clusters.

5.3.6 Contextual distance measure

Contextual distance metric aims at combining multiple context information such as relative time, geographic positioning, and semantic understanding. The metric is a weighted function of all contextual features.

$$dist = SEMANTICNORM * semDist + timespaceDist \tag{5.3}$$

$$timespaceDist = physDist^2 + TIMENORM \cdot \sqrt{temporalDist} \tag{5.4}$$

$$TIMENORM = 1 \tag{5.5}$$

$$SEMANTICNORM = 20 \tag{5.6}$$

5.4 Clustering

Clustering, is the problem of finding groupings of data points, where each point has a set of attributes, and a similarity measure among them. The groupings need to satisfy two conditions:

1. Data points in one cluster are more similar to one another.

2. Data points in separate clusters are less similar to one another.

Every clustering problem contains two subproblems: the definition of clustering metric (which we described in Section 5.3) and the clustering algorithm (which we will intruduce in Section 5.5).

Most of the existing approaches to clustering are based on either probabilistic methods, or distance, and similarity measures (see [32]). Distance-based methods such as k-means analysis, hierarchical clustering [41], and nearest-neighbor clustering [55] use a selected set of words (features) appearing in different parts of the data set. Each such data points can be viewed as a point in a this multi-dimensional space. A detailed background infrmation about clustering is included in Appendix A.

There are a number of problems with clustering in a multi-dimensional space using traditional distance or probability-based methods. First, it is not trivial to define a distance measure in this space. Simple frequency of the occurrence of words is not adequate. Furthermore, some words may occur frequently across documents. Techniques such as TFIDF [72] have been proposed precisely to deal with some of these problems.

Second, the number of all the words in the data set can be very large. Distance-based schemes generally require the calculation of the mean of clusters. If the dimensionality is high, then the calculated mean values do not differ significantly from one cluster to the next. Hence the clustering based on these mean values does not always produce very good clusters. Similarly, probabilistic methods such as Bayesian classification used in AutoClass [21], do not perform well when the size of the feature space is much larger than the size of the sample set. This type of data distribution seems to be characteristic of human speech. Furthermore, the underlying probability models usually assume independence of attributes (features). In many domains, this assumption may be too restrictive. It is possible to reduce the dimensionality by selecting only frequent words from within a context, or to use some other method to extract the salient features of each context. However, the number of features collected using these methods still tends to be very large.

Our proposed clustering algorithms which are described in the next section are designed to efficiently handle very high dimensional spaces, and furthermore.

5.5 Clustering Through Decision Tree Center Convergence

Not satisfied with our initial attempt to generate semantic classes using bottom-up hierarchical clustering we designed a clustering algorithm specifically for semantic clustering of individual words. The algorithm is aware of issues such as multiple word senses, class discontinuities, and universal, non-topical words. The resulting algorithm, which we called Decision Tree Center Convergence (DTCC) clustering is based on Quinlan's ID3 version of decision tree algorithm [69]. However, the original algorithm has been modified in two significant ways:

1. classification to clustering conversion

   The algorithm has been altered to work on unlabeled data

2. single word linguists awareness

   The algorithm is aware of and explores the lexicographical constrains of the data.

In this section we start by a short summary of the ID3 algorithm. Then we introduce the notion of perplexity used as a goodness metric in evaluation of clusters. Next we explain the necessary modifications and present the complete algorithm. Finally we discuss how the algorithm can be modified to incorporate both temporal and spatial information.

5.5.1 ID3 Decision Tree Classification

The goal of ID3 is to build a decision tree for classifying examples as positive or negative instances of a concept. A decision tree is a tree in which each non-leaf node has associated with it an attribute (feature), each leaf node has associated with it a classification (+ or -), and each arc has associated with it one of the possible values of the attribute at the node where the arc is directed from. Supervised learning is used to process a batch of training examples using a preference bias.

The original algorithm called ID3, was developed by Quinlan in 1979 [69]. The algorithm uses a top-down approach to decision tree construction. It recursively selects the “best attribute” to use at the current node in the tree. Once the attribute is selected for the current node, children nodes can be generated, one for each possible value of the selected attribute. The algorithm partitions the data using the values of this attribute. It then assigns subsets of the examples to the appropriate child node. The algorithm is recursive so everything is repeated for each child node until all examples associated with a node are either all positive or all negative.

The ID3 family of decision tree induction algorithms uses information theory to decide which attribute shared by a collection of instances to split the data on next. Attributes are chosen repeatedly in this way until a complete decision tree, one that can classify every input, is obtained. If the data is noisy, some of the original instances may be misclassified. It may be possible to prune the decision tree in order to reduce classification errors in the presence of noisy data.

The speed of this learning algorithm is reasonably high, as is the speed of the resulting decision tree classification system. The algorithm is also easy to generalize. The preference bias used by the algorithm insures that the simplest explanation that is consistent with all observations is the best. More specifically it means the smallest decision tree that correctly classifies all of the training examples is best. Finding the smallest decision tree is commonly unfeasible; however, due to the complexity of the search. Finding the provably smallest decision tree is NP-Hard, so instead of constructing the absolute smallest tree that is consistent with all of the training examples, we construct one that is pretty small. To achieve that use a greedy algorithm (we make locally optimal choices).

5.5.2 Entropy and perplexity measures

To terminate the clustering algorithm we need a way to evaluate each cluster. If the score of a cluster satisfies our judgment property the recursion is terminated. Information theory can be used to compute a cost function such as proportional entropy measure of a cluster.

In information theory, the term entropy refers to the relative degree of randomness. In linguistics entropy is the average surprise value associated with language events:

$$H(Language) = -\sum (\Pr(E_i) lg \Pr(E_i)) \tag{5.7}$$

A form of entropy known as cross-entropy can be used for evaluation purposes of language models. Cross entropy can be calculated by replacing the surprise value in the definition of entropy with an estimate derived from a model (while keeping the correct probability weighting). The resulting quantity is always greater than or equal to the actual entropy:

$$-\sum (\Pr(E_i) \lg \Pr(E_i)) \leq -\sum (\Pr(E_i) \lg \Pr(E_i|Model)) \tag{5.8}$$

The cross entropy can be used as a model evaluator on the assumption that better models will have lower cross entropy. However, this presupposes that the correct probability distribution is know, which is almost never true. The way to get around this problem is to estimate the correct probability distribution using a representative sample of the language in question (a test corpus). The cross entropy can then be estimated by the logprob (LP).

$$LP(Model) = -\frac{l}{n} \sum \lg (\Pr(Corpus|Model)) \tag{5.9}$$

However, a more common evaluator used to measure how surprised is our model M on seeing the actual language is perplexity.

The perplexity (PP) of a language is defined as 2 raised to the power of the entropy:

$$PP(Language) = 2^{H(Language)} \tag{5.10}$$

Again since we do not have access to the correct probability distribution we can use a test text T, and preport its perplexity with regard to the model M. The perplexity (PP) of a model (with respect to a test corpus) is defined as the number two raised to the power of the logprob:

$$PP(Model) = 2^{LP(Model)} \tag{5.11}$$

Perplexity is, crudely speaking, a measure of the size of the set of words from which the next word is chosen given that we observe the history of spoken words. The perplexity of a LM depends on the domain of discourse.

A language with perplexity X has roughly the same difficulty as another language in which every word can be followed by X different words with equal probability.

Perplexity, like entropy and logprob, can be computed per sentence, or per word. The per word perplexity is the more intuitive measure, since it can be interpreted as a weighted average of the branching factor after each word.

5.5.3 Modifying ID3 for clustering

ID3 consists of two major phases:

- building a decision tree
- classification of unlabeled data based on split features from the tree

In the case of clustering, the data we are working with has no preassigned labels. Further words do not have distinctive features that could be used to divide them into meaningful semantic classes. "Feature extraction" for words is discussed in the next section. However, if we assume words do have a continues feature, it will be possible to determine a split point for this feature by comparing entropy calculations for all possible splits. We are searching for a split that will result in the most homogeneous sets. Since the feature is continues it is computationally infeasible to test each point so we need a method that will supply us with a small number of most likely candidates. Once we discover the split point we can divide all data based on that point. The algorithm will run recursively splitting the words into new clusters until a stop condition is reached.

A trivial stop condition would be a 0 entropy measure for a cluster. This would create a clustering tree where the leafs would be all instances of individual tokens. Such tree would be similar to an inverted data structure created by a complete run of a hierarchical clustering algorithm. Thus a comparison of corresponding levels in these two algorithms can be used to judge the effectiveness of the DTCC algorithm.

A more useful stop condition would terminate the algorithm when the words are divided into meaningful topics such as: eating, working, or medical vocabulary. One way to measure the successfulness of the algorithm is to treat it as a language model, and compute the perplexity of each new cluster. If there is a significant increase in perplexity the split should not be executed. To measure the perplexity we need to reserve a test set. Now, we can generate a language model on the remaining cluster, and compute the log probability, which in turn can give us the perplexity for the current cluster. On the condition that we use the same language model type for each evaluation, the resulting perplexity variations will be evaluating changes in goodness of clusters.

