# Disambiguation of Imprecise User Input Through Intelligent Assistive Communication

Karl Wiegand

*Submitted in partial fulfillment of the requirements*
*for the degree of*

Doctor of Philosophy

*to the*

College of Computer and Information Science
Northeastern University
Boston, Massachusetts

December 2014

Karl Wiegand. *Disambiguation of Imprecise User Input Through Intelligent Assistive Communication*. PhD thesis, Northeastern University, Boston, MA, 2014

**NORTHEASTERN UNIVERSITY**
**GRADUATE SCHOOL OF COMPUTER SCIENCE**
**Ph.D. THESIS APPROVAL FORM**

*THESIS TITLE:* Disambiguation of Imprecise User Input Through Intelligent Assistive Communication

*AUTHOR:* Karl Wiegand

*Ph.D. Thesis Approved to complete all degree requirements for the Ph.D. Degree in Computer Science.*

| | |
|---|---|
| *Thesis Advisor* – Rupal Patel, Ph.D. | 12/8/2014 *Date* |
| *Thesis Reader* – Javed Aslam, Ph.D. | 12/5/14 *Date* |
| *Thesis Reader* – David Smith, Ph.D. | 12/5/2014 *Date* |
| *Thesis Reader, External* – Shaun Kane, Ph.D. | 12/5/2014 *Date* |

*GRADUATE SCHOOL APPROVAL:*

| | |
|---|---|
| *Director, Graduate School* | 12/10/2014 *Date* |

*COPY RECEIVED IN GRADUATE SCHOOL OFFICE:*

| | |
|---|---|
| *Recipient's Signature* | 12/10/2014 *Date* |

*Dedicated to everyone who still struggles to communicate.*
*Don't stop trying.*

# ABSTRACT

> *Thesis:* Intelligent interfaces can mitigate the need for linguistically and motorically precise user input to enhance the ease and efficiency of assistive communication.

Augmentative and alternative communication (AAC) systems are used by people with speech impairments severe enough to preclude the use of spoken communication. While communication systems for non-disabled users often implement intelligent prediction, correction, and behavior adaption, current AAC systems are relatively passive conduits for translating user intentions into spoken output. This dissertation seeks to shift the burden of communication from the user to the system by leveraging knowledge of the user's abilities, usage patterns, and contextual needs. The ultimate goal is to create an assistive communication prosthesis that enables users to seamlessly engage in timely and meaningful interactions in educational, vocational, or social settings.

This dissertation makes the following contributions to the advancement of intelligent interfaces for assistive communication, especially in the areas of natural language processing (NLP) and human-computer interaction (HCI):

1. A word-level language model — semantic grams — that bridges the gap between syntax and semantics by leveraging an author's own syntactic delimiters of semantic content. This model is more effective than similar n-gram-based language models for prediction tasks with unusual ordering or syntax.

2. An empirical comparison of contextual language predictors, showing that the use of statistics from a global corpus, such as the New York Times, is sub-optimal. Instead, situational context can provide more accurate background probabilities for pervasive speech and language processing tasks.

3. Results and observations from a touchscreen tablet study with current and potential AAC users, quantifying the challenges faced by people with upper limb motor impairments and showing how they can be addressed through intelligent interfaces.

4. Three user-driven interface designs and prototypes, including an approach to icon-based AAC that can be controlled effectively with a single input signal and leverages semantic frames to accommodate different screen sizes and user abilities.

# PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

K. Wiegand and R. Patel. DigitCHAT: Enabling AAC input at conversational speed. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, pages 343–344, New York, NY, USA, 2014. ACM. doi: 10.1145/2661334.2661422. URL `http://doi.acm.org/10.1145/2661334.2661422`

K. Wiegand. Intelligent assistive communication and the web as a social medium. In *Proceedings of the 11th Web for All Conference*, W4A '14, New York, NY, USA, 2014. ACM. doi: 10.1145/2596695.2596725. URL `http://doi.acm.org/10.1145/2596695.2596725`

K. Wiegand. Semantic disambiguation of non-syntactic and continuous motion text entry for AAC. *ACM SIGACCESS Accessibility and Computing*, 105:38–43, January 2013. doi: 10.1145/2444800.2444808. URL `http://doi.acm.org/10.1145/2444800.2444808`

K. Wiegand and R. Patel. Non-Syntactic word prediction for AAC. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies*, pages 28–36, Montréal, Canada, June 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W12-2905`

K. Wiegand and R. Patel. SymbolPath: A continuous motion overlay module for icon-based assistive communication. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 209–210, New York, NY, USA, 2012. ACM. doi: 10.1145/2384916.2384957. URL `http://doi.acm.org/10.1145/2384916.2384957`

# ACKNOWLEDGMENTS

Obtaining a Ph.D. is a journey. For me, the destination is not nearly as memorable as the route that was traveled: the struggles, the lessons learned, and the friendships forged along the way. Graduating is simply another milestone and the beginning of many new adventures. It is also a good time to pause and reflect. I have been fortunate enough to meet some truly amazing people during my time at Northeastern, and I would like to take a moment to thank them for the lasting impressions they have made on my life:

My advisor, Rupal Patel, to whom I owe so much. She has provided me with countless hours of mentorship and training, and she has guided me around more obstacles than I can remember, often before I even knew they existed. She is an incredible researcher and I am very lucky to have had her in my corner. Without her, this dissertation would not have been possible.

Javed Aslam, a fountain of wisdom and patience, who invited me into his laboratory and donated generously of his time and knowledge. He is a gentleman and a scholar, and a teacher in all senses of the word.

David Smith, who graciously reviewed and helped to improve my work. His seemingly boundless knowledge of linguistics and NLP reminds me that there is always more to learn.

Shaun Kane, an academic powerhouse and one of the nicest, most thoughtful researchers I've ever met. Even after listening to my presentation at the ACM ASSETS 2012 Doctoral Consortium, he still agreed to work with me.

Robert Futrelle, the polymath who took me under his wing, showed me how to immerse myself in NLP, and first introduced me to construction grammar.

Carole Hafner, who knows more about artificial intelligence than I could ever hope to learn, and who generously shared with me her experience and expertise.

Mansoor Pervaiz, whose enthusiasm for learning and helping others is contagious and inspirational. He helped take a number of other duties off my plate as I was finishing this dissertation, and I am very grateful for his time and thoughtfulness.

Kathryn Connaghan, Heather Kember, Brittany Dorer, Nicole Usher, Jonathan Brumberg, Timothy Mills, and all of my current and former colleagues at the Communication Analysis and Design Laboratory, who have provided me with so much feedback and encouragement. It turns out that it's easier to conduct research when you are surrounded by amazing people who are always ready and willing to support you. Go figure.

Sina Bahram, who continually reminds me to think critically and empathize with the user. He introduced me to the field of accessibility research, and he hasn't been able to get rid of me since.

William Furr, with whom I spent many late nights working and brainstorming at the lab. Billy always has a ready ear and an open mind, which are rare and valuable things in this world.

Virgil Pavlu, one of the brightest and most personable mathematicians I have ever met. I admire his ability to carefully consider all aspects of a problem before deciding how to proceed.

Karen O'Brien and Ennio Mingolla, who have fought many good fights on my behalf and put up with more than a few of my poetic and well-timed quips.

Ian Johnson, for being a friendly and supportive classmate when I needed one, and for sticking around just in case I might need one again.

Agnes Chan, Doreen Hodgkin, and Bryan Lackaye, who helped move countless administrative hurdles out of my way. I'm not sure I thanked them often enough, but I hope they know that I am grateful.

Alexis Ortiz, who does his best to pretend that he can actually understand my Spanish, and encourages me to keep trying.

Laura Legendary, a kindred spirit and a constant reminder of what friendship is all about. She multi-tasks on a myriad of different projects and still makes time to ask about your day.

Leisa Quiñones-Oramas, Yara González-Justiniano, and Noemi Santo, who make me smile whenever we meet. They encourage me to sing and dance and drink good wine amongst good friends as often as possible, and I thank them for the opportunities to do so.

Everyone at the Boston Home, especially Robert Scott and Alex Burnham, for taking the time to help me with my research when they certainly didn't have to. I would also like to thank Jack Stuart and the rest of the card sharks for teaching me how to play contract bridge. I bid 4 hearts.

Gordon Chen and John Li, for bringing me into their company and treating me like family. They showed me what it means to work hard and think creatively, and I am forever grateful for their support and their friendship.

Robert Bonds, Paulette Ford, and all of the staff and residents at Camfield Estates, especially the kids. They gave me lots of computers to play with, fed me jambalaya, and made me laugh. Being a part of their community brightens the soul.

Cliff Kushler, for being so positive and enthusiastic about my work, and for inspiring me with his own. He and the Continuous Path Foundation have been very generous in their contributions.

Tricia Stanton, Brenda Wynne, and Kenneth "Butch" Adams, for showering me with hugs, teaching me how to play pool, and helping me remember to have fun. Life is an adventure, and it's definitely more fun when they're around.

María Luisa Martínez, mi amor, who warms my heart every time I think of her. She is an incredibly kind, hard-working, and beautiful woman, and she has supported me in my journey even as she completes her own. I am truly blessed to have her in my life. I don't know if I will ever be able to thank her enough, but I look forward to trying.

My brother, Jean-Paul, who is on his own adventure at the moment, and fighting his own battles. I hope that he knows I am here for him whenever he needs me, and I love him always.

Of course, I wouldn't have even made it past the starting gate if not for my parents, Jamie and Angélique Wiegand. I don't know how they managed to raise me without going crazy, but it couldn't have been easy. And yet, through the thick of it all, they have shown me nothing but love and patience and encouragement. If there's anything good that I might contribute to the world, it will be because of them. Thank you, Mama. Thank you, Papa. I love you both very much.

Karl

Boston, MA
December 2014

# CONTENTS

## LIST OF FIGURES



## LIST OF TABLES

# ACRONYMS

| | |
|---|---|
| AAC | Augmentative and Alternative Communication |
| ALS | Amyotrophic Lateral Sclerosis |
| ASCII | American Standard Code for Information Interchange |
| ASR | Automated Speech Recognition |
| BCI | Brain-Computer Interface |
| BNC | British National Corpus |
| CP | Cerebral Palsy |
| DOM | Day of the Month |
| DOW | Day of the Week |
| EEG | Electroencephalography |
| EM | Expectation-Maximization |
| EMG | Electromyography |
| GPS | Global Positioning Systems |
| HCI | Human-Computer Interaction |
| IPG | Intelligent Parser/Generator |
| ISO | International Standards Organization |
| KL | Kullback-Leibler |
| LDA | Latent Dirichlet Allocation |
| MD | Muscular Dystrophy |
| MRR | Mean Reciprocal Rank |
| MS | Multiple Sclerosis |
| MSA | Multiple System Atrophy |
| NASA | National Aeronautics and Space Administration |

| | |
|---|---|
| ND | Non-Disabled |
| NIST | National Institute of Standards and Technology |
| NLP | Natural Language Processing |
| NLTK | Natural Language Toolkit |
| OLM | Operation-Level Model |
| PLM | Positional Language Model |
| RSVP | Rapid Serial Visual Presentation |
| SBD | Sentence Boundary Detection |
| SD | Standard Deviation |
| SGD | Speech-Generating Device |
| SLP | Speech-Language Pathologist |
| SMCR | Source-Message-Channel-Receiver |
| SMI | Speech and Motor Impairments |
| TLX | Task Load Index |
| TREC | Text REtrieval Conference |
| TTS | Text-to-Speech |
| VOCA | Voice-Output Communication Aid |
| WER | Word-Error Rate |
| WPM | Words Per Minute |
| WWW | World Wide Web |

# 1 INTRODUCTION

## BACKGROUND AND MOTIVATION

There are many ways to model communication. Most modern models of communication are derived from a foundation model called SMCR. Originally intended for use with telecommunications and cryptography, the SMCR model of communication consists of four key components: a Source, a Message, a Channel, and a Receiver [100, 97, 7]. In this type of model, the quality and integrity of a transmitted message can be compromised by distortion to any component, regardless of whether the communication involves face-to-face interaction, telephones, radios, or assistive devices. It is not uncommon for static or background noise, which can be viewed as distortion of the channel, to interrupt a telephone call or radio transmission. With spoken communication, deafness or other hearing impairments could be viewed as distortion to the receiver and pose similar challenges.

For over 2 million Americans with craniofacial deformities or neurological conditions, however, it is the message creation process itself that can be challenging. Natural human speech is not a viable mode of communication for many people who have had a stroke or have cerebral palsy (CP), multiple sclerosis (MS), multiple system atrophy (MSA), or amyotrophic lateral sclerosis (ALS) [70]. Many of these individuals also have physical impairments that limit the use of sign language or written forms of communication [8, 59]. These individuals rely on augmentative and alternative communication (AAC) to interact with the world around them. An estimated 53% of people with CP [41] and 75% of people with ALS [5] use AAC. In the general population, approximately 1 to 15 people of every 1,000 may require AAC at some point in their lives [8, 10, 61]. AAC systems, which range from physical letter boards to dedicated speech output devices, are the primary way for these individuals to convey their thoughts, needs, and desires to those around them.

AAC methods include:

- Unaided techniques, in which the user relies on gestures, facial expressions, vocalizations, or sign languages;
- Low-tech displays or boards, in which the user composes messages by selecting a series of letters or icons; and
- Speech-generating devices (SGDs), sometimes called voice-output communication aids (VOCAs), in which the user's selections on an electronic system are spoken aloud using speech synthesis.

For communication scenarios involving computerized systems, the risks of distortion can be alleviated by endowing components with some level of user-specific, adaptive, or context-aware "intelligence."

Almost every communication technology available today has some level of intelligence: desktop computers, video game consoles, and even Web browsers are now designed to accommodate multiple users with different preferences and capabilities. Similarly, personal mobile devices are increasingly adept at detecting the user's location and time of day in order to provide highly relevant information with minimal prompting. Vocabulary usage and typographical error statistics also aid in accelerating text entry on these systems.

In contrast, current AAC systems are relatively passive conduits for translating user intentions into spoken output. Communicating via these systems is slow and physically demanding because it requires considerable effort to search for, and navigate to, desired items [112]. While frequency statistics and natural language prediction are used in letter-based AAC systems, they are largely absent from icon-based AAC systems, and no commercial devices to date have made use of adaptive or context-sensitive information. Reconceptualizing AAC as an active, adaptive technology that leverages multiple information sources to facilitate and predict user intentions may have a profound impact on the ease, efficiency, and effectiveness of assistive communication.

Current VOCAs can be grouped into two general categories: sub-lemma construction systems and super-lemma construction systems. Sub-lemma construction systems include those that use letter-based approaches, but also those based on phonemes [108], morphemes [4], or any other units of construction that are more linguistically granular than lemmas. Super-lemma construction systems include those that use word-based approaches, but also those that leverage text or images to represent combined lemmas, phrases, or full utterances.

Dominant among sub-lemma construction systems is the letter-based orthographical approach; for the purposes of this document, the term "letter-based" will be used to generally refer to sub-lemma construction methodologies. Similarly, the majority of super-lemma construction systems use icons or symbols, either primarily or as cues to assist in visual search; the term "icon-based" will be used to generally indicate a reliance on images, words, or phrases. "Icons" will also be used to refer to both symbols and words because many systems provide users with the option of displaying any combination of images and associated text labels.

An advantage of letter-based AAC systems is that literate users can theoretically create any possible utterance in the target language; however, these systems can be slow and fatiguing (2 - 5 words per minute) because of the number of selections necessary to complete a message [105]. Additionally, many individuals who have sustained impairments since birth have poor or limited literacy skills [10]. Although icon-based AAC systems are typically not fully generative, they can be advantageous because they have the potential to support faster and more efficient message construction by allowing whole words and phrases to be accessible via a single keystroke [106].

Aided message construction is often an order of magnitude slower than spoken interaction: approximately 15 words per minute (WPM) compared to over 150 WPM for speakers of American English [11, 105, 36]. Thus, icon-based systems are often preferred over letter-based systems for face-to-face conversation and other real-time scenarios to minimize communication delay. They are also useful for non-native speakers and individuals with limited or emerging literacy skills [10], such as young children or those with language impairment due to neurological conditions.

Although the average college student uses approximately 5,000 unique words per day [73], most icon-based AAC systems have much smaller vocabularies, often with several hundred words or phrases [11]. Given that the vocabulary cannot be displayed all at once, typical icon-based AAC systems organize their vocabularies as arrays of icons in hierarchically nested pages categorized according to lexical, semantic, or thematic similarity [69]. Message construction in these systems requires users to complete two major tasks: (1) to search for desired icons by navigating through the available vocabulary, and (2) to select the desired icons. When users have finished composing an utterance, it can be sent to a text-to-speech (TTS) engine for vocalization.

## PROBLEM STATEMENT

Current icon-based AAC systems place the burden of communication on the user and make three fundamental assumptions:

1. **Prescribed Order**: Users will select items in a specific order, such as the syntactically "correct" one;
2. **Intended Set**: Users will select exactly the items that are desired, no fewer or more; and,
3. **Discrete Entry**: Users will make discrete movements or selections, either physically or with a cursor.

### *Prescribed Order*

Current AAC methods passively preserve the order of selected icons in the output, regardless of syntactic accuracy. Thus, if a user selects "hamburger," "eat," "I," and "want," the system would output "hamburger eat I want" rather than the syntactically accurate order for active-voice American English: "I want to eat a hamburger." Detecting semantic ambiguity becomes a problem when icons are selected in an unusual order. While some predicates are non-directional with regard to the subjects and objects that they allow (e.g. "Alice is near Bob" is semantically equivalent to "Bob is near Alice"), some predicates are directional (e.g. "Alice likes Bob") and word order affects meaning.

The Prescribed Order assumption is problematic for a number of reasons. First, there is evidence that users do not always select icons in expected orders [113]. This may be because of motor impairments, which often accompany speech impairments, that prevent users from making complex or repetitive movements. Second, because communication with AAC devices is so much slower than spoken communication, users may maximize speed by constructing simplified or telegraphic utterances [134, 78]. Third, many users may have limited or emerging literacy skills in the target language, making them unfamiliar with all of its syntactic rules. Regardless of the reason, outputting unusual or incomplete utterances has social implications: listeners may have diminished expectations or perceptions of the user's abilities [1].

*Intended Set*

Text entry systems on mobile devices often use dictionary-based approaches to account for scenarios in which the user types fewer or more letters than desired; however, such strategies are often not available for icon-based AAC systems. There are two major issues: (1) subset completion, in which the system suggests additional items after the user has selected a subset of desired icons; and (2) superset pruning, in which the system removes undesirable items that the user may have accidentally selected. Previous efforts in subset completion have either focused on missing function words, such as prepositions and conjunctions [72], or have operated under the Prescribed Order assumption [12, 114, 115]. Although AAC users can manually add and remove icons prior to speech synthesis, automated strategies have not typically been integrated into current devices. The result is that AAC message construction is slow, impeding real-time interaction, and users with fine motor impairments are burdened with the task of trying to avoid accidental selections.

*Discrete Entry*

Current icon-based AAC systems require discrete entry of each desired icon. Movements are often executed via a cursor that is manipulated physically, such as with a finger, hand, or eye; however, research has been conducted on other ways of manipulating an on-screen cursor, including vowel sounds [31, 15] and brain waves [133]. The assumption of Discrete Entry implies that selected icons are important, but the path of the cursor between icons is irrelevant. Recent work in letter-based text entry has explored the use of continuous and relative motion to shift the burden of lexical disambiguation from the user to the system [30, 90]. Several continuous text entry systems have been commercially successful for non-AAC users, especially on mobile platforms [51, 53]. Adapting these techniques for icon-based AAC involves adding semantic components, but may reduce the physical burden faced by users with motor impairments when making navigation and selection movements. Additionally, continuous motion input would support stronger integration with input mechanisms that are naturally continuous, such as vowel sounds, brain waves, or electromuscular signals.

Recent advances in touchscreen sensitivity, brain-computer interfaces, and miniaturized location sensors are just some of the reasons why challenging these three assumptions can allow us to rethink the design of assistive communication technology. This dissertation aims to endow assistive communication systems with enhanced intelligence to support free-order icon selection, unordered prediction and error correction, and continuous motion input. Our approach leverages both semantic knowledge and contextual cues to reduce physical effort and improve the efficiency of message construction.

## OUTLINE

The thesis of this dissertation is:

> Intelligent interfaces can mitigate the need for linguistically and motorically precise user input to enhance the ease and efficiency of assistive communication.

This dissertation makes two types of contributions: theoretical and applied. Part 1 presents the theoretical contributions: algorithms and design approaches that can "mitigate the need for linguistically and motorically precise user input." Chapter 2 presents semantic grams, a language model that can accommodate word prediction without assuming a particular order. Chapter 3 describes how situational context can be used to improve unordered prediction and allow for approximate word selections. In Chapter 4, quantitative results from a user study are analyzed, showing how touchscreen interactions can be improved through personalization.

Part 2 shows how these algorithms and approaches can be applied in practice and describes some of the effects of the current work towards enhancing the "ease and efficiency" of assistive communication. Three prototypes are described, each intended for different categories of AAC users, depending on their language and motor capabilities. Chapter 5 presents RSVP-iconCHAT, a semantic approach to icon-based message construction designed for users with impairments severe enough to necessitate switch interaction. Chapter 6 describes SymbolPath, an AAC system that enables continuous motion, free order, and superset selection of icons for users with mild-to-moderate language and motor impairments. Finally, Chapter 7 presents DigitCHAT, a small-footprint, letter-based AAC system for literate users with minimal upper limb motor impairments that supports AAC input at conversational speeds.

Part I

# THEORY

Summary of contributions in the areas of natural language processing (NLP) and human-computer interaction (HCI):

1. Semantic grams, or sem-grams, an unordered language model that leverages syntactic markers of semantic content for utterance-based, subset-completion tasks.
2. An empirical comparison of contextual language predictors showing that situational context provides more accurate background probabilities for pervasive speech and language processing tasks.
3. Results from a study with current and potential AAC users, quantifying the challenges they face with modern touchscreen technologies and describing how those difficulties can be addressed.

# 2 UNORDERED SELECTION

## 2.1 OVERVIEW

Most icon-based AAC devices require users to formulate messages in syntactic order. Reliance on syntactic ordering, however, may not be appropriate for individuals with limited or emerging linguistic skills. Some of these users may benefit from unordered message formulation accompanied by automatic message expansion to generate syntactically correct messages. Leveraging word prediction to increase communication speed in unordered message formulation, however, requires new methods of prediction. This chapter describes a novel approach to word prediction using semantic grams or "sem-grams," which provide relational information about message components, regardless of word order. Performance of four word-level prediction algorithms, two based on sem-grams and two based on n-grams, are compared on a conversational corpus. Results showed that sem-grams yield accurate word prediction, but lack prediction coverage. Hybrid methods that combine n-gram and sem-gram approaches may be viable for unordered prediction on icon-based AAC devices.

## 2.2 MOTIVATION

Many AAC users with limited or emerging literacy skills use icon-based systems, in which vocabulary items are selected in syntactic order to formulate messages. This syntactic bias stems from historical assumptions that AAC message formulation could be viewed as a corollary to written language, which can be sent directly to speech synthesizers. Selecting vocabulary items serially and in syntactic order can be physically and cognitively arduous depending on the icon organization scheme [112]. Moreover, AAC productions are often syntactically incomplete or incorrect [113], perhaps for efficiency or due to limited linguistic abilities. For many users, unordered vocabulary selection may alleviate the physical and cognitive demands of message formulation and shift the onus of generating syntactically complete and accurate messages onto the AAC device. Although unordered message formulation schemes have been proposed [46, 85], prediction has not been incorporated. This chapter presents an initial step toward text prediction from a set of unordered vocabulary selections.

Rate enhancement is a commonly cited issue in AAC because aided message formulation rates are an order of magnitude slower than spoken interaction [9]. Prediction is a common rate enhancement technique. Text prediction for AAC has primarily focused on well-ordered, syntactic input and has leveraged both semantic characteristics [20, 57, 81] and variations of n-grams [56, 110]. For example, semantic networks and linguistic rules have been used to predict missing function words and apply affixes to content words [72]. The use of n-grams to predict text entry has been extensively studied at both the level of letters [14, 102, 39] and words [12]. For example, memory based language models have been used to predict missing content words using tri-grams [114]. Although some recent work has attempted to loosen syntactic requirements by including either left or right context, some directional context has historically been required [115]. Furthermore, word prediction approaches in AAC have typically been implemented for letter-by-letter message formulation [48, 49, 54, 37]. The contributions in this chapter are fundamentally novel in that: (1) no syntactic order is implied or required during either training or testing, and (2) the prediction is implemented at word level to accommodate icon-based interaction.

Previous work in information retrieval has explored relationships between words with regard to distance [60, 74, 63], grammatical purpose [111, 2], and semantic characteristics [121, 26, 35], particularly for retrieving highly relevant documents or passages. One study in this area resulted in an approach called s-grams, a generalization of n-grams, in which the distance between words directly affects the strength of their semantic relationship [40]. Another approach to predicting semantically related words is to use collocation to indicate topic changes within a moving window of fixed length [71]. Rather than relying on distance to indicate relationship strength, the work in this chapter combines frequency analysis with syntactic indications of semantic coherence.

## 2.3 SEMANTIC GRAMS

Semantic grams, or "sem-grams," provide an alternative approach to quantifying the relationship between co-occurring words. A sem-gram is defined as a multi-set of words that can appear together in a sentence. In English, a sentence is one of the smallest units of language that is typically both coherent, in terms of semantic content, and cohesive, in that the contained semantic content is interrelated. Additionally, because sentences are demarcated with syntactic cues such as punctuation, semantically related items can be efficiently identified using sentence boundary detection [47]. Thus, sem-grams leverage sentence-level co-occurrence to extract semantic content at different levels of granularity, depending on the allowable lengths of multi-sets. Sem-grams can be viewed as non-directional s-grams with a uniform weight applied to all relationships between any words in a given sentence.

In a sentence of length $L$, the number of n-grams of length $n$ is given by the expression $L - n + 3$, which includes the beginning and ending n-grams that contain null elements. By contrast, the number of sem-grams of length $n$ in a sentence of length $L$ is given by the expression $\binom{L}{n}$. Thus, there will typically be many more sem-grams of length $n$ in a single sentence than n-grams of the same length. Unlike n-grams, however, it is unnecessary for sem-grams to contain null elements because a sem-gram of length $S$ with a null element is equivalent to a sem-gram of length $S-1$ without null elements. Sem-grams of length one, containing a single word, are equivalent to the prior probability of that word.

