Augmentative and Alternative Communication with automated vocabularies from photographs

Maurício Fontana de Vargas

School of Information Studies

McGill University

Montreal, Canada

January 2023

A thesis submitted to McGill University in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Abstract

Augmentative and Alternative Communication (AAC) tools can enhance communication for non-speaking individuals, thus offering improved social interaction and independence. While those who experience primarily physical barriers to communication (e.g., people with ALS) can compose complex and nuanced sentences through text-based systems, those with developmental disabilities (e.g., autism, cerebral palsy) or lexical processing impairments currently depend on symbol-based systems that provide word-image pairs in grids grouped by categories. These systems impose meta-linguistic and memory demands on users when finding desired words, and demand the pre-programming of relevant vocabulary into the devices, hindering spontaneous communication and preventing conversation partners from stimulating the use of symbolic language in meaningful moments. Previous studies show that symbolic vocabulary prompted by photographs can alleviate these issues, but no system to date has yet implemented automatic vocabulary generation. This thesis thus presents and explores the first AAC tool able to generate vocabulary symbols automatically from input photographs with the goal of supporting language learning and use for people with communication disabilities. The thesis makes four contributions: i) a novel method that automatically generates vocabulary related to a given photograph, as demonstrated by a thorough performance analysis under different system configurations and a wide range of visual inputs; ii) Click AAC, a novel AAC application that generates situation-specific communication boards formed by a combination of descriptive, narrative, and semantically relevant words and phrases inferred automatically from photographs through the proposed generation method; iii) a nuanced understanding

of how vocabularies generated automatically from photographs can support individuals with complex communication needs in using and learning symbolic AAC, offering insights into the design of automatic vocabulary generation methods and interfaces to better support various goals and scenarios of use, and iv) an ambitious approach for evaluating AAC systems in naturalistic settings with minimal intervention from researchers.

Résumé

Les outils de Communication Améliorée et Alternative (CAA) peuvent améliorer la communication pour les personnes qui ne parlent pas, offrant ainsi une meilleure interaction sociale et une plus grande indépendance. Alors que ceux qui rencontrent principalement des obstacles physiques à la communication (par exemple, les personnes atteintes de SLA) peuvent composer des phrases complexes et nuancées grâce à des systèmes basés sur le texte, ceux qui souffrent de troubles du développement (par exemple, autisme, infirmité motrice cérébrale) ou de déficiences du traitement lexical dépendent actuellement de systèmes basés sur des symboles qui fournissent des paires mot-image dans des grilles regroupées par catégories. Ces systèmes imposent des exigences métalinguistiques et mémorielles aux utilisateurs lorsqu'il s'agit de trouver les mots souhaités, et exigent la pré-programmation du vocabulaire pertinent dans les dispositifs, ce qui entrave la communication spontanée et empêche les interlocuteurs de stimuler l'utilisation du langage symbolique dans les moments significatifs. Des études antérieures montrent que le vocabulaire symbolique suscité par des photographies peut atténuer ces problèmes, mais aucun système à ce jour n'a encore mis en œuvre la génération automatique de vocabulaire. Cette thèse présente et explore donc le premier outil de CAA capable de générer automatiquement des symboles de vocabulaire à partir de photographies, dans le but de soutenir l'apprentissage et l'utilisation du langage pour les personnes ayant des difficultés de communication. La thèse apporte quatre contributions: i) une nouvelle méthode qui génère automatiquement du vocabulaire lié à une photographie donnée, comme le démontre une analyse approfondie des performances sous différentes configurations du système et

une large gamme d'entrées visuelles; ii) Click AAC, une nouvelle application de CAA qui génère des panneaux de communication spécifiques à la situation, formés par une combinaison de mots et de phrases descriptifs, narratifs et sémantiquement pertinents, déduits automatiquement des photographies par la méthode de génération proposée ; iii) une compréhension nuancée de la façon dont les vocabulaires générés automatiquement à partir de photographies peuvent aider les personnes ayant des besoins de communication complexes à utiliser et à apprendre la CAA symbolique, offrant des idées pour la conception de méthodes de génération automatique de vocabulaire et d'interfaces pour mieux soutenir divers objectifs et scénarios d'utilisation, et iv) une approche ambitieuse pour évaluer les systèmes de CAA dans des contextes naturalistes avec une intervention minimale des chercheurs.

Acknowledgments

First, I would like to thank my supervisor, Prof. Dr. Karyn Moffatt and the School of Information Studies (SIS) for the once-in-a-lifetime opportunity to be a PhD student in one of the world's most prestigious universities. Prof. Dr. Karyn Moffatt always provided me encouragement and thoughtful advice, working hard to guarantee the support I needed throughout the entirety of my doctoral studies. I admire her and I am deeply grateful for the opportunity to work with her. I also want to acknowledge the contributions of my advisors Prof. Dr. Ilja Frissen and Prof. Dr. Catherine Guastavino for their feedback and discussions that helped me shape this thesis, in addition to the knowledge I gained in their course lectures.

Thanks to my peers at the ACT Lab and SIS for the valuable feedback at several stages of my research, and the support to deal with the numerous challenges we encounter in academia. In special, Jiamin (Carrie) Dai, Peymon Montazeri, Afroza Sultana, Robert Ferguson, and Aditya Birangal. I would like to extend my thanks to members of the Shared Reality Lab: Prof. Dr. Jeremy Cooperstock, David Marino, Jeffrey Blum, Antoine Weill-Duflos, and Pascal Fortin. It was great learning from and working with you.

My experience at McGill would have not been as enriching without the friends I met through Intramurals Soccer, who quickly became part of my family in Canada. Thanks Constanza (Coni) Martinez-Ramirez, Ingrid Jauffret, Jo Griffin, Yassemine Khawajkie, Cristiana Spinelli, and Camilo Gómez for all the fun and the happy moments we shared.

For the last six years, old friendships have got stronger and helped forming the solid founda-

tion I needed to face tough and uncertain times. Thanks Gustavo Zucco, Luísa Fontoura, Kévin Serre, Albino Szesz Junior, Henrique Zanetti, and Julia Coelho for all the encouragement and the unforgettable stories added to our histories.

Thanks to the most amazing person I have ever met; whose love, support, and comprehension have been unconditional for over a decade. It has been a delight to walk though life's maze holding your hand, Joana Tomaz Tancredo. My journey in academia would not be possible without you. I am forever grateful for what we built together and for what is yet to come.

Everything I have accomplished in life would not be possible without the support of my family. Neri and Beatriz, thanks for raising me in an environment where care, dedication, and love always thrived. Thank you for igniting my curiosity in the world and for allowing me to take risks while pointing me to the right direction. To my sister, Camila, I wish I could express to you the importance you have in my life ♥. Conducting the research presented in this thesis was heavily influenced by the experience of being your brother. Your kindness, happiness, and innocence in its purest form have always inspired me and shown me what is really important in our lives. I hope this work contributes to your community a bit of what you have contributed to me.

Finally, I would like to acknowledge that this research was funded by the Fonds de Recherche du Québec - Nature et Technologies (FRQNT), the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chairs Program (CRC), AGE-WELL NCE (Canada's technology and aging network), and the Microsoft AI for Accessibility Grant. I am thankful for Canada's and Québec's government and taxpayers for providing me these funds.

Contents

List of Figures

List of Tables

List of Acronyms

AAC Augmentative and Alternative Communication

ALS Amyotrophic Lateral Sclerosis

ASD Austism Spectrum Disorder

BLE Bluetooth Low Energy

HCI Human-Computer Interaction

NLG Natural Language Generation

NLP Natural Language Processing

RFID Radio Frequency Identification

SNUG Spontaneaous Novel Utterance Generation

VSD Visual Scene Display

Chapter 1

Introduction

One uniquely human ability is communication through a shared and complex language [7]. Communicating is fundamental not only for the exchange of information and achievement of basic needs, such as requesting and sharing information, but mostly importantly for connecting to others, building relationships, and creating and participating in communities. Indeed, the Universal Declaration of Human Rights entitles every human being "to seek, receive and impart information and ideas through any media and regardless of frontiers".

However, people with various disabilities often encounter major barriers to effective communication because their abilities and preferences are not entirely accepted and supported by the environment, or they lack access to services required to achieve their communication potential. These people with complex communication needs often rely on Augmentative and Alternative Communication (AAC)—a variety of unaided strategies such as gestures, sign language, and body expression, or the use of tools ranging from paper documents composed of images, drawings, or

photographs to speech-generating applications and devices. While those who experience primarily physical barriers to communication (e.g., people with ALS) can compose complex and nuanced sentences through text-based systems, those with lexical processing impairments that limit their ability to spell out words (e.g., adults with aphasia[1]) or developmental disabilities (e.g., ASD, cerebral palsy) depend on symbol-based systems that display word-image pairs in grids grouped by category (e.g., actions, sports, locations), or visual scene displays that associate vocabulary with a contextually-rich photograph (e.g., playing soccer at the park).