5.5.4 Modifying ID3 for word clustering

As mentioned earlier the algorithm requires feature points that can be used for splits of the data set. Word sets are usually treated as very high dimensional spaces (where each token is a different dimension), such treatment does not lend itself to decision tree type space divisions. Instead we propose to first reduce the dimensionality of the space by projecting all dimensions onto a bounded 2 dimensional plane. This projection is accomplished by choosing two cluster centers, and by projecting all remaining words through their distance from the cluster centers. This projection can further be viewed as one-dimensional ordering

```
ChooseCenterClusters(tokens: center pool)
  FOR Each word i in the pool
    FOR Each word j = i+1 in the pool
     FOR Each word k = j+1 in the pool
     find closest(i,j,k)
     create LanguageModel(i,j,k);
  FOR Each LanguageModel i
   FOR Each LanguageModel j = i+1
   find furtherst(i,j)
   leftCenter = i
   rightCenter = j
  RETURN centerOf(leftCenter) and centerOf(rightCenter)
END
```

Figure 5-1: The center slection pseudocode for the DTCC clustering algorithm.

if we normalize each norm by the center distances.

Choosing the cluster centers is a multipart process:

1. the $n\%$ of highest frequency tokens is selected from the data training corpus

2. the half of these tokens that carries least information (has lowest entropy based on general word frequency for English taken from the Brown corpus) is discarded (only high frequency, and high entropy words are used for center selection)

3. these tokens are arranged into three word clusters using hierarchical clustering

4. the two furthest clusters are selected

5. the geometric centers for these clusters are used as categorization features for the original cluster

Figure 5-1 shows a pseudocode version of the algorithm.

Given these feature centers all words in the current cluster are divided based on their proximity to each center. The distance used is the WordNet distance 5.3.3 scaled by the frequency ratio of the two centers.

When $n\%$ of tokens is less than 9, then one word clusters are chosen for the center splitting. When $n\%$ of tokens is less than 3, then the 3 most frequent words are chosen and none are rejected for low entropy. These are boundary cases that will almost never occur in a real working system. They will arise only given a very small cluster sizes. In most cases the algorithm will have terminated long before such cluster sizes are reached.

5.5.5 Cluster evaluation

Each cluster is evaluated by calculating a perplexity measure for its n-gram language model. A non tainted test set is used for this calculation. If the relative per word cluster cross-perplexity increases by more than $x\%$ the algorithm reverts to the previous cluster. A small increase in perplexity is allowed in order to avoid converging on local minima.

5.5.6 The DTCC clustering algorithm

DTCC is a recursive algorithm that starts with a data set of words that need to be clustered, and an n-gram language model. The language model contains all individual words from the data set, their frequencies within the set, and their entropies calculated based on the Brown Data Corpus [34]. The level just signifies the current depth of the tree. The algorithm selects the $n\%$ of most frequent words within the current cluster, making sure that at least 2 words are selected. Then if there are enough words, only half of the words with the highest entropies is selected. From within these words we can find the anchor points as shown by the center selecting algorithm 5-1. Now we can cluster all remaining words based on their proximity to either the left, or the right anchors. If necessary the algorithm can be called recursively until each cluster contains only one type of tokens. Figure 5-2 shows the complete algorithm.

5.5.7 Extension of DTCC for spatial and temporal clustering

The algorithm can be easily augmented to include splits for spatial and temporal clustering. We can either alternate between the tree features or we can combine all three features and employ a composite distance measure for the DTCC algorithm. In our implementation we augmented the distance formula by including the temporal and spatial information, as shown in Equation 4.3.

```
DTCC ( words_cluster: data_set, tokens: language_model, level)
return cluster_tree
  IF  only one type of tokens
   OR perplexity increases
  THEN  RETURN new Cluster based on language_model

  center_pool = n% of language_model

  IF center_pool < 2
  THEN set the center pool to 3 or size of language_model

  center_pool = top half with highest entropy but at least top 2
  ChooseCenterPoints(center_pool)

  left = left center
  right = right center
  entrop_l, entrop_r = entropies for left and right
      calculated based on Brown frequencies
  scaleL = entropL/entropR
  scaleR = 1/scaleL
  FOR ALL data_set
    dLeft = distance of current word from left
    dLeft = dLeft + 0.1*dLeft*multL
    dRight = distance of current word from right
    dRight = dRight + 0.1*dRight*multR

    IF (dLeft <= dRight)
      add current word to leftCluster
    ELSE
      add current word to rightCluster

  clusterer(leftCluster, left tokens, level+1)
  clusterer(rightCluster, right tokens, level+1)
END
```

Figure 5-2: The DTCC clustering algorithm.

Chapter 6

Data

In this chapter we will explain the data collection process: how and where to collect data that is useful for evaluation of the system. The communication aid should be tested on conversations that are representative of conversations that would be led by a user. It is also necessary to collect the dialogs with their corresponding contextual information such as exact time and day, and general positioning coordinates that would be used to establish location.

A small data sample has been collected for this project, used solemnly as proof of concept, not for testing purposes. The conversations were recorded within different rooms in the house and such locations as stores and buses. This data sample is very small; it includes only about 12 hours of recording time over 6 different days. Due to local laws and privacy issues most conversations were conducted with the same individual. Few were conducted with willing individuals from the service industry (bus drivers and store clerks). The data was collected using a laptop that was carried in a backpack. This chapter provides a more detailed of the information on the equipment and the actual sample data that has been collected for the project.

6.1 Collection Process

The intention of the prediction algorithm is only to mimic not to surpass human performance. Thus, the training and the testing data needs to be composed out of real life interactions. For this purpose every day conversations have been collected through the use of sound recording device, and correlated with simultaneously collected global positioning data. The information within both files is also time stamped. This data can be used to establish both spatial and temporal correlations between words and word clusters.

6.1.1 Equipment

The data was collected using a global positioning (GPS) monitoring system, and a microfon. Both devices at all times were attached to a laptop computer where the data was immediately stored in the appropriate formats described in Section 6.1.3. The laptop was carried in a backpack for outdoor locations or placed near a window for indoor locations. This was necessary in order to collect usefull GPS information.

6.1.2 Locations

The locations where data was selected were specifically chosen to be representive for viroments where the aid is most likely to be used.

These locations are:

1. bathroom (about 4 hours of data)
2. kitchen (about 3 hours of data)
3. bedroom (about 2 hours of data)
4. food store (about 1.5 hours of data)
5. general store (about 2 hours of data)
6. bus (about 1.5 hours of data)

6.1.3 Data format

The collected data is stored using very simple formats.

Conversation Files

Conversation – data_x.txt files contain a direct transcription of the data speech. The file is split into 30 second fragmets. This facilitates simple temporal allignment with the positioning data.

A same from the data file would look like this:

- …
- 27:00 I will take one, red chocolate chip ahoy. Different cheeses.
- 27:30 Where is the goat, the goat one should be here. You said there is goat one on sale. They have pepper goat cheese. Is this the one?
- 28:00 Do you want to buy more pasta? We are low on pasta. There is some pasta on sale. There is some pasta on sale. There is lot of varieties. OK, OK, I did not here you.
- 28:30
- 29:00 Only spaghetti and ziti, ziti seems to be on sale. They do not have... Uuu turkey spam.
- …

GPS Files

The absolute collection time and the global positioning information for each data (data_x.txt) file is stored in a cooresponding gps (data_gps_x.txt) Dialog files are transcribed in 30 second increments. The data_gps files have been collected in 10 second increments so there are 3 gps/time points for each phrase.

This is an example of a sample entry in the gps file:

- 2001/09/30 12:13:29 N4223610 W0710546 25+000

1. 2001 is the year, 09 is the month and 30 is the day on which the data has been collected.

2. 12:13:29 is the time (time is given in 24 hour days), so 12 means 12 pm and 24 means 12 am

3. N means northern hemisphere, 4223610 is the geographic coordinate

4. W means western hemisphere, 0710546 is the geographic coordinate

5. 25 is the number of feet above see level and +000 is the error rate

   Currently the height is not considered in the temporal calculations.

Day of week can be extracted based on the actual date.

Sometimes the gps/date point looks like this:

- 2001/09/30 10:15:48 ________ ________ ______

  This signifies that no gps data was available at that time (this is usually due to the system being inside a building). In these cases the best option is to average the last available data point with the next available one. If no new points are avaliable, the estimate should be made based on the last avaliabe sygnal.

6.2 Data Corpus

6.2.1 Brown corpus

World Edition of the Brown Corpus to train and test our word-prediction algorithms. The Brown Corpus (The Standard Corpus of Present-Day Edited American English) [33], [34] is a corpus of 1 million words of written American English printed in the year 1961. It was the first corpus to be put on computer medium and is the most analyzed corpus of English to date. It consists of 500 written American English texts of 2,000 words apiece, selected to represent diverse genres of written American language. There are two main sections: Informative Prose and Imaginative Prose. Genres represented include newspaper reportage, press editorials, memoirs, religion, science fiction, detective fiction, and romance novels (excluding drama and fiction with more than 50% dialog). This corpus of running text is available for academic research. The large size, high percantage of dialogs and wide specturn of topic representation made this corpus very appropriate for our use.