## 2.4 PREDICTION ALGORITHMS

Unordered word prediction poses the following problem: given a set of existing words $E$ that have already been selected by a user and a set of candidate words $C$ that the user may select from, which candidate word $c \in C$ is the user most likely to select in order to complete the message? As an initial step toward addressing this problem, the following four algorithms, two based on sem-grams and two based on n-grams, are presented:

**S1: NAIVE BAYESIAN SEM-GRAMS** Given existing words $E$, rank all candidate words $c \in C$ in descending order of probability according to:

$$P(c|E) = P(c) \prod_{w \in E} P(w|c) \tag{1}$$

S1 is a modification of the Bayesian ranking of sem-grams in that it assumes independence of existing words to each other, conditional on the given candidate word. Using true Bayesian probabilities for sem-grams, the probability of a candidate word would look like the following for each $P(c|E)$, given $w \in E$ and $|E| = 3$:

$$\frac{P(c)P(w_1|c, w_2, w_3)P(w_2|c, w_3)P(w_3|c)}{P(w_1, w_2, w_3)} \tag{2}$$

The exact form of this equation depends on the ordering branch chosen, but it also requires joint probabilities for sem-grams of different lengths. By assuming conditional independence of the existing words to each other, S1 only requires sem-grams of length 2.

**S2: INDEPENDENT SEM-GRAMS** Given existing words $E$, rank all candidate words $c \in C$ in descending order of probability according to:

$$P(c|E) = \prod_{w \in E} P(w, c) \tag{3}$$

The approach of S2 is a "hand of cards" approach that treats the message formulation task as a random drawing of sem-grams from a pool of available sem-grams. While the formula above is specified for sem-grams of length 2, it can be extended to support sem-grams of any length.

**N1: NAIVE BAYESIAN N-GRAMS** Given existing words E, rank all candidate words $c \in C$ in descending order of probability according to:

$$P(c|E) = P(c) \underset{w \in E}{\quad\quad} P(w|c) \tag{4}$$

N1 is a copy of S1, except that the definition of the joint probability $P(w, c)$ includes the counts for n-grams that contain both $w$ and $c$, regardless of order. This algorithm was designed to compare whether the information provided by n-grams can be used to approximate the information provided by sem-grams. N1 assigns high ranks to candidate words that are likely to appear adjacent to all other words in the sentence.

**N2: APPLIED N-GRAMS** Given existing words E, rank all candidate words $c \in C$ in descending order of probability according to:

$$P(c|E) = \sum_{w \in E} P(w, c) \tag{5}$$

N2 is designed to leverage the strength of n-grams and rank candidate words based on the probability of them appearing adjacent to any of the existing words. N2 uses the same definition of joint probability as N1, where $P(w, c)$ includes the counts for n-grams that contain both $w$ and $c$, regardless of order.

### 2.5 CORPUS SELECTION AND PREPARATION

Given the lack of large corpora of AAC message formulations [55], approximations have often been used [109, 116]. The Blog Authorship Corpus [96] was selected because it is freely available and tends to be written in an informal style that emulates conversational speech. The corpus is both large and diverse, comprising over 140 million words written by 19,320 bloggers in August 2004. The bloggers ranged in age from 13 - 48 and were equally divided between males and females.

To prepare the corpus, all blog posts were extracted as ASCII text. Every blog post was split into sentences using the PunktSentenceTokenizer [47] of the Natural Language Toolkit (NLTK) [13] and then split into words using the following regular expression:

```
\w+(\w*([\-\'\.]\w+)*)*
```

English stop words were removed according to a popular list [89] and remaining words were stemmed using the NLTK's PorterStemmer, which is a modified implementation of the original Porter stemming algorithm [87]. Finally, all stemmed words were examined for membership in a stemmed American-English dictionary [120]. Any stemmed words not found in the dictionary were removed to further constrain the vocabulary and account for spelling errors and nonsensical text.

The corpus was then randomly split into a training and testing set, with 80% of the authors (15,451) being placed in the training set and 20% of the authors (3,871) being placed in the testing set. The training set comprised over 7 million sentences written by 7,682 males and 7,768 females with a combined average age of 22 years. All n-gram and sem-gram statistics, with plus-one smoothing, were gathered using only sentences in the training set and both n-grams and sem-grams were limited to a word length of 2 (bi-grams).

## 2.6 EVALUATION

Testing was conducted on 2,000 sentences that were randomly selected from the test corpus. The same processing steps used during training were performed on the test sentences: stop words were removed, the remaining words were stemmed, and all stems not in the dictionary were filtered out. To avoid run-on sentences and sentence boundary detection errors, all test sentences were also truncated to a maximum of 20 words. The words in each test sentence were then shuffled and one word was removed at random and designated as the target word. Each of the four algorithms were provided the shuffled words as input; as output, each algorithm attempted to identify the target word by generating a ranked list of candidates.

In addition to the shuffled set of input words, each algorithm required a seed list of candidate words. Ideally, all known words in the corpus would be used as candidate words. To constrain the computational requirements, the two algorithms based on n-grams (N1 and N2) were provided with the list of most frequently co-occurring words that appeared as n-grams with any of the set of input words, limited to the top 10 n-grams for a given input word. Similarly, each sem-gram algorithm (S1 and S2) received a list of most frequently co-occurring words that appeared as sem-grams with any of the set of input words, limited to the top 10 sem-grams for a given input word. With a limit of 19 input words (20 minus the target word), each algorithm received at most 190 unique candidate words to rank.

Two primary metrics were used to quantify the performance of each algorithm: (1) a boolean value that was true if the output list contained the target word in any position, indicating that the target word had been successfully predicted; (2) if the algorithm successfully predicted the target word, the algorithm received a positive integer score corresponding to the position of the target word in the output list, with lower scores indicating more accurate prediction. For example, if an algorithm suggested the target word as the first item in its ranked list, it received a score of 1; if it suggested the target word as the second item in its ranked list, it received a score of 2. The output lists of each algorithm were truncated to the first 100 items; thus, if an algorithm's output list contained the target word in a position after 100, it was marked as failing to predict the target word.

Table 1: Summary of N-Grams vs. Sem-Grams

| | N1 | N2 | S1 | S2 |
|---|---|---|---|---|
| **Sentences** | 2000 | 2000 | 2000 | 2000 |
| **# Predicted** | 647 | 649 | 435 | 435 |
| **% Predicted** | 32% | 32% | 22% | 22% |
| **Avg Score** | 16.26 | 19.70 | 9.04 | 12.67 |
| **Score SD** | 16.15 | 19.65 | 7.40 | 10.04 |
| **MRR** | 0.0643 | 0.0412 | 0.0737 | 0.0556 |
| **MRR SD** | 0.1667 | 0.2018 | 0.1335 | 0.1746 |

A third metric, Mean Reciprocal Rank (MRR), was used to obtain an overall picture of each algorithm's performance by merging accuracy and coverage. MRR is the average of reciprocal ranks (i.e. scores) over all evaluated sentences, where *rank* corresponded to the position of the target word in the output list, and was calculated as:

$$\mathrm{MRR} = \frac{1}{2000} \sum_{i=1}^{2000} \frac{1}{\mathrm{rank}_i} \tag{6}$$

If the target word did not appear in the first 100 items, a reciprocal rank value of zero was used.

## 2.7 RESULTS

The n-gram algorithms successfully predicted 32% of the 2,000 test sentences while the sem-gram algorithms successfully predicted 22% (Table 3). Although both n-gram algorithms performed similarly, N1 consistently predicted the target word more accurately than N2. On average, N1 suggested the target word as the 16th word in its ranked list, where N2 suggested the target word as the 20th word in its list. While the sem-gram algorithms predicted fewer sentences than the n-gram algorithms, they were almost twice as accurate on sentences that they did predict. On average, S1 suggested the target word as the 9th word in its ranked list; for S2, the target word was the 13th item. The results by MRR were similar: S1 outperformed N1, while S2 outperformed N2; however, N1 did show better performance than S2 by this metric.

To further compare the effectiveness of sem-grams and n-grams, sentences were grouped according to their input length, from 1 to 19 words, and statistics were gathered for each algorithm on each sentence length (Table 2). For test sentences in which the algorithms were only given a single input word, both n-gram algorithms ranked the target word at least one full ranking higher than either sem-gram algorithm, thus giving more accurate predictions. For all other sentence lengths, the sem-gram algorithms were more accurate. Between the n-

Table 2: N-Grams vs. Sem-Grams by Input Sentence Length

| # Words | N1 % | N1 Avg | S1 % | S1 Avg | N2 % | N2 Avg | S2 % | S2 Avg |
|---|---|---|---|---|---|---|---|---|
| **1** | 20.88% | 3.44 | 12.05% | 4.47 | 20.88% | 3.42 | 12.05% | 4.47 |
| **2** | 26.55% | 6.07 | 19.47% | 5.89 | 26.55% | 6.32 | 19.47% | 6.23 |
| **3** | 22.22% | 7.64 | 16.89% | 6.87 | 22.22% | 9.82 | 16.89% | 9.84 |
| **4** | 32.11% | 10.46 | 22.94% | 7.62 | 32.11% | 11.91 | 22.94% | 9.94 |
| **5** | 31.25% | 12.13 | 21.88% | 6.14 | 31.25% | 14.02 | 21.88% | 9.14 |
| **6** | 38.18% | 15.25 | 26.67% | 8.75 | 38.18% | 17.68 | 26.67% | 12.11 |
| **7** | 42.86% | 16.17 | 29.46% | 9.52 | 42.86% | 21.77 | 29.46% | 12.73 |
| **8** | 39.60% | 18.08 | 25.74% | 11.15 | 39.60% | 22.00 | 25.74% | 15.73 |
| **9** | 29.11% | 19.13 | 20.25% | 11.31 | 29.11% | 23.48 | 20.25% | 17.88 |
| **10** | 44.74% | 24.47 | 35.53% | 10.52 | 44.74% | 23.56 | 35.53% | 16.22 |
| **11** | 38.46% | 28.55 | 26.92% | 15.21 | 38.46% | 26.80 | 26.92% | 17.93 |
| **12** | 46.00% | 23.39 | 14.00% | 13.71 | 46.00% | 41.26 | 14.00% | 9.14 |
| **13** | 38.46% | 24.47 | 25.64% | 14.30 | 38.46% | 34.07 | 25.64% | 15.90 |
| **14** | 29.41% | 26.30 | 14.71% | 10.80 | 29.41% | 39.10 | 14.71% | 26.20 |
| **15** | 46.67% | 32.14 | 20.00% | 16.17 | 46.67% | 36.79 | 20.00% | 15.17 |
| **16** | 47.62% | 25.70 | 28.57% | 12.83 | 47.62% | 30.50 | 28.57% | 12.67 |
| **17** | 53.85% | 23.14 | 38.46% | 12.20 | 53.85% | 35.14 | 38.46% | 21.40 |
| **18** | 40.95% | 38.35 | 25.71% | 13.56 | 42.86% | 43.07 | 25.71% | 25.11 |
| **19** | 38.46% | 23.80 | 38.46% | 11.00 | 38.46% | 52.40 | 38.46% | 32.00 |

Note: % = Prediction coverage; Avg = Average prediction score.

gram algorithms, N1 consistently predicted the target word more accurately and more often than N2. Similarly, S1 consistently predicted the target word more accurately and more often than S2.

For every input sentence length greater than one, S1 outperformed N1 in all gathered metrics. When comparing the prediction accuracy of N1 and S1, S1's prediction accuracy was also more stable, with N1's prediction accuracy continuing to degrade as the length of the input sentence increased (Figure 1). Note that lower values represent earlier prediction and thus higher prediction accuracy.

## 2.8 Discussion

Message formulation using AAC devices has historically relied on selection of letters or words (icons) in syntactic order. This chapter aimed to facilitate unordered vocabulary selection through the use of text prediction. Results indicate that word prediction for unordered message formulation is viable using statistical approaches. Although the n-gram algorithms predicted a larger number of test sentences than the sem-gram algorithms, evaluation of the ranked output indicated that the sem-gram approaches were more accurate. Because n-grams assume that adjacent words are strongly related, it was expected that n-grams would provide more accurate prediction for shorter sentences; however, this advantage was not maintained as

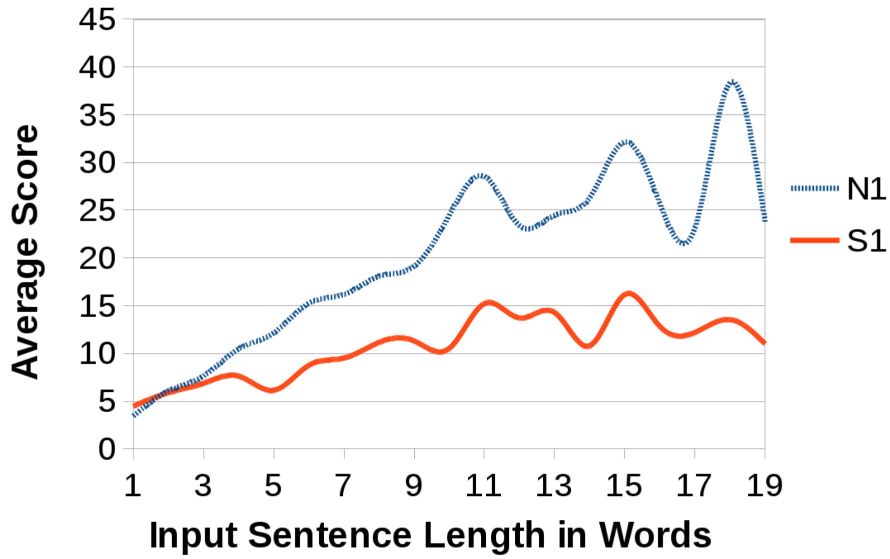



Figure 1: Accuracy of N-Grams and Sem-Grams by Sentence Length

sentence length increased beyond two words. Prediction accuracy is likely to be more important in AAC devices because the cognitive demands of choosing from prediction lists can sometimes outweigh rate enhancements [48, 49].

The use of bi-grams may have resulted in poor accuracy of the n-gram algorithms because there were many more sem-grams than n-grams of length 2. Increasing n-gram length, up to a cardinality equal to the number of sem-grams of length 2, could allow n-gram algorithms to potentially match or surpass the prediction accuracy of sem-grams. For unordered word prediction, however, this larger set of n-grams would need to be indexed in an order-independent manner which would further increase computational demands. Prediction lags are unlikely to be tolerated by users as they engage in interactive tasks [37].

Of the two n-gram algorithms, N1 outperformed N2 on both prediction coverage and accuracy. It was hypothesized, however, that N2 would yield more accurate predictions because the target word was defined to be adjacent to at least one of the input words. It was expected that N1 would unfairly reward candidate words that had appeared adjacent to each input word in the training set, while punishing more desirable candidate words that had not appeared adjacent to some of the input words. Perhaps this bias was not evident in the current corpus because plus-one smoothing removed all zero probabilities for adjacency likelihoods. Additionally, N1 may have been more successful because it favored candidates that were related to all input words rather than candidates that were strongly related to just a subset of the input words.

Despite the encouraging prediction coverage of n-grams and the prediction accuracy of sem-grams, approximately two-thirds of the test sentences were not predicted by any of the algorithms. One possible explanation may relate to the decision to seed each algorithm with only the top 10 most frequent words that co-occurred with each input word. Ideally, each algorithm would have considered all words in the vocabulary as candidate words; however, because there were almost 100,000 unique stems in the vocabulary, the computational requirements were prohibitive for this initial implementation. The decision to use this sparse seeding strategy in combination with a prediction rank cut-off of 100 is also likely to be responsible for the extremely low MRR scores that we observed. Relatively high standard deviations (SDs) for the MRR scores make these values especially difficult to compare with similar research results. An open empirical question is whether increasing the seed values to include a larger set of co-occurring words would result in greater prediction coverage and potentially more comparable MRR scores. It should be noted, however, that while seeding sem-grams with more candidate words may improve prediction coverage, it is unlikely to increase prediction accuracy for the n-gram approaches.

Icon-based AAC devices typically have vocabularies with much fewer than 100,000 words, which may negate the need for seeding candidate words. For example, two commonly used icon sets, the Widgit Symbol Set and the Mayer-Johnson Picture Communication Symbol collection, each contain approximately 11,000 icons [122]. While a large dictionary was used to provide a conservative estimate of prediction performance, it is possible that using a smaller and more representative AAC vocabulary would improve prediction coverage and accuracy. Restricting vocabulary size would also reduce computational demands, making it more feasible to use all vocabulary words as candidates.

## 2.9 SUMMARY

Semantic grams, or sem-grams, provide a promising approach to word prediction for AAC users who may benefit from unordered message formulation. Sem-grams make use of co-occurrence between words within a sentence to improve prediction accuracy. While n-grams have historically provided a strong foundation for word prediction in letter-by-letter systems, results indicate that they can also be used for unordered word prediction, although they are not as accurate as sem-grams. A hybrid approach that seeds both types of algorithms with a superset of candidate words and merges the prediction lists may simultaneously exhibit the wide prediction coverage of n-grams and the high prediction accuracy of sem-grams. Such a hybrid approach could enable unordered message formulation on icon-based AAC devices.

Extensions of this work may be possible using the breadth of information available within well-documented and comprehensive corpora. For example, while the Blog Authorship Corpus included age and gender information about each blogger, this information was not used in the present study. To tailor prediction to individual users, it may be possible to limit the available vocabulary and gram-based statistics to information gathered from users of similar age and gender. This may improve prediction accuracy for both n-gram and sem-gram algorithms, as well as provide an approach to designing icon-based AAC devices that can evolve and adapt to users as their needs and abilities mature, potentially even suggesting new vocabulary words as the users age.

# 3

# CONTEXTUAL PREDICTION

## 3.1 OVERVIEW

Speech and language technologies benefit from contextual awareness, especially in the area of assistive communication. Historically, the use of context in language prediction and disambiguation tasks has primarily focused on words in the same utterance. With the rapid advancement of mobile technologies, however, both written and spoken language interaction can be augmented with information from other types of contextual predictors, such as calendar date and geographical location. It is generally agreed that knowledge of a specific author's language patterns provides the most predictive power; however, author-specific corpora are not always available, especially for users with speech and motor impairments. While some research has been conducted in the area of linguistic content analysis by gender, age, and other demographic variables, there has been little work on quantifying the effects of context on language prediction. We compared seven contextual cues (age, gender, day of the week, day of the month, month, city, and state) within two different corpora in order to determine the most reliable predictors of language usage when author-specific training samples are not available. Across three different metrics, we show that contextually sensitive language distributions can often provide more useful predictions and significantly more accurate reflections of reality than global distributions; however, we also show that contextual cues must be chosen carefully to avoid introducing noise and degrading the performance of standard similarity measures. The work in this chapter has implications for speech and language processing systems, especially personalized assistive communication, that rely on statistical distributions to predict or disambiguate intended word usage.

## 3.2 MOTIVATION

Nearly every language-related assistive technology, from AAC to automated speech recognition (ASR), relies on probability distributions of word usage [42, 43, 82, 131]. These distributions, sometimes called prior or background probabilities, are used to ensure effective prediction or disambiguation of utterances. The most common sources for these distributions are currently large collections of global text, such as the New York Times [94] or Google's N-Gram corpus [75]; however, many systems use the same global distribution for all users, regardless of the user's demographics or situational context.

As mobile technologies become more advanced and social networks more pervasive, users are increasingly interacting via spoken or written language in a variety of situations [36, 38]. Sensors have also become smaller and more powerful, allowing mobile devices to gather multidimensional data. Understanding language usage patterns is essential to enhancing the ease and efficiency of communication on these various platforms.

While "context" has traditionally been used to refer to words within a target utterance, or surrounding utterances, ubiquitous sensor technologies now provide a richer set of contextual cues. For example, Global Positioning Systems (GPS) can provide location and an internal clock or synchronization system can provide the date and time, all potential predictors of language usage. Most mobile devices also include a configuration process in which the user is prompted for demographic information, such as age and gender. Many mobile devices even have the ability to associate with preconfigured user accounts, which may contain much more detailed information, such as the topography of the user's social network, the user's interests and hobbies, education, and career or job title. All of this information can be viewed as context that provides insight into that user's potential language usage patterns.

There has been extensive prior work on comparing language usage patterns across socio-economic backgrounds [23], as well as age and gender groups [98]. People who share similar personality traits have been found to share similar word usage patterns, particularly within social networks [135]. Prior work has also noted that a person's vocabulary is strongly correlated to location of use [84, 45]. There is even a National Institute of Standards and Technology (NIST) Text REtrieval Conference (TREC) Contextual Suggestion Track, first offered in 2012, in which search terms are augmented with contextual information, including location and season [19]. The current work extends this line of inquiry to understand the relative contribution of each contextual cue, using raw feature counts and vocabulary distributions, across two different corpora to determine the most reliable predictors of language usage. The resultant list of prioritized contextual cues could be applied to enhance vocabulary prediction or disambiguation in assistive communication systems or other pervasive speech and language technology.

## 3.3 APPROACH

We examined several contextual categories that can be derived from current mobile devices during initialization or via sensor technology, such as known attributes of the author and the current location. We further divided these categories into seven contextual language predictors for a given utterance or writing sample: the author's age, gender, city, state (geographical), as well as the calendar month, day of the month, and day of the week. In addition to examining the contribution of each predictor independently, all available combinations of predictors were also compared. In this chapter, the term "predictors" is used interchangeably with "contextual categories" or "contextual cues," and the "context" of a message or utterance is the set of values for each predictor when the message or utterance was constructed.

Predictive power was estimated by comparing word-level unigram distributions, as provided by predictor combinations, to actual distributions of spoken or written content. Word-level unigrams were used, rather than higher-level n-grams or skip-grams, to obtain low-level semantic granularity and provide a foundation for future studies with more complex language models. More computationally intense approaches should only improve upon the performance of these unigram baselines.

## 3.4 CORPORA

The lack of representative corpora in the field of assistive communication technology is well documented [109]. Additionally, existing corpora are often not tagged with contextual information. To support reproducible results, we chose to analyze two freely available corpora, primarily in English, that contain overlapping contextual information: the Blog Authorship corpus and the Yelp Academic Dataset. The Blog Authorship corpus is a collection of over 680,000 blog posts from the Blogger.com website [96]; the Yelp dataset contains over 330,000 narrative reviews of 250 businesses in 16 different states [136].

Each corpus was processed to create a mapping between every word, converted to lowercase, with the contexts in which it was used and the number of times it was used in each context. Words were parsed using the Penn Treebank tokenizer from version 3.0 of the NLTK [13]. No stemming was performed in order to empirically discover language patterns that might have been hidden by conflated conjugations or unusual spellings. A minimal list of 34 stop words was used, consisting primarily of articles, prepositions, and corpus-specific placeholders:

> *a, about, an, and, are, as, at, be, by, com, for, from, how, http, in, is, it, of, on, or, that, the, this, to, urllink, was, what, when, where, who, why, will, with, www*

Common stop words such as "like" were not removed because there are well-understood language patterns that make extensive use of certain stop words, including "like" and "so" [132, 22]. Although the removal of stop words is common in information retrieval [129], it can sometimes be a confounding factor [24, 137]. In this study, stop words were removed primarily to reduce the burden of data storage and computational complexity, but it is important to note its potential effects.

## 3.5 EVALUATION

Because both corpora were collected under different usage scenarios, which could be viewed as higher-order contexts, each corpus was analyzed separately. For each corpus, a feature list was generated, consist-

Table 3: Summary of Processed Corpora

| Attribute | Blog Authorship | Yelp |
|---|---|---|
| Authors | 19,320 | 130,850 |
| Features | 525,253 | 134,199 |

Table 4: Unique, Non-Empty Predictor Values per Processed Corpus

| Predictor | Blog Authorship | Yelp |
|---|---|---|
| Age | 26 | - |
| Gender | 2 | - |
| Day of the Week (DOW) | 7 | 7 |
| Day of the Month (DOM) | 31 | 31 |
| Month | 12 | 12 |
| City | - | 119 |
| State | - | 16 |

Note: A hyphen indicates that the predictor was not available in the corpus.

ing of all vocabulary words used more than once in that corpus. Each corpus was randomly divided into 10 groups of authors and remaining authors were discarded (Table 3). One fold (10%) was analyzed from each of the corpora: 1,932 authors from the Blog Authorship corpus and 13,085 authors from the Yelp corpus. For each analyzed author in each corpus, all non-empty combinations of all values for each available predictor were used as target distributions: age, gender, day of the week, day of the month, month, and city/state. To further lower computational complexity and account for intra-predictor similarity, values for "day of the month" were combined into 4 groups (1 - 8, 9 - 16, 17 - 24, 25 - 31), roughly corresponding to "week of the month."

There was a theoretical upper limit of $26 \times 2 \times 7 \times 4 \times 12 \times 119 = 2,079,168$ possible contexts for each author; however, because age and gender were generally constant for each author, and other predictors often had only a few possible values for each author, the average number of unique contexts was approximately 18 per author in the Blog Authorship corpus and 4 per author in the Yelp dataset.

For each context per author, the counts of each feature formed a vocabulary distribution within that context. This distribution was the target distribution. Predicted distributions were obtained for every available combination of predictors in the corpus (Table 4). Thus, $2^5 = 32$ predictor combinations were compared in each corpus for an overall total comparison of 48 combinations. For every predictor combination, the predicted distribution was taken over all authors in the other 9 folds in the corpus: there was no intersection between training and test authorship.

Table 5: Example Combinations for a Hypothetical Target Distribution

| Predictors | Description |
|---|---|
| Age | The vocabulary distribution of all 23-year-olds in the non-target 9 folds. |
| Gender | The vocabulary distribution of all females in the non-target 9 folds. |
| Age + Gender | The vocabulary distribution of all 23-year-old females in the non-target 9 folds. |
| Gender + City | The distribution, from the non-target 9 folds, of all words written in Seattle by female authors. |
| DOM | The distribution of words written between the 25th and 31st days of the month, inclusive, by all authors in the non-target 9 folds. |

For example, suppose the Yelp corpus had contained a hypothetical author Alice, a 23-year-old female. One target distribution might have been the vocabulary used by Alice on Monday, July 25th, 2011, when Alice reviewed a restaurant in Seattle, Washington, USA. For this distribution, the context would have been:

Age = 23
Gender = Female
DOW = Monday
DOM = 25 - 31
Month = July
City = Seattle
State = Washington

Some example predictor combinations for the target distribution specified by this context are described in Table 5. The target distribution would have been compared to the distribution from every predictor combination, and this process repeated for all of Alice's possible target distributions.