Although current AAC tools provide opportunities for improved communication and quality of life, many challenges still hinder their mass acceptance and adoption [9, 2, 8, 5]. AAC users, their families, and speech-language professionals raise several issues, including i) extremely slow communication [2, 3, 12], ii) difficulties navigating symbolic vocabulary to find desired words [1, 13, 10], and iii) time and effort needed to program the tools with relevant vocabulary [1, 4].

To address these problems, researchers have explored the use of contextual information to tailor communication support according to user's immediate needs, aiming to reduce the time and effort required to program and access relevant vocabulary. In text-based AAC devices, researchers have mostly employed language models trained with data from conversations on a specific topic or in a certain location, improving standard next letter or word prediction and consequently reducing the number of keystrokes needed to write messages. On the other hand, prediction of vocabulary for symbol-based systems is largely unexplored. Current commercial devices and previous research have employed contextual information mostly for providing vocabulary that was manually pre-

[1] a language disorder affecting lexical and semantic processing that is mostly often caused by a stroke [6].

assigned to categories representing locations, conversation topics, or partners.

Automatic generation of vocabulary has been explored only for limited use cases. For example, one set of researchers [11] proposed a system that creates short narratives about students' activities at school based on data collected through dedicated sensors and a domain model of objects, people, and locations within the school, in addition to the students timetables. Another study proposed algorithms to retrieve relevant vocabulary from websites and Wikipedia pages associated with the user's location or current conversation topic, which limits its applicability to contexts for which internet-accessible corpora are likely to exist (e.g., retail location). To date, there has been no research on the creation of AAC tools that automatically generate vocabulary for use in a broad variety of communication contexts. The design of such tools, both in terms of generation methods and interactive interfaces, and the factors of the dynamics between individuals with complex communication needs, their conversation partners, and automated language support are unexplored. Consequently, the exact kind of support and how these tool could be integrated into real-life settings is unknown.

1.1 Scope

This thesis recognizes that contextual information can provide the means to improve Augmentative and Alternative Communication and poses the central hypothesis that **photographs are a rich and sufficient source of information that can be used to generate contextually relevant vocabulary in an automated manner, without restricting scenarios of use or demanding actions alien to**

everyday life. Such an approach should help make AAC tools more user-friendly and increase the opportunities for users and family members to adopt symbolic AAC. Towards this objective, this thesis proposes the application of artificial intelligence techniques, such as computer vision and natural language processing (NLP), in tandem with Human-Computer Interaction (HCI) methods, to the AAC field for the provision of automated symbolic communication. The thesis expands on existing NLP, HCI, and AAC literature to answer the following research questions:

1. How can vocabulary related to situations depicted in photographs be generated with the goal of supporting symbolic AAC?

2. How can an interactive mobile application be designed to support AAC users capturing personally relevant situations and objects and access symbolic vocabulary relevant to the scene photographed?

3. How can vocabularies automatically generated from photographs support people with complex communication needs, their caregivers, and speech language professionals during their naturally occurring activities?

1.2 Overview

This thesis presents answers to the aforementioned research questions through the introduction of original findings, natural language generation methods, and interactive systems, and by applying creative and ambitious research methodologies. This overview summarizes the content of each chapter.

Chapter 2 presents relevant background literature on i) underlying aspects of Augmentative and Alternative Communication, including the purposes of communication interactions, communicative competence for AAC users, factors related to AAC acceptance and vocabulary selection and usage.; ii) novel techniques for improving AAC tools through context-awareness, and finally, iii) methodologies for studying the effectiveness of those techniques. Later chapters introduce additional work or elaborate on work briefly discussed in this chapter when relevant to the specific topics discussed.

Chapter 3 introduces a novel method for the automated generation of storytelling vocabulary from photographs for use in AAC systems. In addition to being the first method that extends the concepts directly depicted in the photo with a broader vocabulary of commonly used words to create narrative sentences about the scene, results from the benchmark experiment demonstrate that the method is able to generate significantly more relevant vocabulary than AAC devices often offer (i.e., the most common English words).

Chapter 4 explores how to integrate the method introduced in Chapter 2 to automatically generate context-related vocabulary in an interactive mobile application, Click AAC. The chapter analyzes semi-structured interviews with AAC professionals who used the app with their clients with complex communication needs in naturalistic school and speech-language therapy settings. The research findings demonstrate that the immediacy of vocabulary reduces conversation partners' workload, opens up opportunities for AAC stimulation, and facilitates symbolic understanding and sentence construction. In addition to being the first study in which a novel AAC system was deployed "in the wild," with no intervention on how or when participants should use it, this chapter

offer insights into the design of such tools, including the vocabulary generation method and the user interface.

Chapter 5 discusses how the methodologies, systems, and findings introduced in this thesis advance research on state-of-the-art AAC tools and are envisioned to impact future work in the field. The chapter also discusses the practical challenges in conducting research in the intersection of system design, human-computer interaction, and accessibility fields, and how they are approached within the scope of this thesis.

Chapter 6 concludes the thesis by restating its main findings and discussing future research directions enabled by its findings and systems.

Note: Chapters 3–4 consist of published manuscripts. A statement of each co-author's contribution is presented in Section 1.4.

1.3 Summary of Contributions

All elements of this thesis are original scholarship and contribute to advancing research in the fields of AAC, natural language processing, human-computer interaction, and accessibility.

This thesis makes the following empirical, technical, and methodological contributions:

1. A novel method that generates vocabulary automatically from a user's photographs to support autobiographical storytelling, demonstrating how it performs under different combinations of the system's controllable parameters and a wide range of input photographs.

2. Click AAC, a novel AAC application that generates situation-specific communication boards

organized in a Visual Scene Display (VSD)-like layout and formed by a combination of descriptive, narrative, and semantically relevant words and phrases inferred automatically from photographs through the novel generation method proposed.

3. Evidence on how vocabulary automatically generated from photographs can support end users and speech language professionals in their naturalistic environments, and insights into the design of automatic vocabulary generation methods and interactive interfaces to provide adequate support during communication in those naturalistic settings.

4. An approach for evaluating AAC systems, consisting of the system distribution to AAC professionals, who then selected end users for trials and performed the assessment using their own expertise without any researchers' intervention, and reported their findings through online interviews.

1.4 Contribution of Authors

I, Mauricio Fontana de Vargas, was the primary author in the two published papers forming Chapters 3 and 4.

In Chapter 3, I was responsible for the conceptualization, design, and implementation of the generation method and the quantitative evaluation experiment. I wrote all sections of the paper.

In Chapter 4, I was responsible for the conceptualization, design, implementation, and deployment of the mobile application. I also coded the transcripts of all interviews and analyzed the data to find common themes and interpret their meanings, which resulted in the first version of the the-

matic analysis. Jiamin Dai actively reviewed the thematic analysis' codes, themes, interpretations, and the paper's drafts, often offering valuable insights.

Prof. Dr. Karyn Moffatt edited the two manuscripts, reviewed the thematic analysis, and supervised the research at a high level, providing crucial guidance on uncertain moments and when facing roadblocks.

Bibliography

[1] Rita L Bailey, Howard P ParetteJr, Julia B Stoner, Maureen E Angell, and Kathleen Carroll. Family members' perceptions of augmentative and alternative communication device use. *Language, Speech, and Hearing Services in Schools*, 37(1), 2006.

[2] Susan Baxter, Pam Enderby, Philippa Evans, and Simon Judge. Barriers and facilitators to the use of high-technology augmentative and alternative communication devices: a systematic review and qualitative synthesis. *International Journal of Language & Communication Disorders*, 47(2):115–129, 2012.

[3] David R Beukelman, Pat Mirenda, et al. *Augmentative and alternative communication*. Paul H. Brookes Baltimore, 1998.

[4] Filip Bircanin, Bernd Ploderer, Laurianne Sitbon, Andrew A Bayor, and Margot Brereton. Challenges and opportunities in using augmentative and alternative communication (AAC) technologies: Design considerations for adults with severe disabilities. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, pages 184–196, 2019.

[5] Melanie Fried-Oken, David R Beukelman, and Karen Hux. Current and future AAC research considerations for adults with acquired cognitive and communication impairments. *Assistive Technology*, 24(1):56–66, 2012.

[6] Kathryn L Garrett. Adults with severe aphasia. In David R Beukelman and Pat Mirenda, editors, *Augmentative and alternative communication for children and adults with complex communication needs*, pages 467–504. Paul H. Brookes, Baltimore, 2005.

[7] Charles F Hockett. Animal" languages" and human language. *Human Biology*, 31(1):32–39, 1959.

[8] Janice Light and David McNaughton. Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication?, 2014.

[9] Shelley K Lund and Janice Light. Long-term outcomes for individuals who use augmentative and alternative communication: Part iii–contributing factors. *Augmentative and Alternative Communication*, 23(4):323–335, 2007.

[10] Sonya Nikolova, Marilyn Tremaine, and Perry R 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*, pages 155–162, 2010.

[11] Nava Tintarev, Ehud Reiter, Rolf Black, Annalu Waller, and Joe Reddington. Personal storytelling: Using natural language generation for children with complex communication needs, in the wild.... *International Journal of Human-Computer Studies*, 92:1–16, 2016.