6.2.2 Statistics on the data corpus

The data collected for this project consists of over fiveteen thousand words. The size of the vocabulary is over eleven hundred after stop words prooning. Table 6.1 gives more detail information about each data category.

6.2.3 Filtered data set

Since the contextual data was collected from actual speech and not from usage of the communication aid, we decided to run it through a vocabulary filter. The vocabulary filter removes all words that are not in the core vocabulary for the communication aid. Since the clustering, and prediction algorithms use only semantic, and contextual information removing some words should not have an effect on average performance. It is important to remember that if we were working at the level of comple phrases or doing syntactic word prediction, removing some word could change the resulting prediction rates.

6.2.4 Sample clustering

For initial estimation of the parameters of the clustering algorithm we constructed training and testing samples based on small, hand selected sets of data points. Appendix C shows each step of the Decission Tree Center Convergence (DTCC) algorithm. For comparison the same data has also been clustered using simple hierarchical clustering. It is important to notice that hierarchical clustering seems to be performing significantly worse than DTCC on sets that contain more than two classes of words. DTCC divides the data using binary splits so the minimum number of splits for *n* classes would be the ceiling of n/2.

6.2.5 All data clustering

When running the clustering algorithm on all the data it is important to notice that the algorithm could converge on a local minimum if we require a monotonic decrease in perplexity as a condition for a split.

	total word #	total token #	# after filtering	top content words
bathroom	6582	712	1872	water need want bath think know shower lotion use hot take hair done bathtub put shampoo conditioner bathroom face nice
kitchen	3418	514	1259	bacon want think bread need put guess cheese spoons get ready table muffins lots cup frying nice coffee know half
bedroom	1759	374	728	think wear want go sleep going get light night take see come guess noise good wake please hours pillows downstairs
food store	3073	563	985	get sale want think guess lets good ones know need take go stuff look right try broccoli lettuce price got
general store	2932	542	888	want think ones lets go guess know nice look cute dollar big see stuff dollars store right lots try need
bus	746	246	376	guess bus people hour right half work hear class problems recording know go need data long rare stops bike buses
all	18510	1655	5380	want think need know guess nice eggs water put use take sale go good ones lets right bath lots bacon

Table 6.1: Statistics on the collected data.

Chapter 7

Evaluation of the Use of Contextual Information for Category Generation

In this Chapter we will evaluate the Decision Tree Clustering Algorithm as a tool for generation of contextual categories. We will use perplexity calculations to judge the predictive value of each cluster. It is also important to consider the number and size of clusters, since the categories need to be usable through the communication aid interface. First we will concentrate on generation of contextual categories through purely semantic clustering. Finally, we will show how incorporation of additional information such as time of day and geographical position can improve the predictive value of the clusters.

7.1 Semantic Clustering

Since, we working within the constraints of a communication aid, we can assume simple dialog principles when trying to predict the next word for the user. Dialogs usually tend to follow a topic. What this means, is that if we can establish the topic of the most recent sentence, and we guess that the topic of the next sentence will be the same, we will be right most of the time.

This topic coherence automatically suggests a clustering method. We can establish the current topic by comparing the meaning of the most recent sentences to previous sentences that have been clustered into coherent topics. To for the meaning clusters we need to use a version of semantic clustering. The algorithm that we used was the earlier introduced DTCC clustering algorithm.

We will now look at statistics generated by the DTCC algorithm on the real life dialog data collected for the project.

7.2 Preliminary Evaluation of Perplexity Changes

Initially we run the clustering algorithm on relatively small sets of data. The sets ranged from one topic conversations to three topic conversations. The intention of this preliminary evaluation was to establish how well the algorithm can handle multi-topic conversations in relation to one topic ones. The results (shown in Table 7.2 suggests that the algorithm handles well multi-topic conversations. With a large enough sample we small numbers of clusters with similar perplexity values across different numbers of topics.

The Table shows also the relation between different data sizes. Small data sets have less variance, thus they should have lower perplexity. The algorithm behaves as expected. Clusters generated from small data sets have lower perplexity values. However, for sufficiently large data sets (data sets with more than 3000 elements) perplexity variations due to size are minimal.

	one topic		two topics		three topics		four topics	
Words	*clusters*	*perplex.*	*clusters*	*perplex.*	*clusters*	*perplex.*	*clusters*	*perplex.*
50	2	8	2	0.5	2	0.5	2	1
100	5	1	3	9.82	4	1	5	1
500	8	8.37	5	4.20	6	8	5	4
1000	5	23.61	10	18.4	8	19.75	6	21.83
3000	2	70.29	2	81.75	3	90.66	3	67.33

Table 7.1: Preliminary clustering information. *clusters* denotes the number of resulting clusters and *perplex.* denotes the average perplexity value for the final clustering. The table shows variations over the number of topics present in the data and the size of the data cluster.

7.2.1 Algorithmic parameters

Having established that the algorithm works with different numbers of topics it was necessary to evaluate the general robustness of the algorithm. We decided to run the DTCC clustering on the contextual real life data and estimate the best parameters. We tried varying the perplexity tolerance constant on the terminating condition. Small tolerance requires each new cluster to have a strongly monotonically decreasing perplexity, large tolerance allows for small perplexity variations. Large tolerance might be useful for very large and divers clusters when the few initial best splits might actually increase the perplexity. Also, looser constraints on perplexity allow for finer clustering, which could be useful for large topics.

The second parameter that we evaluated was the size of Center selection pool for the cluster convergence part of the algorithm. Small pools might miss crucial content words. However, large pools are very expensive. Also, if a pool is too large extraneous information is introduced into the selection process, which might cause the center selection to become more random.

Table 7.2.1 shows few sample points in the parameter space. Small variations of parameters have no significant effect on the average cluster perplexity. This suggests that the algorithm is robust to small variations. The optimal parameters across all the data were found to be: 0.90 perplexity tolerance and 35% of data for center selection.

7.2.2 Power word (topic specific) clustering vs. core word clustering

It is important to notice that since we are using semantic distance the algorithm will perform much better on data with high content of topic specific words. This behavior is illustrated in Table 7.2.2. Applying a generous stop word filter, and removing all non-noun parts of speech will result in an easy to cluster data set. Unfortunately, we do not want to limit our users to only power words. As mentioned earlier 80% of human speech uses only core vocabulary. So it is necessary to find a method that will properly categorize all words, including those that are non topic specific. Also, semantic clustering by its nature will necessarily always cluster different instances of the same word together. This behavior is

	small tolerance		large tolerance		small center pool	
Topics	clusters	avg. perp.	clusters	avg. perp.	clusters	avg. perp.
bathroom	2	93.82	2	82.95	4	114.80
bedroom	3	40.22	3	38.26	6	43.99
kitchen	3	73.33	2	78.67	3	68.14
store	2	55.65	4	42.00	4	34.18
food store	3	61.33	3	46.45	3	51.02
bus	3	10.57	3	17.01	3	16.60
all	56	86.46	2	117.32	22	97.33

Table 7.2: Parameter variance shows no strong effects on the resulting clusters. The first column was run with 0.85 (p) perplexity tolerance and 33% of data center selection pool (n), the results in the second column are for p = 1.15 and n = 33% and the last column is for p = 1 and n = 20%

often less than optimal. For example, we would like to have access to the word *water* both within a kitchen environment and within a bathroom environment.

	one topic		two topics		three topics	
Power vs. Core	clusters	avg. perp.	clusters	avg. perp.	clusters	avg. perp.
power only	4	3.68	5	1.09	6	2.03
normal usage	8	8.37	5	9.20	8	12.75

Table 7.3: Power words carry more information so they are easier to cluster.

7.2.3 Core lexicon clustering

Filtering out core or stop words might not be desired. On the contrary, filtering out words not in the lexicon is very beneficial. Words that are not in the lexicon cannot be displayed in the form of individual icons. Using letters to display the word is costly and does not provide significant access improvement since all letters are already contained within one category. Further, words that are not in the lexicon are fringe words that are used very rarely, so even if such a word was used recently, it is very unlikely that it will be used again soon.

7.2.4 All data clustering

Table 7.2.4 shows average statistics for semantic clustering of the complete data set.

Topics	number of clusters	average perplexity
bathroom	2	93.82
bedroom	3	40.22
kitchen	3	73.33
store	2	55.65
food store	3	61.33
bus	3	10.57
all	22	97.33

Table 7.4: This table shows semantic clustering statistics on the complete data set.

7.3 Combining Semantic Clustering with Spatial and Temporal Clustering

Since it is apparent that semantic clustering alone is not sufficient for formation of highly predictive categories we decided to combine the semantic information with contextual information about each word.

7.4 Preliminary Evaluation of Perplexity Changes

Table 7.4 shows the cluster statistics for the DTCC algorithm with the use of time and space information for distance calculation. This clustering was done on the same data that was used to generate the perplexity value in Table 7.2. A comparison of the two tables makes it evident that contextual significantly improves the predictive value of the new categories.

	one topic		two topics		three topics	
Words	clusters	avg. perp.	clusters	avg. perp.	clusters	avg. perp.
50	2	4.00	2	4.25	2	4.99
100	3	9.50	2	4.25	3	7.33
500	3	29.49	4	25.01	8	13.50
1000	13	27.50	12	21.49	9	31.99

Table 7.5: Contextualized version of Table 7.2.