Each corresponding target vector $A$ and predictor vector $B$, consisting of raw feature counts, was additively smoothed (uniform plus-one) and then normalized to create a corresponding target distribution $P$ and predictor distribution $Q$. These feature vectors and distributions were compared using the following metrics:

KULLBACK-LEIBLER DIVERGENCE A non-symmetric measurement of the bits of information lost when using distribution $Q$ to approximate "true" distribution $P$ [52], Kullback-Leibler (KL) divergence measures the directed divergence between distributions under the assumption that each distribution sums to 1 and there is absolute continuity ($Q(i) = 0 \Rightarrow P(i) = 0$). Both of these assumptions were fulfilled by the additive smoothing, giving:

$$D_{KL}(P\|Q) = \sum_{i=1}^{n} \log_2 \left( \frac{P(i)}{Q(i)} \right) P(i) \tag{7}$$

KL divergence is well-suited to evaluating statistical language models: it is strongly related to perplexity and additively equivalent to cross-entropy [107].

COSINE SIMILARITY Used extensively in information retrieval and recommender systems, cosine similarity is related to Pearson's product-moment correlation coefficient $r$, can be readily calculated, and has well-understood predictive properties [93, 103]. Cosine similarity treats each of the $n$ features as an axis in high-dimensional positive space and measures the cosine of the angle $\theta$ between vectors $A$ and $B$:

$$\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{8}$$

PRECISION AT 20 (PREC@20) Commonly used in information retrieval [88], Prec@20 measures the percentage of relevant items returned within a rank cut-off of 20. Precision is a useful measurement for prediction because, especially with written tasks, message composition is usually augmented by a small number of the most likely suggestions. In the current work, Prec@20 is calculated by assuming that the top 20 most likely words in $A$ are relevant. Thus, for $M$ and $N$ as the sets of 20 most likely words from $A$ and $B$, respectively:

$$\text{Prec@20} = \frac{|M \cap N|}{20} \tag{9}$$

## 3.6 RESULTS

Table 6 presents the results for predictor combinations in increasing order of mean KL divergence, with lower divergence indicating better performance (i.e. stronger similarity and higher prediction accuracy). Results for predictor combinations in decreasing order of mean cosine similarity and precision are presented in Table 7 and Table 8. In addition to relative rank position and SD, "Count" indicates the number of times the predictor was compared to a target distribution; variation among counts are related to the availability of contextual cues and non-empty predictor combinations in each corpus. All numbers shown are rounded to four decimal places. The empty predictor combination, labeled as "No Context," represents the global vocabulary distribution over the non-target folds and indicates that no contextual information was used.

Combinations of predictors that make use of both city and state ("City+State") are not included in the results because cities are generally unique to a state. We exclude the possibility of similarly named cities having related influences, so in terms of information content, knowledge of the current state does not add further information if the current city has already been determined.

The best performing predictor combination by KL divergence made use of the author's location and all available information about the calendar date ("DOW+DOM+Month+City"). The second best combination by KL divergence replaced location with the author's age and gender and performed only slightly worse. The non-contextual predictor performed second-worst, surpassed in inaccuracy only by the Gender predictor. The ranking pattern displayed by the KL divergence metric shows that, in general, using more contextual cues results in predictor distributions that more accurately reflect true distributions. For example, the predictor combination leveraging both calendar month and geographical state ("Month+State") performed better than either predictor separately.

Predictor rankings by cosine similarity and precision were highly correlated; however, because few stop words were removed, it is possible that the top 20 words in each feature vector contained other high-frequency grammatical terms. By cosine similarity, the best performing predictor combination leveraged the author's gender and the calendar date of message construction ("Gender+DOM+Month"). By precision, the best predictor combination used the author's age and the calendar month ("Age+Month").

In general, predictor combinations that outperformed the non-contextual predictor by cosine similarity and precision leveraged demographic information about the author. Unlike the rankings by KL divergence, however, the predictor combination with demographics and calendar date ("Age+Gender+DOW+DOM+Month") did not perform as well as either the Age or Gender predictors individually. The worst predictor combinations by cosine similarity were those that leveraged location information, such as City or State. Interestingly, the best predictor combination by KL divergence ("DOW+DOM+Month+City") was the worst predictor combination by cosine similarity and precision.

Overall, the non-contextual predictor, representing the global vocabulary distribution across the non-target 9 folds for each corpus, had decidedly sub-par performance. It was 47th out of 48 by KL divergence, 31st out of 48 by cosine similarity, and 27th out of 48 by precision (prec@20).

## 3.7 DISCUSSION

Differences between the rank orderings by KL divergence and cosine similarity were to be expected because of the different aims and behaviors of each metric. KL divergence measures information loss due to the use of an approximate probability distribution; cosine similar-

Table 6: Best Predictor Combinations by KL Divergence

| **Predictor Combination** | **Count** | **KL Divergence** | | |
|---|---|---|---|---|
| | | *Rank* | *Mean* | *SD* |
| DOW+DOM+Month+City | 48996 | 1 | 0.0331 | 0.0290 |
| Age+Gender+DOW+DOM+Month | 34053 | 2 | 0.0457 | 0.0577 |
| Age+DOW+DOM+Month | 34257 | 3 | 0.0913 | 0.1029 |
| DOW+DOM+Month+State | 49644 | 4 | 0.1233 | 0.1039 |
| DOW+Month+City | 49494 | 5 | 0.1382 | 0.1074 |
| Age+Gender+DOW+Month | 34170 | 6 | 0.1575 | 0.1395 |
| DOM+Month+City | 49588 | 7 | 0.2372 | 0.1726 |
| Age+Gender+DOW+DOM | 34314 | 8 | 0.2680 | 0.1552 |
| Age+Gender+DOM+Month | 34182 | 9 | 0.2791 | 0.2424 |
| Age+DOW+Month | 34308 | 10 | 0.2944 | 0.2385 |
| DOW+DOM+City | 49626 | 11 | 0.3786 | 0.2536 |
| DOW+Month+State | 49708 | 12 | 0.4265 | 0.3163 |
| Gender+DOW+DOM+Month | 34320 | 13 | 0.4824 | 0.3403 |
| Age+DOM+Month | 34314 | 14 | 0.4913 | 0.3859 |
| Age+DOW+DOM | 34314 | 15 | 0.4922 | 0.2536 |
| DOW+DOM+Month | 84031 | 16 | 0.6021 | 0.2523 |
| DOM+Month+State | 49712 | 17 | 0.6538 | 0.4518 |
| Month+City | 49679 | 18 | 0.7323 | 0.4213 |
| Age+Gender+Month | 34221 | 19 | 0.7710 | 0.5134 |
| Age+Gender+DOW | 34314 | 20 | 0.8283 | 0.3644 |
| DOW+DOM+State | 49710 | 21 | 0.9299 | 0.5874 |
| DOW+City | 49695 | 22 | 1.0426 | 0.5337 |
| Gender+DOW+Month | 34323 | 23 | 1.1803 | 0.5785 |
| Age+Gender+DOM | 34314 | 24 | 1.1854 | 0.4863 |
| Age+Month | 34317 | 25 | 1.1890 | 0.6959 |
| Age+DOW | 34314 | 26 | 1.3016 | 0.5014 |
| DOM+City | 49702 | 27 | 1.4370 | 0.6483 |
| Month+State | 49714 | 28 | 1.4888 | 0.7904 |
| DOW+Month | 84037 | 29 | 1.4956 | 0.3500 |
| Gender+DOM+Month | 34323 | 30 | 1.6390 | 0.7516 |
| Age+DOM | 34314 | 31 | 1.7436 | 0.6139 |
| Gender+DOW+DOM | 34320 | 32 | 1.7510 | 0.2473 |
| DOW+State | 49711 | 33 | 1.8972 | 0.8795 |
| DOM+Month | 84038 | 34 | 1.9852 | 0.4095 |
| DOM+State | 49713 | 35 | 2.3530 | 0.9293 |
| Age+Gender | 34314 | 36 | 2.3682 | 0.7049 |
| DOW+DOM | 84036 | 37 | 2.4074 | 0.1546 |
| City | 49714 | 38 | 2.5801 | 0.8278 |
| Gender+Month | 34323 | 39 | 2.8122 | 0.8261 |
| Age | 34317 | 40 | 3.0170 | 0.7381 |
| Gender+DOW | 34323 | 41 | 3.0752 | 0.1621 |
| Month | 84040 | 42 | 3.1959 | 0.3626 |
| State | 49714 | 43 | 3.4373 | 0.8353 |
| Gender+DOM | 34323 | 44 | 3.5635 | 0.1806 |
| DOW | 84037 | 45 | 3.5817 | 0.0926 |
| DOM | 84039 | 46 | 3.9456 | 0.0975 |
| (No Context) | 84040 | 47 | 4.4164 | 0.0814 |
| Gender | 34323 | 48 | 4.4743 | 0.0952 |

Table 7: Best Predictor Combinations by Cosine Similarity

| Predictor Combination | Count | Cosine Similarity | | |
|---|---|---|---|---|
| | | *Rank* | *Mean* | *SD* |
| Gender+DOM+Month | 34323 | 1 | 0.5623 | 0.2220 |
| Gender+Month | 34323 | 2 | 0.5620 | 0.2218 |
| Age+Month | 34317 | 3 | 0.5615 | 0.2225 |
| Age+Gender | 34314 | 4 | 0.5614 | 0.2226 |
| Gender+DOW+Month | 34323 | 5 | 0.5614 | 0.2216 |
| Age | 34317 | 6 | 0.5612 | 0.2236 |
| Age+DOM | 34314 | 7 | 0.5608 | 0.2231 |
| Gender+DOW+DOM+Month | 34320 | 8 | 0.5607 | 0.2217 |
| Gender+DOM | 34323 | 9 | 0.5607 | 0.2222 |
| Gender+DOW | 34323 | 10 | 0.5606 | 0.2227 |
| Gender | 34323 | 11 | 0.5606 | 0.2227 |
| Age+Gender+DOM | 34314 | 12 | 0.5605 | 0.2220 |
| Gender+DOW+DOM | 34320 | 13 | 0.5605 | 0.2220 |
| Age+DOW | 34314 | 14 | 0.5604 | 0.2235 |
| Age+Gender+DOW | 34314 | 15 | 0.5600 | 0.2224 |
| Age+DOM+Month | 34314 | 16 | 0.5597 | 0.2225 |
| Age+Gender+Month | 34221 | 17 | 0.5595 | 0.2214 |
| Age+DOW+DOM | 34314 | 18 | 0.5586 | 0.2226 |
| Age+DOW+Month | 34308 | 19 | 0.5575 | 0.2217 |
| Age+Gender+DOW+DOM | 34314 | 20 | 0.5564 | 0.2216 |
| Age+Gender+DOM+Month | 34182 | 21 | 0.5562 | 0.2213 |
| Age+Gender+DOW+Month | 34170 | 22 | 0.5523 | 0.2207 |
| Age+DOW+DOM+Month | 34257 | 23 | 0.5490 | 0.2214 |
| Age+Gender+DOW+DOM+Month | 34053 | 24 | 0.5375 | 0.2213 |
| DOM+Month | 84038 | 25 | 0.4754 | 0.1927 |
| Month | 84040 | 26 | 0.4754 | 0.1927 |
| DOW+Month | 84037 | 27 | 0.4753 | 0.1926 |
| DOW | 84037 | 28 | 0.4752 | 0.1927 |
| DOW+DOM | 84036 | 29 | 0.4752 | 0.1926 |
| DOM | 84039 | 30 | 0.4752 | 0.1926 |
| (No Context) | 84040 | 31 | 0.4751 | 0.1927 |
| DOW+DOM+Month | 84031 | 32 | 0.4747 | 0.1924 |
| City | 49714 | 33 | 0.4576 | 0.1788 |
| DOM+City | 49702 | 34 | 0.4569 | 0.1787 |
| State | 49714 | 35 | 0.4565 | 0.1792 |
| DOM+State | 49713 | 36 | 0.4564 | 0.1792 |
| DOW+City | 49695 | 37 | 0.4564 | 0.1787 |
| DOW+State | 49711 | 38 | 0.4563 | 0.1791 |
| Month+State | 49714 | 39 | 0.4561 | 0.1791 |
| Month+City | 49679 | 40 | 0.4555 | 0.1786 |
| DOW+DOM+State | 49710 | 41 | 0.4554 | 0.1788 |
| DOM+Month+State | 49712 | 42 | 0.4544 | 0.1786 |
| DOW+DOM+City | 49626 | 43 | 0.4533 | 0.1783 |
| DOW+Month+State | 49708 | 44 | 0.4530 | 0.1781 |
| DOM+Month+City | 49588 | 45 | 0.4506 | 0.1781 |
| DOW+Month+City | 49494 | 46 | 0.4469 | 0.1775 |
| DOW+DOM+Month+State | 49644 | 47 | 0.4433 | 0.1758 |
| DOW+DOM+Month+City | 48996 | 48 | 0.4267 | 0.1748 |

Table 8: Best Predictor Combinations by Precision

| Predictor Combination | Count | Precision @ 20 | | |
|---|---|---|---|---|
| | | *Rank* | *Mean* | *SD* |
| Age+Month | 34317 | 1 | 0.3285 | 0.1635 |
| Gender+DOM+Month | 34323 | 2 | 0.3279 | 0.1640 |
| Gender+DOW+DOM+Month | 34320 | 3 | 0.3278 | 0.1634 |
| Gender+Month | 34323 | 4 | 0.3277 | 0.1642 |
| Gender+DOW+Month | 34323 | 5 | 0.3274 | 0.1639 |
| Age+Gender+Month | 34221 | 6 | 0.3274 | 0.1630 |
| Age+DOM+Month | 34314 | 7 | 0.3272 | 0.1629 |
| Age+DOM | 34314 | 8 | 0.3269 | 0.1623 |
| Age+Gender | 34314 | 9 | 0.3268 | 0.1630 |
| Age+Gender+DOM | 34314 | 10 | 0.3267 | 0.1624 |
| Age | 34317 | 11 | 0.3265 | 0.1616 |
| Age+DOW | 34314 | 12 | 0.3264 | 0.1626 |
| Gender | 34323 | 13 | 0.3263 | 0.1623 |
| Age+DOW+Month | 34308 | 14 | 0.3260 | 0.1625 |
| Gender+DOW+DOM | 34320 | 15 | 0.3259 | 0.1624 |
| Gender+DOM | 34323 | 16 | 0.3259 | 0.1626 |
| Gender+DOW | 34323 | 17 | 0.3258 | 0.1630 |
| Age+DOW+DOM | 34314 | 18 | 0.3256 | 0.1618 |
| Age+Gender+DOM+Month | 34182 | 19 | 0.3254 | 0.1619 |
| Age+Gender+DOW | 34314 | 20 | 0.3253 | 0.1621 |
| Age+Gender+DOW+DOM | 34314 | 21 | 0.3248 | 0.1613 |
| Age+Gender+DOW+Month | 34170 | 22 | 0.3233 | 0.1614 |
| Age+DOW+DOM+Month | 34257 | 23 | 0.3207 | 0.1613 |
| Age+Gender+DOW+DOM+Month | 34053 | 24 | 0.3135 | 0.1595 |
| Month | 84040 | 25 | 0.2757 | 0.1337 |
| DOM+Month | 84038 | 26 | 0.2757 | 0.1338 |
| (No Context) | 84040 | 27 | 0.2756 | 0.1334 |
| DOM | 84039 | 28 | 0.2753 | 0.1331 |
| DOW+Month | 84037 | 29 | 0.2748 | 0.1339 |
| DOW+DOM | 84036 | 30 | 0.2746 | 0.1332 |
| DOW | 84037 | 31 | 0.2745 | 0.1331 |
| DOW+DOM+Month | 84031 | 32 | 0.2744 | 0.1339 |
| State | 49714 | 33 | 0.2636 | 0.1235 |
| DOM+State | 49713 | 34 | 0.2633 | 0.1235 |
| City | 49714 | 35 | 0.2632 | 0.1238 |
| Month+State | 49714 | 36 | 0.2632 | 0.1232 |
| DOM+City | 49702 | 37 | 0.2626 | 0.1236 |
| DOW+State | 49711 | 38 | 0.2624 | 0.1231 |
| DOW+City | 49695 | 39 | 0.2624 | 0.1235 |
| Month+City | 49679 | 40 | 0.2619 | 0.1233 |
| DOW+DOM+State | 49710 | 41 | 0.2618 | 0.1228 |
| DOM+Month+State | 49712 | 42 | 0.2614 | 0.1227 |
| DOW+Month+State | 49708 | 43 | 0.2605 | 0.1223 |
| DOW+DOM+City | 49626 | 44 | 0.2603 | 0.1229 |
| DOM+Month+City | 49588 | 45 | 0.2588 | 0.1227 |
| DOW+Month+City | 49494 | 46 | 0.2563 | 0.1218 |
| DOW+DOM+Month+State | 49644 | 47 | 0.2546 | 0.1207 |
| DOW+DOM+Month+City | 48996 | 48 | 0.2457 | 0.1200 |

ity measures congruency of subject matter through the angular orientation of term vectors. KL divergence has the effect of penalizing predictor distributions that have many differences among low-frequency terms. Cosine similarity, however, has the opposite characteristic: vectors with strong similarity along a few axes can score more highly than vectors with lower similarity across more axes, sometimes due to sparsity and called "chance correlation" [95, 32].

Our results show that the effectiveness of many predictor combinations depends on the metric used; however, it is interesting to note that there are some predictor combinations that retain their accuracy. In particular, combining demographic information about the author with calendar information resulted in predictor combinations that were both strongly representative of the true distributions (i.e. low KL divergence) and good predictors by cosine similarity. For example, the combinations of "Age+Month," "Gender+DOM+Month," and "Age+Gender+DOM" maintained higher rank positions than the non-contextual predictor across all three metrics.

If term frequencies are assumed to follow Zipfian distributions, then there is strong similarity in the distribution of high-frequency words among authors of the same age or gender (i.e. high cosine similarity), but this similarity seems to diminish in the long tail (i.e. high KL divergence). This behavior means that adding more context may not always be beneficial, depending on the task. While adding more contextual cues generally seems to bring the predictor distribution closer to the true distribution, it may introduce noise that can confuse some common measurement techniques, such as cosine similarity. This situation is intrinsically similar to the statistical challenges associated with machine learning in high dimensionality [101, 25]. As dimensionality increases, classification often becomes more difficult due to diverging spectra and relative sample size. Dimension reduction and feature selection become vital to decrease misclassification errors.

The generally poor performance of the non-contextual predictor would seem to suggest that global distributions, such as might be obtained from Google's N-Gram corpus, may not always be appropriate as background probabilities, especially in scenarios where the composition of lower frequency terms is important. This chapter did not examine the probability distributions of bigrams or trigrams, however, so there may be a stronger relationship between the global distributions of higher-level n-grams or skip-grams.

Location-based predictors, such as city and state, and date metrics, such as day of the week or day of the month, appeared to provide more realistic probability distributions by KL divergence. While global vocabulary distributions may be sub-par, this result suggests that using word probabilities from a significantly different location, such as from the Wall Street Journal [86] or New York Times if the author is not from New York, may be counter-productive. Instead, it could be more appropriate to use distributions from a publishing source more geographically related to the author, such as a local newspaper, and perhaps even sub-divide the vocabulary into distributions for each month or day of the week. From a socio-linguistic perspective, city and state-wide word distributions would be strongly related because of shared regional characteristics, including local events, sports teams, political candidates, companies, weather, and geographical landmarks. These results also suggest that, given an author's language sample, it might be possible to examine the vocabulary distribution's low-frequency words to obtain a "signature" and discern where and when the author created the content.

A confounding factor in the Yelp corpus is the fact that there is a higher order context: Yelp reviews are primarily about businesses, which may change infrequently. So, it is possible that the low KL divergence of the location-based predictors is due to the natural groupings of different types of businesses in different cities or states, and thus different language necessary to describe them. For example, if there are a lot of Italian restaurants in a particular city, it is likely that descriptions may differ from a city in which there are many Mexican restaurants. This would seem to be confirmed by the TREC contextual challenge: in 2012, multiple contextual categories were provided, but the 2013 competition only provided location and search terms [21], perhaps indicating that location was one of the more dominant predictors.

We attempted to be pragmatic in our analysis of the corpora; however, it is possible that we may have discovered stronger similarities if we had implemented stemming and stop words. The lack of stemming and minimal usage of stop word removal resulted in discovering several word tokens that we were not originally aware of, such as various types of "smiley faces" (e.g. ^_^). Our data also confirmed the lack of punctuation as discovered by related work [104, 62]; instead, we saw extensive use of ellipses or no punctuation at all.

The results of this study should be considered within the context of the corpora examined. In particular, the two corpora examined in the study were diverse: the Blog Authorship and Yelp corpora were written for conceptually different audiences and for unspecified reasons. It is a well-understood phenomenon that people speak and write differently for different audiences and purposes [6], so even if there had been shared authors within these corpora, it is possible that their vocabulary distributions would have been different in each context. Values within "day of the month" were also artificially conflated into contiguous groups; it is possible that values within these categories are related across boundaries or in different ways that were not explored.

Additionally, not all of the contextual categories were provided by both corpora. Author demographics were only provided by the Blog Authorship corpus and location was only provided by the Yelp corpus. It might have been possible to infer gender from the names of the authors [17, 77], but that would have introduced more potential confounds and still not allowed for age comparisons. The locations available from the Yelp dataset also only describe the businesses that were reviewed, not the locations or origins of the reviewers. It is possible that people are more likely to write reviews for businesses close to where they live or frequently travel, and thus it could be assumed that authors have some relationship with the locations, but that theory could not be definitively established by the data provided.

There are higher level contexts that we did not explore, such as might be discovered by techniques like Latent Dirichlet Allocation (LDA), and some of the corpora provided further contextual information that we did not make use of in this study. The Blog Authorship corpus, for example, provides self-stated "career." It is possible that people in the same career share much of the same vocabulary, particularly on weekdays and during normal work hours. In combination with date and time-related predictors, this contextual category might have performed very well, however that possibility was not explored in this chapter. Additionally, authors in the Yelp corpus were tagged with a hyperlink leading to an "author information" page, from which further information about the author could potentially be found via data scraping and parsing of the website.

We chose to use freely available corpora to encourage reproduction and analysis of this work; however, it would be good to confirm our results with full ten-fold analysis or by running a similar study on larger corpora with more uniform contextual tagging. Although it is difficult to obtain and distribute such data due to licensing restrictions and privacy concerns, social networks like Twitter, Facebook, or Google+ could provide further evidence of the relative usefulness of different contextual language predictors.

## 3.8 summary

This chapter focused on the relative contributions of various contextual cues for predicting language usage. We compared global and contextual language distributions against true distributions to learn which combinations of contextual categories provided the closest approximations and most accurate predictions. The major findings are summarized as follows:

1. The non-contextual predictor, representing a global language distribution, was not among the best performing predictors by any of the three metrics.

2. For tasks that benefit from highly accurate probability distributions, such as speaker identification and other "fingerprinting" type tasks, our results suggest that training data should be categorized according to calendar date and further subdivided, if possible, based on author demographics (e.g. age and gender) or geographical region (e.g. city and state).

3. For tasks aimed at classification, clustering, and disambiguation, our results suggest that it is still beneficial to leverage calendar date and author demographics; however, variability in regional language patterns may reduce the efficacy of some standard similarity metrics.

In general, the inclusion of more context improved the similarity between predictor distributions and true distributions. Even information as categorically broad as the city, state, or day of the week was enough to significantly increase accuracy. It remains unclear whether more granular levels of location, such as might be provided by higher resolution GPS, would result in further improvements or diminishing returns. Although neither of the corpora in the current study allowed for combinations of age, gender, and city, it would also be worth examining whether more accuracy could be gained by combining author demographics with location.

The results of the present study have implications for speech recognition systems and assistive communication devices that benefit from language prediction to reduce fatigue and improve the efficiency of message formulation. This chapter provides a prioritized list of contextual predictors to help in addressing the cold-start problem and provide reasonable performance while user-specific data is gathered. Our results suggest that usage patterns based on author attributes and calendar information are highly informative and should be preferred over many other sources. Regional vocabulary patterns can also be more informative than global statistics, but should be considered carefully based on the specific task. Given that many newer systems are deployed on mobile platforms, it may be beneficial to leverage embedded sensor information relating to location and the individual's monthly or weekly schedule. These findings point to the importance of gathering an individual's date-specific and location-specific vocabulary distributions using continuous data collection and active learning, especially for pervasive assistive technology.

# 4 PERSONALIZED INTERACTION

## 4.1 OVERVIEW

Although an increasing number of new AAC systems are being designed for use with touchscreen technologies, such as Android or iOS tablets, many AAC users have concomitant upper limb motor impairments (e.g. tremors, spasms, or reduced mobility) that make using standard touchscreen technology difficult and frustrating. The work in this chapter explores alternative approaches to the standard "lift-move-touch" interaction sequence on current touchscreens. To help improve the accessibility of touchscreen technologies, we studied the movement patterns of 15 individuals with progressive neurological disorders and upper limb motor impairments. This chapter presents the quantitative results of our study, observations of functional compensation patterns, and the personal feedback from study participants. The results of this work are an evidence-based roadmap towards more personalized and adaptive touchscreen interfaces for current and potential AAC users.

## 4.2 MOTIVATION

Touchscreen technologies have rapidly increased in sensitivity and availability over the last decade. Most modern mobile devices are touchscreen systems that support multiple simultaneous touches and gestural interactions. The accessibility of these devices, however, has not kept pace with the general technological improvements. Accessibility features for people with upper limb motor impairments, in particular, are often limited to switch control, stored gestures, and adjustment of click timing. Users are often programmatically prohibited from toggling or adjusting sliding functionality. Users are also prevented from modifying the location, size, shape, and orientation of many buttons and toolbars.