[12] Keith Trnka, Debra Yarrington, John McCaw, Kathleen F McCoy, and Christopher Pennington. The effects of word prediction on communication rate for AAC. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, pages 173–176, 2007.

[13] Mieke van de Sandt-Koenderman. High-tech AAC and aphasia: Widening horizons? *Aphasiology*, 18(3):245–263, 2004.

Chapter 2

Background

This literature review chapter lays the foundation for my thesis by exploring broad aspects of the use, design, and evaluation of AAC tools. First, basic notions regarding Augmentative and Alternative Communication are discussed, such as the purposes of communication interactions, the definition of communicative competence of AAC users, and factors related to AAC acceptance (Fig. 2.1 presents an overview of the themes discussed). Then, we present prior research on novel techniques for AAC tools aimed at facilitating communication. Finally, we discuss different methodologies used for the evaluation of these techniques.

While this chapter takes a broader view of literature, later chapters introduce additional work when relevant to the specific topics discussed.

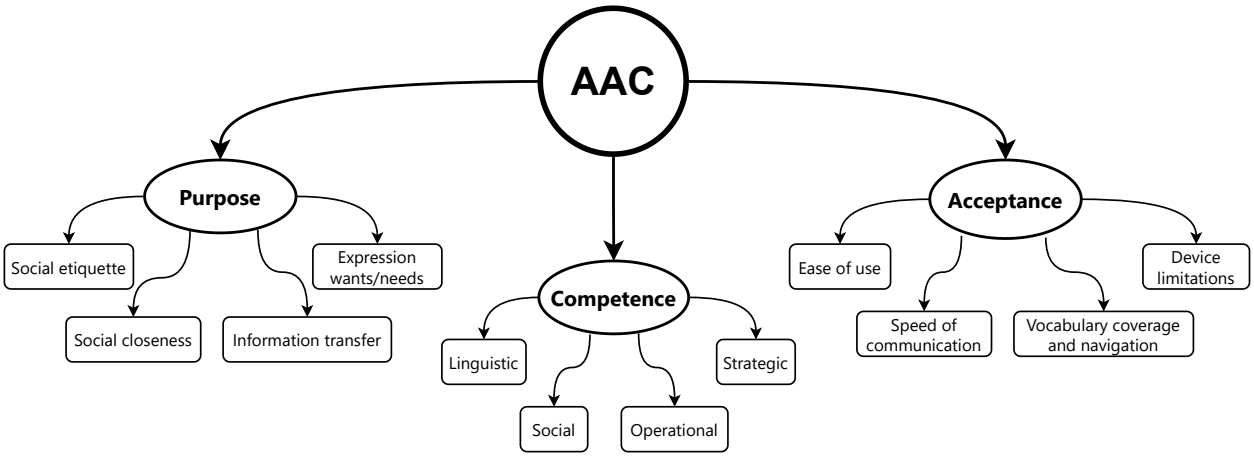


Figure 2.1 Basic notions of Augmentative and Alternative Communication (AAC)

2.1 Augmentative and Alternative Communication

2.1.1 Purposes of communication interactions

The overarching goal of AAC interventions is to improve individuals' communication abilities, supporting effective and efficient participation in various personally relevant interactions. These interactions can be classified in two main groups according to their goals: transactional and interpersonal. The former refers to interactions concerned with the content of exchanges and focuses on the outcomes external to these interactions, e.g., when the individual benefits from something or retrieves information from someone [21]. The latter concerns the interaction itself, focusing on its personal and social aspects. Examples range from immediate enjoyment of social interactions to long-term effects on self-esteem, social relationships, and independence [49].

To properly enhance communication across a wide range of life-situations, AAC tools must cope with the particular requirements and characteristics inherent to those different types of inter-

actions. To better explain the characteristics of communicative interactions, it is possible to classify them based on their social purpose (the first two with transactional goals and the remaining with interpersonal goals), resulting in four categories: (i) expression of needs/wants, (ii) information transfer, (iii) social closeness, and (iv) social etiquette [28].

Expression of needs/wants: The goal in this type of interaction is to control the behavior of the communication partner to provide a desired object or perform an action. Since the focus in such interactions is to fulfill a certain need, the duration of the interactions is limited and the vocabulary used is highly predictable. Both the content and rate of communication are important, and communication breakdowns are usually not tolerated. To enable users to engage in this type of interaction, AAC systems should support the generation of accurate messages at a fast rate, taking advantage of the highly predictable vocabulary used in this type of interaction. Examples that illustrate these characteristics include ordering food in a restaurant and requesting someone to turn on the air conditioning.

Information transfer: The goal in these interactions is to support the exchange of ideas, thoughts, and experiences. Consequently, interactions may be lengthy, and the content and speed of communication are crucial factors in maintaining conversations. Since messages exchanged are complex and difficult to predict, AAC systems must provide the means for composing novel sentences that allow the user to discuss a wide range of topics. Examples of this kind of interaction include a student arriving home from school and telling their parents about their day, a person answering questions in a job interview, and a patient describing their medical symptoms to a doctor.

Social closeness: Communication in this type of interaction is performed to establish, maintain

and/or develop social engagement. The interaction itself and the feelings caused by it are the most important aspects, and thus content, rate of communication, and accuracy of messages are not critical. Examples of people interacting this way include romantic partners expressing amorous feelings and teenagers chatting on the bus to school.

Social etiquette: The goal of the fourth type of interaction is to conform to social conventions of politeness, such as greeting a cashier at the checkout counter and politely conversing with strangers at a bus stop. Interactions are brief, and the vocabulary used is even more limited and predictable than in the interactions expressing needs and wants.

2.1.2 Communicative competence for AAC users

Participating effectively and efficiently in all these types of interaction is not a simple process for people with complex communication needs. It depends on many combined factors, including the user's unique capabilities, communication environment, and the support provided by AAC tools. According to the Communicative Competence Model [29, 31], communicative competence for AAC users requires knowledge, judgment, and skills in four different domains: linguistic, operational, social, and strategic, in addition to a range of psychosocial factors that help them to overcome the numerous constraints and challenges faced when trying to communicate. These four domains entail various opportunities and challenges in terms of AAC device design to improve overall communicative competence according to the unique requirements imposed by different disabilities, as elucidated below.

Linguistic Competence

Linguistic competence is the ability to understand or access the language spoken and written by the people of the community in which the individual lives. This domain includes the linguistic aspects of the individual's AAC system, such as the symbols used to represent vocabulary, as well as the semantic and syntactic facets needed to formulate meaningful messages.

Linguistic competence widely varies among people with complex communication needs. For example, people with communication difficulties resulting from impaired movements of the muscles used for speech production (i.e., dysarthria) may have relatively preserved cognitive and language skills, and can therefore read and compose meaningful written messages. Since these individuals do not require alternative forms for representing vocabulary, they can rely on text-based AAC devices used to create messages by typing the words. Text entry is supported by standard word prediction techniques, such as the ones often used on smartphones. On the other hand, individuals with developmental disabilities (e.g., autism, cerebral palsy) or language disorders (e.g., aphasia) often have reduced abilities in several linguistic and communicative modalities, including speaking, auditory comprehension, reading, and writing. For example, individuals with Wernicke's (receptive) aphasia have no impairments related to language syntax, being able to produce connected speech with complex syntactical forms. However, people with this condition present variable damages in brain areas important for processing the semantics of language, having lost the ability to comprehend speech and attribute meaning to words and sentences. Thus, although they are able to produce sentences at a normal rate and rhythm, their speech is hard to follow,

consisting of seemingly random sentences containing irrelevant and/or made up words and missing important words. On the other hand, individuals with Broca's (expressive) aphasia present good receptive communication skills, but have highly limited or absent grammar, with difficulties combining linguistic elements. Sentences produced are usually brief, containing mainly nouns [3].

To support the improvement of linguistic competence in people with difficulties with written language, AAC devices must present alternative forms of vocabulary representation and provide strategies for composing messages other than spelling or combining written words. For example, AAC devices for people with developmental disabilities or aphasia often support the construction of messages through meaningful images, picture, and animations representing vocabulary items. Providing access to complete utterances and support for creating messages with proper syntax are other important features needed to support the achievement of linguistic competence.

Social Competence

Social competence involves the skills needed for appropriate social interactions—an important issue for people with the autism spectrum disorder (ASD), who present profound social communication deficits due to ASD's impairment of the nature of communication as a social mediator [53], and for those with severe intellectual disabilities. To support the achievement of social competence, AAC devices must provide the means to improve both sociolinguistic and sociorelational skills. Sociolinguistic skills refer to the pragmatic aspects of communication such as discourse skills (e.g., taking turns, initiating and terminating conversations) and the ability to express communicative functions (e.g., requesting, rejecting). Sociorelational skills are the ones needed to

develop effective relationships, such active participation in interactions, personal interest in others, and responsiveness to partners. Thus, AAC devices must address the various social demands of different interactions and contexts. For example, the messages that the AAC device provides for greeting a teacher should be different from those used for greeting a close friend or a family member. In addition, AAC devices should be designed using strategies that allow the user to focus on the conversation partner instead of the device.