7.4.1 Power word clustering vs. Core lexicon clustering

The use of contextual information further allows us to successfully reintroduce core words into the clustering algorithm. Also, since each word is marked with its time and place stamp, it is now possible for the same word to appear in different clusters. Table 7.6 is the corresponding contexualized version of Table 7.2.2.

	one topic		two topics		three topics	
Power vs. Core	clusters	avg. perp.	clusters	avg. perp.	clusters	avg. perp.
power only	3	4.68	4	1.19	5	2.03
normal usage	4	5.33	2	4.25	4	6.33

Table 7.6: Power words are no longer essential in order to form a small number of highly predictive classes.

7.4.2 All data clustering

Table 7.7 shows statistical information on the use of contextual information to cluster the complete data set.

Topics	number of clusters	average perplexity
bathroom	7	74.53
bedroom	4	40.82
kitchen	5	44.22
store	4	32.33
food store	3	41.59
bus	2	8.57
all	14	84.75

Table 7.7: The use of time and space information in conjunction with semantic value results in smaller number of highly predictive clusters.

Chapter 8

Contributions and Conclusions

8.1 Contributions

8.1.1 Designed and implemented a communication aid

In this thesis we proposed and developed a prototype version of a communication aid for people with severe speech disabilities. We concentrated on users who do not have the ability to access the most common alternative modes of communication, such as typing or writing. The communication aid is a software program that can be run on almost all personal computers, laptops, and some pda's. The designed interface is a dynamic version of a semantically organized picture array. The user constructs a sentence by accessing icons through direct selection via a touch screen. Consecutively chosen images create a sentence by completing a semantic frame.

8.1.2 Introduced semantic frames into a hierarchical lexical system

The notion of semantic frames is based on the idea of thematic roles. In particular, a sentence is strongly constrained by its verb or verb noun pair. For example, if we would like to construct a sentence where the action is *giving*, a minimalist version of the sentence will necessarily contain a giver, a receiver, and an object being transferred.

The semantic frames can be used for word prediction by constraining the lexicon on the basis of the current slot label and the realizations of the previous slots. More importantly, explicit use of semantic frames eases the process of sentence creation for the user. The individual does not need to know or follow grammatical rules; rather, he or she can concentrate on selecting the words that exactly represent the message they are trying to communicate.

8.1.3 Improved lexical prediction through the use of contextual information

Even though semantic frames limit somewhat the appropriate lexicon, the number of valid slot realizations can still be extremely large. We decided to further constrain the suggested lexicon by using easily available contextual information, such as time of day, recent conversational history, and absolute geographical positioning.

We collected a representative sample of contextually tagged conversations. These were recorded based on unscripted interaction of one individual with other people. The environments chosen for data collection were diverse and included places like a kitchen, a bathroom, a store, or a bedroom.

8.1.4 Developed a clustering algorithm for contextually annotated words

In order to fully benefit from the additional context information our interface allowed, we devised a novel clustering algorithm: Decision Tree Center Convergence (DTCC) clustering. The algorithm is loosely based on Decision Tree classification paradigm (ID3); however, it is specifically adapted for clustering. The DTCC algorithm further addresses the issue of sparse, high dimensionality of words by projecting them around cluster centers onto a one-and-a-half dimensional bounded plane.

Initial statistics on the available data suggest that the use of contextual information offers a significant improvement over purely semantic clustering methods. The number of contextual clusters is smaller; thus, resulting in more coherent complete topical categories.

8.1.5 Proposed an adaptive system based on use of machine learning techniques

Finally, we proposed a way to modify the communication aid in order to make it adaptive. As the user communicates with the help of the device, the aid could learn individual-specific contextual categories, and provide the user with accurate personalized predictions.

Chapter 9

Future Extensions

This Thesis describes first steps towards incorporating contextual information into lexical prediction. There are many issues that still need to be resolved before the aid can successfully be used outside the academic environment.

9.1 Combining Clustering with the Interface

Further research should include full incorporation of the clustering algorithm (as a primary word prediction algorithm) into the communication aid software. For this to be accomplished, the aid needs to keep a complete contextually annotated usage database, i.e. each time the user invokes a message, that message would be added into the database. The database, in turn, would be used to generate personalized contextual clusters that would be meaningful to and for the specific individual.

In addition, the database could be used for cluster selection. This can be achieved by using it as a communication memory with some sort of pre-programmed decay factor. Utterances (and their contexts) that have been produced most recently would have the strongest influence on the contextual cluster selection. If no utterances were produced within the time decay limit, a purely spatio-temporal template would be used. In the few cases where this method is not viable yet, the contextual category will be loaded with context-ignorant high-frequency words appropriate to the individual's personal usage history.

9.2 HCI Issues with Dynamic Interfaces

As mentioned earlier, in Chapter 3, dynamic display systems tend to be disorienting and difficult to use. The time requirement is imposed by the necessity to search and navigate through the system in order to locate relevant vocabulary [77]. Our proposed solution to this problem was providing access to a cross-categorical set of most likely words. However, accurate prediction is often very difficult and sometimes impossible, especially with a novice user and a device that has not yet learned individual specific situations.

Features inherent in dynamic displays carry cognitive, motor, visual-perceptual and learning loads that challenge some users. Moreover, the problem of repeated visual refocusing, interjected into the language act, further increases the load. For example, moving within and among screens to construct messages requires high level of visual attention and an elevated degree of decision-making. This means that, although memory demands are reduced, full automation in the message generation process may be difficult to achieve.

The use of a computer as a communication aid offers memory and general computational power. However, off-the-shelf computers are not designed to be mounted on wheelchairs and/or used as communication aids. Therefore, a more specialized version of a minimal computer might offer a better choice.

Furthermore, even with the assumption that adequate word prediction can be achieved, it is still necessary to devise a model of how this information should be incorporated into the existing interface. In the current version of the project we decided to generate a special contextual category that is dynamically updated with the environment. The user could always attempt to use a suggested word, but he or she would still have access to all existing, familiar categories.

Another way to introduce the prediction information would be to change the current image presentation (or layout) regardless of the displayed category. One potential way to modify the existing presentation would be to include a small set of generic ”most likely” words with the current category. It is also possible to keep the information and update only the layout of images: icons within a category would be presented in their predictive value order.

9.3 Better Icons

There are many products on the AAC market that use pictures to convey meaning, and we are beginning to see the emergence of additional capabilities in a subset of these products. Product developers currently have very little knowledge about the manner in which individuals with disabilities operate with two-dimensional representations of concepts. However, researchers rely on what they know about language acquisition and language behavior to provide the AAC field with a scaffold for exploration of the image usage design issues.

Recently, one very active area of research has concentrated on the use of animation. Current iconic systems are very poor at representing actions in two-dimensional forms. Researchers are trying to determine the relative efficiency of a number of approaches for representing movement, including static pictures, video, and animated pictures. The results will provide guidance to those selecting and customizing AAC systems, as well as to manufacturers who are trying to make their products maximally responsive to the needs of people who rely on picture-based AAC systems.

Future work might aim to investigate as well the ease with which individuals understand and use multi-modal, picture-based representations of actions. Some approaches to representing action that could be considered include: static photographs with dis-equilibrium cues, static line drawings with dis-equilibrium cues, static line drawings with visual marking cues, video segments depicting action, and animated line drawings. These should be compared to measures of the subjects' comprehension of the actions when presented via a live model. It is important to examine the effect of the representation type on both learnability and language performance.

Research is also currently carried out on the design of a new static picture language. Sheel Dande at the MIT Media Laboratory is working on a drawing tool that will facilitate global design of communication symbols. He has implemented a simple, yet powerful line drawing tool that allows the user to design and quickly create personalized icons. The symbols are intended to be easily combinable into meaningful messages. Sheel Dande is hoping that through an arbitration of icon synching, users will be able to converge on a pictorial, culture independent core communication set. Hypothetically, such a set would most likely permit visual and semantic combination of the system. Even more importantly, such a system would be globally transparent.

9.4 Grammatical Correction

Even though the semantic frames facilitate grammatical sentence construction, they are very limited and do not handle any complex grammar rules. In a future version of the system, developers should aim to provide the user with automatic grammar correction through compansion.

The Compansion technique is a syntactically and semantically based AAC method that generates complete well-formed sentences from a user input of uninflected content words. It has applications both as a rate enhancement technique for people who use word or picture based systems (it reduces the number of keystrokes and the number of word forms needed), and as a tool for people learning a language who do not have good grammatical skills yet [24].

Compansion can handle declarative sentences that include multiple adjectives, prepositional phrases, possessive noun phrases, direct and indirect objects, and some verbal clauses (e.g., "I want to go to the store"). It can also generate grammatical questions, imperatives, complex verb tenses, and do structures. In addition, Compansion can be augmented to understand some metaphorical verb uses [25].

9.5 Spatial Syntax for Semantic Frames

We believe that the semantic frame system will ease and possibly accelerate message construction. The current template method requires no grammatical competence and conveys only directly selected information. Nevertheless, the frame system still imposes linear message construction.