It is difficult to quantify the touchscreen usage patterns, needs, and behavioral compensation of people with upper limb motor impairments because of the diversity of the population and available devices; however, there is a growing body of research in this area. Button sizes and spacing effects in layout-specific selection tasks have been compared between users with and without motor impairments [18], and researchers have obtained basic usage patterns through surveys [44] or by watching online videos [3]. Related work has demonstrated the potential advantages of swabbing (i.e. sliding) as a selection technique [117] and explored the effects of form factor on pointing tasks [28]. No study to date, however, has examined the combined effects of handedness and motor impairment on full-screen touch interactions.

Table 9: Attributes of Participants in Motor Movement Study

| Gender | Handedness | Stylus or Finger | Speech Impairment | Motor Impairment |
|---|---|---|---|---|
| M | Left | Finger | Mild | Moderate |
| F | Right | Finger | Mild | Mild |
| F | Right | Stylus | Mild | Mild |
| F | Left | Stylus | Mild | Moderate |
| F | MS-Left | Finger | Mild | Moderate |
| F | Right | Finger | Moderate | Moderate |
| F | Right | Finger | Moderate | Moderate |
| F | MS-Left | Stylus | Mild | Mild |
| F | Left | Stylus | Mild | Mild |
| M | Right | Finger | Severe | Severe |
| M | Right | Finger | Non-Speaking | Severe |
| F | MS-Left | Finger | Moderate | Mild |
| M | Right | Stylus | Mild | Mild |
| F | Left | Stylus | Mild | Moderate |
| M | Right | Finger | Non-Speaking | Severe |

In this chapter, we present the results of a controlled study comparing the touch behavior of left-handed and right-handed subjects with upper limb motor impairments as they performed full-screen tapping and sliding tasks.

## 4.3 APPROACH

We conducted a motor skills assessment study in which participants were asked to play a touchscreen game, called "MoGUI" (Motor Optimization GUI), that involved popping animated balloons by touching them. We recruited 15 adults (10 females and 5 males) from The Boston Home, a residential facility for people with progressive neurological diseases, especially MS, MSA, and muscular dystrophy (MD). All participants used wheelchairs and had some level of speech and motor impairment (Table 9). Participants were screened by a speech-language pathologist (SLP) to verify that they were current or potential AAC users, but had adequate hearing, vision, and cognitive abilities to fully consent and complete the tasks. The SLP also categorized their impairments and verified that all participants could interact with a touchscreen computer using their fingers, hands, or a stylus. They had a combined average age of 56 years, with a minimum of 35 and a maximum of 71. Seven of the participants were left-handed: four naturally and three due to MS. The remaining participants were right-handed.

The tablet computer used in the study was an Asus Transformer TF101 with a 10.1-inch diagonal display size at 1280x764 pixel resolution, running Android 4.4.2 with default settings. All participants were familiar with touchscreen tablets, but only 8 of them used one on a regular basis. Of these 8 participants, 7 used iPads and 1 used

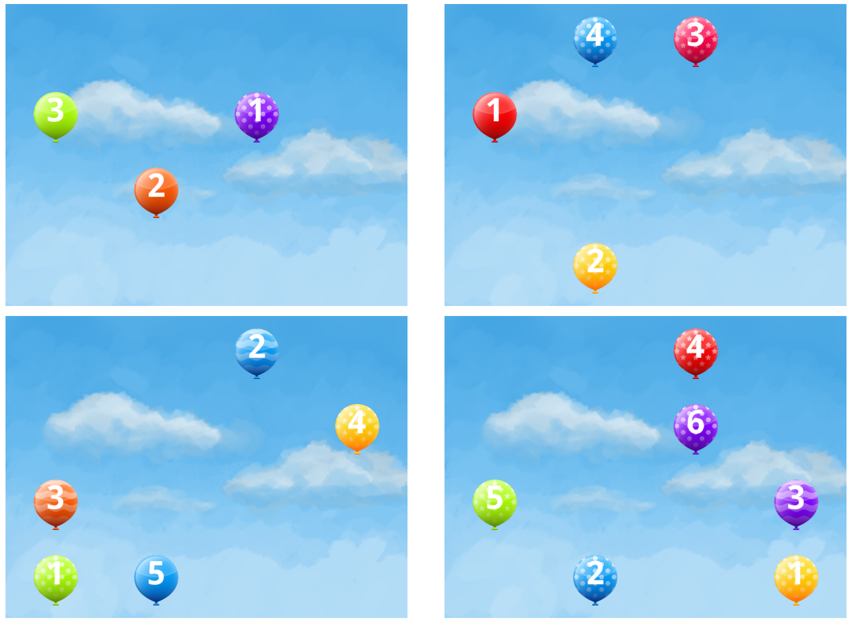



Figure 2: Screenshots of MoGUI

a Kindle. Although 6 of the participants indicated that they wanted to use a stylus, 4 of these participants had difficulties opening their fingers and requested that the stylus be placed in their hands. For these participants, the stylus served to separate their hand from the screen to prevent accidental touches from the non-pointing portions of their hands, such as their palms or knuckles.

When prompted for the most comfortable position to place the tablet, such that they could physically touch all areas of the screen, 9 users requested that the tablet be placed on a table in front of them at approximately a 45-degree angle. One participant asked for the tablet to be placed flat on the table. Two participants regularly used their tablets with wheelchair desk-mounts and requested that the study tablet be positioned in the same way. Two other participants requested that the tablet be placed in their laps; one of these participants requested that the tablet lie flat and the other requested that it be propped towards him with a rolled up towel. The last participant held the tablet against her body with one arm and used her other arm for interaction.

Each participant provided data during two sessions, separated by at least one full day of rest. Each session was approximately 30 - 45 minutes long and consisted of 10 levels, with 3 rounds per level. During each round, a series of balloon-shaped targets were displayed, labeled with consecutive numbers (Figure 2). One balloon was shown for each round during level one, two balloons were shown for each round during level two, three balloons during level three, etc. Thus, each session required a participant to touch 165 targets. Users were asked to touch each target balloon in ascending numerical order. Target balloons were 256x256 pixels in size and were randomly generated in one of 16 locations on the screen using a 4x4 grid. As soon as a balloon target was touched, it popped and disappeared from the screen; balloons touched out of order did not pop.

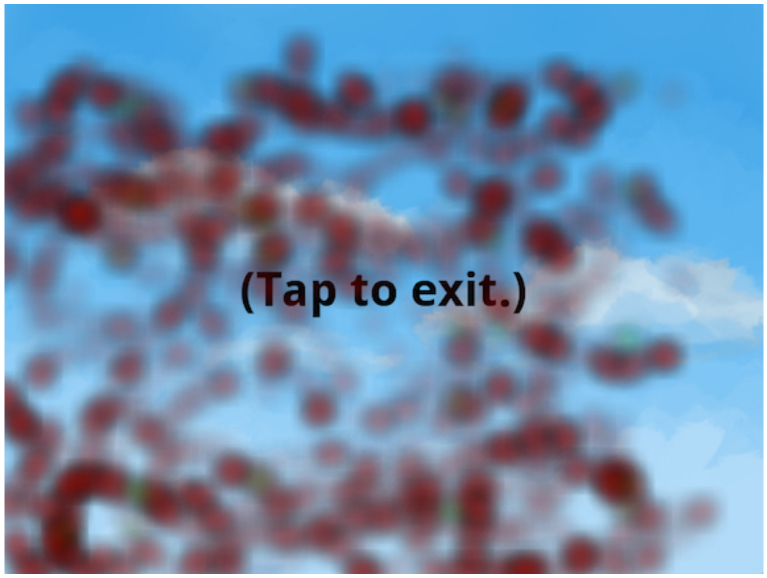



Figure 3: Example Interaction Heatmap Generated by MoGUI

In one of the two order-balanced sessions, participants were asked to use discrete movements (i.e. tapping or pointing) and avoid touching the screen except to hit a target. In the other session, participants were asked to use continuous motion (i.e. sliding or goal-crossing) and avoid disconnecting from the screen as much as possible. Users were offered a stylus, but were also allowed to use their fingers. Users were encouraged to hit all targets as quickly as possible, but also to rest whenever necessary.

All interactions with the touchscreen were recorded by the system, including the `on_touch_down`, `on_touch_up`, and `on_touch_move` events. After both sessions were completed, study participants were asked the following questions:

1. Did you find any areas of the screen easier or more difficult to reach than others?

2. Did you prefer tapping, sliding, a combination of both, or neither?

3. Were the balloon targets too big or too small?

4. What would you like to see improved in touchscreen tablets?

For the severely dysarthric and non-speaking participants, the questions were rephrased as multiple "yes or no" questions and combined with pointing:

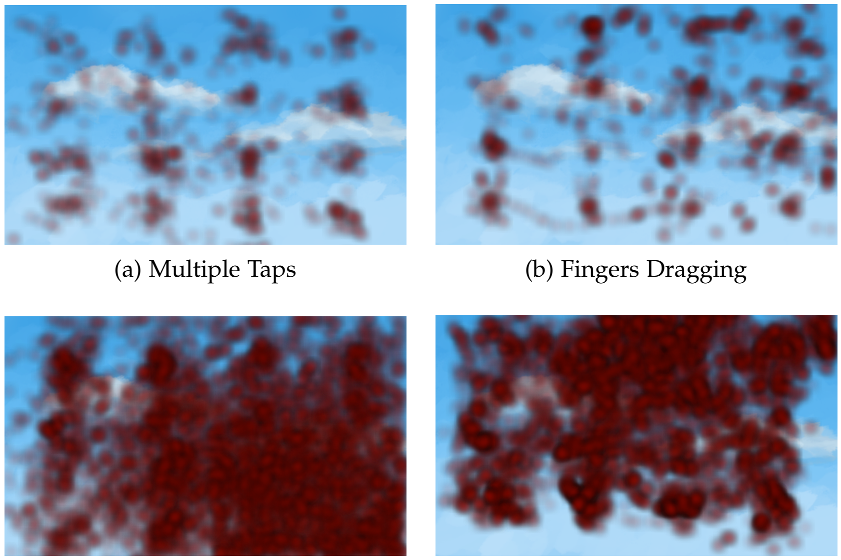
1. Was this area of the screen easy for you to reach?

2. Was this area of the screen difficult for you to reach?

3. Did you like tapping?

4. Did you like sliding?

5. Were the balloons too big?



(a) Multiple Taps (b) Fingers Dragging

(c) Hand Resting (d) Thumb Usage

Figure 4: Example Variability of Non-Target Tapping

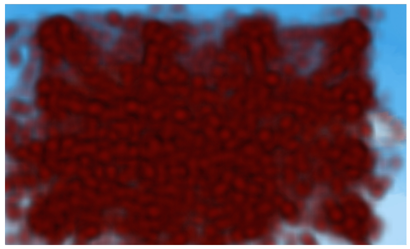
6. Were the balloons too small?
7. Would you like to use a touchscreen tablet in the future?
8. Would touchscreen tablets need to be changed before you could use them?

Study participants were also shown the resulting "heat maps" generated by MoGUI (Figure 3), depicting their touch interactions with the screen during each session.

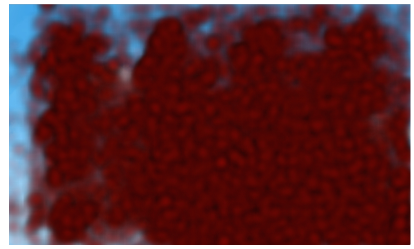
## 4.4 RESULTS

We observed high variability in motor profiles between participants, especially with regard to which locations on the screen had the highest accuracy or fewest misses (Figure 4) and which areas were fastest or easiest to reach (Figure 6). There were also significant differences between left-handed and right-handed participants; however, it is important to remember that 3 of the 7 left-handed participants were right-handed prior to the onset of MS. For left-handed participants, there were numerous accidental touches, often from other fingers or knuckles, on both the bottom and left sides of the screen (Figure 5); for right-handed participants, these accidental touches occurred on the bottom and right sides of the screen. The average speed-to-target of left-handed participants was 365 pixels per second compared to 429 pixels per second for right-handed participants. For each handedness, there were significant time delays when reaching for targets on the far side of the screen: there was approximately a one second difference, on average, between touching targets on the nearest versus the farthest side of the screen.

There were also significant differences in the average directional speeds for each handedness: for left-handed users, moving down and



(a) Left-Handed Users



(b) Right-Handed Users

Figure 5: Non-Target Touches by Handedness

to the left was approximately 1.5 times faster than moving up and to the right (Figure 7); for right-handed users, moving down and to the right was almost 4 times faster than moving up and to the left. In general, participants were much faster while returning their hands and arms to a natural resting position than moving away from that resting position.

There were speed differences between discrete movements and continuous motion in this study, regardless of participant handedness; however, they were not significant. The average speed-to-target of all participants while sliding was 407 pixels per second compared to 392 pixels per second while tapping. It is important to note, however, that we saw numerous accidental slides during designated tapping sessions, and vice versa. During designated tapping sessions, the average participant tapped 84% of the targets and slid into 16% of the targets; during designed sliding sessions, the average participant slid into 57% of the targets and tapped the remaining 43%. This behavior appeared to be caused by physical issues rather than confusion: we observed problems with friction, finger humidity, tremors, and spasms.

## 4.5 FEEDBACK AND OBSERVATIONS

At the end of the study, 3 participants said that they preferred tapping the screen, 5 participants preferring sliding over the screen, 5 participants preferred a combination of both input methods, and the remaining 2 participants had no preference. Ten participants mentioned that sliding required planning out a path ahead of time and required lifting your hand or arm to see the screen. Out of all the participants, 8 pointed out that sliding felt "faster" and "easier," but only for short distances. Over longer distances, participants said that there problems with skin friction and difficulties with stylus pressure. Additionally, this study involved arbitrary targets in random locations, which is fundamentally different than a user interface that is primarily static and can be learned over time.

One participant alternated hands numerous times during the experiment and explained that interacting with a touchscreen required him to use his shoulder muscles, which fatigued rapidly. Another participant rested frequently (approximately 1 of every 5 minutes), but

| | | | |
|---|---|---|---|
| 382 | 285 | 343 | 348 |
| 372 | 317 | 302 | 359 |
| 374 | 300 | 388 | 371 |
| 395 | 395 | 557 | 412 |

(a) Left-Handed Users

| | | | |
|---|---|---|---|
| 305 | 630 | 269 | 307 |
| 302 | 455 | 282 | 384 |
| 342 | 458 | 438 | 831 |
| 315 | 740 | 420 | 356 |

(b) Right-Handed Users

Figure 6: Mean Speeds-to-Target by Handedness (Pixels/Second)

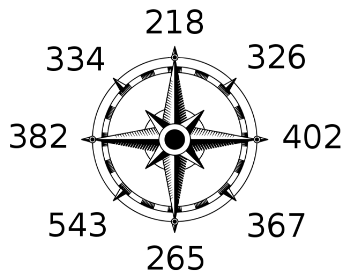
explained that it was because of his eyes, not due to upper limb fatigue. For this individual, focusing his vision to read numbers and moving his eyes to search for items on the screen was extremely tiring.

Certain areas of the screen were especially susceptible to accidental touches. In particular, the Android Action Bar, statically bound to the bottom of the screen, caused significant issues for 8 participants and resulted in repeated triggering of Google Now or window management functionality. Rather than starting their finger or stylus at the physical margin of the tablet and moving upwards, as they attempted to do naturally, participants were forced to try to control their arms enough to touch the middle of the screen and move downwards. Although there were similar problems when participants tried interacting with the top of the screen, these problems were observed less frequently.

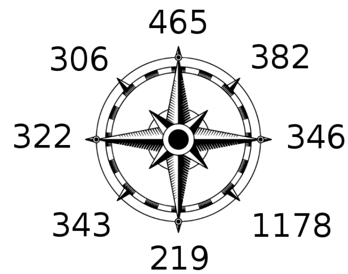
We observed a number of unusual hand positions. One participant was a former athlete; because of his larger physical size and musculature, tremors and spasms were especially severe in his upper limbs. For this participant, the tablet was placed on a table immediately in front of him, tilted at a 45-degree angle. The participant rested his entire hand on top of the tablet, and used his thumb to touch the screen; this interaction method made it very difficult to the participant to touch items at the bottom edge and lower corners of the screen. To reach these items, the participant needed to push the tablet farther away and rest his hand on the table, then try to raise his fingers upwards. This movement often resulted in hitting the Android Action Bar instead of the targets, activating Google Now functionality or switching between available windows.

Another participant requested that the tablet be placed in her lap, but not propped up and tilted towards her. Instead, the tablet rested on her thighs and occasionally shifted in angle to tilt slightly away from her, triggering the auto-rotate functionality and flipping the screen upside-down. This participant indicated that the position was how she normally interacted with her iPad, so auto-rotate was disabled in order to support her preferences.

Anecdotally, two participants mentioned that they would appreciate more confirmation dialogs on their tablets. Due to tremors and spasms, they said that they often activate a feature or perform an ac-




(a) Left-Handed Users




(b) Right-Handed Users

Figure 7: Mean Directional Speeds by Handedness (Pixels/Second)

tion by accident. They acknowledged that non-disabled users would probably find such repeated confirmation dialogs very annoying, but they would be valuable for users with motor impairments.

One of the non-speaking participants used a letter-based AAC system that primarily consisted of a QWERTY keyboard with word prediction. During the consent process with this participant, we made several observations about how he used his system. Because of severe motor impairments, this participant often made multiple accidental taps on each letter. He also missed the screen occasionally, perhaps due to vision impairments, when attempting to touch a button, resulting in omitted characters. This individual also rarely used the space bar to separate words, perhaps to increase communication speed, and never used the TTS functionality. Instead, conversation partners looked over his shoulder at the tablet screen and watched for confirmation while trying to guess his intended utterances.

## 4.6 SUMMARY

Current accessibility techniques group users with motor impairments together and assume uniform interaction over the entire touch surface. It is understood that users with motor impairments have different touchscreen behavior than non-disabled users; however, there is further diversity within the population of users with motor impairments. Our results show that functional compensation and attributes such as handedness have significant effects within this population.

For right-handed users, both the upper left and lower right corners of the screen required significantly more time and effort to reach; for left-handed users, this difficulty was associated with the upper right and lower left corners of the screen. Additionally, all study participants had significant amounts of unintentional interaction near the edges of the screen closest to their primary hand: for right-handed users, this was the right and bottom edges of the screen; for left-handed users, this was the left and bottom edges of the screen. Unfortunately, current touchscreen interfaces have essential system functionality located in almost all of these areas, especially the top and bottom edges of the screen. Because these system functions are often activated by sliding gestures, our results show that they are easily triggered by users with motor impairments.

Simply shifting the positioning of the tablet relative to the user is probably insufficient: users would be unable to reach all parts of the screen. Although some of our results could suggest that the study tablet was too large, users may benefit from a "safety margin" around their particular device, or on configurable sides, to mitigate accidental touches. System functionality should also be customizable in style and location: statically binding behavior to the top or bottom of the screen can be problematic for many users.

There appeared to be an optimal movement area for each user, strongly correlated to both handedness and tablet positioning. For most users, the shape of this area was an arc, approximately 3 - 4 inches wide, that could be found by fixating the user's elbow and rotating his or her hand across the screen. We observed reduced performance when attempting to hit targets outside of this arc, possibly because our study participants all used wheelchairs and positioned their elbows on the arm rests. Hitting targets outside of these areas required users to depart from a comfortable, homeostatic position in order to lift their elbows off of the chair.

There may be tangible advantages to departing from grid-based button positions and statically located system functionality, especially for users with upper-limb motor impairments. For optimal performance, or even acceptable performance in many cases, touchscreen interfaces may benefit by allowing users to relocate buttons away from the screen edges and closer to optimal touch areas. Users should also be allowed to toggle or relocate sliding and swiping gestures, especially for essential system functionality. Finally, the results of our study suggest that sliding to arbitrary targets is not significantly faster than tapping; however, users with motor impairments may benefit from systems that are able to combine the benefits of both movement styles, especially given that many users encounter difficulties when restricted to only one interaction technique.

# Part II

# APPLICATION

Summary of interface designs and prototypes:

1. RSVP-iconCHAT, a semantic approach to icon-based message construction that requires only a single input signal.
2. SymbolPath, an icon-based AAC system that supports continuous motion input and unordered superset selection.
3. DigitCHAT, a low-footprint, letter-based AAC system that supports utterance generation at conversational speeds.

# 5 RSVP-ICONCHAT

## 5.1 Overview

Many individuals with especially severe mobility and language constraints require icon-based AAC systems controlled by switches. Single-switch AAC systems are typically simplifications of multi-array AAC systems that share elements of the same interface layout, but support some form of scanning, such as linear or row-column. From a development standpoint, this conversion technique makes it easy to transform almost any AAC system into a single-switch system; however, it means that most of these systems were originally designed for users with much greater mobility. The purpose of this work was to design an icon-based AAC interface specifically for use with binary signals, such as switches. A usability study was conducted with both non-disabled adults as well as adults with speech and mobility impairments to determine performance bounds and observe individual use cases. Results indicated similar learning curves for both groups and promising performance characteristics for the target population. These results have immediate applications to the design of icon-based AAC and implications for mobile, icon-mediated communication platforms.

## 5.2 Motivation

Depending upon their mobility impairments and language constraints, many AAC users require icon-based systems controlled by switches [134]. Current single-switch AAC systems are typically simplifications of multi-array AAC systems and display a complex array of vocabulary on the screen, organized into a navigation hierarchy based on categories. To increase the size of the vocabulary on these systems, the screen size must be increased, the button sizes must decrease, or the navigation hierarchy must become more complex. Each of these approaches has its limitations. An increased screen size reduces mobility of the system, smaller buttons are more difficult to view, and complicated navigation hierarchies require more time and effort to find the target button and increase the likelihood of confusing the user. Additionally, when these interfaces are used with scanning, users must visually locate their target button from among the many options on the screen. People who use single-switch AAC systems often have extremely limited physical mobility or control, making it difficult to repeatedly perform the necessary head, neck, or eye movements when attempting to locate target items [78].

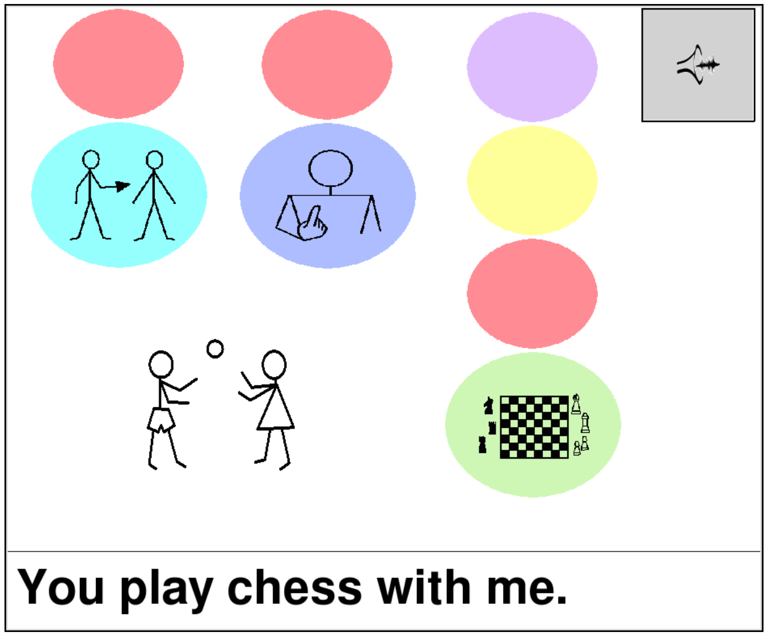



Figure 8: Functional Elements of the RSVP-iconCHAT Interface

## 5.3 APPROACH

Our single-switch AAC interface, called RSVP-iconCHAT, aims to minimize the amount of head, neck, and eye movements required to efficiently control the system. RSVP-iconCHAT was designed to be robust enough to function with a brain-computer interface (BCI), as well as conventional access methods, such as sip-and-puff devices, eye-blink detectors, surface electromyography (EMG), or physical switches. To that end, we leverage a technique called rapid serial visual presentation (RSVP), in which the user fixates on a relatively stable location while different images are displayed in that location, one at a time. RSVP originates from the field of psychology and has been used successfully to control letter-based AAC systems [83].

To leverage RSVP, our interface focuses the user's attention on the message being constructed instead of displaying all of the available vocabulary. To demarcate different visual fixation areas, messages are represented as semantic frames. Semantic frames are a product of case grammar theory, which asserts that the main action, or verb, is the central component of a message [27]. Each message can be expressed as a formulaic frame for which certain semantic roles are understood and expected. For example, the frame for the verb "to give" might require, at a minimum, an actor that does the giving, a participant that receives the gift, and an object that can be given or received. Semantic frames, such as obtained from WordNet [76], can be used to constrain relevant roles for a given action, and these roles can then be populated with appropriate concepts to generate complete utterances.

In RSVP-iconCHAT, each message is subdivided into semantic roles (e.g. actor, action, participant, and object) and applicable vocabulary options are displayed using RSVP within each semantic role (Fig-

Figure 9: Subset of Single-Action Picture Cards

ure 8). This design uncouples required screen real estate and physical movement from the vocabulary size and instead ties them to the length of the desired message. For more advanced or more mobile users, the number of available semantic roles can be increased, enabling users to create longer and more complex messages; for beginning users, or those with severely reduced mobility, the number of roles can be decreased to enable the creation of simpler messages with the same vocabulary.

To construct a message using the RSVP-iconCHAT approach, users first select the desired verb or action. Once an action has been selected, the corresponding semantic frame is displayed with semantic roles such as actor, actor modifier, participant, participant modifier, object, object modifier, quantity, and possessives. These roles are displayed as a set of fillable slots that are spatially organized around the verb. Each semantic role is then highlighted sequentially. Once a role has been selected, icons that can fulfill that role are displayed via RSVP and users can select a desired icon to populate the role. After an icon has been selected for a given semantic role, other roles are highlighted sequentially to allow users to populate as many roles as desired, and in any order. Articles (e.g. "a," "an," "the") and prepositions (e.g. "in," "of," "to") are automatically inserted to efficiently generate grammatically complete messages. At any point during message construction, users can select the "command field" to perform conversational actions (e.g. "speak" or "clear" the current message). Selecting a "speak" command, for example, might send the message to an integrated TTS system, clear the current message, and prompt the user to begin constructing a new message.