Strategic Competence

Strategic competence refers to a range of compensatory strategies needed to overcome the substantial barriers imposed by society and the inherent limitations of AAC systems. Strategies include resolving communication breakdowns, interacting with people unfamiliar with AAC systems, and compensating for the slow communication rate with different modes of communication. People with locked-in syndrome or other severe physical conditions have their strategic competence severely diminished by the impossibility of using body gestures to communicate, in contrast to people with aphasia or other acquired communication disorders who do not present major physical impairments. AAC systems should provide the means to reduce the extent of strategic competence needed by enabling communications with as few as possible breakdowns, which requires appropriate vocabulary coverage and fast message generation.

Operational Competence

Operational competence refers to the technical skills needed to operate an AAC device. It involves the skills needed to properly navigate and select desired vocabulary items through some selection method (e.g., direct selection in a touch screen, eye gaze, single switch) and to maintain the device fully operational, i.e., updating the vocabulary, protecting the device against damage, and repairing it when needed. Given that operational competence is directly related to the devices operation, this domain is the most influenced by the design of AAC tools and thus it plays a fundamental role in the acceptance of AAC.

Summary of communicative competences and corresponding recommendations for AAC

Table 2.1 presents a summary of recommendations for AAC devices for each communicative competence.

Competence	Recommendation for AAC devices
Linguistic	Alternate forms of vocabulary representation (e.g., images, animations); complete sentences ; syntax scaffolding
Social	Facilitated and personalized support for initiating and terminating conversations, requesting, rejecting, and talking about personally-relevant topics
Strategic	Mitigate communication breakdowns; extensive vocabulary coverage, fast message generation
Operational	User-friendly interface; facilitated navigation; availability of technical support

Table 2.1 Summary of communicative competences and corresponding recommendations for AAC devices

2.1.3 Factors related to AAC acceptance

Although Augmentative and Alternative Communication provides opportunities for improved communication and quality of life, the pattern of AAC use is not always optimal, i.e., the devices are not always used spontaneously and at every opportunity. Johnston et al. [24] identified three patterns in which AAC devices are not optimally used by beginner communicators: (i) the person has an AAC system but does not use it; (ii) the person has an AAC system but communication partners do not participate actively in their conversations; (iii) the person uses alternative strategies for communication that are contextually inappropriate.

AAC device abandonment and avoidance happen for various reasons, ranging from dissatisfaction with the device itself to changes in user needs and priorities. In an attempt to explore and clarify the many factors influencing AAC acceptance, Lasker and Bedrosian [27] applied the Matching Person and Technology (MPT) model [43] to the AAC realm, resulting in the AAC Acceptance Model. This model is composed of three components: (i) the milieu, referring to the environment in which the device is used, the attitude of communication partners, and the funding options for acquiring the necessary technology; (ii) the features of the person using the device, including impairment level, age, personality, and skills, and (iii) the technology-related aspects, such as reliability, ease of use and programming, generated voice quality, and cost.

When examining research on the factors impacting the provision and use of AAC devices, we can clearly identify a set of common themes among all settings, ranging from studies focusing on children with developmental disabilities to studies investigating older adults with acquired impair-

ments, such as aphasia. These common themes are discussed next.

Ease of use

The amount of physical and cognitive effort needed to operate and maintain the device appears as a crucial element for enhancing an AAC user's experience and consequently their acceptance of AAC systems. The time involved in programming (e.g., selecting vocabulary, adjusting layouts) and the responsibility for programming the devices are common concerns indicated by family members of children with multiple disabilities who use high-tech AAC devices [4], as well by special education teachers, and speech-language pathologists working as members of an AAC team [5]. The perceived complexity of the communication systems confuses and intimidates families when programming the devices, and prevents their children from using these tools to their full potential [23]. In addition, one third of parents and caregivers of children with communication-related disabilities interviewed by Meder and Wegner [36] pointed to ease of use as the characteristic that mainly guided and influenced them when purchasing their mobile-based AAC devices.

Vocabulary coverage and navigation

The language content included within a symbol- or utterance-based AAC device is of great importance for enabling users to achieve their communication needs and to participate actively in conversations across different settings. However, the range of vocabulary found in traditional AAC devices is often limited and supports the individual to engage only in a limited number of routinized conversations, even requiring other strategies for communication in certain situations [23].

Families of young AAC users often express the need for access to a vocabulary that would enable their children to carry on conversations and increase their interactions with communication partners [4].

The processing and storage capabilities of modern AAC devices offer opportunities for addressing the vocabulary limitation issue. Current high-tech AAC devices are able to store and display numerous vocabulary items, including complete phrases and long sentences, in written or symbolic forms. However, having extensive language content creates a major challenge in the design of AAC devices: how to organize and display the vocabulary items in a manner that provides easy and fast access to words desired during conversations [18].

Since most users of symbol-based AAC devices have great difficulty with written language, alphabetical organization is not possible; instead, vocabulary items are usually organized in a static hierarchy of categories (e.g., "food" → "breakfast" → "croissant") and presented using a grid-based display layout. This strategy does not reflect common usage and is not meaningful for people with intellectual disabilities and people with aphasia, who often find it difficult to relate things with categories, and therefore increases the number of keystrokes and the cognitive load during navigation [38]. For example, to compose a simple statement such as "The croissants in France are amazing", the user would need to drill down into a food category for the "croissant", drill into a places category to find "France" and drill into an adjective category for "amazing", with additional steps needed to fill in the verbs and articles. Even with this issue, clinicians and language pathologists tend to choose traditional grid displays, mostly because they have been available for a long time, and a large research base involves interventions using this kind of display [47].

Speed of communication

Symbol-based AAC's navigation issues and text-based AAC's access difficulties create one of the most critical barriers to AAC use: the extremely low rate of communication. Users of AAC devices typically communicate at rates of 15 words per minute while natural speakers without disabilities can produce 150 to 250 words per minute. This enormous difference between communication rates shortens and delays communication acts while hindering interpersonal interactions, as highlighted by a wide range of users in the literature [6]. Adults with cerebral palsy, for example, demonstrated their frustration regarding the slow communication rate using AAC and identified it as a major challenge in using such devices [12]. Young adults with complex communication needs interviewed by Lund and Light [33] also expressed the need for faster and more accessible AAC technology. Children and adults relying on AAC for communication also expressed the slowness of communication aids as a major problem in a study by Hodge [23].

Therefore, to support operational competence, AAC devices must not only provide a vast vocabulary but also present it in a way that facilitates access across a wide range of daily-life situations, from ordering food in a restaurant to expressing feelings to a partner. Some authors [21, 24] argue that AAC devices should contain vocabulary adapted according to the most important factors (e.g., speed, accuracy) for the user's situation. For example, the breadth of the available vocabulary in one's device may be reduced to facilitate vocabulary navigation, consequently improving communication speed and accuracy when the user is ordering food in a fast-food restaurant. Alternatively, in a situation where a person with aphasia uses an AAC device to interact with family

members at home, the number of available conversation topics would be maximized because, for the family member, listening to a complete and detailed story about the person's day is more important than the speed of communication.

Device-specific limitations

The reliability of AAC devices is often cited as a barrier to AAC acceptance. Issues related to hardware limitations, such as battery running out faster than expected and low volume levels when playing generated messages were some of the problems mentioned by participants in a study by Cooper et al. [11].The frustration with the time needed for repairing devices and when systems were not available or not working properly were also indicated by participants in the study by Dattilo et al. [12]. In addition, the portability of AAC devices is often a source of frustration, as raised by caregivers of young users who reported issues regarding device weight and mounting options [4] and by AAC team members that highlighted the need for devices that are practical to carry around, weigh less and are not as cumbersome [5].

2.1.4 Vocabulary selection and usage

As discussed in the previous section, the quality of vocabulary available on an AAC device has a great impact on its usability. The vocabulary selection process requires careful consideration of the specific topics and words to enable the AAC user to engage in different types of interactions and natural communication exchanges. Due to the great variability in the capabilities and needs of AAC users, vocabulary selection is a complex and dynamic process, depending on several aspects,

including the individual's cognitive limitations, literacy skills, and communicative context.

A first step for providing such optimal vocabulary relies on understanding the patterns of vocabulary use across the general population during everyday life. Several researchers have investigated vocabulary use patterns in terms of frequency and commonality of word use, as well as the topics discussed. Researchers recorded participants' routine conversations using a portable audio recorder, then transcribe the audio, and finally analyze the transcripts using statistical tools. Most of the studies in this realm required participants to wear an audio recorder for several hours a day (4-6 hours) across a range of activities and communication partners, such as watching TV at home with family, eating breakfast in a restaurant, visiting friends, and talking on the telephone with a stranger. The main difference in these studies were the populations being investigated. Elderly women were the focus in Stuart et al. [45]'s study, while two cohorts of older adults (65 to 74 years old and 75 to 85 years old) participated in the study by Stuart et al. [46], preschoolers in Fallon et al. [17], and students between 7 and 14 years in Boenisch and Soto [9].