In the future, bearing this flaw in mind, one could design and incorporate into the aid non-linear semantic frames. Slots within a frame would be positioned relative to their thematic roles and not relative to parts of speech. Sheel 9.3 has done some preliminary work on the specifications for a simple generative module. Such a module can be used for creating natural language representations of the visual sentence. He is exploring the use of his drawing program for generation of meaning-carrying pictographic sentences through position sensitive placement of icons. In this system the meaning and grammatical realization of an icon would be influenced by its position relative to other icons. Sheel is aiming to develop a set of general positioning rules that would be transparent to the user, and in the same time result in a simple to draw, yet visually pleasing construction.

9.6 Usability Testing

Finally, it is necessary to conduct both small- and large-scale usability tests in order to evaluate the merits offered by contextual prediction. The data that we collected offers an indication of how well the system might perform, but the estimate is only crude.

The first step would involve close work with a small set of users. "User feedback" would allow us to make necessary adjustments to the existing features and possibly to add new features before attempting large scale testing.

After the interface "debugging" stage a full-scale usability test would be conducted. A group of users will be asked to substitute our product in place of the existing AAC technology that they have been using so far. This would provide an opportunity to measure learning curves and establish a more precise estimate on prediction improvement. For a between subject design, in an administrator blinded study, four groups of users would be provided with four versions of the system with:

1. no word prediction and no semantic frames
2. semantic frames but no word prediction
3. word prediction but no semantic frames
4. both word prediction and semantic frames

For a within subject design each user would be asked to disable both or one of the features. These studies would not only reveal if the aid offers improvement over existing devices, but could also allow us to estimate the absolute communication rate improvements of both features.

Appendix A

Background on Clustering Methods

Cluster analysis is a technique for grouping data, and for finding structures in the data. The most common application of clustering methods is to partition a data set into clusters or classes, where similar data are assigned to the same cluster whereas dissimilar data should belong to different clusters. In this project we will use clustering to establish a set of contextually similar words; meaning words that would likely be used within the same context.

In this appendix we will provide an introduction to clustering and mention background research and existing methods.

Figure A-1 presents the most common clustering methods within relation to other methods. Each branch specifies a more precise classification method. This presentation was done in [22] and is based on the work of Williams [82]. A summary of each method follows the figure. Clustering is just uninformed case of classification.

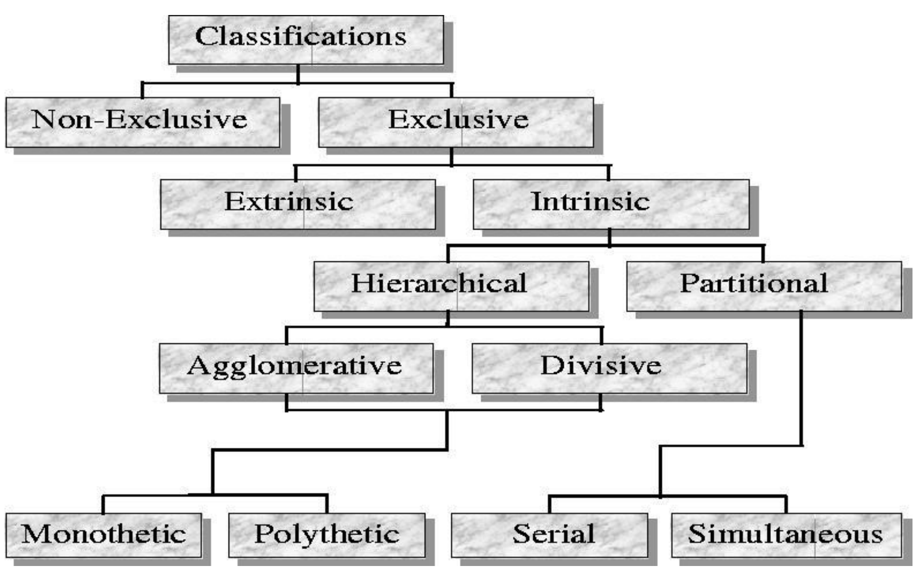

Figure A-1: Hierarchy of clustering

A.1 Exclusive

Within an exclusive clustering process, the data point can belong only to one specific cluster. If the data point is between two clusters, the algorithm has to determine to which group it should belong and cannot use probabilities or grade of membership for each cluster like in fuzzy or clump clustering (see [8] and [28]).

A.2 Non-Exclusive

A data point can be assigned to several clusters, allowing an overlapping of clusters. This can be done as described above using probabilities, grades, or a list of clusters the data point belongs to. Examples of work in this field include [74] and [42].

A.3 Intrinsic

The intrinsic clustering process does not know any cluster in the data set a priori, but uses only information that is in the data set itself. The resulting clusters are based on the relations between the attributes. Dubes and Jain ([27]) claim that intrinsic classification is the essence of cluster analysis and this might be the reason it is the method that is mostly discussed in literature.

A.4 Extrinsic

This clustering method uses the data set and its category labels on the data points to perform a preclustering. An example in [27] uses health information of smokers and non-smokers to find a way of discriminating them. Compared to this, an intrinsic clustering would cluster the data and afterwards analyze if smoking was a cause for a disease.

A.5 Hierarchical

Hierarchical clustering transforms a set of data points with a given measurement for similarity into a sequence of nested partitions. This can be divided into two different directions. We can either start with each data point as a single cluster and merge in each step two of them together (agglomerative, bottom-up) or start with all data point in one cluster and divide one cluster in two clusters in each step (divisive, top-down). One drawback of this method is that each step is definitive and can not be discarded or changed later, e.g., a merge can not be redone if it will be obvious at a later point that other clusters would have merged in a better way [44].

Examples of both methods include:

1. Top-down (splitting)

   e.g., hierarchical mixture models

2. Bottom-up (merging)

   e.g., hierarchical agglomerative clustering

Given a set of N items to be clustered, and an NxN distance (or similarity) matrix, the basic process of hierarchical [45] clustering consists of the following steps:

1. Start by assigning each item to its own cluster, so that if you have N items, you now have N clusters, each containing just one item. The distances or similarities between the objects are known and defined. Most hierarchical methods store the distances in a matrix D of dimension .

2. Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one less cluster. You can do that by finding the maximum in the matrix D

3. After the two clusters i and j have been merged into one (ij) cluster the similarity measure between the new cluster and the old clusters needs to be recomputed.

4. Repeat steps 2 and 3 until all items are clustered into a single cluster of size N.

Before we explain the different versions of this method, there are some general properties that should be mentioned. Hierarchical clustering is not robust and does not protect against outliers and therefore results can be influenced by extraneous data or noise in the data set. A misclassification in an early stage can happen and the final result needs tp be checked for obvious cases of misgrouping. Finally, the algorithm is nondeterministic. There is the possibility of having the same minimum distance between different clusters and we have to choose the next step arbitrarily. With different choices the algorithm will results in different dentograms, a graphical presentation of the results in form of a tree, and the user should therefore use more than one way of deciding which of the clusters to merge first.

The different versions of the algorithm differ in Step 3. For example, in single-link clustering (also called the connectedness or minimum method), we consider the distance between one cluster, and another cluster to be equal to the shortest distance from any member of one cluster to any member of the other cluster. In complete-link clustering (also called the diameter or maximum method), we consider the distance between one cluster, and another cluster to be equal to the longest distance from any member of one cluster to any member of the other cluster. In average-link clustering, we consider the distance between one cluster, and another cluster to be equal to the average distance from any member of one cluster to any member of the other cluster [17]. Now we will describe each variation in more detail.

- Single Linkage

  The single link method is probably the best known of the hierarchical methods, and operates by joining, at each step, the two most similar objects, which are not yet in the same cluster. The name single link thus refers to the joining of pairs of clusters by the single shortest link between them.

  single link method

- Similarity:

  Join the most similar pair of objects that are not yet in the same cluster. Distance between 2 clusters is the distance between the closest pair of points, each of which is in one of the two clusters.

- Type of clusters:

  Long straggly clusters, chains, ellipsoidal

- Time:

  Usually $O(N^2)$ though it can range from $O(N \log N)$ to $O(N^5)$

- Space:

  $O(N)$

- Advantages:

  Theoretical properties, efficient implementations, widely used. No cluster centroid or representative required, so no need arises to recalculate the similarity matrix.

- Disadvantages:

  Unsuitable for isolating spherical or poorly separated clusters

* Complete Linkage

  The complete link method is similar to the single link method except that it uses the least similar pair between two clusters to determine the inter- cluster similarity (so that every cluster member is more like the furthest member of its own cluster than the furthest item in any other cluster). This method is characterized by small, tightly bound clusters.

  complete link method

  - Similarity:

    Join least similar pair between each of two clusters

  - Type of clusters:

    All entries in a cluster are linked to one another within some minimum similarity, so have small, tightly bound clusters.

  - Time:

    Voorhees alg. worst case is $O(N^3)$ but sparse matrices require much less

  - Space:

    Voorhees alg. worst case is $O(N^2)$ but sparse matrices require much less

  - Advantages:

    Good results from (Voorhees) comparative studies.

  - Disadvantages:

    Difficult to apply to large data sets since most efficient algorithm is general HACM using stored data or stored matrix approach.