## 5.4 METHOD

We conducted a usability study involving a constrained message elicitation task for the purpose of exploring how potential users would interact with and respond to the interface. After a brief demonstration and training period, participants were shown a series of 30 picture scenes depicting simple actions (e.g. a boy drinking milk, a man combing his hair, and a woman reading a book) and asked to use our

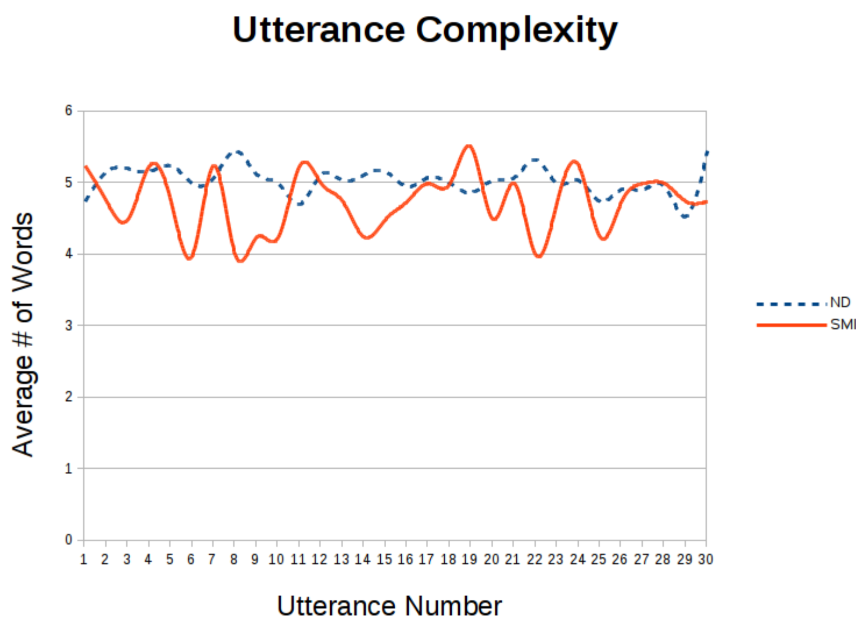



Figure 10: Mean Sentence Lengths with RSVP-iconCHAT

prototype RSVP-iconCHAT interface to create a sentence describing each scene (Figure 9). The order of the picture scenes was randomized across participants in order to observe behavior as users became more familiar with the system. Participants were directed to construct sentences that were as detailed as necessary such that, if the picture cards were shown to another person, that person would be able to match the appropriate description with the scene.

Each experimental session was conducted in one 60 - 90 minute block per participant, and all sessions were conducted in a sound-treated acoustic booth. Each participant was seated in a chair, or personal wheelchair, facing a computer screen. The space bar of a standard QWERTY keyboard was designated as the switch mechanism, and the RSVP process was configured to show images in alphabetical order using a timing mechanism with a starting speed of 700 milliseconds per image; however, participants could increase or decrease the speed in increments of 100 milliseconds per image. Participants were encouraged to change the RSVP speed whenever and however they preferred, either by pressing the up and down arrows on the keyboard or by requesting it verbally. The icon set, or vocabulary size, consisted of 106 items preselected for their relevance to the picture scenes and tagged within each of 8 possible semantic roles. After each session, participants were asked to provide qualitative feedback via an informal interview.

Two groups of users were recruited: non-disabled (ND) users to provide a theoretical upper bound on performance, and users with speech and motor impairments (SMI) to provide a realistic evaluation from the target population. For the group of ND users, we recruited 24 English-speaking adults from the greater Boston area, with no declared speech, language, hearing, or cognitive impairments (10 males and 14 females; mean age 24 years; age range 19 - 43). On average, each of these participants had approximately 3 years of formal education following high school and spent approximately 11 hours per week using a computer. The ND users did not have prior exposure or experience with AAC devices.

For the group of users with SMI, we recruited 4 additional English-speaking adults from the greater Boston area (2 males and 2 females; mean age 41; age range 33 - 56). On average, each participant had approximately 4 years of formal education following high school and spent approximately 15 hours per week using a computer. Two of these participants (P1 and P2) had mild motor impairments; two (P3 and P4) had moderate-to-severe motor impairments. All of these participants used wheelchairs, except for P1 who used a walker. P1 and P2 had experience with AAC devices, but used unaided communication on a normal basis. P3 used both unaided communication and switch-based AAC. P4 was unable to use existing AAC systems and required the assistance of a caregiver to communicate.

## 5.5 results

Theoretically, the open-ended design of the task allowed for the possibility of users creating nonsensical sentences; however, in practice, there were no such instances. Because our prototype implementation required that every sentence contain at least a verb, the short possible sentence was one word in length. On average, both the ND participants and the participants with SMI created sentences consisting of 5 words, excluding articles and prepositions that were automatically inserted by the system (Figure 10). Thus, users selected a verb and an average of 4 additional icons to construct descriptions of each picture scene. In fact, the participants with SMI created slightly more complex sentences, up to 6 additional words, on at least 2 occasions throughout the study.

In terms of message construction speed, both groups of users showed similar learning curves, with the ND group achieving a final speed approximately 1.5 times faster than the group with SMI (Figure 11). The average time for constructing each of the last five sentences was 70 seconds for the ND users and 107 seconds for the users with SMI.

If users populated a semantic role more than once, even if they selected the same icon or cleared the role of any value, it was considered a self-correction. This metric was used to probe fatigue and learnability of the system. On average, ND users changed or deleted 1 word per sentence before submission, compared to an average of 2 word changes or deletions per sentence for the participants with SMI (Figure 12).

During the study, ND users adjusted the RSVP speed an average of 10 times per sentence (Figure 13), returning to an average ending speed of approximately 700 milliseconds per image (Figure 14). In contrast, users with SMI adjusted the RSVP speed an average of 9 times per sentence for the first 5 sentences and an average of once

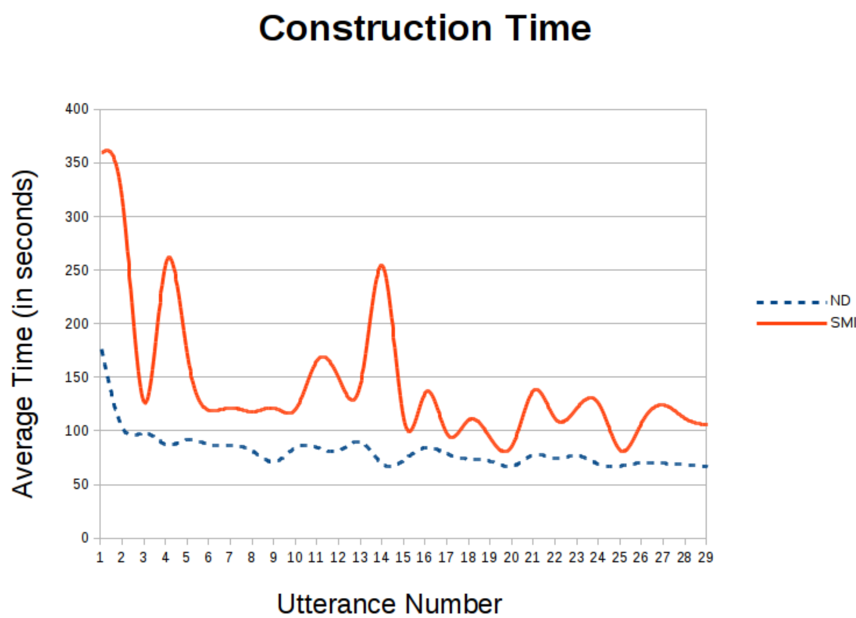



Figure 11: Mean Message Construction Times with RSVP-iconCHAT

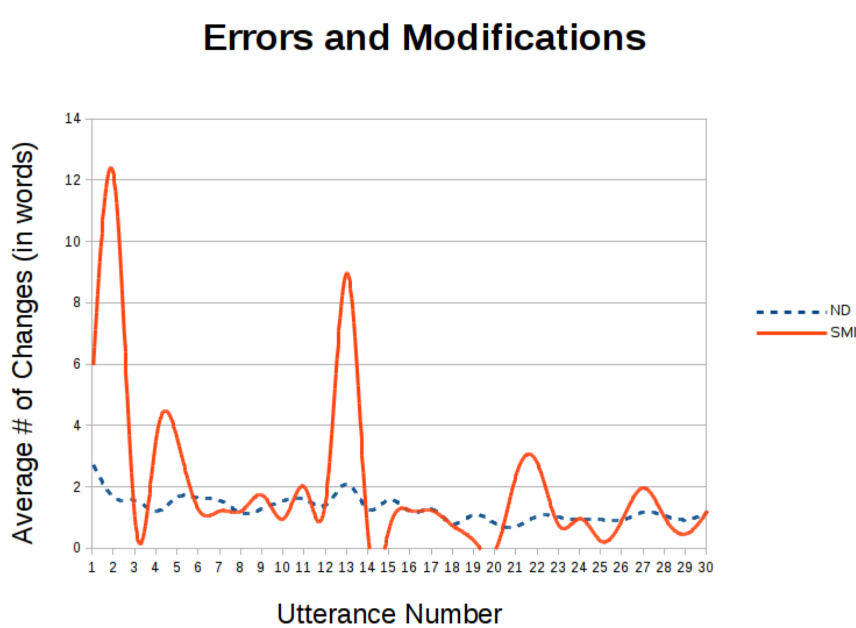



Figure 12: Mean Number of Self-Corrections with RSVP-iconCHAT

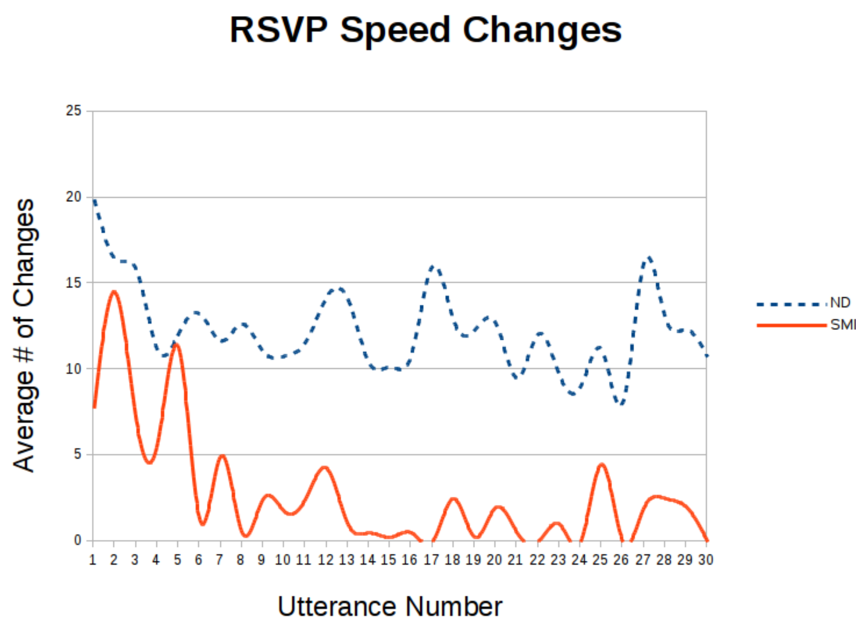



Figure 13: Mean Number of RSVP Adjustments with RSVP-iconCHAT

per sentence for the remaining 25 sentences, returning to an average ending speed of 1200 milliseconds per image.

## 5.6 DISCUSSION

This study examined user behavior while composing messages with the RSVP-iconCHAT interface and a single switch mechanism. The aim was to assess the learnability and ease-of-use of the system. Icon-based message construction via RSVP proved to be learnable within less than 30 minutes for both user groups. Users were able to construct messages of 4 - 7 words in approximately 1 minute, which is faster than some traditional letter-based systems [134], but users were unable to surpass the performance of conventional icon-based systems [78].

The results of our study suggest that expressiveness and generativity are not necessarily compromised by limiting selection tasks to a single key. In fact, both user groups constructed relevant sentences that were an average of 5 selected words in length. Examples of constructed messages included: "an old woman knitting a sweater," "a small child drawing a house," and "a man talking on a blue telephone with his friend." Although this study did not replicate the social pressures of realistic conversation rates, these sentences are longer and more complete than the simple 2 - 3 word sequences documented using some traditional icon-based systems [113].

Frequently changed RSVP speed throughout the course of the study suggests that users may have been unsatisfied with a constant presentation speed and may have wanted to skip ahead to specific roles or icons. For ND users, the average of 10 changes per sentence suggests that participants increased the speed 5 times and then decreased the

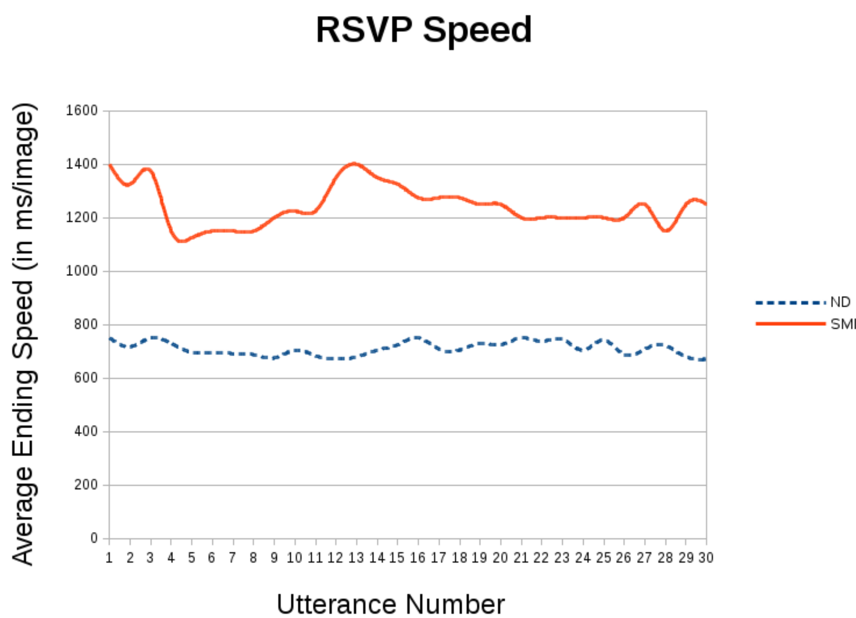



Figure 14: Mean Ending RSVP Speeds with RSVP-iconCHAT

speed 5 times, possibly to skip through a large number of undesirable words; however, this behavior was not displayed by the users with SMI (Figure 13). Although it is possible that the users with SMI found a comfortable speed within the first few sentences, it may have also required too much effort to change the RSVP speed more often, especially for those with moderate-to-severe motor impairments.

Users with SMI appeared to prefer an RSVP speed approximately 1.7 times slower than ND users, yet it is possible that they may be comfortable with faster RSVP speeds for other input modalities. For example, two users (P3 and P4) indicated they could have constructed messages more quickly if the interface were integrated with a sip-and-puff device. Given that both ND participants and participants with SMI converged to consistent ending RSVP speeds, their respective presentation rates (Figure 14) may be appropriate defaults for physical input modalities, such as button presses.

Two of the participants with SMI (P1 and P2) explored almost the entire vocabulary approximately halfway through the experiment. Additionally, a spike in self-correction for these users, at approximately sentence 13 (Figure 12), may indicate that they were exploring more expressive possibilities and testing the boundaries for sentence complexity. This phenomenon was not observed with ND users, possibly indicating different preferences between the two groups when familiarizing themselves with new communication interfaces. Self-corrections may also be explained by mistaken selection of a word due to slow motor movement, which would have been a sustained problem for the users with SMI, even as familiarity with the system increased.

While quantitative measurements of fatigue or cognitive load were not collected, qualitative feedback indicated that ND users felt “fidgety" and "impatient" at having to wait for a desired icon to be displayed, but almost all users commented that the interface was "simple" and "easy to use." One user with SMI (P1) also expressed impatience at having to wait for the target icon; however, the other three users with SMI did not indicate any similar frustration. All 28 participants noticed and favorably commented on the fact that the RSVP-iconCHAT approach did not require them to capitalize words, conjugate verbs, or provide articles and prepositions. Two of the users with SMI (P1 and P3) remarked that they had not seen an existing AAC system with similar functionality, and several ND users asked if there were a way to enable this feature in their current mobile devices.

## 5.7 SUMMARY

Many individuals with severe speech and motor impairments use icon-based AAC systems with switches; however, these systems often require larger screens, use complex navigation hierarchies, or necessitate repetitive head, neck, and eye movements. We aimed to design an alternative to conventional icon-based AAC systems that would require less screen real estate, yet still be easy to navigate and allow for sufficiently large vocabularies. RSVP was leveraged to display icons and reduce the required motor control to a single action. Furthermore, RSVP was combined with semantic frames to segment the screen into multiple fields and place the burden of search on the system rather than the user. By organizing vocabulary into semantic roles, rather than lexical categories, the display requirements of this approach are not tied to vocabulary size, but to the number of semantic roles necessary to construct a desired message.

The usability study suggests that an RSVP approach to icon-based message construction is a viable option for users with severe speech and motor impairments. Given that both cohorts of study participants were unfamiliar with the RSVP-iconCHAT approach, their performance should be considered as a reasonable lower bound that can be expected to improve with practice.

The RSVP-iconCHAT design has important implications for mobile devices that have small screens and a limited number of buttons. Depending on the complexity of the desired message, the number of semantic roles can be chosen to match the available display space of a given mobile device. All search and prediction tasks can be delegated to the system, requiring only a single reliable selection mechanism for control. While the minimal control requirements are a single binary signal, as in the conducted usability study, control over the RSVP process can be expanded to include directional control of the display sequence or even the ability to modify RSVP speed.

Our prototype implementation of RSVP-iconCHAT accepts keyboard entry or mouse clicks, but the design can be configured to work with eye blinks, muscle twitches, brain waves, or any other input that can be discretized into binary form. This interface is potentially beneficial for users with profound impairments, such as those with locked-in syndrome who require EMG or BCI solutions. Because many EMG and BCI systems provide a single output signal, and RSVP-iconCHAT requires only a single input signal, integrating such signaling methods is feasible and likely to be successful. Once the need for a voluntary motor response is removed, natural language processing and machine learning could be used to dynamically reorder the sequences of suggested semantic roles and associated icons, further increasing communication speed.

# 6 SYMBOLPATH

## 6.1 OVERVIEW

Icon-based AAC systems typically present users with arrays of icons that are sequentially selected to construct utterances, which are then spoken aloud using TTS. For touch-screen devices, users must lift their finger or hand to select individual icons and avoid selecting multiple icons at once. Because many individuals with severe speech impairments have concomitant limb impairments, repetitive and precise movements can be slow and effortful. The work in this chapter aims to enhance message formulation ease and speed by using continuous motion icon selection rather than discrete input. SymbolPath is an overlay module that can be integrated with existing icon-based AAC systems to enable continuous motion icon selection. Message formulation using SymbolPath consists of drawing a continuous path through a set of desired icons. The system then determines the most likely subset of desired icons on that path and rearranges them to form a meaningful and grammatical sentence.

## 6.2 MOTIVATION

Many individuals with speech impairments severe enough to preclude spoken communication also have accompanying limb impairments that must be considered when designing assistive communication interfaces [70, 59]. Icon-based AAC systems offer the potential for faster and less effortful message formulation compared to letter-based systems [106] and thus are often used by individuals with compromised motor function; however, manual methods of icon selection on current icon-based AAC devices require precise and discrete movements that hinder communication rate and ease. Additionally, the complex and repetitive nature of discrete movements can further contribute to fatigue. Several letter-based approaches to continuous selection have demonstrated commercial success (e.g. Swype, SlideIT, TouchPal, and ShapeWriter [51]), but no such approaches currently exist for word-based or icon-based formulation. This project aims to enhance message formulation ease and communication rate by combining continuous motion icon selection with a free-order language model.

## 6.3 IMPLEMENTATION

SymbolPath is implemented in Python as an overlay module for traditional icon-based AAC systems. A simple single-layer array serves as the interface for the current work. The top row is dedicated to

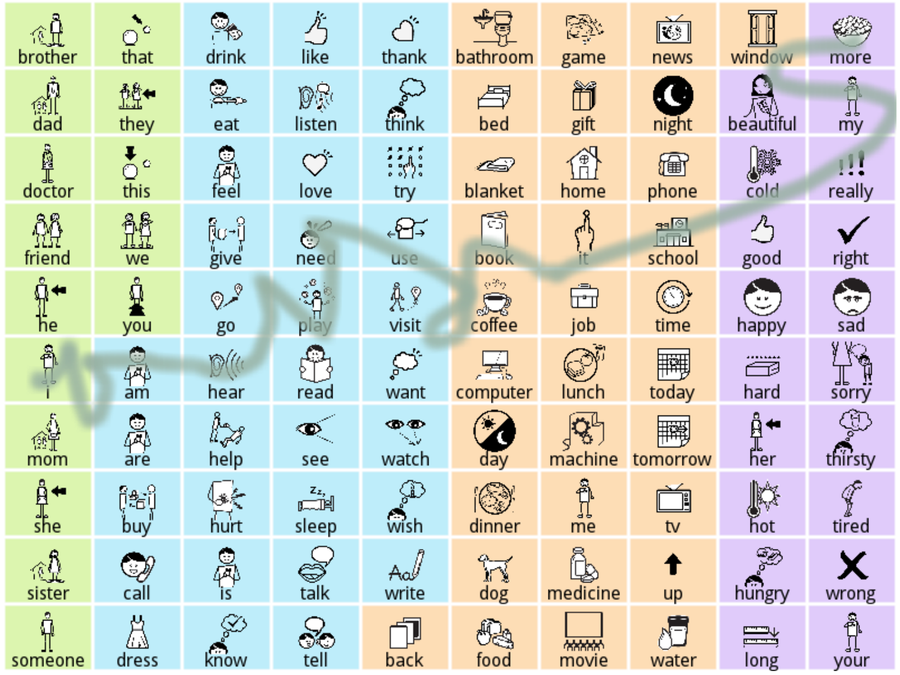



Figure 15: Construction of "I Need More Coffee" with SymbolPath

displaying the message being formulated and the remainder of the interface is arranged as a grid of candidate icons (Figure 15). Icons are grouped based on lexical roles: actors, verbs, objects, and modifers. Icon groups are color coded and arranged from left to right to mirror the subject-verb-object syntax in English. To formulate a message, users create a continuous path through a set of desired icons. To further reduce the physical demands of message formulation, the order of icons on the path is not constrained by syntax: users can select icons in close physical proximity rather than in syntactical order. The only requirement is that a continuous path be drawn through all desired items without breaking contact with the interface. During message formulation, the treaded path is displayed for feedback. Once the user breaks the path or enters the message formulation window, the language module attempts to concatenate a meaningful and syntactically accurate utterance from the set of selected icons. For example, a user might create a path traversing through the target words "need," "coffee," "I," and "more," as well as intermediary icons (e.g. "give," "book," and "window"), which are then pruned and reordered to generate the syntactically complete and most likely message "I need more coffee." The text-to-speech synthesizer then voices the message. SymbolPath is compatible with any input modality that can provide a continuously varying analog signal such as a stylus, mouse, joystick, or laser pointer.

Two major issues need to be resolved in order to enable continuous motion icon selection: (1) superset pruning, because the user's path may include both target elements and bystander elements, and this superset must be pruned to yield the most likely desired candidates; and (2) syntactic reordering, because the user may have selected icons

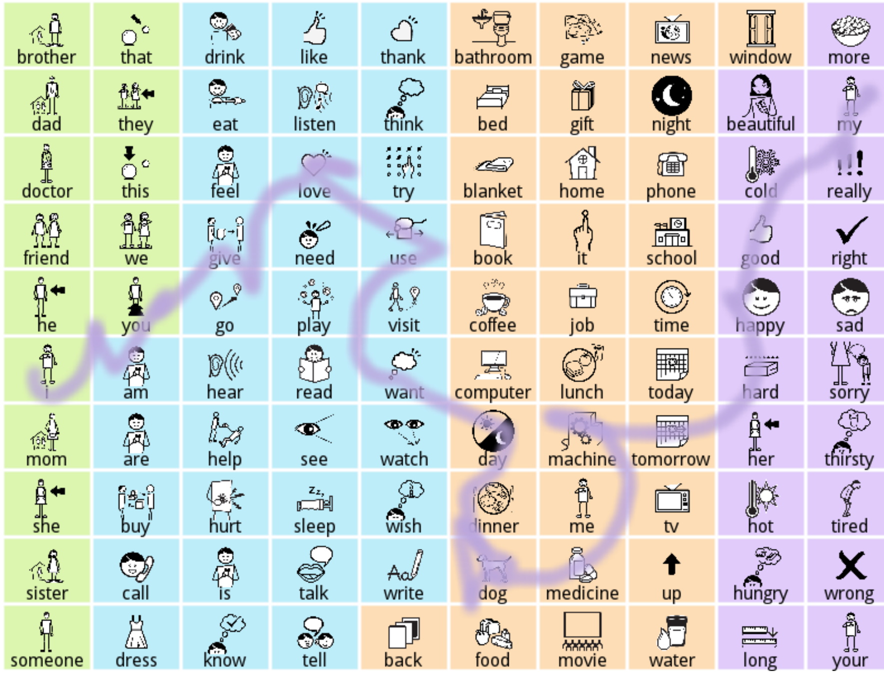



Figure 16: Construction of "I Love My Dog" with SymbolPath

in an unordered way and the system must reorder those icons in the proper syntax of the target language.

Semantic disambiguation is required for situations in which removing or reordering words could dramatically alter the meaning of the potential message. SymbolPath relies on a combination of semantic frames, semantic grams, and physical characteristics of the path to generate a prioritized list of potential utterances. Although the demonstration version automatically selects the most likely utterance to enhance communication speed, it can also display the list of potential utterances for user verification prior to speech generation.

### 6.3.1 *Semantic Frames*

Fundamental to the design and functionality of SymbolPath is the use of semantic frames [27], in which the predicate or verb of an utterance is the central element of a frame that can be filled by a set of relational items [85]. Thus, SymbolPath generates syntactically complete utterances by relying on the semantic frames of predicates in the selected path. Because each icon group is associated with a set of possible syntactic and semantic roles, the superset of selected icons is pruned by assessing subset probabilities within a given semantic frame. This approach provides a rudimentary solution to the issue of syntactic reordering, but does not address the issue of semantic disambiguation, especially with regard to assigning statistical values to potential utterances.