Despite the great variety in participant's characteristics, the results of these studies were similar and showed that a relatively small number of words accounted for a relatively large portion of communication. The 25 most frequent occurred words represented 46%–54% of communication samples, while the top 100 represented 71%, and the top 250 words represented 80%–89%. In addition, the vocabulary usage was similar across all participants: 65-99 of most frequent words were used by all participants in the same study. Regarding the topics discussed by the participants, Stuart et al. [46] found that conversations about family life were more frequent for the youngest cohort, while conversations about social networks and close friends were more common between

the oldest cohort.

Other researchers have explored whether the vocabulary use patterns are related to the location where the communication happens. Marvin et al. [34] examined the conversation patterns of typically-developing kindergarten children in different locations and times. After recording the speech of ten children at home and school and analyzing the transcribed data, they found that approximately one third of the words were produced only at home, one third only at preschool, and the other one third were used across both locations. Similarly, Patel and Radhakrishnan [41] investigated the spoken corpus of one adult without disabilities across eight locations (bookstore, research lab, classroom, clothing store, electronics store, grocery store, kitchen, lab meeting) and after applying data mining algorithms, concluded that the words most frequently used were different in each location.

As discussed in the previous section, the success of an AAC device depends not only on its ability to provide the right vocabulary content to the user, but also on it presenting this vocabulary efficiently and accurately such that it can be located during communication. With this in mind, researchers also started exploring whether the standard vocabulary arrangements (organized taxonomically) in AAC devices truly reflect the cognitive organization of people who require AAC. Fallon et al. [16], for example, asked twenty young children with no disabilities to organize a set of images representing familiar vocabulary items in any way they wanted. Results showed that 93% of the vocabulary concepts were organized using schematic grouping and only 7% were arranged following the classic taxonomic structure. In the schematic organization, vocabulary items are grouped together in scenes or script that describe the typical sequence of events in common situa-

tions (e.g., dining at a restaurant, going to school) “because they have a function in that scene and they are related to each other as parts within a functional whole” [37]. Thus, their findings strongly support the idea that the intrinsic lexical vocabulary of children users of AAC tools is temporally and spatially organized, which is not reproduced by the standard vocabulary organization on most of the current AAC devices.

The combination of the discussed findings provides some practical implications for the design of vocabularies for AAC devices. First, the existence of a group of words that are commonly used by a variety of individuals and represent a substantial proportion of the vocabulary produced during face to face interaction implies that these core words should be an integral part of any AAC system and should be available to the user in a straight-forward manner. Second, although the core vocabulary allows flexibility across most situations and is able to meet individuals’ needs most of the times, many words produced by natural speakers were not present in the core vocabulary. This indicates that AAC users must have access to specific words and messages that are associated with the individuals’ unique interests, as current events or specific topics. Therefore, fringe vocabulary must be predicted by sources other than statistically derived core lists. Another implication is that specific vocabulary may be arranged into categories according to the conversational topic in order to promote initiation of communication interactions and facilitate the change of topics in the middle of a conversation. Third, the fact that fringe vocabulary may change substantially as the user faces different situations throughout the day suggests that the device should provide, besides the core vocabulary, the words and phrases most likely to be used in the user’s current situation to allow the user to engage in timely and relevant conversations. Finally, considerations regarding how the

vocabulary should be presented to the user in terms of visualization and navigation schemes are of extreme importance in order to reduce the cognitive effort when navigating and searching for vocabulary items.

2.2 Improving AAC devices through context-awareness

The capabilities of modern mobile devices has created great opportunities to improve communicative competence of people with complex communication needs through context-aware computing—the ability of an application to adapt its functionality based on situational and environmental information sensed with the goal of providing tailored support to the immediate users' needs [1]. Examples of contextual information include location, activity, time of the day, user identity, communication partner, and user emotions. Next are discussed different approaches used for designing text-based and symbol-based AAC systems (or components) aimed at improving users' communicative competence.

2.2.1 Contextually organized vocabulary

One of the most common approaches in the literature is the use of contextual information to retrieve vocabulary that was manually pre-assigned to specific categories. There are two main factors between works following this approach: the person responsible for selecting and categorizing the vocabulary into context-related categories and what contextual information was used by the system. The vocabulary used in the AAC systems was selected and organized into user-created

categories in the studies by Kane et al. [25] and Epp et al. [15]. On the other hand, vocabulary was selected and organized into fixed categories by a special education teacher and by a speech therapist in the work from Park et al. [40] and by the school staff in Chan et al. [10].

All works relying on this approach used geographic location as the basic contextual information to recommend vocabulary items. While some studies [15, 40] used only location, some also provided categorization based on the user's conversation partner [25] and goals [26]; interesting approaches that enable the provision of different adaptations for when the user is at the same place. Indeed, participants with aphasia were excited about the idea of AAC devices able to provide content relevant to their friends at the aphasia center [25], and the system proposed by Kim et al. [26] was able to suggest relevant pictographic cards to describe symptoms of a specific medical condition (e.g., running nose) when the user was at a otolaryngologist (i.e., a doctor specialized in the treatment of ear, nose, throat, head and neck disorders).

Whereas most works in the literature use device's built-in GPS to detect user's location, a few recent studies are exploring alternative positioning systems that are able to detect the micro-location of people and objects within indoor environments, such as RFID sensors and BLE beacons. Chan et al. [10], for example, aimed to enhance an AAC solution with ranging and micro-location detection features to reduce the user's cognitive effort when interacting with the user interface. They developed a mobile application that automatically determines the current user's location within a school and displays only the picture cards associated to the corresponding location. These picture cards were chosen by the school staff and were already being used by the students in the current paper-based AAC tool. Only a preliminary evaluation was conducted, where a student

was able to select a relevant symbol (a glass of water) to indicate that he was thirsty.

The approach used in these works is a first step to provide supporting evidence regarding the capabilities of context-aware computing for improving communicative competence of AAC device users. However, the burden of manually choosing the vocabulary and programming the device imposes extra effort on the user and/or caregivers that may counterbalance the positive outcomes of using context-awareness. In addition, this approach does not scale to unexpected situations or unplanned locations. This is particularly problematic for those who require a large vocabulary to attend their communication expectations (e.g., adults with acquired disabilities and relatively preserved intellectual abilities), and for those learning the association between real world concepts and vocabulary symbols (e.g., children with autism).

2.2.2 Prediction based on usage patterns

Another possible strategy to improve communication performance is to provide easy access to vocabulary items predicted to be useful in the current user context based on past vocabulary usage patterns in the same context, represented in form of a context-specific language model. Generally speaking, a language model represents the probability of a word to occur immediately after a given sequence of words. For example, after the sequence “like to eat”, an effective language model would assign high probabilities for words related to food, such as ”hamburger” and ”pizza” and low scores to words that never appear after that sequence, as “table”. Context-specific language models work in the same way, but are designed to represent only vocabulary patterns associated with specific locations, activities, or situations. AAC solutions proposed in the literature usually

benefit from context-specific language models by detecting user context, usually location, activity, or conversation partner and running a prediction mechanism using the associated specific language model in order to suggest the most relevant words while the user composes the messages to communicate.

Patel and Radhakrishnan [41] and Garcia et al. [19], for example, selected location as contextual information to tailor the vocabulary based on previous usage patterns. Higginbotham et al. [22] used a different strategy and implemented a word prediction mechanism for AAC aimed at providing task-specific vocabulary. To create the language models, they recruited twenty fluent English-speakers with no disabilities, divided in ten pairs, and recorded their communication interactions through AAC devices while performing three tasks (talk about a piece of text, draw a route in a map, arrange pieces of a tangram puzzle).

Although language models have been successfully used in many applications such as in the predictive text tools found on current mobile devices, their efficacy highly depends on the quality of the language corpora used for training. This is an important limitation considering the current unavailability of appropriate corpora for AAC; most language models in the literature are trained using corpora that are not representative of AAC use, but rather are trained on newspapers, books, and notebooks. Recorded conversations of non-disabled people are also problematic because they might not represent natural conversation due the fact the participant knows ahead of time that their conversation is going to be recorded. In fact, results from Garcia et al. [19] demonstrated the limitations of such a strategy. In their evaluations, the keystroke saving rate and the communication rate in words per minute were not significantly improved using the proposed location-specific

language models trained on corpora not representative of AAC when compared to an all-purpose language model. Context priming based on statistical language models also had a marginally significant effect on keystroke savings that were not transformed into higher level measures of rate, task performance, or user perceptions in the evaluation performed by Higginbotham et al. [22].