- Average Linkage

  The group average method relies on the average value of the pair wise within a cluster, rather than the maximum or minimum similarity as with the single link or the complete link methods. Since all objects in a cluster contribute to the inter-cluster similarity, each object is on average more like every other member of its own cluster than the objects in any other cluster.

  group average link method

  - Similarity:
    Use the average value of the pairwise links within a cluster, based upon all objects in the cluster. Save space and time if use inner product of 2 (appropriately weighted) vectors — see Voorhees alg. below.
  - Type of clusters:
    Intermediate in tightness between single link and complete link
  - Time:
    $O(N^2)$
  - Space:
    $O(N)$
  - Advantages:
    Ranked well in evaluation studies
  - Disadvantages:
    Expensive for large collections

- Wards

  Ward's method is a minimum variance method.

  - Similarity:
    Join the cluster pair whose merger minimizes the increase in the total within-group error sum of squares, based on Euclidean distance between centroids
  - Type of clusters:
    Homogeneous clusters in a symmetric hierarchy; cluster centroid nicely characterized by its center of gravity
  - Time:
    $O(N^2)$
  - Space:
    $O(N)$ — carries out agglomerations in restricted spatial regions
  - Advantages:
    Good at recovering cluster structure, yields unique and exact hierarchy
  - Disadvantages:
    Sensitive to outliers, poor at recovering elongated clusters

- RNN

  We can apply a reciprocal nearest neighbor (RNN) algorithm, since for any point or cluster there exists a chain of nearest neighbors (NNs) that ends with a pair that are each others' NN.

A.6 Non-Hierarchical = Partitional

The result is one single clustering of the data set in a given number of clusters. Another name is partitional clustering. With partitional clustering, the final solution with g clusters is constructed in one step and does not produce a hierarchy. This is either done by starting from scratch and constructing a new solution or by using a valid solution as a starting point for improvements. Therefore, there must be a criterion to determine the quality of the solution. This can be done e.g., by minimizing the sum of squared distances within the clusters or maximizing the between cluster distance (see [36]). Most algorithms use the Euclidean distance of data points.

The partitioning methods generally result in a set of M clusters, each object belonging to one cluster. Each cluster may be represented by a centroid or a cluster representative; this is some sort of summary description of all the objects contained in a cluster. The precise form of this description will depend on the type of the object which is being clustered. In case where real-valued data is available, the arithmetic mean of the attribute vectors for all objects within a cluster provides an appropriate representative; alternative types of centroid may be required in other cases. If the number of the clusters is large, the centroids can be further clustered to produces hierarchy within a dataset.

There are two common partitioning methods:

- Single-pass methods

  Single Pass is a very simple partition method, the single pass method creates a partitioned dataset as follows:

  1. Make the first object the centroid for the first cluster.
  2. For the next object, calculate the similarity, S, with each existing cluster centroid, using some similarity coefficient.
  3. If the highest calculated S is greater than some specified threshold value, add the object to the corresponding cluster and re-determine the centroid; otherwise, use the object to initiate a new cluster. If any objects remain to be clustered, return to step 2.

  As its name implies, this method requires only one pass through the dataset.

  - Similarity:

    Computed between input and all representatives of existing clusters
  - Example:

    Cover coefficient algorithm of Can et al.: Select set of documents as cluster seeds; assign each document to cluster that maximally covers it

- Time:
  $O(N \log N)$
- Space:
  $O(M)$
- Advantages:
  Simple, requiring only one pass through data; may be useful as starting point for reallocation methods
- Disadvantages:
  Produces large clusters early in process; clusters formed are dependent on order of input data

- Reallocation methods

  This is the algorithm for the Reallocation method:

  1. Select M cluster representatives or centroids
  2. Assign each document to the most similar centroid
  3. Recalculate the centroid for each cluster
  4. Repeat steps 2 and 3 until there is little change in cluster membership from pass to pass

  - Similarity:
    Allows various definitions of similarity / cluster cohesion
  - Type of clusters:
    Refinements of initial input clustering
  - Time:
    $O(MN)$
  - Space:
    $O(M + N)$
  - Advantages:
    Allows processing of larger data sets than other methods
  - Disadvantages:
    Can take a long time to converge to a solution, depending on the appropriateness of the reallocation criteria to the structure of the data

A.7 Divisive/Agglomerative

Describes the direction of the hierarchical clustering. The clustering can start with each point being a cluster and joining two clusters to a new one in each step (agglomerative) or seeing the data set as one cluster in the beginning and dividing a cluster in two new ones in each step (divisive).

A.8 Monothetic/Polythetic

If the clustering is done by using only one attribute of the objects at a time, it is called a monothetic clustering. A polythetic clustering would use all attributes at the same time. Within a hierarchical clustering, a monothetic clustering can be used with a changing attribute after each step.

A.9 Serial/Simultaneous

A serial clustering would assign one object after each other, whereas a simultaneous clustering is assigning all objects at the same time.

A.10 Probabilistic clustering

Methods belonging to this type of clustering methods assume a statistical model instead of a predefined metric as before. A statistical model can use a mixture likelihood. Here, we will introduce probabilistic clustering using the Expectation-Maximization (EM) algorithm ([29]).

The EM-Algorithm was first introduced by Dempster, Laird, and Rubin ([26]) and is used to maximize the mixture likelihood. The algorithm can be split in two parts, the estimation of the probabilities for a data point to be in a cluster and the maximization of the mixture likelihood. The algorithm starts with any initial values and continues with its iterations as long as the parameter estimates differ a certain amount from one iteration to the next. The result is the means and covariance matrices for the clusters.

Appendix B

Core lexicon

I	you	he	she	they	hot
eggs	sale	their	may	like	know
want	use	bath	bacon	lots	states
years	stuff	shower	go	day	lotion
american	man	get	bathtub	dont	see
cheese	good	men	united	water	shampoo
differ	conditioner	bathroom	take	government	didnt
came	spell	said	way	people	business
think	long	little	number	president	social
right	national	soap	given	towel	cute
each	father	spoons	look	could	church
development	family	members	put	time	bubble
shelf	brush	god	asked	muffins	service
cup	year	sorry	went	york	human
ease	world	form	powder	action	sentence
local	important	eyes	him	juice	mountain
definitely	watch	college	days	information	bubbles
plain	frying	political	going	bird	high
anyway	family	pose	school	federal	scrub
rinse	known	numeral	available	towels	economic
smells	dirty	individual	knifes	areas	life
society	tea	group	community	act	butter
court	want	mushrooms	future	laundry	department
need	pan	policy	big	times	plates
verb	sing	students	ill	new	broccoli

vowel	means	bus	nations	university	education
thanks	military	recording	govern	washing	things
rate	quarter	fly	washington	cry	dish
night	washer	evidence	find	lettuce	noun
tax	war	point	various	order	work
orange	peace	contain	schools	teach	set
cans	lines	children	skip	months	ocean
situation	personal	increase	back	making	tail
america	face	living	city	longer	multiply
shave	early	boy	sponge	private	comb
away	greater	decide	needed	cloths	thing
expected	socks	secretary	that	ribs	values
following	basis	pressure	hungry	stead	dishwasher
union	interest	laugh	spirit	required	moved
container	support	ones	equate	return	conditions
tire	attention	distant	costs	began	stage
paint	forces	language	couldnt	grand	wave
prefer	pasta	power	young	beans	report
hours	land	forgot	sail	cookies	cents
vary	settle	added	peak	followed	question
bigger	research	stuck	including	study	pulp
music	simply	divide	amount	syllable	garlic
certain	sudden	left	cant	fried	developed
committee	reached	represent	north	defense	sour
equipment	shown	pizza	hunt	religious	probable
sunny	central	got	girl	friends	sort
received	fraction	terms	medical	administration	money
computer	lone	place	peeling	mount	thermal
done	meeting	open	walked	joy	dressed
instrument	half	futon	west	training	foreign
basket	cow	congress	mean	nuts	police
present	county	field	book	international	yeast
growth	having	week	girls	england	wasnt
live	suddenly	clothe	omelette	love	kind
issue	hall	meet	countries	root	considered
getting	idea	youre	bag	working	solve
push	purpose	paragraph	cases	late	labor
entire	told	brand	describe	weeks	flavored

results	production	burn	goat	near	small
william	better	involved	voice	shaws	earlier
increased	cheep	knowledge	copy	effort	phrase
miles	christian	sand	paid	morning	club
ideas	bill	medium	street	certainly	excite
addition	points	ear	industrial	low	moral
directly	boys	decided	home	reading	say
questions	observe	tub	statement	due	consonant
rinsing	programs	moisturizing	methods	probably	services
toothpaste	fire	ago	organ	brought	thought
force	member	section	southern	herbal	cloud
physical	essence	understand	western	bond	population
complete	strength	climb	finishing	clear	stood
river	concerned	feels	volume	hot	plan
district	come	earth	appeared	tangled	merely
iron	direction	trial	crease	lobster	sales
barbecue	continued	melody	common	literature	today
row	white	hampshire	association	exact	met
position	close	army	systems	shout	generally
changes	seed	provided	furniture	join	toilet