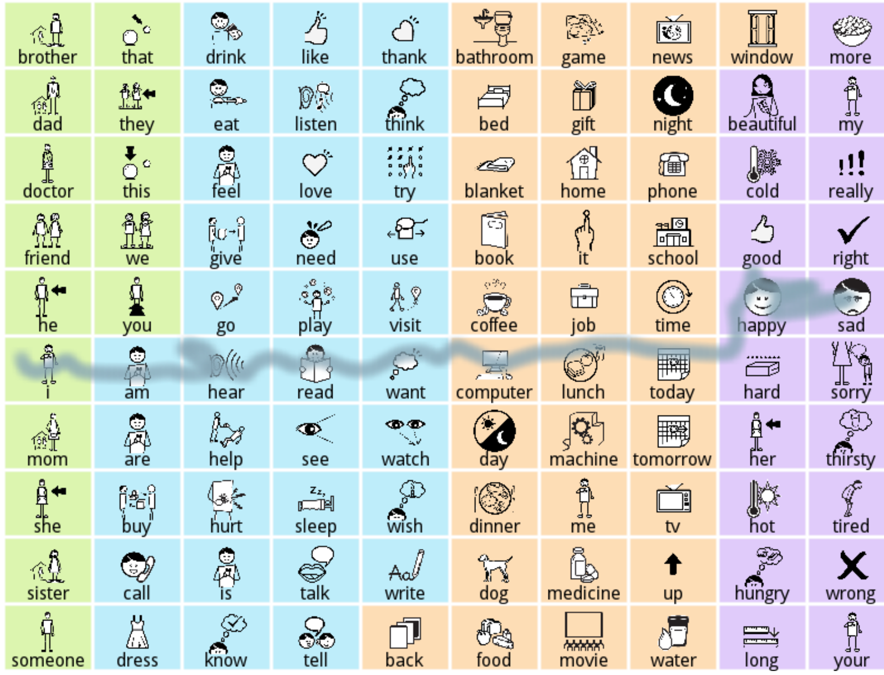



Figure 17: Construction of "I Am Happy" with SymbolPath

### 6.3.2 *Semantic Grams*

To prioritize the list of potential utterances, SymbolPath leverages prior work in the areas of subset completion and non-syntactic prediction [125]. Specifically, semantic grams, or sem-grams, are used to assign each potential utterance a value that corresponds to the probability of that combination of words appearing in a sentence together, regardless of order. Semantic ambiguity is not a concern because lexical roles are specified for each icon based on its grouping.

### 6.3.3 *Path Characteristics*

In addition to the probabilities of each potential utterance based on its semantic coherence, the physical characteristics of the path are also considered. Once the list of potential utterances has been prioritized semantically, the rankings are adjusted based on the two-dimensional collision space of the continuous motion path and each icon's surface area. Icons that collided with a larger area of the user's drawn path are assigned a greater likelihood than icons that were only marginally on the drawn path.

## 6.4 DISCUSSION

SymbolPath does not currently support complex utterances that contain multiple verbs (e.g. "I like to play baseball"), utterances that contain multiple actors and participants (e.g. "I like to play chess with my brother"), or utterances that make extensive use of modifiers (e.g. "I really drank that huge soda too quickly"). Although many of these situations can be supported through the use of semantic tagging, the goal of SymbolPath is to incorporate automated solutions to these problems. One potential approach is to supplement sem-gram statistics with corpus-based frame statistics in order to determine probabilities for each semantic frame and its arguments. While large corpora of AAC messages are unavailable, there have been recent efforts to simulate corpora that may be useful for obtaining such frame statistics [116]. Additionally, each user's message formulation history may be used to automatically refine the language model between sessions. Future work on SymbolPath may also include smoothing of the physical path to accommodate users with hand or arm tremors, as well as a calibration mode to detect each user's movement preferences and adjust the path's physical characteristics accordingly.

# 7 DIGITCHAT

## 7.1 OVERVIEW

While almost all AAC users have speech impairments that preclude the use of verbal communication, they may also have varying levels of vision or motor impairments, perhaps due to age or the particular nature of their disorder. Speed, expressiveness, and ease of communication are key factors in choosing an appropriate system; however, there are social considerations that are often overlooked. AAC systems are increasingly being used on mobile devices with smaller screens, in part because ambulatory AAC users may feel uncomfortable carrying around large or unusual machines. DigitCHAT is a prototype AAC system designed for fast and expressive communication by literate AAC users with minimal upper limb motor impairments. DigitCHAT's interface was designed to be used discretely on a mobile phone and supports continuous motion input using a small set of visually separated buttons.

## 7.2 MOTIVATION

Depending upon the nature of their disorders, many AAC users may have accompanying motor impairments, such as tremors or reduced mobility of their hands and arms [36, 59]. They may also have vision impairments that make it difficult to see small font sizes. AAC systems that operate on small mobile devices often use on-screen keyboards that were not designed for people with bigger hands or people with upper limb motor impairments. These keyboards usually occupy less than half of the available screen real estate and have small buttons that are positioned adjacent to each other. Many elderly users, much less users with diagnosed motor or vision impairments, have difficulty with these keyboards because of the button and font sizes [79]. Additionally, these systems tend to focus on the creation and use of stored utterances instead of real-time composition, unnecessarily reducing the flexibility of conversation. The work in this chapter aims to address the need for an AAC system that can be used at conversational speeds on a small-screen mobile device by ambulatory users with mild upper limb motor impairments. Although intended for AAC users, this prototype system also has potential for non-AAC users who may be temporarily unable to use their voices.

## 7.3 APPROACH

DigitCHAT enables rapid, face-to-face communication via small touchscreen devices, such as mobile phones. The system uses large buttons

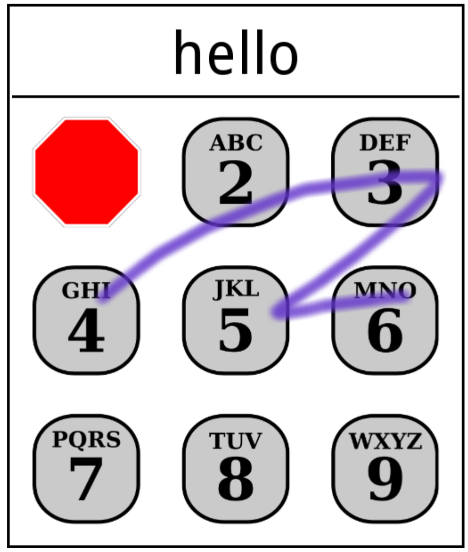



Figure 18: Example Path for "Hello" in DigitCHAT

to assist users who may have difficulty making precise movements. Buttons are visually separated to maximize visibility when a user naturally obscures part of the screen by touching it. Additionally, the interface is organized as a telephone number pad to provide familiarity, especially for older users, and reduce the amount of time required to learn the layout.

To further increase communication speed and assist users with upper limb motor impairments, DigitCHAT supports two types of input: mixed and continuous. In mixed mode, users can provide a combination of discrete taps or non-contiguous path segments to specify the desired word. At any given time, the most likely word is displayed at the top of the screen, and can be selected by tapping it or ending a path on it. In continuous mode, users draw a single line through all desired buttons. As the user's finger or stylus moves over the screen, the most likely word is displayed at the top of the screen. When the user disconnects from the screen's surface, the most likely unigram is spoken aloud immediately. Users can cancel the current path, without speaking the displayed word, by ending on the Stop or Cancel sign. With many current AAC systems, listeners must wait while the user composes a complete utterance [58, 130]. This waiting period places increased pressure on the AAC user to generate utterances as quickly as possible, creating uncomfortable silences and often encouraging the use of telegraphic utterances. By automatically speaking each word as it is completed, DigitCHAT can significantly reduce these gaps and facilitate conversational turn-taking.

DigitCHAT uses a predetermined vocabulary and dictionary-based implementation with unigram statistics based on the Crowdsourced AAC-Like Corpus [116]. Every word in the dictionary is converted into a physical path traversing a standard telephone number pad. The width of this path is incrementally varied up to the size of a standard button and all collisions are recorded as possible sequences that a user might take to specify a given word. These paths are then reverse-

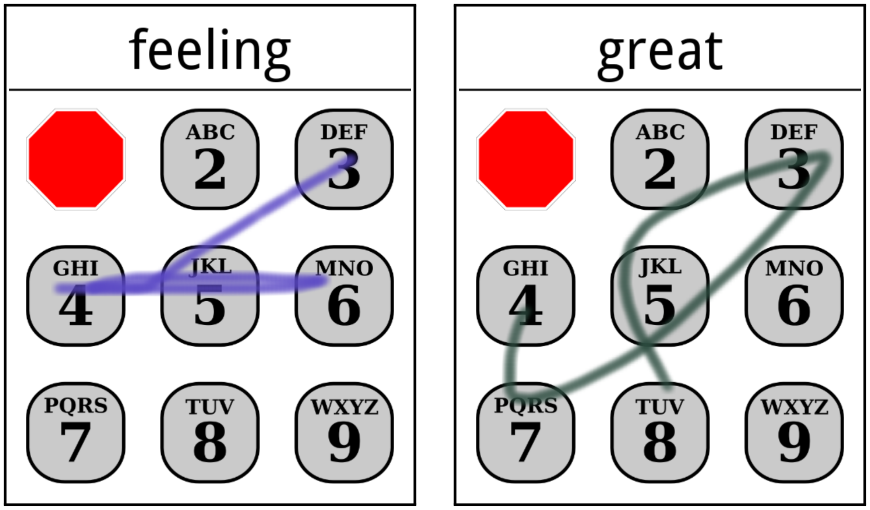



Figure 19: Example Sequence for "Feeling Great" in DigitCHAT

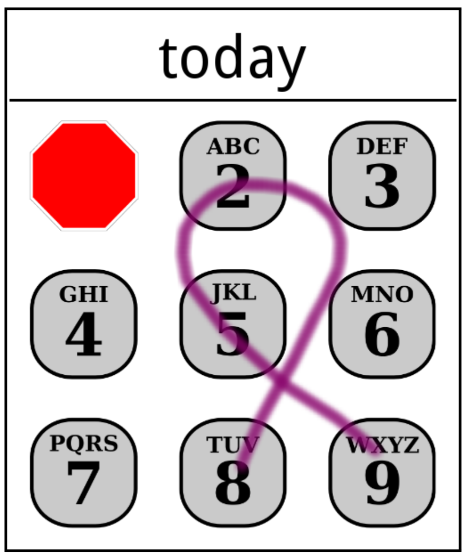



Figure 20: Example Path for "Today" in DigitCHAT

indexed, so that DigitCHAT can look up the user's provided path and retrieve the set of words, with unigram probabilities, that the path could indicate.

Words that share the same sequence of buttons are sometimes called "textonyms." For example, the words "bat" and "cat" are textonyms because they are both specified with the discrete numeric sequence 2-8 or the continuous motion sequence 2-5-8. DigitCHAT implements two approaches to resolving textonyms. In the first approach, the most likely textonym is displayed and users can scribble over the last button in their numeric sequence to rotate through possible textonyms. Users can disconnect from the screen to speak the displayed word aloud or end their scribbling on the Stop sign to cancel the utterance. In the second approach, DigitCHAT implements basic learning and remembers the user's preferred textonym for any given path.

## 7.4 DISCUSSION

We have made DigitCHAT freely available for Android devices on the Google Play Store in order to gauge interest and elicit feedback. Thus far, DigitCHAT has undergone two design and development iterations based on suggestions from users in the target population. In addition to informal feedback from ad-hoc testers, we have received narrative emails from three users and are preparing for a formal study with participants at a clinical facility that serves individuals with chronic neuromotor disorders. A common request, which has since been implemented, was to allow cancellation in order to prevent unexpected or undesirable words from being spoken. We have created several user-configurable options, such as the movement threshold for textonym rotation, but it may be possible to implement a learning algorithm to discover the ideal values for these settings automatically. While the current version of DigitCHAT relies primarily on unigram statistics, we intend to look at potential improvements from using skip-grams or implementing phrasal prediction. We are also experimenting with different methods to efficiently add and remove words from the dictionary to allow for full vocabulary customization.

# 8 CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation aimed to transform AAC systems from passive conduits into active, intelligent communication aids. The long-term vision is to create AAC systems that leverage user-specific information, adapt to a user's behaviors and capabilities, and observe and act on situational context. Towards that end, this dissertation identifies and addresses three problematic assumptions commonly made by most icon-based AAC systems, as well as many letter-based AAC systems: Prescribed Order, Intended Set, and Discrete Entry.

Research into semantic frame theory has become increasingly active over the last few years, and this field has shown promise for use with message construction systems. This dissertation contributes a simple and fast language model, semantic grams, that is specifically designed for use with semantic frames. Semantic grams offer a flexible approach to predictive language modeling that does not require users to make selections in a particular order. Combining semantic frames with semantic grams enables free-order message construction for AAC, and this approach may continue to benefit from future advancements in semantic frame theory.

Context-specific language distributions can be used to help compensate for relaxed linguistic structure, enable more permissive input, and further improve the performance of semantic grams. This dissertation contributes a prioritized list of contextual language predictors based on an empirical analysis of word distributions. Leveraging context can improve unordered prediction and error correction and enable AAC systems to seamlessly adapt to different situations and environments.

Modern touchscreen technologies are challenging for many people with upper limb motor impairments; however, this dissertation shows that systematic approaches can be used to address these challenges. In particular, there are motor profile patterns even among highly diverse users, and static improvements can be made to accommodate these patterns. Mixed-mode or continuous motion input can be offered as useful alternatives to discrete input. Additionally, this dissertation shows that diagnostic games can be used to obtain quantitative motor profiles in order to personalize touchscreen technologies to the capabilities and preferences of an individual user.

Finally, this dissertation demonstrates that these algorithms and design approaches can be combined in different ways to create new, more intelligent interfaces: interfaces that can mitigate the need for linguistically and motorically precise user input to enhance the ease and efficiency of assistive communication.

# Part III

# APPENDIX

# A

# LITERATURE REVIEW

E. Alant, K. Uys, and K. Tönsing. Communication, language, and literacy learning in children with developmental disabilities. In *Treating Childhood Psychopathology and Developmental Disabilities*, pages 373–399. Springer New York, 2009. doi: 10.1007/978-0-387-09530-1_12. URL `http://dx.doi.org/10.1007/978-0-387-09530-1_12`

> How children with disabilities learn about language and how to communicate, particularly with regard to: (1) the impact of specific impairments on information processing and the symbol interpretation, and (2) the role of socio-cultural factors on learning; describes several studies in which children who use AAC were perceived by family and teachers as being less capable of participating in activities and less able to relate to stories than other children.

J. Allan and H. Raghavan. Using part-of-speech patterns to reduce query ambiguity. In *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval*, SIGIR '02, pages 307–314, New York, NY, USA, 2002. ACM. ISBN 1-58113-561-0. doi: 10.1145/564376.564430. URL `http://dx.doi.org/10.1145/564376.564430`

> Semantic disambiguation of one-word queries using clarification questions, and corresponding answers, based on statistical language modeling of nearby words based on their parts-of-speech.

B. R. Baker. Using images to generate speech. *BYTE*, 11(3):160–168, March 1986. ISSN 0360-5280. URL `http://portal.acm.org/citation.cfm?id=5882`

> Encoding technique called semantic compaction, or Minspeak, in which a sequence of multiple icons, some representing morphemes, maps to a long phrase or utterance.

L. Ball, D. Beukelman, and G. Pattee. Acceptance of augmentative and alternative communication technology by persons with amyotrophic lateral sclerosis. *Augmentative and Alternative Communication*, 20:113–123, 2004

> Reviews literature for AAC usage by people with ALS and suggests that approximately 25% of people with ALS reject AAC interventions; presents a case study of 50 people with ALS, in which 96% accepted AAC technology, with 4% rejecting it; describes given reasons for rejection.

D. K. Berlo. *The process of communication: An introduction to theory and practice*. Holt, Rinehart and Winston, 1960

> Continuation of work on the Shannon-Weaver and SMCR models of communication to incorporate contextual information, such as shared cultural background or experiences.

D. Beukelman, R. Jones, and M. Rowan. Frequency of word usage by nondisabled peers in integrated preschool classrooms. *Augmentative and Alternative Communication*, 5(4):243–248, January 1989. doi: 10.1080/07434618912331275296. URL `http://dx.doi.org/10.1080/07434618912331275296`

> Case study showing that non-AAC users, both children and adults, use a small number of words frequently; encourages the core-fringe organization of AAC vocabulary.

D. Beukelman and B. Ansel. Research priorities in augmentative and alternative communication. *Augmentative and Alternative Communication*, 11(2):131–134, January 1995. doi: 10.1080/07434619512331277229. URL `http://dx.doi.org/10.1080/07434619512331277229`

> Examines research priorities in AAC; claims field needs to evaluate the effects of AAC on communication development and create tools or strategies to aid in evaluation; provides a population estimate for potential AAC users of between 0.1% and 1.5%.

D. Beukelman and P. Mirenda. *Augmentative and alternative communication: Management of severe communication disorders in children and adults*. Paul H. Brookes, Baltimore, 1998

> Review of AAC methodologies and systems for various demographics; provides measurements and estimates of aided message construction rates that are less than 20 WPM.

D. Beukelman and P. Mirenda. *Augmentative and Alternative Communication: Supporting Children and Adults With Complex Communication Needs*. Paul H. Brookes Publishing Co., 2006

> Review of AAC methodologies and systems for various demographics; shows that icon-based AAC systems are often preferred for face-to-face communication and by users with language or literacy challenges, such as children and second-language learners.

S. Bickel, P. Haider, and T. Scheffer. Predicting sentences using n-gram language models. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 193–200, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics. doi: 10.3115/1220575.1220600. URL `http://dx.doi.org/10.3115/1220575.1220600`

> Word-based n-gram prediction that uses tab-complete to activate phrasal additions.

A. C. Broerse and E. J. Zwaan. The information value of initial letters in the identification of words. *Journal of Verbal Learning and Verbal Behavior*, 5(5):441–446, October 1966. ISSN 00225371. doi: 10.1016/s0022-5371(66)80058-0. URL `http://dx.doi.org/10.1016/s0022-5371(66)80058-0`

> Letter-based n-gram prediction showing that initial letters of a word have more information content than later letters in the word.

J. Brumberg, A. Nieto-Castanon, P. Kennedy, and F. Guenther. Brain-computer interfaces for speech communication. *Speech Communication*, 52(4):367–379, April 2010. ISSN 0167-6393. doi: 10.1016/j.specom.2010.01.001. URL `http://dx.doi.org/10.1016/j.specom.2010.01.001`

> Reviews silent-speech AAC; demonstrates 2D cursor-control using BCI-EEG tied to the speech production areas of the brain.

E. Bustamante and R. Spain. Measurement invariance of the nasa TLX. *Human Factors and Ergonomics Society*, 52(19):1522–1526, 2008

> Review of the NASA TLX system for measuring mental workload; presents results of a case study with 200 participants showing that TLX lacks scalar invariance; shows that mean TLX scores are not easily compared with invariance measurements; provides an argument for using the shortened "Raw TLX" method, without pairwise comparisons.

P. Demasco and K. McCoy. Generating text from compressed input: an intelligent interface for people with severe motor impairments. *Communications of the ACM*, 35(5):68–78, May 1992. ISSN 0001-0782. doi: 10.1145/129875.129881. URL `http://dx.doi.org/10.1145/129875.129881`

> "Sentence companion" technique, in which users only select content words (semantically salient words) and the system expands the message into its syntactically correct form; this technique relies on the content words being selected in syntactic order (no reordering is performed) and only expands function words; uses hand-coded syntactic rules and semantic labels.

H. Fang and C. Zhai. Semantic term matching in axiomatic approaches to information retrieval. In *Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval*, SIGIR '06, pages 115–122, New York, NY, USA, 2006. ACM. ISBN 1-59593-369-7. doi: 10.1145/1148170.1148193. URL `http://dx.doi.org/10.1145/1148170.1148193`

> Query approach for determining document relevance based on semantic term matching; presents an information-theoretic approach that determines semantically similar

terms by looking at pre-judged relevant documents and tagging non-query words that have high mutual information; uses those similar terms, and their similarity distances, to reweight potentially relevant documents during query time.

C. J. Fillmore. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(Origins and Evolution of Language and Speech):20–32, October 1976. ISSN 1749-6632. doi: 10.1111/j.1749-6632.1976.tb25467.x. URL http://dx.doi.org/10.1111/j.1749-6632.1976.tb25467.x

Seminal work on frame semantics, an extension of case grammar, that describes semantic frames as a collection of facts that specify "characteristic features, attributes, and functions of a denotatum, and its characteristic interactions with things necessarily or typically associated with it;" argues that one cannot understand a word without understanding its semantic frames and the associated semantic roles of each frame; relates linguistic semantics to knowledge.

M. Goel, J. Wobbrock, and S. Patel. GripSense: using built-in sensors to detect hand posture and pressure on commodity mobile phones. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 545–554, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1580-7. doi: 10.1145/2380116.2380184. URL http://dx.doi.org/10.1145/2380116.2380184

Approach for determining which of 4 different hand postures someone is using to hold their mobile phone; requires only the built-in sensors of a standard mobile phone (e.g. accelerometer, inertial sensors, vibration motor, and touchscreen); detects one of 3 different levels of pressure being used on the touchscreen.

D. Goldberg. *Unistrokes for computerized interpretation of handwriting (US Patent #5596656)*. US Patent and Trademark Office, January 1997. URL http://www.patentlens.net/patentlens/patent/US_5596656/en/

Patent of a method for classifying continuous strokes, referred to as "unistrokes," into sets of characters from a predetermined alphabet; allows for multiple characters to be drawn in a single, unbroken stroke; useful for converting handwritten text on a touchscreen into digitized text.

F. Guenther, J. Brumberg, E. Wright, A. Nieto-Castanon, J. Tourville, M. Panko, R. Law, S. Siebert, J. Bartels, D. Andreasen, P. Ehirim, H. Mao, and P. Kennedy. A wireless Brain-Machine interface for Real-Time speech synthesis. *PLoS ONE*, 4(12), December 2009. doi: 10.1371/journal.pone.0008218. URL http://dx.doi.org/10.1371/journal.pone.0008218

Case study of a patient with locked-in syndrome in which a wireless BCI was used to monitor attempts to produce speech; attempted speech was converted into synthesized vowel sounds with a feedback delay of 50 milliseconds.

J. P. Hansen, H. Lund, H. Aoki, and K. Itoh. Gaze communication systems for people with ALS. In *ALS Communication Workshop, Yokohama, Japan*, pages 35–38, December 2006. URL http://www.cogain.org/w/images/a/a2/ALS_Workshop_Yokohama2006.pdf

System description of GazeTalk, a multi-lingual typing system based on eye tracking and dwell-time selection; intended for users with ALS.

S. Hart and L. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human mental workload*, 1(3):139–183, 1988

Results of multiple experiments on the effectiveness of NASA's initial TLX system; proposes a multi-dimensional rating scale for six subjective, workload-related factors to quantify total workload.

R. Hemayati, W. Meng, and C. Yu. Semantic-based grouping of search engine results using WordNet. In *Proceedings of the joint 9th Asia-Pacific web and 8th international conference on web-age information management conference on Advances in data and web management*, APWeb/WAIM'07, pages 678–686, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-72483-4. URL http://portal.acm.org/citation.cfm?id=1769795

Approach for grouping semantically similar search engine results, primarily to increase diversity, by using synsets from WordNet; synsets are combined into super-synsets; current approach only works for single-term queries.

J. Higginbotham, H. Shane, S. Russell, and K. Caves. Access to AAC: Present, past, and future. *Augmentative and Alternative Communication*, 23(3):243–257, January 2007. ISSN 0743-4618. doi: 10.1080/07434610701571058. URL http://dx.doi.org/10.1080/07434610701571058

Survey of AAC system types, design factors, and emerging technologies and approaches; reviews several studies to provide communication rate estimates of 3 - 7 WPM for scanning systems and 5 - 10 WPM for eye-tracking systems.

J. Higginbotham, A. Bisantz, M. Sunm, K. Adams, and F. Yik. The effect of context priming and task type on augmentative communication performance. *Augmentative and Alternative Communication*, 25(1): 19–31, 2009

Case study in which an AAC device was, or was not, primed with task-specific vocabularies and tested on non-AAC users; contextual priming had a small but significant

effect on keystroke savings; higher level measurements of communication rate, task performance, and user perceptions suggested keystroke savings that were not seen in experiments; keystroke-based measurements may not be predictive of task-level performance.

Y. How and M. Kan. Optimizing predictive text entry for short message service on mobile phones. In *Human Computer Interfaces International (HCII 05)*, 2005. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.638`

Remapping of letters to the 9-button telephone keypad that is optimized for language patterns from a corpus of text messages; optimization by genetic algorithms was based partly on an operation-level model (OLM) of required time to perform certain physical movements on the keypad.

A. Järvelin, A. Järvelin, and K. Järvelin. s-grams: Defining generalized n-grams for information retrieval. *Information Processing & Management*, 43(4):1005–1019, July 2007. ISSN 03064573. doi: 10.1016/j.ipm.2006.09.016. URL `http://dx.doi.org/10.1016/j.ipm.2006.09.016`

Enhanced definitions for using s-grams, where the "s" stands for "skip," as a generalization of n-grams in which a number of characters (for letter-based s-grams) or words (for word-based n-grams) are skipped to form the gram; n-grams can be considered s-grams with a skip value of zero, thus requiring adjacency; presents an enhancement to Jaccard distance that is more sensitive to gram counts.

A. Jinks and B. Sinteff. Consumer response to AAC devices: Acquisition, training, use, and satisfaction. *Augmentative and Alternative Communication*, 10(3):184–190, January 1994. doi: 10.1080/07434619412331276890. URL `http://dx.doi.org/10.1080/07434619412331276890`

Case study that surveyed former patients of an AAC rehabilitation center; there were 76 respondents between 3 and 79 years of age; 71% of respondents received devices and 81% of those devices were taxpayer-funded; 53% of respondents with CP continued using their AAC devices.

S. K. Kane, B. L. Church, K. Althoff, and D. McCall. What we talk about: designing a context-aware communication tool for people with aphasia. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '12, pages 49–56, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1321-6. doi: 10.1145/2384916.2384926. URL `http://dx.doi.org/10.1145/2384916.2384926`

Context-aware, adaptive AAC system for people with aphasia; context includes location (via GPS and 802.11), user (via front camera), and conversation partner (via rear camera).