2.2.3 Prediction based on Internet corpora

The major drawbacks in the works discussed so far is that they require either manual selection and programming of vocabulary, or a large and adequate communication corpora collected in advance. In one of the few works aimed to provide an automatically generated location-specific vocabulary to AAC users, Demmans Epp et al. [13] proposed and evaluated the use of four algorithms to retrieve relevant vocabulary from Internet-based corpora. All algorithms relied on obtaining the raw HTML data from a specific corpus' website and processing the textual content to generate collections of vocabulary items. The difference between the four algorithms were the corpora used: websites associated with a location or theme; reviews and comments from multiple review websites; dictionary websites; and Wikipedia pages associated with a topic or location along the University of South Florida's Free Association Database

The authors evaluated the algorithms through a discourse completion study where participants were required to select words generated by the algorithms to respond to situations occurring in four different contexts: at a movie theatre, at a restaurant, talking about illness, and shopping. If a desired word was not present in the location-specific vocabulary, participants could use words

from a general purpose vocabulary. The results indicated that even though the algorithms did not generate a large number of vocabulary items (20.3 - 34.0 per location), a considerable proportion of the items (12-45%) was used by the participants, demonstrating that the algorithms were able to provide contextually-relevant vocabulary. Despite the high use of location-specific vocabulary, most of the vocabulary used in the tasks was from the general purpose vocabulary, indicating that specific vocabularies are not able to cover all vocabulary needs and should be combined with a core vocabulary to address users' needs.

2.2.4 Prediction based on partner's speech

Another type of contextual information that may be used to predict vocabulary is the speech of the person talking to the AAC user. Wisenburn and Higginbotham [54] applied automatic speech recognition and NLP techniques to identify noun phrases spoken by the AAC partner. These noun phrases were then used to produce scripted messages or were combined with typed text, enabling faster communication in future conversations. In the evaluation process, participants achieved a communication rate 36% higher and produced more utterances when using the proposed system in tasks involving an interview and a conversation about a given topic. Although their results are positive, it is important to highlight that the noun phrase identification accuracy is extremely low due to the poor automatic speech recognition achieved with audio recorded in natural environments. In fact, identification recall and precision in this work were estimated as only 53% and 69%, respectively, despite the recording taking place in a quiet laboratory room. Thus, for this approach to have practical applicability for people with higher demands of communication, speech recognition

systems will need to be improved. However, this approach may be useful for those with severe conditions, precluded from participating in society and for which most communication occurs in their homes or health facilities.

2.2.5 Automatically generated utterances

Works discussed so far focused on spontaneous novel utterance generation (SNUG), where messages are constructed by writing individual words or selecting and combining pictograms representing words. While this approach allows flexibility of speech, it may be not optimal for transactional interactions and especially, for people with major physical difficulties interacting with the device (e.g., multiple sclerosis) and severely impaired linguistic competence, as those with absent grammar (e.g., Broca's aphasia) or intellectual impairments (e.g., cerebral palsy). Utterance-based AAC systems can support these people's communication by providing access to pre-stored utterances but lack flexibility, requiring users (or their caregivers) to predict their communication needs and to manually program the authored sentences into the device, and present the same navigational issue as SNUG AAC devices when containing a large number of utterances. One approach for addressing these issues is the automatic generation of utterances based on users' activities or conversational topics of interest.

Giving that conversational narratives are crucial to social engagement and social engagement is one of the ultimate goals of AAC, researchers have developed prototype systems aimed to provide structured personal narratives using automatically generated utterances reflecting past user activities or events. In one of the most complete works in the domain, evolved throughout several years,

authors [42, 8, 48] developed a system that creates personal narratives about a child's school activities from multiple sources of data, as pictures, voice recordings, and sensors detecting the child's proximity to objects, people, or locations. The authors developed three algorithms to segment raw collected data into meaningful events , each one relying on different information to identify the boundaries of an event.

Although parents and staff involved in their research stated that communication at home was improved and participants demonstrated enthusiasm sharing stories, important issues and limitations in the proposed system were identified. First, the narrative software used to communicate the stories generated by the system could not be used independently by staff or parents, requiring the researchers to set it up before its use. Second, one of the participants stated that they did not enjoy using the prototype because the wrong stories were generated. This issue arises due the system's dependency on the quality and amount of data collected in school, which was not done automatically nor efficiently. School staff and family members were required to record voice messages and take pictures every time something interesting happened. The technology used to track children's location required a RFID reader device to be carried with the student all the time, which would not be possible if the student were not using a wheelchair. Also, staff members were required to carry a RFID card and swipe it against a reader every time a relevant interaction between the staff and the student happens. Similarly, to record students' interactions with objects in the classrooms, staff members had to swipe a specific card associated with the object used in the interaction. Therefore, this research highlights the need for a more accurate approach for tracking users' locations and activities that is also transparent to the user and to the people who interact with them. It is

also important to note that, even with an optimal tracking solution, the natural language generation techniques used to create the utterances based on the collected data play a fundamental role in the usability and adequacy of storytelling devices.

2.2.6 Script-based vocabulary organization

A possible strategy for organizing pre-stored utterances to reduce cognitive workload relies on the use of scripts. A script captures the essence of a particular situation and represents the typical sequence of events encountered, allowing people to understand what is happening and what will happen in that situation.

The concept of scripts can be applied to AAC devices through visual scenes that describe everyday situations and provide information about the environment where the situation occurs (e.g., people, objects, actions, location) and associating relevant utterances to those scenes. Thus, when AAC device users face one of these situations, they can access relevant vocabulary in an easy and rapid manner. This is a particularly interesting approach for those with difficulties in understanding abstract concepts and in relating words with categories. In addition, visual scenes establish a shared communication space between AAC device users and their communication partners, and thus can provide the means for addressing the pragmatic challenges faced by, for example, people with aphasia and on the autism spectrum disorder [7].

A different approach, not relying on visual scenes, was proposed in the preliminary work described by McCoy et al. [35]. The main differences are the higher amount of utterances available in each script and the use of a hierarchy composed of categories to represent the scenes within a

script and the available vocabulary in each scene instead of images. Navigation through the scenes is accomplished by either selecting a tab representing a scene or by selecting an utterance within a non-current scene. An interesting feature that enhances utterances' flexibility is the use of slots that can be filled with a number of options relevant in that utterance. For instance, after selecting the utterance "I'll have the nachos", the user would be prompted with other food options to be used in the place of 'nachos'. Unfortunately, the system was not implemented and no follow-up research was reported at the present. However, the design proposed in this work offers interesting strategies that may serve as grounds for future development of AAC systems.

The biggest limitation in the discussed works is that they requires the user, caregiver, or researcher to manually construct the scripts' scenes and to predict communication needs in order to create utterances related with those scenes. Although the authors argue that this task may be facilitated through a user-friendly authoring system designed specifically for creating scripts, the amount of content that needs to be created to properly cover the most basic situations in daily life cannot be fitted into their designs, i.e., another mechanism would be needed to present such amount of vocabulary without overwhelming the user. Therefore, the trade-off between vocabulary coverage and vocabulary flexibility is still a challenge that has not been fully solved.

2.3 Methodologies for studying the effectiveness of AAC systems

When looking at previous research where the design of AAC devices was evaluated, it can be clearly noted that there has been a lack of a consistent experimental methodology. This lack

of pattern is understandable due the unique aspects inherent in working with such a unique and diverse population——the high diversity of individuals' capabilities and needs implies a high diversity of system's designs, which consequently requires a high variety of evaluation strategies and the measurement of different variables during and after the system's development.

To compare the different methodologies applied for assessing the effectiveness of AAC systems, previous research can be classified into two groups according to their research goals. The first group involves works focused on assessing a specific characteristic involved in the human-AAC device interaction: the ability of facilitating access to relevant vocabulary. On the other hand, works in the second group aims to investigate the effectiveness and the efficiency of AAC systems in relation to broader aspects involved in communication interactions, such as communication partner's perceptions and engagement in conversation. Table 2.2 and Table 2.3 present an overview of the methodologies in these two groups.

2.3.1 Studies assessing techniques for facilitating access to relevant vocabulary

These studies are mostly interested in quantitatively comparing communication performance achieved with both a standard, untailored vocabulary, and a novel method. While successful AAC interventions depends on the acquisition of linguistic, operational, social, and strategic skills, in addition to a range of psychosocial factors, it is also important to gather evidence about specific aspects of the human-device interaction that are relevant for gaining communication competence according to the stakeholders' perspectives [30]. Since the availability of relevant vocabulary and efficient vocabulary navigation is a crucial demand on AAC use [12, 33] this approach is likely to prove

Ref.	Setting	Participants	Procedure	Data Collection	Quant. Data	Qual. Data
[54]	Lab	34 w/o disabilities	Conversation; Interview	Logs	WPM; KSR	N/A
[55]	Lab	34 w/o disabilities	Conversation; Interview	Questionnaire, interview	Speed and quality of communication; Impression of relevacy, appropriateness, usefulness	Overall xp.
[22]	Lab	48 w/o disabilities	Tasks	Logs	Task completion, correctness, time; num. words, WPM, KSR; User satisfaction	N/A
[38]	Lab	20 w aphasia	Word Guessing	Logs; Questionnaire; Observation	Number of selection; Navigation pah, Overall experience	N/A
[13]	Lab	16 w/o disabilities	Discourse completion task	Logs	Words used; Goal achieved	N/A
[19]	Sim.	N/A	Pre-defined sentences	Logs	WPM, KSR	N/A

WPM: Words per minute KSR: Keystroke saving rate

Table 2.2 Summary of experimental designs evaluating the ability of facilitating access to relevant vocabulary

beneficial.