Appendix C

Clustering examples

C.1 DTCC clustering

```
Level 0: [egg, bacon, egg, milk, water, bread, butter,
bread, drink, candy, egg, water, water, shower, bath, bath,
soap, shampoo, conditioner, shower, bath, tub, sleep, bed,
pillow, sleep, dream, bed, night, blanket, sleep, blanket,
bed, night, dream, ]
----------------------------------------------------
Level 1: [egg, bacon, egg, milk, water, bread, butter,
bread, drink, candy, egg, water, water, ]
Level 1: [shower, bath, bath, soap, shampoo, conditioner,
shower, bath, tub, sleep, bed, pillow, sleep, dream, bed,
night, blanket, sleep, blanket, bed, night, dream, ]
----------------------------------------------------
Level 2: [egg, egg, milk, egg, ]
Level 2: [bacon, water, bread, butter, bread, drink,
candy, water, water, ]
Level 2: [shower, bath, bath, soap, shampoo, conditioner,
shower, bath, tub, bed, pillow, bed, blanket, blanket, bed,
]
Level 2: [sleep, sleep, dream, night, sleep, night,
dream, ]
----------------------------------------------------
Level 3: [egg, egg, egg, ]
Level 3: [milk, ]
Level 3: [bacon, ]
Level 3: [water, bread, butter, bread, drink, candy,
water, water, ]
```

```
Level 3: [shower, shower, pillow, blanket, blanket, ]
Level 3: [bath, bath, soap, shampoo, conditioner, bath,
tub, bed, bed, bed, ]
Level 3: [sleep, sleep, night, sleep, night, ]
Level 3: [dream, dream, ]
----------------------------------------------------
Level 4: [bread, bread, ]
Level 4: [water, butter, drink, candy, water, water, ]
Level 4: [shower, shower, ]
Level 4: [pillow, blanket, blanket, ]
Level 4: [bath, bath, conditioner, bath, tub, ]
Level 4: [soap, shampoo, bed, bed, bed, ]
Level 4: [sleep, sleep, sleep, ]
Level 4: [night, night, ]
----------------------------------------------------
Level 5: [butter, ]
Level 5: [water, drink, candy, water, water, ]
Level 5: [bath, bath, bath, tub, ]
Level 5: [conditioner, ]
Level 5: [shampoo, ]
Level 5: [soap, bed, bed, bed, ]
----------------------------------------------------
Level 6: [candy, ]
Level 6: [water, drink, water, water, ]
Level 6: [soap, ]
Level 6: [bed, bed, bed, ]
----------------------------------------------------
Level 7: [drink, ]
Level 7: [water, water, water, ]
```

C.2 Hierarchical clustering

```
Level: 0        bacon,
Level: 0        milk,
Level: 0        butter,
Level: 0        candy,
Level: 0        soap,
Level: 0        shampoo,
Level: 0        conditioner,
Level: 0        pillow,
Level: 0        bread, bread,
Level: 0        egg, egg, egg,
```

```
Level: 0        shower, shower,
Level: 0        dream, dream,
Level: 0        night, night,
Level: 0        blanket, blanket,
Level: 0        sleep, sleep, sleep,
Level: 0        bed, bed, bed,
Level: 0        water, water, drink, water,
Level: 0        tub, bath, bath, bath,
---------------------------------------
Level: 1        bacon,
Level: 1        candy,
Level: 1        conditioner,
Level: 1        pillow,
Level: 1        bread, bread,
Level: 1        shower, shower,
Level: 1        dream, dream,
Level: 1        night, night,
Level: 1        blanket, blanket,
Level: 1        sleep, sleep, sleep,
Level: 1        bed, bed, bed,
Level: 1        water, water, drink, water,
Level: 1        tub, bath, bath, bath,
Level: 1        shampoo, soap,
Level: 1        butter, milk, egg, egg, egg,
---------------------------------------
Level: 2        bacon,
Level: 2        candy,
Level: 2        pillow,
Level: 2        dream, dream,
Level: 2        shampoo, soap,
Level: 2        water, water, drink, water, conditioner,
Level: 2        butter, milk, egg, egg, egg, bread, bread,
Level: 2        tub, bath, bath, bath, shower, shower,
Level: 2        sleep, sleep, sleep, night, night,
Level: 2        bed, bed, bed, blanket, blanket,
---------------------------------------
Level: 4        candy,
Level: 4        pillow,
Level: 4        dream, dream,
Level: 4        sleep, sleep, sleep, night, night,
Level: 4        butter, milk, egg, egg, egg, bread, bread,
bacon,
Level: 4        tub, bath, bath, bath, shower, shower,
shampoo, soap,
Level: 4        bed, bed, bed, blanket, blanket, water,
```

```
water, drink, water, conditioner,
---------------------------------------
Level: 5        pillow,
Level: 5        dream, dream,
Level: 5        sleep, sleep, sleep, night, night,
Level: 5        butter, milk, egg, egg, egg, bread, bread,
bacon, candy,
Level: 5        bed, bed, bed, blanket, blanket, water,
water, drink, water, conditioner, tub, bath, bath, bath,
shower, shower, shampoo, soap,
---------------------------------------
Level: 6        dream, dream,
Level: 6        sleep, sleep, sleep, night, night,
Level: 6        butter, milk, egg, egg, egg, bread, bread,
bacon, candy,
Level: 6        bed, bed, bed, blanket, blanket, water,
water, drink, water, conditioner, tub, bath, bath, bath,
shower, shower, shampoo, soap, pillow,
---------------------------------------
Level: 8        dream, dream,
Level: 8        sleep, sleep, sleep, night, night,
Level: 8        bed, bed, bed, blanket, blanket, water,
water, drink, water, conditioner, tub, bath, bath, bath,
shower, shower, shampoo, soap, pillow, butter, milk, egg,
egg, egg, bread, bread, bacon, candy,
---------------------------------------
Level: 13       sleep, sleep, sleep, night, night, dream,
dream, bed, bed, bed, blanket, blanket, water, water, drink,
water, conditioner, tub, bath, bath, bath, shower, shower,
shampoo, soap, pillow, butter, milk, egg, egg, egg, bread,
bread, bacon, candy,
---------------------------------------
```

Bibliography

[1] T. Agate A. Dougherty. Iconic encoding with picture communication symbols. In *CSUN Conference*, 2001.

[2] J. Allen, S. Hunnicutt, and D. Klatt. *From Text to Speech: The MITalk System*. Cambridge University Press, Inc, 1987.

[3] N. Alm. Approaches to predicting utterances using conversational modelling. Workshop on NLP and AAC: Abstracts of presentations.

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

[5] B. Baker. Using images to generate speech. *Byte*, 11:160–168, 1986.

[6] B. R. Baker, K. Hill, and R. Devylder. Core vocabulary is the same across environments. In *Proceedings of the Technology and Persons with Disabilities Conference*, 2000.

[7] W. Beattie. ord prediction by word association. Workshop on NLP and AAC: Abstracts of presentations.

[8] R.E. Bellmann, R. Kalaba, and L.A. Zadeh. Abstraction and paterrn classification. *Journal of Mathematical Analysis and Applications*, 13:1–7, 1966.

[9] D. Beukelman and P. Mirenda. *Augmentative and Alternative Communication*. Paul H. Brookes Publishing Co., 1992.

[10] D. Beukelman and P. Mirenda. *Augmentative and alternative communication: Management of severe communication disorders in children and adults (2nd ed.)*. Baltimore, Maryland, USA: Paul H. Brookes Publishing Company, 1998.

[11] D. R. Beukelman and P. Mirenda. *Augmentative and Alternative Communication Management of Severe Communication Disorders in Children and Adults*. Paul H. Brookes Publishing Co., 1992.

[12] S. Blackstone. The medium is the message. *Augmentative Communication News*, 3:2:1–3, 1990.

[13] S. Blackstone. Thinking a little harder about communication displays. *Augmentative Communication News*, 6:1–4, 1993.

[14] S. Blackstone. Dynamic displays. 7:2, 1994.

[15] K. Bloomberg and H. Johnson. A statewide demographic survey of people with severe communication impairments. *AAC Augmentative and Alternative Communication*, 6:50–60, 1990.

[16] K. Bloomberg, G. R. Karlan, and L. L. Lloyd. The comparative translucency of initial lexical items represented in five graphic symbol systems and sets. *Journal of Speech and Hearing Research*, 33:717–725, 1990.

[17] S. P. Borgatti. How to explain hierarchical clustering. *Connections*, 17(2):78–80, 1994.

[18] G. E. Briggs and J. C. Naylor. Relative efficiency of several training methods as a function of transfer task complexity. *Journal of Experimental Psychology*, 64:505–512, 1962.

[19] P. F. Brown, S.A. DellaPietra, V.J. DellaPietra, J.C. Lai, and R.L. Mercer. An estimate of an upper bound for the entropy of english. *Computational Linguistics*, 18(1):31–40, 1992.

[20] A. Budanitsky and G. Hirst. Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures. In *In Proceedings of the NAACL-2001 Workshop on WordNet and Other Lexical Resources*. sponsored by the Association for Computational Linguistics, 2001.

[21] P. Cheeseman and J. Stutz. *Baysian classification (autoclass): Theory and results*, pages 153–180. AAAI/MIT Press, 1996. In U.M. Fayyad, G. Piatetsky- Shapiro, P. Smith and R. Uthurusamy, editors.