G. Karberis and G. Kouroupetroglou. Transforming spontaneous telegraphic language to Well-Formed greek sentences for alternative and augmentative communication. In *Proceedings of the Second Hellenic Conference on AI: Methods and Applications of Artificial Intelligence*, SETN '02, pages 155–166, London, UK, UK, 2002. Springer-Verlag. ISBN 3-540-43472-0. URL `http://portal.acm.org/citation.cfm?id=645861.670294`

> Telegraphic-to-Full Sentence (TtFS) module that converts telegraphic input (compressed, incomplete, grammatically and syntactically ill-formed) into full sentences that are grammatically and syntactically correct; designed for the Greek language; uses hand-coded information for each word in the lexicon; assumes that all content words are provided and basic order is correct for active voice Greek (subject then object).

T. Kiss and J. Strunk. Unsupervised multilingual sentence boundary detection. *Computational Linguistics*, 32(4):485–525, December 2006. ISSN 0891-2017. doi: 10.1162/coli.2006.32.4.485. URL `http://dx.doi.org/10.1162/coli.2006.32.4.485`

> Punkt sentence boundary detection (SBD), a language-independent and unsupervised approach using only criteria that are independent of context; especially relies on detection of abbreviations; defines abbreviations as tight collocations of truncated words with internal or final periods.

H. Koester and S. Levine. Effect of a word prediction feature on user performance. *Augmentative and Alternative Communication*, 12(3):155–168, 1996

> Case study of 14 people who regularly use AAC, including mouth-stick systems, and the effects of word prediction on their typing speed for a "copy phrase" task; demonstrated that the increased cognitive and perceptual costs for using prediction can overwhelm keystroke gains; number of keystrokes was reduced, but keystroke time increased.

H. Koester and S. Levine. Keystroke-level models for user performance with word prediction. *Augmentative and Alternative Communication*, 13(4), 1997

> Examines the estimated/theoretical and empirical performance improvements of different keystroke savings models; presents four primary factors in correlating actual and estimated performance improvements from word prediction: average number of searches per character, keystroke savings, keypress time, and prediction-list search time.

P. O. Kristensson and S. Zhai. SHARK2: a large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, pages 43–52, New York, NY, USA, 2004. ACM. ISBN 1-58113-957-8. doi: 10.1145/1029632.1029640. URL http://dx.doi.org/10.1145/1029632.1029640

Shorthand Aided Rapid Keyboarding (SHARK), that uses sokgraphs (shorthand defined on a keyboard as a graph) to recognize movement patterns on any letter-based keyboard from a precalculated mapping of patterns to words; tested vocabulary of roughly 20,000 sokgraphs; uses both shape and location to define the patterns.

P. O. Kristensson and K. Vertanen. The potential of dwell-free eye-typing for fast assistive gaze communication. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 241–244, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1221-9. doi: 10.1145/2168556.2168605. URL http://dx.doi.org/10.1145/2168556.2168605

Application of continuous motion typing (e.g. Swype, ShapeWriter, T9 Trace) for letter-based systems using eye-tracking; shows improvement of 20 WPM for dwell-time eye-typing to over 40 WPM for dwell-free eye-typing; uses a theoretically perfect recognizer to give upper bound estimates.

C. Kushler and R. Marsden. *System and method for continuous stroke word-based text input (US Patent #7453439)*. US Patent and Trademark Office, November 2008. URL http://www.patentlens.net/patentlens/patent/US_7453439/en/

Patent of a continuous motion typing system for letter-based keyboards that uses a dictionary approach with gestural thresholds (e.g. path angles) to perform lexical disambiguation on a superset of letters.

G. Lesher, B. Moulton, and J. Higginbotham. Techniques for augmenting scanning communication. *Augmentative and Alternative Communication*, 14(2):81–101, January 1998. doi: 10.1080/07434619812331278236. URL http://dx.doi.org/10.1080/07434619812331278236

Comparison of 14 different switch-based scanning techniques for AAC, including both letter and word prediction; the best character and word-based prediction techniques each provided about 40% switch savings.

G. Lesher and C. Sanelli. A Web-Based system for autonomous text corpus generation. In *Proceedings of ISAAC*, 2000. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.15.9386

Automatic generation of large corpora by crawling the World Wide Web (WWW) and automatically tagging text blocks with information about genre, style, or education level.

G. Lesher and G. Rinkus. Domain-Specific word prediction for augmentative communication. In *Proceedings of the RESNA 2002 Annual Conference*, 2002

Method of deriving word prediction models from domain-specific corpora; case study of using telephone transcripts from the Switchboard Corpus to generate models for 20 different topic domains; shows benefit of using domain-specific models; mentions idea of dynamically switching between statistical models, but does not present an approach.

J. Li and G. Hirst. Semantic knowledge in word completion. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, Assets '05, pages 121–128, New York, NY, USA, 2005. ACM. ISBN 1-59593-159-7. doi: 10.1145/1090785.1090809. URL `http://dx.doi.org/10.1145/1090785.1090809`

Approach to word prediction that merges n-gram probabilities with semantic knowledge based on pointwise mutual information, with a Lesk-like filter, of co-occurring words in the British National Corpus (BNC); shows prediction improvement as a keystroke savings of 14% for completion of nouns.

J. Light, D. Beukelman, and J. Reichle. *Communicative competence for individuals who use AAC: From research to effective practice*. Paul H. Brookes Publishing Co., 2003

Book about AAC research, systems, and therapeutic practices; surveys studies and demographic information showing that many people with motor-speech impairments also have upper limb impairments that prevent the use of sign language or standard QWERTY keyboards.

C. Lin and E. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 71–78, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. doi: 10.3115/1073445.1073465. URL `http://dx.doi.org/10.3115/1073445.1073465`

Comparison of human evaluation and n-gram co-occurrence for determining how similar machine-produced texts are to human-produced texts for automatic summarization and translation tasks; shows that n-gram co-occurrence is a useful metric that could save on human evaluation efforts.

G. Lindsay, J. Dockrell, M. Desforges, J. Law, and N. Peacey. Meeting the needs of children and young people with speech, language and communication difficulties. *International Journal of Language & Communication Disorders*, 45(4):448–460, July 2010. doi: 10.3109/13682820903165693. URL `http://dx.doi.org/10.3109/13682820903165693`

Six case studies consisting of interviews with program managers, teachers, and specialists in speech, language, and communication therapy centers across England; showed a lack of consistency in services and approaches; 7% of children entering school had significant speech and language difficulties; 1% of children had severe and complex communication needs.

Y. Lv and C. Zhai. Positional language models for information retrieval. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 299–306, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. doi: 10.1145/1571941.1571994. URL `http://dx.doi.org/10.1145/1571941.1571994`

Combination of query-term proximity measurements and passage retrieval into a positional language model (PLM); breaks documents into "soft" passages based on term clustering and density for each word position; presents 4 proximity-based density functions to estimate PLMs, with the Gaussian density kernel and Dirchlet smoothing performing the best.

I. S. MacKenzie and R. W. Soukoreff. Text entry for mobile computing: Models and methods, theory and practice. *Human-computer Interaction*, 17:147–198, 2002. doi: 10.1207/S15327051HCI172\&3_2. URL `http://dx.doi.org/10.1207/S15327051HCI172&3_2`

Survey of mobile, letter-based text entry techniques and combinations of Fitts' law with language models; identifies primary optimization techniques as language prediction and movement minimization; shows that corpora often do not represent user language.

I. S. MacKenzie and R. W. Soukoreff. Phrase sets for evaluating text entry techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 754–755, New York, NY, USA, 2003. ACM. ISBN 1-58113-637-4. doi: 10.1145/765891.765971. URL `http://dx.doi.org/10.1145/765891.765971`

Collection of 500 phrases for use in testing letter-based text entry systems; ecologically validated to represent letter and word frequencies in English.

I. S. MacKenzie and T. Felzer. SAK: Scanning ambiguous keyboard for efficient one-key text entry. *ACM Transactions on Computer-Human Interaction*, 17(3), July 2010. ISSN 1073-0516. doi: 10.1145/1806923.1806925. URL `http://dx.doi.org/10.1145/1806923.1806925`

A one-key scanning technique that uses an ambiguous, letter-based keyboard followed by word selection; letter-selection step scans over an alphabet split into 3 keys plus a space; word-selection step uses word prediction based on frequency; used able-bodied participants.

B. MacWhinney. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum, 2000

Book describing the CHILDES project, which collects conversational interactions from children, their caregivers and siblings, as well as bilingual children, second-language learners, and children with various types of language disabilities.

B. MacWhinney. The TalkBank project. creating and digitizing language corpora: Volume 1, synchronic databases, 2007

Chapter of a book describing the TalkBank project, which collects corpora of first language acquisition, second language acquisition, conversation analysis, classroom discourse, and aphasic language; CHILDES is a sub-project of TalkBank.

C. Marvin, D. Beukelman, and D. Bilyeu. Vocabulary-use patterns in preschool children: Effects of context and time sampling. *Augmentative and Alternative Communication*, 10(4):224–236, January 1994. doi: 10.1080/07434619412331276930. URL `http://dx.doi.org/10.1080/07434619412331276930`

Case study of vocabulary usage by 10 non-disabled, preschool-aged children; similar vocabulary was used at home and at school; reviews vocabulary designs of AAC systems; suggests core vocabulary is 20% of total and structure (function) words are 2% of total.

J. Matas, P. Mathy-Laikko, D. Beukelman, and K. Legresley. Identifying the nonspeaking population: a demographic study. *Augmentative and Alternative Communication*, 1(1):17–31, December 1985. doi: 10.1080/07434618512331273491. URL `http://dx.doi.org/10.1080/07434618512331273491`

Two studies in Washington state to determine the size, characteristics, and intervention needs of school-age nonspeaking students; estimated 3 - 5 potential AAC users per 1,000 students.

J. Matiasek and M. Baroni. Exploiting long distance collocational relations in predictive typing. In *Proceedings of the 2003 EACL Workshop on Language Modeling for Text Entry Methods*, TextEntry '03, pages 1–8, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. URL `http://portal.acm.org/citation.cfm?id=1628196`

Collocation-based word prediction using mutual information within a fixed window size of 50 words; only used pairs of semantically related words (300,000 pairs); hand-tuned and fixed weighting for frequencies of unigrams, bigrams, and collocations.

K. McCoy, C. Pennington, and A. Badman. Companion: From research prototype to practical integration. *Natural Language Engineering*, 4(01):73–95, 1998. URL http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=48437

Continuation of work on "Compansion" technique that expands telegraphic input by adding function words and conjugation to an in-order stream of content words based on hand-coded rules; presents plans for practical implementation of an intelligent parser/generator (IPG).

M. Mehl, S. Vazire, N. Ramírez-Esparza, R. Slatcher, and J. Pennebaker. Are women really more talkative than men? *Science*, 317(5834):82, July 2007. ISSN 1095-9203. doi: 10.1126/science.1139940. URL http://dx.doi.org/10.1126/science.1139940

Longitudinal study of 400 college students in the U.S. and Mexico that sampled their speech using electronically activated recording (EAR) devices; showed that, on average, both men and women speak about 16,000 words per day, of which 32% (about 5,000) were unique words.

K. R. Muller and B. Blankertz. Toward noninvasive brain-computer interfaces. *Signal Processing Magazine, IEEE*, 23(5), 2006. ISSN 1053-5888. doi: 10.1109/msp.2006.1708426. URL http://dx.doi.org/10.1109/msp.2006.1708426

Hex-o-Spell BCI system for letter-based typing that uses surface-level electroencephalography (EEG) and 2-signal motor imagery; typing rate was 2 - 8 characters per minute and called "world-class spelling speed" for a BCI system on untrained users.

J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 249–256, New York, NY, USA, 1990. ACM. doi: 10.1145/97243.97281. URL http://doi.acm.org/10.1145/97243.97281

Efficiency and accuracy of heuristic evaluation of user interfaces using Molich and Nielsen's "9 Heuristics;" shows that 3 - 5 people is the optimal group size for heuristic evaluations.

S. Nikolova, M. Tremaine, and P. Cook. Click on bake to get cookies: guiding word-finding with semantic associations. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 155–162, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-881-0. doi: 10.1145/1878803.1878832. URL http://dx.doi.org/10.1145/1878803.1878832

Visual Vocabulary for Aphasia (ViVA), a semantically organized vocabulary network based on Lingraphica's vocabulary structure, intended for AAC systems used by people with aphasia.

R. Patel, S. Pilato, and D. Roy. Beyond linear syntax: An Image-Oriented communication aid. *Journal of Assistive Technology Outcomes and Benefits*, 1:57–66, 2004

> IconCHAT, an icon-based AAC system that uses case grammar to allow for verb-first, non-linear message construction.

M. F. Porter. An algorithm for suffix stripping. In K. S. Jones and P. Willett, editors, *Readings in information retrieval*, pages 313–316. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997. ISBN 1-55860-454-5. URL `http://portal.acm.org/citation.cfm?id=275705`

> Porter stemming algorithm that uses suffix-stripping techniques to approximate lemmatization for the English language.

D. Rashid and N. Smith. Relative keyboard input system. In *Proceedings of the 13th international conference on Intelligent user interfaces*, IUI '08, pages 397–400, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-987-6. doi: 10.1145/1378773.1378839. URL `http://dx.doi.org/10.1145/1378773.1378839`

> Letter-based input system that uses a blank touch-screen and touch-typing; single word input sequences are disambiguated by the relative positioning of touch locations and a dictionary.

B. Roark, J. D. Villiers, C. Gibbons, and M. F. Oken. Scanning methods and language modeling for binary switch typing. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, SLPAT '10, pages 28–36, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL `http://portal.acm.org/citation.cfm?id=1867754`

> Comparison of row-column scanning, Huffman scanning, and RSVP for letter-based AAC; shows that language modelling is a big factor in accuracy and speed of Huffman approach; RSVP was slower than row-column scanning with the same language model.

B. C. Roy, M. C. Frank, and D. Roy. Relating activity contexts to early word learning in dense longitudinal data. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, 2012

> Results from the Human Speechome corpus showing that a child's word acquisition and usage is contextually related to location and activity.

J. Schler, M. Koppel, S. Argamon, and J. Pennebaker. Effects of age and gender on blogging. In *Proc. of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, March 2006

Analysis of 700,000 posts by 20,000 bloggers over a month shows that there are differences in writing style and content between males, females, and writers of different ages; age and gender can be predicted given writing samples; creation of the freely available Blog Authorship Corpus.

W. Schramm. How communication works. the process and effects of mass communication. *Urbana: University of Illinois Press*, 1954

Continuation of work on the Shannon-Weaver model of communication to incorporate full reciprocity (i.e. feedback) via effects and interaction; reformulation as SMCR.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, July 1948. doi: 10.1145/584091.584093. URL http://dx.doi.org/10.1145/584091.584093

Seminal work on information theory; models information with a logarithmic measure, usually binary digits (bits); models communication with 5 components: an information source, a transmitter, a channel, a receiver, and a destination.

C. E. Shannon and W. Weaver. The mathematical theory of communication. *University of Illinois Press*, 19(7):1, 1949

Formalization, clarification, and repackaging of work on information and communication theory; asserts definitions for formal problems of communication.

C. Suen. n-Gram statistics for natural language understanding and text processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):164–172, April 1979. ISSN 0162-8828. doi: 10.1109/tpami.1979.4766902. URL http://dx.doi.org/10.1109/tpami.1979.4766902

Representative work on letter-based n-grams in natural language processing; analyzes a corpus of 1 million words and presents frequency statistics.

J. Todman, N. Alm, and L. Elder. Computer-aided conversation: A prototype system for nonspeaking people with physical disabilities. *Applied Psycholinguistics*, 15(01):45–73, 1994. doi: 10.1017/s0142716400006974. URL http://dx.doi.org/10.1017/s0142716400006974

Simulated results of using an utterance-based AAC system that follows predicted sequences of speech acts; given a specific topic, conversations progressed quickly and were recoverable.

J. Todman. Rate and quality of conversations using a text-storage AAC system: Single-case training study. *Augmentative and Alternative Communication*, pages 164–179, September 2000. ISSN 0743-4618. doi: 10.1080/07434610012331279024. URL http://dx.doi.org/10.1080/07434610012331279024

TALK system that incorporates multiple AAC features, such as labels that hide shuffled selections (e.g. "Hi," "Hello," "Hi there"), subsequent moves based on turns, holistic phrases, advance planning based on conversational progress, and feedback utterances (e.g. "uh-huh"); results of a single-user case study showed communication rates of 30 WPM.

H. Trinh, A. Waller, K. Vertanen, P. A. Kristensson, and V. Hanson. iSCAN: a phoneme-based predictive communication aid for nonspeaking individuals. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '12, pages 57–64, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1321-6. doi: 10.1145/2384916.2384927. URL `http://dx.doi.org/10.1145/2384916.2384927`

Predictive AAC system based on 42 phonemes (17 vowels and 25 consonants); uses mixture model with 6-gram phonemes and 3-gram words; tested on 16 able-bodied participants and 1 cerebral palsied participant.

K. Trnka, D. Yarrington, K. McCoy, and C. Pennington. Topic modeling in fringe word prediction for AAC. In *Proceedings of the 11th international conference on Intelligent user interfaces*, IUI '06, pages 276–278, New York, NY, USA, 2006. ACM. ISBN 1-59593-287-9. doi: 10.1145/1111449.1111509. URL `http://dx.doi.org/10.1145/1111449.1111509`

Comparison of two topic-modelling algorithms for letter-based prediction, using trigrams, of fringe words for AAC.

K. Trnka and K. McCoy. Corpus studies in word prediction. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 195–202, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-573-1. doi: 10.1145/1296843.1296877. URL `http://dx.doi.org/10.1145/1296843.1296877`

Survey of corpus studies for AAC; shows that no large, authentic AAC corpora exist; shows that approximations, such as with out-of-domain data, are useful even with advanced language modeling techniques, such as topic modeling.

E. Tzoukermann, J. Klavans, and C. Jacquemin. Effective use of natural language processing techniques for automatic conflation of multi-word terms: the role of derivational morphology, part of speech tagging, and shallow parsing. *SIGIR Forum*, 31(SI):148–155, July 1997. ISSN 0163-5840. doi: 10.1145/278459.258554. URL `http://dx.doi.org/10.1145/278459.258554`

Description of system that uses NLP techniques, especially derivational morphology and phrasal relations, to determine semantically related terms for information retrieval.

O. Udwin and W. Yule. Augmentative communication systems taught to cerebral palsied children - a longitudinal study:I. the acquisition of signs and symbols, and syntactic aspects of their use over time. *British Journal of Disorders of Communication*, 25(3):295–309, January 1990. doi: 10.3109/13682829009011979. URL http://dx.doi.org/10.3109/13682829009011979

> Year-long observation of conversational behavior by 40 cerebral-palsied children with language impairments showed that 80% of utterances could be labelled using just 4 categories; sign and symbol AAC systems being used were severely restrictive.

H. Van Balkom and M. Welle Donker-Gimbrere. A psycholinguistic approach to graphic language use. *Augmentative and alternative communication: European Perspectives*, pages 153–170, 1996

> Examination of language production behaviors in users of graphical AAC systems shows shorter and less complete narratives, single-word utterances, and shorter average sentences.

A. Van Den Bosch. Scalable classification-based word prediction and confusible correction. *Traitement Automatique des Langues*, 46(2):39–63, 2006

> Application of a IGTree, a decision-tree algorithm for multi-label classification, to word prediction suggests that prediction accuracy increases at a log-linear rate with more training data; discarding low-frequency words from training data (i.e. the long tail) does not improve results; left-context-only prediction is not as good as left-and-right-context prediction.

A. Van Den Bosch and P. Berck. Memory-based machine translation and language modeling. In *The Prague Bulletin of Mathematical Linguistics*, 2009. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.189.5165

> Memory-based machine translation (MBMT) that maps all possible trigram translations in a source language to trigrams in a target language; full sentence translation exploits overlaps.

K. Vertanen and P. O. Kristensson. The imagination of crowds: Conversational AAC language modeling using crowdsourcing and large data sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 700–711. ACL, 2011

> Creation of a fake AAC corpus by crowdsourcing non-AAC users and asking what they would say if they were AAC users; seed corpus (6,000 utterances) amplified by searching Twitter, Usenet, and other corpora for utterances with low word-error rate (WER) and low cross-entropy per word.

T. Walsh. Utterance-based systems: organization and design of AAC interfaces. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 327–328, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-881-0. doi: 10.1145/1878803.1878895. URL http://dx.doi.org/10.1145/1878803.1878895

> Guided, full-utterance AAC system that uses activity contexts and limited word substitution with optional keyboard input and button relabelling.

T. Wandmacher and J. Antoine. Training language models without appropriate language resources: Experiments with an AAC system for disabled people. In *Proceedings of LREC*, 2006. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.62.1815

> Comparison of three techniques to reduce the dependence of statistical language models on their training resources: cache model that augments probabilities of last N inserted words, a user dictionary, and interpolation between a base model and a dynamic user model; the dynamic user model worked best (trigrams with a linear interpolation and EM-like weighting).

S. J. Westerman and T. Cribbin. Mapping semantic information in virtual space: dimensions, variance and individual differences. *International Journal of Human-Computer Studies*, 53(5):765–787, November 2000. ISSN 10715819. doi: 10.1006/ijhc.2000.0417. URL http://dx.doi.org/10.1006/ijhc.2000.0417

> Comparison of 2D and 3D organization of semantic information for manual search; suggests that the amount of additional semantic content in 3D representation is unlikely to be worth the additional cognitive demands of a third dimension.

J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, June 2002. ISSN 13882457. doi: 10.1016/s1388-2457(02)00057-3. URL http://dx.doi.org/10.1016/s1388-2457(02)00057-3

> Survey of BCI systems for communication and control operations; current systems have maximum information transfer rates of 10 - 25 bits per minute.

J. Wolpaw. Brain-computer interfaces (BCIs) for communication and control. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 1–2, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-573-1. doi: 10.1145/1296843.1296845. URL http://dx.doi.org/10.1145/1296843.1296845

> Surface-level EEG, P300 BCI system being distributed for testing in user homes.

X. Zhang and S. MacKenzie. Evaluating eye tracking with ISO 9241 - part 9. In *Proceedings of the 12th international conference on Human-computer interaction: intelligent multimodal interaction environments*, HCI'07, pages 779–788, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-73108-5. URL http://portal.acm.org/citation.cfm?id=1769678

First evaluation of eye-tracking techniques using the evaluation standard in International Standards Organization (ISO) 9241-9; comparison of three techniques: long dwell-time, short dwell-time, and keypress during fixation; keypress during fixation was best with throughput of 3.8 bits of information per second compared to 4.7 bits for a standard computer mouse.

## REFERENCES


[1] E. Alant, K. Uys, and K. Tönsing. Communication, language, and literacy learning in children with developmental disabilities. In *Treating Childhood Psychopathology and Developmental Disabilities*, pages 373–399. Springer New York, 2009. doi: 10.1007/978-0-387-09530-1_12. URL http://dx.doi.org/10.1007/978-0-387-09530-1_12.

[2] J. Allan and H. Raghavan. Using part-of-speech patterns to reduce query ambiguity. In *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval*, SIGIR '02, pages 307–314, New York, NY, USA, 2002. ACM. ISBN 1-58113-561-0. doi: 10.1145/564376.564430. URL http://dx.doi.org/10.1145/564376.564430.

[3] L. Anthony, Y. Kim, and L. Findlater. Analyzing user-generated youtube videos to understand touchscreen use by people with motor impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1223–1232, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2466158. URL http://dx.doi.org/10.1145/2470654.2466158.

[4] B. R. Baker. Using images to generate speech. *BYTE*, 11(3):160–168, March 1986. ISSN 0360-5280. URL http://portal.acm.org/citation.cfm?id=5882.

[5] L. Ball, D. Beukelman, and G. Pattee. Acceptance of augmentative and alternative communication technology by persons with amyotrophic lateral sclerosis. *Augmentative and Alternative Communication*, 20:113–123, 2004.

[6] A. Bell. Language style as audience design. *Language in society*, 13(2):145–204, 1984.

[7] D. K. Berlo. *The process of communication: An introduction to theory and practice*. Holt, Rinehart and Winston, 1960.

[8] D. Beukelman and B. Ansel. Research priorities in augmentative and alternative communication. *Augmentative and Alternative Communication*, 11(2):131–134, January 1995. doi: 10.1080/07434619512331277229. URL http://dx.doi.org/10.1080/07434619512331277229.

[9] D. Beukelman and P. Mirenda. *Augmentative and alternative communication: Management of severe communication disorders in children and adults*. Paul H. Brookes, Baltimore, 1998.

[10] D. Beukelman and P. Mirenda. *Augmentative and Alternative Communication: Supporting Children and Adults With Complex Communication Needs*. Paul H. Brookes Publishing Co., 2006.

[11] D. Beukelman, R. Jones, and M. Rowan. Frequency of word usage by nondisabled peers in integrated preschool classrooms. *Augmentative and Alternative Communication*, 5(4):243–248, January 1989. doi: 10.1080/07434618912331275296. URL http://dx.doi.org/10.1080/07434618912331275296.

[12] S. Bickel, P. Haider, and T. Scheffer. Predicting sentences using n-gram language models. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 193–200, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics. doi: 10.3115/1220575.1220600. URL http://dx.doi.org/10.3115/1220575.1220600.

[13] S. Bird, E. Klein, and E. Loper. *Natural Language Processing with Python*. O'Reilly Media, 1 edition, July 2009. ISBN 0596516495. URL http://www.worldcat.org/isbn/0596516495.

[14] A. C. Broerse and E. J. Zwaan. The information value of initial letters in the identification of words. *Journal of Verbal Learning and Verbal Behavior*, 5(5):441–446, October 1966. ISSN 00225371. doi: 10.1016/s0022-5371(66)80058-0. URL http://dx.doi.org/10.1016/s0022-5371(66)80058-0.

[15] J. Brumberg, A. Nieto-Castanon, P. Kennedy, and F. Guenther. Brain-computer interfaces for speech communication. *Speech Communication*, 52(4):367–379, April 2010. ISSN 0167-6393. doi: 10.1016/j.specom.2010.01.001. URL http://dx.doi.org/10.1016/j.specom.2010.01.001.

[16] E. Bustamante and R. Spain. Measurement invariance of the nasa TLX. *Human Factors and Ergonomics Society*, 52(19):1522–1526, 2008.

[17] K. W. Cassidy, M. H. Kelly, and L. J. Sharoni. Inferring gender from name phonology. *Journal of Experimental Psychology*, 128 (3):362–381, 1999.