Methodologies applied

In general, the evaluation in these studies was based on controlled experiments held in laboratory settings trying to mimic daily life interactions taking place under different locations or contexts (e.g., different conversation partners). In these experiments, participants were encouraged to communicate using vocabulary provided both by a non-tailored vocabulary and by a novel, context-tailored vocabulary as they were participating in such concrete situations. Most studies in this group recruited only participants who were fluent English speakers and without communication

Ref.	Assessment Goal	Setting	Participants	Procedure	Data Collection	Quant. Data	Qual. Data
[14]	Match device with comm. demands	Fast food restaurant	2 researchers acting	Task	Observation	Time spent, number of clarifications	N/A
[51]	Participation in conversation	Daily lives	3 aphasia	Conversation	Video recording	Classification codes	Description of communicative behavior
[15]	Comm. support	Coffee shop	1 aphasia	Task	Observation	N/A	Task completion
[52]	Usage and resposse to head-worn vocab prompts	lab; market	14 aphasia	Task conversation	Logs; audio video recording; observation; interview	Task completion, num of words, time, num of touch events	Overall xp.
[48]	Independent use; support for narrative	School	2 children w disabilities	Story telling	Semi-structured interview; questionnaire	N/A	Independent use; support for narrative

Table 2.3 Summary of experimental designs assessing broader aspects of communication interactions

and cognitive disabilities, with the exception of Nikolova et al. [38] who recruited people with aphasia. Demmans Epp et al. [13], for example, tried to simulate the natural communication environment using a series of discourse completion tasks in four different contexts: at a movie theatre, at a restaurant, talking about illness, and shopping. In these tasks, participants were fluent English communicators and were required to communicate using vocabulary generated by four different algorithms to make a request or to express a level of agreement with the partner and to correct a misunderstanding in case of a disagreement.

The communication interactions were not limited to discourse completion tasks; [54, 55] used conversational tasks in which participants were requested to converse on a particular topic (e.g., fa-

vorite foods, vacation, movies) and also to interview a communication partner about a given topic; Higginbotham et al. [22] requested participants to use the vocabularies available to accomplish specific tasks, such as reading a text about a baseball game and talking about it to a partner; participants in the study by Nikolova et al. [38] were requested to guess missing words in a sentence and find those words in the vocabularies.

In contrast to the aforementioned studies, Garcia et al. [19] conducted the evaluation of two location-aware pictogram prediction algorithms through computer simulations. They asked non-disabled individuals to compose pictogram sentences that they judged to be useful for three different locations (classroom, cafeteria, and home) and ran the simulations to calculate the optimal number of keystrokes needed to select the words in the generated corpus using three vocabulary prediction mechanisms.

Data collected in these studies were mostly quantitative measures of communication performance, such as communication rate in words per minute, number or words used, and keystroke saving rate (i.e., the reduction in number of selections needed to select vocabulary items comparing to a baseline vocabulary). For example, the dependent variables in Demmans Epp et al.'s study [13] were the AAC device's vocabulary coverage and achievement of communicative goal, which were conceptualized as the number of words from the participant's response that were not provided by the vocabulary, and the device's ability to provide the words needed to achieve the participant's goal even if not all desired words were provided, respectively. Studies where participants had to communicate to achieve a particular goal also collected task-performance metrics, such as the time needed to finish the task, correctness of task, and number of tasks successfully

completed. Data collection in most of the works of this group was accomplished through logs automatically collected from the devices; the exception across the literature is the study by Wisenburn and Higginbotham [54], who explored the subjective impressions regarding the efficacy of Converser, an AAC application that uses natural language processing to assist in communication. Efficacy was conceptualized as speed and quality of communication and operationalized using a Likert-scale questionnaire adapted from the literature. An additional questionnaire was applied to gather participants' impressions of relevancy, appropriateness, and overall usefulness of Converser and three open-ended questions were used to elicit general comments about the application.

Critique of methodologies applied

The use of a laboratory setting offers compelling benefits that support the approach adopted in these studies, such as greater experiment control, reduction of confound factor, and ease of recruiting and managing participants during the experiments. Still, it is important to notice that these benefits are valid because these studies were concerned with only one aspect of the human-device interaction (i.e., the techniques' ability to provide relevant vocabulary) and not with obtaining a holistic understanding of an AAC device's use—which would involve many other factors in addition to the access to relevant vocabulary, such as the attitude of communication partners, user's impairment level and skills, and many other technology-related aspects like display layout and selection method.

Another benefit of the methodology applied in these works is the ability of providing an overall description of how the communication performance is affected by the different vocabularies

provided, allowing the use of statistical techniques for comparing and for inferring additional conclusions. For example, the inferential statistical analysis conducted by [22] was able to indicate that typical device-level measures of AAC performance (e.g., keystroke saving rate) may be not directly related with task-level performance. In addition, researchers [19] were able to show, using statistical techniques, that communication performance achieved with their proposed pictogram prediction outperformed the baseline vocabulary in conditions where users reused more than 50% of their sentences, but that there was no significant difference under low sentence reuse conditions.

The main issue of simulated interactions in laboratory settings is the difficulty in guaranteeing that these interactions truly reflects typical interaction in the daily lives of AAC device users. In addition, the type of interaction chosen has great influence on the communication performance due to the very particular characteristics of the different types of communicative interactions, such as the required speed of communication and the predictability of vocabulary used. For example, the context-tailored vocabularies proposed by Higginbotham et al. [22] were evaluated based on three tasks very far from daily life experiences encountered by people relying on AAC devices: (i) reading a text about a baseball game and talking about it to a partner; (ii) describing a route to a partner who had to draw it in a map; (iii) exchanging instructions with a partner to assemble a tangram puzzle. It is hard to argue that findings drawn from these interactions can be generalized to other scenarios relevant for people with complex communication needs, such as talking with a doctor or ordering food.

The use of simulated interactions is delicate even when interactions are properly chosen, as in the work by Wisenburn and Higginbotham [54, 55], where participants were requested to converse

on common daily topics such as favorite hobbies, places visited, and favorite type of music. Many authors argue that the communicative behaviors in contrived interactions within laboratory settings are not similar to the communicative demands of real life and consequently, performance measured in these simulated contexts may not be representative of performance in the real world [30]. In addition, evaluations should not be restricted to the support of expression of needs and wants and the exchange of information, but focus on the full breadth of communication goals, including the development of social relationships—which cannot be assessed in simulated interactions in laboratory experiments.

Although the use of participants with no disabilities in AAC research is a point of great controversy in the field and there is not yet sufficient evidence supporting the generalization of findings obtained from the typical population to people with complex communication needs, several appealing arguments favorable to the use of participants without disabilities can be found in the literature. First, recruiting individuals with communication disorders is naturally challenging due the low-incidence of such population and the difficulties of accessing people that are already excluded from participation in society. Researchers need to allocate extra time and effort not only to reach potential participants, but also to overcome the complications when communicating with them and to guarantee their accessibility to the research sites and their comfort during experiments. Second, AAC users are a quite heterogeneous population with a wide range of cognitive, physical, and communicative capabilities and consequently different communication solutions needs. The combination of these two factors lead to small and non-representative samples in most research in the field, which contributes to the production of findings that are difficult to interpret and

generalize [44]. External validity is compromised even in research with high number of participants and statistically significant findings due to the low probability of matching other individuals' performance-related characteristics with the study participants.

Participation of people without disabilities is pertinent and may be the best option when studying specific aspects of the user-device interaction, especially in early stages of investigation [20], which was the case of the studies in this group. For example, if a number of AAC systems are being evaluated in terms of the time needed to find and select vocabulary items on their interfaces, the performance of participants with disabilities would be influenced by a wider range of factors, such as their motor and vision capabilities, and would be more susceptible to intermittent fluctuations due to physical and cognitive fatigue. On the other hand, the performance of participants without disabilities would be more consistent and would provide an overall measure that could be compared to determine the best interface design among the ones under study. Then, a second study involving AAC users could be conducted under comparable conditions to investigate in more depth the aspects underlying the level achieved by end users.

It is also important to remember that AAC users' ultimate goal is to participate in all kinds of daily life situations and therefore most communication occurs between the user and a typical speaker partner. This means that researchers need to understand how the interactions and the overall communication are influenced by the use of another method of communication (AAC) in addition to the natural speech. This will only be accomplished after having a base for reference constructed with the participation of people without disabilities in research evaluating AAC techniques. Alant et al. [2] claims that the use of both types of participants in research would collabo-

rate for the field growth and would facilitate the understanding of conditions in which it is possible to generalize from typical communicators and the reasons that allow or not this generalization—which would also contribute to generalizations within AAC users.