[22] H. T. Clifford and W. Stephenson. *An Indroduction to Numerical Classification*. Academic Press, 1975.

[23] I. Dagan, L. Lee, and F Pereira. Similarity-based models of cooccurrence probabilities. *Machine Learning*, 34(1-3):43–69, 1999.

[24] P. Demasco, K. F. McCoy, M. Jones, C. Pennington, and S. Snyder. Companion: A technique that applies natural language processing to augmentative communication. *Rehabilitation R and D Progress Reports / Journal of Rehabilitation Research and Development*, 29:180–181, 1991.

[25] P. Demasco, K. F. McCoy, M. Jones, C. Pennington, and G. Stum. Rate enhancement through sentence compansion. *Rehabilitation R and D Progress Reports / Journal of Rehabilitation Research and Development*, 28:197, 1990.

[26] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(B):1–38, 1977.

[27] R. Dubes and A.K. Jain. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, New Jersey, 1988.

[28] J. Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4:95–104, 1974.

[29] B.S. Everitt. *Cluster Analysis*. Halsted Press, third edition, 1993.

[30] C. Fellbaum and editor. *WordNet: An Electronic Lexical Database*. The MIT Press, 1998.

[31] C. Fillmore. *A case for case*, pages 1–88. Rinehart and Winston, New York, 1968.

[32] W. B. Frakes and R. Baeza-Yates. *Information Retrieval Data Structures and Algorithms*. Prentice Hall, 1992.

[33] W. N. Francis. *A tagged corpus: Problems and prospects*. London: Longman, 1980.

[34] W. N. Francis and H. Kucera. *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin, 1982.

[35] C. Goodenough-Trepagnier. Design goals for augmentative communication. *Assistive Technology*, 6.1:3–9, 1994.

[36] A.D. Gordon and J.T. Henderson. An algorithm for euclidean sum of squares classification. *Biometrics*, 33:355–362, 1977.

[37] J. Hersh. *The Tyranny of the Keyboard*. January 1998.

[38] K. J. Hill and B. A. Romich. A rate index for augmentative and alternative communication. *International Journal of Speech Technology*, 5(1):57–64, 2002.

[39] M. B. Huer. Examining perceptions of graphic symbols across cultures: Preliminary study of the impact of culture/ethnicity. *AAC Augmentative and Alternative Communication*, 16:180–185, 2000.

[40] S. Hunnicutt, A. Carlberger, and J. Carlberger. A probabilistic model using syntactic and semantic features for a word prediction system. Workshop on NLP and AAC: Abstracts of presentations.

[41] A.K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.

[42] N. Jardine. Towards a general theory of clustering. *Biometrics*, 25:609–610, 1969.

[43] R. Mayer Johnson. Picture communication symbols, 1981.

[44] R.W. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, New Jersey, 1998.

[45] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 2:241–254, 1967.

[46] B. H. Kantowitz and R. D. Sorkin. *Human Factors: Understanding People-System Relationships*. New York: John Wiley and Sons., 1983.

[47] H. Kucera and W. N. Francis. *Computational analysis of present-day American English*. Providence: Brown University Press, 1967.

[48] L. M. Lafontaine and F. DeRuyter. The nonspeaking cerebral palsied: A clinical and demographic database report. *AAC Augmentative and Alternative Communication*, 3:153–162, 1987.

[49] N. E. Lane. *Skill Acquisition Rates and Patterns: Issues and Training Implications*. Springer-Verlag, 1987.

[50] S. Langer and M. Hickey. Automatic message indexing and full text retrieval for a communication aid. Workshop on NLP and AAC: Abstracts of presentations.

[51] C. Leacock and M. Chodorow. *Combining local context and WordNet similarity for word sense identi-fication*, pages 265–283.

[52] B. Levin. *English Verb Classes and Verb Alternations: A Preliminary Investigation*. University of Chicago Press, 1993.

[53] H. Li and N. Abe. Word clustering and disambiguation based on co-occurrence data. In *In Proceedings of the 17th Internation Conference on Computational Linguistics*, pages 749–755, 1998.

[54] J. Light and P. Lindsay. Cognitive science and augmentative and alternative communication. *AAC Augmentative and Alternative Communication*, 7(3):186–203, 1991.

[55] S.Y. Lu and K.S. Fu. A sentence-to-sentence clustering procedure for pattern analysis. *IEEE Transactions on Systems, Man and Cybernetics*, 8:381–389, 1978.

[56] J. Matas, P. Mathy-Laikko, D. Beukelman, and K. Legresley. Identifying the nonspeaking population: A demographic study. *AAC Augmentative and Alternative Communication*, 1:17–31, 1985.

[57] A. Meyers, C. Macleod, and R. Grishman. Comlex syntax 2.0 manual for tagged entries. Technical report, proteus project, 1995.

[58] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. Five papers on wordnet, 1991.

[59] P. Mirenda and P. A. Locke. A comparison of symbol transparency in nonspeaking persons with intellectual disabilities. *Journal of Speech and Hearing Disorders*, 54:131–140, 1989.

[60] M. Mizuko. Transparency and ease of learning of symbols represented by blissymbols, pcs and picsyms. *AAC Augmentative and Alternative Communication*, 3:129–136, 1987.

[61] T. Nantais, F. Shein, and J. Treviranus. A predictive selection technique for single-digit typists. *IEEE Transactions on Rehabilitation Engineering*, 2(3):130–136, 1994.

[62] C. K. Ogden. *Basic English: A General Introduction with Rules and Grammar*. 1930.

[63] American Speech-Language-Hearing Association Amyotrophic Lateral Sclerosis Society Brain Injury Association Center on Disability, Health Communication Aid Manufacturers Association Communication Independence for the Neurologically Impaired International Society for Augmentative, Alternative Communication National Association of Protection, Advocacy Systems National Multiple Sclerosis Society RESNA Sunrise Medical United Cerebral Palsy Associations United States Society for Augmentative, and Alternative Communication. Formal request for national coverage decision for augmentative and alternative communication devices, December 1999.

[64] J. Owens and T. Caithness. Future directions for core vocabulary selection and key word signing. 1996.

[65] M. Palmer. Consistent criteria for sense distinctions. submitted to Computers and the Humanities, special issue on SENSEVAL, 1999.

[66] M. Palmer and Z. Wu. Verb semantics for english-chinese translation. *Machine Translation Journal*, 1995.

[67] F. Pereira, N. Tishby, and L. Lee. Distributional clustering of english words. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, 1993*, 1993.

[68] A. M. Peshkovskij. *Russkij sintaksis v nauchnom osveshchenii*. Moskva: Gosudarstvnnoe pedagogicheskoe izdatel'stvo, 1931.

[69] J. Ross Quinlan. *Discovering rules by induction from large collections of examples*. Edinburgh Univ. Press, 1995. D. Michie, editor.

[70] P. Resnick. Using information content to evaluate semantic similarity in a taxonomy. In *In Proceedings of the Fourteenth International Joint Conference on Articial Intelligence*, pages 448–453, 1995.

[71] B. Romich, G. Vanderheiden, and K. Hill. *Augmentative communication*, pages 101–122. Boca Raton, FL: CRC Press., 2000.

[72] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.

[73] R. W. Schvaneveldt, F. T. Durso, B. R., and Mukherji. Semantic distance effects in categorization tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 8:1–15, 1982.

[74] R.N. Shepard and P. Arabie. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86:87–123, 1979.

[75] B. Sinteff. Applying vocabulary search strategies in augmentative communication: Dynavox and dynamyte products. In *Proceedings of the Technology and Persons with Disabilities Conference*, 1998.

[76] L. Talmy. Path to realization: A typology of event conflation. In *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, pages 480–510, 1991. General Session and Parasession on the Grammar of Event Structure.

[77] G. Tatenhove. Vocabulary versatility for the person who is nonspeaking. *Communicating Together*, 4:19–20, 1986.

[78] J. Todman. Word prediction by word association. Presentation at ISAAC.

[79] J. Treviranus and L. Norris. Predictive programs: Writing tools for severely physically disabled students. In *In Proceedings of the 10th Annual Conference on Rehabilitation Technology*, pages 130–132. RESNA, 1987.

[80] P. B. Vanderheyden, C. A. Pennington, D. M. Peischl, W. M. McKnitt, K. F. McCoy, and Demasco. Developing aac systems that model intelligent partner interactions: Methodological considerations. In *In M. Binion (Ed.), Proceedings of the RESNA '94 Annual Conference*, pages 126–128. RESNA Press., 1994.

[81] M. Walker and A. Armfield. What is the makaton vocabulary? *Special Education: Forward Trends*, 8 (3):19–20, 1981.

[82] W. T Williams. Principles of clustering. *Annual Review of Ecology and Systematics*, 2:303–326, 1971.

[83] W. Woltosz. Dynamic graphic displays. 7, 3:15–17, 1993.

[84] M. E. J. Wood and E. Lewis. A syntactic word prediction system for disabled people. In *Proceedings of the 1996 Autumn Conference on Speech and Hearing*, volume 18:9, pages 315–322. Institute of Acoustics, 1996.