[18] K. B. Chen, A. B. Savage, A. O. Chourasia, D. A. Wiegmann, and M. E. Sesto. Touch screen performance by individuals with and without motor control disabilities. *Applied ergonomics*, 44(2): 297–302, 2013.

[19] A. Dean-Hall, C. L. A. Clarke, J. Kamps, P. Thomas, and E. Voorhees. Overview of the TREC 2012 contextual suggestion track. In *21st Text REtrieval Conference, Gaithersburg, Maryland*, 2012.

[20] P. Demasco and K. McCoy. Generating text from compressed input: an intelligent interface for people with severe motor

impairments. *Communications of the ACM*, 35(5):68–78, May 1992. ISSN 0001-0782. doi: 10.1145/129875.129881. URL http://dx.doi.org/10.1145/129875.129881.

[21] G. Drosatos, G. Stamatelatos, A. Arampatzis, and P. S. Efraimidis. DUTH at TREC 2013 contextual suggestion track. In *National Institute of Standards and Technology (NIST) Text REtrieval Conference (TREC)*, December 2013.

[22] P. Eckert and S. McConnell-Ginet. *Language and gender*. Cambridge University Press, 2013.

[23] P. Eckert and J. R. Rickford. *Style and sociolinguistic variation*. Cambridge University Press, 2001.

[24] I. A. El-Khair. Effects of stop words elimination for arabic information retrieval: a comparative study. *International Journal of Computing & Information Sciences*, 4(3):119–133, 2006.

[25] J. Fan and R. Li. Statistical challenges with high dimensionality: feature selection in knowledge discovery. In *Proceedings of the International Congress of Mathematicians Madrid, August 22-30, 2006*, pages 595–622, 2007.

[26] H. Fang and C. Zhai. Semantic term matching in axiomatic approaches to information retrieval. In *Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval*, SIGIR '06, pages 115–122, New York, NY, USA, 2006. ACM. ISBN 1-59593-369-7. doi: 10.1145/1148170.1148193. URL http://dx.doi.org/10.1145/1148170.1148193.

[27] C. J. Fillmore. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(Origins and Evolution of Language and Speech):20–32, October 1976. ISSN 1749-6632. doi: 10.1111/j.1749-6632.1976.tb25467.x. URL http://dx.doi.org/10.1111/j.1749-6632.1976.tb25467.x.

[28] J. Gilliot, G. Casiez, and N. Roussel. Impact of form factors and input conditions on absolute indirect-touch pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 723–732, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2556997. URL http://dx.doi.org/10.1145/2556288.2556997.

[29] M. Goel, J. Wobbrock, and S. Patel. GripSense: using built-in sensors to detect hand posture and pressure on commodity mobile phones. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 545–554, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1580-7. doi: 10.1145/2380116.2380184. URL http://dx.doi.org/10.1145/2380116.2380184.

[30] D. Goldberg. *Unistrokes for computerized interpretation of handwriting (US Patent #5596656)*. US Patent and Trademark Office, January 1997. URL http://www.patentlens.net/patentlens/patent/US_5596656/en/.

[31] F. Guenther, J. Brumberg, E. Wright, A. Nieto-Castanon, J. Tourville, M. Panko, R. Law, S. Siebert, J. Bartels, D. Andreasen, P. Ehirim, H. Mao, and P. Kennedy. A wireless Brain-Machine interface for Real-Time speech synthesis. *PLoS ONE*, 4(12), December 2009. doi: 10.1371/journal.pone.0008218. URL http://dx.doi.org/10.1371/journal.pone.0008218.

[32] G. Guo, J. Zhang, and N. Yorke-Smith. A novel bayesian similarity measure for recommender systems. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*, 2013.

[33] J. P. Hansen, H. Lund, H. Aoki, and K. Itoh. Gaze communication systems for people with ALS. In *ALS Communication Workshop, Yokohama, Japan*, pages 35–38, December 2006. URL http://www.cogain.org/w/images/a/a2/ALS_Workshop_Yokohama2006.pdf.

[34] S. Hart and L. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human mental workload*, 1(3):139–183, 1988.

[35] R. Hemayati, W. Meng, and C. Yu. Semantic-based grouping of search engine results using WordNet. In *Proceedings of the joint 9th Asia-Pacific web and 8th international conference on web-age information management conference on Advances in data and web management*, APWeb/WAIM'07, pages 678–686, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-72483-4. URL http://portal.acm.org/citation.cfm?id=1769795.

[36] J. Higginbotham, H. Shane, S. Russell, and K. Caves. Access to AAC: Present, past, and future. *Augmentative and Alternative Communication*, 23(3):243–257, January 2007. ISSN 0743-4618. doi: 10.1080/07434610701571058. URL http://dx.doi.org/10.1080/07434610701571058.

[37] J. Higginbotham, A. Bisantz, M. Sunm, K. Adams, and F. Yik. The effect of context priming and task type on augmentative communication performance. *Augmentative and Alternative Communication*, 25(1):19–31, 2009.

[38] T. Holtgraves. Text messaging, personality, and the social context. *Journal of Research in Personality*, 45(1): 92–99, 2011. doi: http://dx.doi.org/10.1016/j.jrp.2010.11.015. URL http://www.sciencedirect.com/science/article/pii/S0092656610001698.

[39] Y. How and M. Kan. Optimizing predictive text entry for short message service on mobile phones. In *Human Computer Interfaces International (HCII 05)*, 2005. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.638.

[40] A. Järvelin, A. Järvelin, and K. Järvelin. s-grams: Defining generalized n-grams for information retrieval. *Information Processing & Management*, 43(4):1005–1019, July 2007. ISSN 03064573. doi: 10.1016/j.ipm.2006.09.016. URL http://dx.doi.org/10.1016/j.ipm.2006.09.016.

[41] A. Jinks and B. Sinteff. Consumer response to AAC devices: Acquisition, training, use, and satisfaction. *Augmentative and Alternative Communication*, 10(3):184–190, January 1994. doi: 10.1080/07434619412331276890. URL http://dx.doi.org/10.1080/07434619412331276890.

[42] D. Jurafsky, J. H. Martin, A. Kehler, K. Vander Linden, and N. Ward. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*, volume 2. MIT Press, 2000.

[43] M. Kamvar and S. Baluja. Deciphering trends in mobile search. *Computer*, 40(8):58–62, 2007.

[44] S. K. Kane, C. Jayant, J. Wobbrock, and R. E. Ladner. Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 115–122. ACM, 2009.

[45] S. K. Kane, B. L. Church, K. Althoff, and D. McCall. What we talk about: designing a context-aware communication tool for people with aphasia. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '12, pages 49–56, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1321-6. doi: 10.1145/2384916.2384926. URL http://dx.doi.org/10.1145/2384916.2384926.

[46] G. Karberis and G. Kouroupetroglou. Transforming spontaneous telegraphic language to Well-Formed greek sentences for alternative and augmentative communication. In *Proceedings of the Second Hellenic Conference on AI: Methods and Applications of Artificial Intelligence*, SETN '02, pages 155–166, London, UK, UK, 2002. Springer-Verlag. ISBN 3-540-43472-0. URL http://portal.acm.org/citation.cfm?id=645861.670294.

[47] T. Kiss and J. Strunk. Unsupervised multilingual sentence boundary detection. *Computational Linguistics*, 32(4):485–525, December 2006. ISSN 0891-2017. doi: 10.1162/coli.2006.32.4.485. URL http://dx.doi.org/10.1162/coli.2006.32.4.485.

[48] H. Koester and S. Levine. Effect of a word prediction feature on user performance. *Augmentative and Alternative Communication*, 12(3):155–168, 1996.

[49] H. Koester and S. Levine. Keystroke-level models for user performance with word prediction. *Augmentative and Alternative Communication*, 13(4), 1997.

[50] P. O. Kristensson and K. Vertanen. The potential of dwell-free eye-typing for fast assistive gaze communication. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 241–244, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1221-9. doi: 10.1145/2168556.2168605. URL http://dx.doi.org/10.1145/2168556.2168605.

[51] P. O. Kristensson and S. Zhai. SHARK2: a large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, pages 43–52, New York, NY, USA, 2004. ACM. ISBN 1-58113-957-8. doi: 10.1145/1029632.1029640. URL http://dx.doi.org/10.1145/1029632.1029640.

[52] S. Kullback. The Kullback-Leibler distance. *The American Statistician*, 41(4):340–341, 1987.

[53] C. Kushler and R. Marsden. *System and method for continuous stroke word-based text input (US Patent #7453439)*. US Patent and Trademark Office, November 2008. URL http://www.patentlens.net/patentlens/patent/US_7453439/en/.

[54] G. Lesher and G. Rinkus. Domain-Specific word prediction for augmentative communication. In *Proceedings of the RESNA 2002 Annual Conference*, 2002.

[55] G. Lesher and C. Sanelli. A Web-Based system for autonomous text corpus generation. In *Proceedings of ISAAC*, 2000. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.15.9386.

[56] G. Lesher, B. Moulton, and J. Higginbotham. Techniques for augmenting scanning communication. *Augmentative and Alternative Communication*, 14(2):81–101, January 1998. doi: 10.1080/07434619812331278236. URL http://dx.doi.org/10.1080/07434619812331278236.

[57] J. Li and G. Hirst. Semantic knowledge in word completion. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, Assets '05, pages 121–128, New York, NY, USA, 2005. ACM. ISBN 1-59593-159-7. doi: 10.1145/1090785.1090809. URL http://dx.doi.org/10.1145/1090785.1090809.

[58] J. Light and K. Drager. AAC technologies for young children with complex communication needs: State of the science and

future research directions. *Augmentative and alternative communication*, 23(3):204–216, 2007.

[59] J. Light, D. Beukelman, and J. Reichle. *Communicative competence for individuals who use AAC: From research to effective practice*. Paul H. Brookes Publishing Co., 2003.

[60] C. Lin and E. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 71–78, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. doi: 10.3115/1073445.1073465. URL http://dx.doi.org/10.3115/1073445.1073465.

[61] G. Lindsay, J. Dockrell, M. Desforges, J. Law, and N. Peacey. Meeting the needs of children and young people with speech, language and communication difficulties. *International Journal of Language & Communication Disorders*, 45(4):448–460, July 2010. doi: 10.3109/13682820903165693. URL http://dx.doi.org/10.3109/13682820903165693.

[62] R. Ling and N. S. Baron. Text messaging and IM linguistic comparison of american college data. *Journal of Language and Social Psychology*, 26(3):291–298, 2007.

[63] Y. Lv and C. Zhai. Positional language models for information retrieval. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 299–306, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. doi: 10.1145/1571941.1571994. URL http://dx.doi.org/10.1145/1571941.1571994.

[64] I. S. MacKenzie and T. Felzer. SAK: Scanning ambiguous keyboard for efficient one-key text entry. *ACM Transactions on Computer-Human Interaction*, 17(3), July 2010. ISSN 1073-0516. doi: 10.1145/1806923.1806925. URL http://dx.doi.org/10.1145/1806923.1806925.

[65] I. S. MacKenzie and R. W. Soukoreff. Text entry for mobile computing: Models and methods, theory and practice. *Human-computer Interaction*, 17:147–198, 2002. doi: 10.1207/S15327051HCI172\&3_2. URL http://dx.doi.org/10.1207/S15327051HCI172&3_2.

[66] I. S. MacKenzie and R. W. Soukoreff. Phrase sets for evaluating text entry techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 754–755, New York, NY, USA, 2003. ACM. ISBN 1-58113-637-4. doi: 10.1145/765891.765971. URL http://dx.doi.org/10.1145/765891.765971.

[67] B. MacWhinney. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum, 2000.

[68] B. MacWhinney. The TalkBank project. creating and digitizing language corpora: Volume 1, synchronic databases, 2007.

[69] C. Marvin, D. Beukelman, and D. Bilyeu. Vocabulary-use patterns in preschool children: Effects of context and time sampling. *Augmentative and Alternative Communication*, 10(4):224–236, January 1994. doi: 10.1080/07434619412331276930. URL http://dx.doi.org/10.1080/07434619412331276930.

[70] J. Matas, P. Mathy-Laikko, D. Beukelman, and K. Legresley. Identifying the nonspeaking population: a demographic study. *Augmentative and Alternative Communication*, 1(1):17–31, December 1985. doi: 10.1080/07434618512331273491. URL http://dx.doi.org/10.1080/07434618512331273491.

[71] J. Matiasek and M. Baroni. Exploiting long distance collocational relations in predictive typing. In *Proceedings of the 2003 EACL Workshop on Language Modeling for Text Entry Methods*, TextEntry '03, pages 1–8, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. URL http://portal.acm.org/citation.cfm?id=1628196.

[72] K. McCoy, C. Pennington, and A. Badman. Compansion: From research prototype to practical integration. *Natural Language Engineering*, 4(01):73–95, 1998. URL http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=48437.

[73] M. Mehl, S. Vazire, N. Ramírez-Esparza, R. Slatcher, and J. Pennebaker. Are women really more talkative than men? *Science*, 317(5834):82, July 2007. ISSN 1095-9203. doi: 10.1126/science.1139940. URL http://dx.doi.org/10.1126/science.1139940.

[74] D. Metzler and W. B. Croft. A markov random field model for term dependencies. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 472–479. ACM, 2005.

[75] J. B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2011.

[76] G. A. Miller. WordNet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[77] A. Mislove, S. Lehmann, Y. Ahn, J. Onnela, and J. N. Rosenquist. Understanding the demographics of twitter users. In *ICWSM*, 2011.

[78] K. R. Muller and B. Blankertz. Toward noninvasive brain-computer interfaces. *Signal Processing Magazine, IEEE*, 23(5), 2006. ISSN 1053-5888. doi: 10.1109/msp.2006.1708426. URL http://dx.doi.org/10.1109/msp.2006.1708426.

[79] H. Nicolau and J. Jorge. Elderly text-entry performance on touchscreens. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 127–134, New York, NY, USA, 2012. ACM. doi: 10.1145/2384916.2384939. URL http://doi.acm.org/10.1145/2384916.2384939.

[80] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 249–256, New York, NY, USA, 1990. ACM. doi: 10.1145/97243.97281. URL http://doi.acm.org/10.1145/97243.97281.

[81] S. Nikolova, M. Tremaine, and P. Cook. Click on bake to get cookies: guiding word-finding with semantic associations. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 155–162, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-881-0. doi: 10.1145/1878803.1878832. URL http://dx.doi.org/10.1145/1878803.1878832.

[82] D. Norris and J. M. McQueen. Shortlist b: a bayesian model of continuous speech recognition. *Psychological Review*, 115(2):357, 2008.

[83] U. Orhan, K. E. Hild, D. Erdogmus, B. Roark, B. Oken, and M. Fried-Oken. RSVP keyboard: An EEG based typing interface. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 645–648. IEEE, 2012.

[84] R. Patel and R. Radhakrishnan. Enhancing access to situational vocabulary by leveraging geographic context. *Assistive Technology Outcomes and Benefits*, 4(1):99–114, 2007.

[85] R. Patel, S. Pilato, and D. Roy. Beyond linear syntax: An Image-Oriented communication aid. *Journal of Assistive Technology Outcomes and Benefits*, 1:57–66, 2004.

[86] D. B. Paul and J. M. Baker. The design for the wall street journal-based CSR corpus. In *Proceedings of the workshop on Speech and Natural Language*, pages 357–362. Association for Computational Linguistics, 1992.

[87] M. F. Porter. An algorithm for suffix stripping. In K. S. Jones and P. Willett, editors, *Readings in information retrieval*, pages 313–316. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997. ISBN 1-55860-454-5. URL http://portal.acm.org/citation.cfm?id=275705.

[88] D. M. W. Powers. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

[89] Ranks. English stopwords (http://www.ranks.nl), March 2012. URL http://www.ranks.nl/resources/stopwords.html.

[90] D. Rashid and N. Smith. Relative keyboard input system. In *Proceedings of the 13th international conference on Intelligent user interfaces*, IUI '08, pages 397–400, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-987-6. doi: 10.1145/1378773.1378839. URL http://dx.doi.org/10.1145/1378773.1378839.

[91] B. Roark, J. D. Villiers, C. Gibbons, and M. F. Oken. Scanning methods and language modeling for binary switch typing. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, SLPAT '10, pages 28–36, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. URL http://portal.acm.org/citation.cfm?id=1867754.

[92] B. C. Roy, M. C. Frank, and D. Roy. Relating activity contexts to early word learning in dense longitudinal data. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, 2012.

[93] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, November 1975. ISSN 0001-0782. doi: 10.1145/361219.361220. URL http://dx.doi.org/10.1145/361219.361220.

[94] E. Sandhaus. The new york times annotated corpus. *Linguistic Data Consortium, Philadelphia*, 6(12), 2008.

[95] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[96] J. Schler, M. Koppel, S. Argamon, and J. Pennebaker. Effects of age and gender on blogging. In *Proc. of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, March 2006.

[97] W. Schramm. How communication works. the process and effects of mass communication. *Urbana: University of Illinois Press*, 1954.

[98] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar. Personality, gender, and age in the language of social media: The Open-Vocabulary approach. *PLoS ONE*, 8(9), September 2013. doi: 10.1371/journal.pone.0073791. URL http://dx.doi.org/10.1371/journal.pone.0073791.

[99] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, July 1948. doi: 10.1145/584091.584093. URL http://dx.doi.org/10.1145/584091.584093.

[100] C. E. Shannon and W. Weaver. The mathematical theory of communication. *University of Illinois Press*, 19(7):1, 1949.

[101] M. Steinbach, L. Ertöz, and V. Kumar. The challenges of clustering high dimensional data. In *New Directions in Statistical Physics*, pages 273–309. Springer, 2004.

[102] C. Suen. n-Gram statistics for natural language understanding and text processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):164–172, April 1979. ISSN 0162-8828. doi: 10.1109/tpami.1979.4766902. URL http://dx.doi.org/10.1109/tpami.1979.4766902.

[103] S. Tata and J. M. Patel. Estimating the selectivity of TF-IDF based cosine similarity predicates. *ACM SIGMOD Record*, 36(2): 7–12, June 2007. ISSN 0163-5808. doi: 10.1145/1328854.1328855. URL http://dx.doi.org/10.1145/1328854.1328855.

[104] C. Thurlow and A. Brown. Generation txt? the sociolinguistics of young people's text-messaging. *Discourse analysis online*, 1(1): 30, 2003.

[105] J. Todman. Rate and quality of conversations using a text-storage AAC system: Single-case training study. *Augmentative and Alternative Communication*, pages 164–179, September 2000. ISSN 0743-4618. doi: 10.1080/07434610012331279024. URL http://dx.doi.org/10.1080/07434610012331279024.

[106] J. Todman, N. Alm, and L. Elder. Computer-aided conversation: A prototype system for nonspeaking people with physical disabilities. *Applied Psycholinguistics*, 15(01):45–73, 1994. doi: 10.1017/s0142716400006974. URL http://dx.doi.org/10.1017/s0142716400006974.

[107] T. Tomokiyo and M. Hurst. A language model approach to keyphrase extraction. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18*, MWE '03, pages 33–40, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. doi: 10.3115/1119282.1119287. URL http://dx.doi.org/10.3115/1119282.1119287.

[108] H. Trinh, A. Waller, K. Vertanen, P. A. Kristensson, and V. Hanson. iSCAN: a phoneme-based predictive communication aid for nonspeaking individuals. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '12, pages 57–64, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1321-6. doi: 10.1145/2384916.2384927. URL http://dx.doi.org/10.1145/2384916.2384927.

[109] K. Trnka and K. McCoy. Corpus studies in word prediction. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 195–202, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-573-1. doi:

10.1145/1296843.1296877. URL http://dx.doi.org/10.1145/1296843.1296877.

[110] K. Trnka, D. Yarrington, K. McCoy, and C. Pennington. Topic modeling in fringe word prediction for AAC. In *Proceedings of the 11th international conference on Intelligent user interfaces*, IUI '06, pages 276–278, New York, NY, USA, 2006. ACM. ISBN 1-59593-287-9. doi: 10.1145/1111449.1111509. URL http://dx.doi.org/10.1145/1111449.1111509.

[111] E. Tzoukermann, J. Klavans, and C. Jacquemin. Effective use of natural language processing techniques for automatic conflation of multi-word terms: the role of derivational morphology, part of speech tagging, and shallow parsing. *SIGIR Forum*, 31 (SI):148–155, July 1997. ISSN 0163-5840. doi: 10.1145/278459.258554. URL http://dx.doi.org/10.1145/278459.258554.

[112] O. Udwin and W. Yule. Augmentative communication systems taught to cerebral palsied children - a longitudinal study:I. the acquisition of signs and symbols, and syntactic aspects of their use over time. *British Journal of Disorders of Communication*, 25(3): 295–309, January 1990. doi: 10.3109/13682829009011979. URL http://dx.doi.org/10.3109/13682829009011979.

[113] H. Van Balkom and M. Welle Donker-Gimbrere. A psycholinguistic approach to graphic language use. *Augmentative and alternative communication: European Perspectives*, pages 153–170, 1996.

[114] A. Van Den Bosch. Scalable classification-based word prediction and confusible correction. *Traitement Automatique des Langues*, 46(2):39–63, 2006.

[115] A. Van Den Bosch and P. Berck. Memory-based machine translation and language modeling. In *The Prague Bulletin of Mathematical Linguistics*, 2009. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.189.5165.

[116] K. Vertanen and P. O. Kristensson. The imagination of crowds: Conversational AAC language modeling using crowdsourcing and large data sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 700–711. ACL, 2011.

[117] C. Wacharamanotham, J. Hurtmanns, A. Mertens, M. Kronenbuerger, C. Schlick, and J. Borchers. Evaluating swabbing: A touchscreen input method for elderly users with tremor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 623–626, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979031. URL http://dx.doi.org/10.1145/1978942.1979031.

[118] T. Walsh. Utterance-based systems: organization and design of AAC interfaces. In *Proceedings of the 12th international ACM*

*SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 327–328, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-881-0. doi: 10.1145/1878803.1878895. URL http://dx.doi.org/10.1145/1878803.1878895.

[119] T. Wandmacher and J. Antoine. Training language models without appropriate language resources: Experiments with an AAC system for disabled people. In *Proceedings of LREC*, 2006. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.62.1815.

[120] G. Ward. Moby word list: American english (http://www.gutenberg.org/ebooks/3201). Public domain in the USA, 2002.

[121] S. J. Westerman and T. Cribbin. Mapping semantic information in virtual space: dimensions, variance and individual differences. *International Journal of Human-Computer Studies*, 53(5): 765–787, November 2000. ISSN 10715819. doi: 10.1006/ijhc.2000.0417. URL http://dx.doi.org/10.1006/ijhc.2000.0417.

[122] Widgit. About symbols (http://www.widgit.com), March 2012.

[123] K. Wiegand. Semantic disambiguation of non-syntactic and continuous motion text entry for AAC. *ACM SIGACCESS Accessibility and Computing*, 105:38–43, January 2013. doi: 10.1145/2444800.2444808. URL http://doi.acm.org/10.1145/2444800.2444808.

[124] K. Wiegand. Intelligent assistive communication and the web as a social medium. In *Proceedings of the 11th Web for All Conference*, W4A '14, New York, NY, USA, 2014. ACM. doi: 10.1145/2596695.2596725. URL http://doi.acm.org/10.1145/2596695.2596725.

[125] K. Wiegand and R. Patel. Non-Syntactic word prediction for AAC. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies*, pages 28–36, Montréal, Canada, June 2012. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-2905.

[126] K. Wiegand and R. Patel. SymbolPath: A continuous motion overlay module for icon-based assistive communication. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 209–210, New York, NY, USA, 2012. ACM. doi: 10.1145/2384916.2384957. URL http://doi.acm.org/10.1145/2384916.2384957.

[127] K. Wiegand and R. Patel. DigitCHAT: Enabling AAC input at conversational speed. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, pages 343–344, New York, NY, USA, 2014. ACM. doi: 10.1145/2661334.2661422. URL http://doi.acm.org/10.1145/2661334.2661422.

[128] Karl Wiegand. *Disambiguation of Imprecise User Input Through Intelligent Assistive Communication*. PhD thesis, Northeastern University, Boston, MA, 2014.

[129] W. J. Wilbur and K. Sirotkin. The automatic identification of stop words. *Journal of information science*, 18(1):45–55, 1992.

[130] B. Wisenburn and J. J. Higginbotham. An AAC application using speaking partner speech recognition to automatically produce contextually relevant utterances: Objective results. *Augmentative and alternative communication*, 24(2):100–109, January 2008. ISSN 1477-3848. doi: 10.1080/07434610701740448. URL http://dx.doi.org/10.1080/07434610701740448.

[131] B. Wisenburn and J. J. Higginbotham. Participant evaluations of rate and communication efficacy of an AAC application using natural language processing. *Augmentative and alternative communication*, 25(2):78–89, June 2009. ISSN 1477-3848. doi: 10.1080/07434610902739876. URL http://dx.doi.org/10.1080/07434610902739876.

[132] L. Wolk, N. B. Abdelli-Beruh, and D. Slavin. Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, 26(3):e111–e116, 2012. doi: http://dx.doi.org/10.1016/j.jvoice.2011.04.007. URL http://www.sciencedirect.com/science/article/pii/S0892199711000701.

[133] J. Wolpaw. Brain-computer interfaces (BCIs) for communication and control. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 1–2, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-573-1. doi: 10.1145/1296843.1296845. URL http://dx.doi.org/10.1145/1296843.1296845.

[134] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, June 2002. ISSN 13882457. doi: 10.1016/s1388-2457(02)00057-3. URL http://dx.doi.org/10.1016/s1388-2457(02)00057-3.

[135] T. Yarkoni. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44(3):363–373, 2010. doi: http://dx.doi.org/10.1016/j.jrp.2010.04.001. URL http://www.sciencedirect.com/science/article/pii/S0092656610000541.

[136] Yelp. Yelp's academic dataset (https://www.yelp.com/academic_dataset), March 2013.

[137] A. N. K. Zaman, P. Matsakis, and C. Brown. Evaluation of stop word lists in text retrieval using latent semantic indexing. In *Digital Information Management (ICDIM), 2011 Sixth International Conference on*, pages 133–136. IEEE, 2011.

[138] X. Zhang and S. MacKenzie. Evaluating eye tracking with ISO 9241 - part 9. In *Proceedings of the 12th international conference on Human-computer interaction: intelligent multimodal interaction environments*, HCI'07, pages 779–788, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-73108-5. URL http://portal.acm.org/citation.cfm?id=1769678.