2.3.2 Studies investigating broader aspects of communication interactions

A methodological issue that has been extensively discussed in the literature is the need for a more comprehensive and holistic view in AAC research and interventions. To participate effectively in society and achieve their educational, vocational, health, social, and personal goals, individuals relying on AAC must develop communicative competence comprised of a combination of linguistic, operational, social, and strategic skills, in addition to a range of psychosocial factors that help them to overcome the numerous constraints and challenges faced when trying to communicate. This idea is consistent with the International Classification of Functioning, Disability, and Health (ICF) [39], which claims that disability is a complex phenomenon that involves both intrinsic characteristics of the individual and factors associated with the social context in which the individual lives. Thus, AAC intervention and research should not investigate the acquisition of specific skills in isolation, but rather focus on the individual's participation in natural environments (e.g., home, community, work) [32].

While there is a need for a more comprehensive and holistic view in AAC research, measuring the multidimensional changes caused by interventions is a complicated process not solved yet. According to Light [30], it is critical to look not only at the immediate effect of the intervention on a specific skill or behavior, but also to evaluate the effect on performance in varied situations

in the real world over the long term. The main difficulty is how to link specific changes, such as the development and expansion of one's vocabulary in comprehension and production, to broader changes in how this person is perceived by their communication partners and how this will impact him/her participation in society [44].

Two works illustrate well designed approaches that attempted to investigate broader changes on the user's participation in communicative interactions caused by the use of a particular AAC device. These studies did not investigate the acquisition of specific skills in isolation, but rather focus on the individual's participation in natural, uncontrolled environments as an attempt to obtain a holistic view of the impact of AAC devices use. To accomplish that, researchers selected tasks representative of daily situations faced by any person, including those with complex communication needs, and used not only interactions for expressing needs—as most research in the field—but also for information transferring and for developing social closeness. In addition, evaluations covered a broader range of communication aspects critical to the success of AAC systems (i.e., conversational control, focus on communication partner, engagement in conversation), in contrast to the studies discussed in the first group that were focused on assessing only communication performance metrics (e.g., speed of conversation, vocabulary coverage) when using different vocabularies.

Waller et al. [51] were interested in assessing how a novel communication system designed for adults with aphasia (TalksBac) could improve participation in conversation in comparison to unaided AAC. Researchers investigated whether TalksBac was able to redress the skewed conversational control balance inherent to aphasiac conversations (i.e., the non-aphasic usually controls

the conversation) by increasing the proportion of the conversation in which the aphasiac partner initiates and elaborates on topics. Interactions with familiar partners and unfamiliar partners using both AAC strategies were video recorded in participants' own homes or in a day centre, where participants were suggested to discuss recent events with familiar partners and to find out about each other with unfamiliar partners. Analysis was accomplished by coding conversations using twelve classification codes representing the degree to which participants initiated new topics and elaborated on the topic under discussion.

Williams et al. [52] were interested in how the use of a head-worn AAC device providing vocabulary prompts during conversations could support the user to maintain focus on the conversation partner, which would lead to a greater engagement in communicative interactions. They assessed usage and response to head-worn vocabulary prompts during conversation by requesting participants to use the device in brief conversations on a familiar topic (e.g., what they would like to do this weekend) and in two conversational tasks at a market place with unfamiliar store clerks: (i) ask for an item that would be difficult to find (a pumpkin), and (ii) ask whether a particular product (a muffin) contained an allergen (nuts). Data collected covered a broad number of aspects: device's logged data and Likert scale feedback were able to provide a quantitative sense of performance and allow the comparison to other evaluation contexts (e.g., different tasks, different devices); observational data provided task performance and additional evidence regarding the device's usability, such as whether the participant said the target vocabulary word or whether the device created disruptions during conversation; semi-structured interviews captured relevant themes across the participants that may have not been perceived only by observing and that help to

explain the quantitative results. For example, some participants stated that the navigation mechanism was problematic and distracted them from speaking, and some highlighted the importance of audio for feeling in control during a conversation.

Besides being concerned with broader aspects of communication and attempting to perform a holistic evaluation, these two studies present another strength: they performed a detailed functional assessment of participants' abilities, creating a rich profile for each participant. To support replication of findings and facilitate the understanding of the reasons behind an intervention failure or success, AAC research should always include complete and detailed participants' descriptions, including demographic information, cognitive and linguistic skills, experience using AAC, and intervention history [44]. In the study by Williams et al. [52], participants were screened by a licensed speech-language pathologist using the Communication Activities of Daily Living (CADL-2), to assess the impact of impairment on daily communication, and the Western Aphasia Battery (WAB) to assess the type and degree of aphasia. In the work by Waller [50], participants' comprehension, expression and communication abilities were assessed before and after the 9-months intervention period using standardized tests to ensure that underlying difficulties had remained unchanged.

Some other studies also investigated the impact of different AAC systems on other aspects related with communication, but applied a narrower range of interaction types in their evaluations. An example is one of the studies reported by Tintarev et al. [48], which focused on assessing whether a novel AAC software for automatic generation of story-telling narratives could be used independently in the school and how personal narratives would change using the system in comparison with other methods for narrative support. To assess independent use, researchers observed

participants and collected feedback from all individuals involved in the evaluation (i.e., children, school staff, family members) through questionnaires. Support for narrative was assessed through questionnaires and semi-structured interviews, also conducted with all individuals involved.

Another example was reported by Doss et al. [14], who were interested in understanding the matching process between the type of AAC system used to communicate with unfamiliar partners and the communicative demands imposed by social environments. They compared the efficiency and the effectiveness of two communication devices for purchasing long and short orders at fast food restaurants. In their experiments, one researcher acted as the AAC user in a wheelchair while a researcher partner accompanied him/her to observe the interaction and collect relevant data, i.e., the time spent to complete the task and the number of clarifications requested by the restaurant's clerk or others in the restaurant. The procedure was conducted in the same manner at fifty-six restaurants: the companion positioned the AAC user at the counter and stood behind the wheelchair collecting data while the AAC user communicated with the clerk exclusively through one of the AAC devices being evaluated. The task (ordering food) was considered completed if (i) the clerk asked if the order was complete, (ii) the clerk left the register, (iii) the clerk initiated money exchange, or (iv) the clerk asked for additional information (e.g., "For here or to go"). A request for clarification was considered as when the clerk asked for repetition, stated that they did not understand, requested information that had already been given, or asked assistance from the companion.

Despite the age of this research, the experimental design adopted offers a few strengths. First, the biases that could have been created due to the participation of an individual without disabilities

were diminished by the explicitness of the conditions implemented, such as the use of a wheelchair in order to foster the clerk's perception that the AAC user had a disability. Second, researchers conducted a large number of interactions in total, but restricted the number of interactions in each restaurant to only one to ensure that clerks would not become familiar with the AAC systems used. Finally, the researcher's conceptualization of communication effectiveness and efficiency supported the social validity of the study, i.e., the completion of an order and the time needed to complete an order are important factors in the daily living of individuals with communication disorders.

While the aforementioned studies in this group were able to connect the effects of the measured variables to the gain of at least one competence (e.g., operational, social) required to achieve overall communicative competence, and consequently to improve participation in society, most related works assessed only very specific aspects using non-standard measuring strategies during preliminary stages of investigation or reported only the system's descriptions and design issue discussions. Evaluations performed in these works were extremely simple and did not provide evidence in terms of communication performance or users' impressions, nor comparisons with other works. For example, Chan et al. [10] assessed whether a student was able to select a relevant symbol (a glass of water) to indicate that he was thirsty; Epp et al. [15] reported that the proposed AAC system (Marcopolo) was successfully used in a cafeteria to order tea, but no performance measures were reported; Kane et al. [25] only reported that participants prefer their system (TalkAbout) over current AAC devices and that they would use it.

In this thesis, we tackle the methodological limitations previously discussed by reducing the

interference caused by the study in the routine activities of end users with communication disabilities. Instead of creating a certain task for participants to complete, we distributed our application to AAC professionals, who then selected end users for trials and performed the assessment using their own expertise without any researchers' intervention. Participants used the app in their routine school and therapy activities, as opportunities and necessity arose—just as other forms of AAC are used.

Bibliography

[1] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *International symposium on handheld and ubiquitous computing*, pages 304–307. Springer, 1999.

[2] Erna Alant, Juan Bornman, and Lyle L Lloyd. Issues in AAC research: How much do we really understand? *Disability and rehabilitation*, 28(3):143–150, 2006.

[3] Alfredo Ardila. A proposed reinterpretation and reclassification of aphasic syndromes. *Aphasiology*, 24(3):363–394, 2010.

[4] Rita L Bailey, Howard P ParetteJr, Julia B Stoner, Maureen E Angell, and Kathleen Carroll. Family members' perceptions of augmentative and alternative communication device use. *Language, Speech, and Hearing Services in Schools*, 37(1), 2006.

[5] Rita L Bailey, Julie B Stoner, Howard P Parette Jr, and Maureen E Angell. AAC team perceptions: Augmentative and alternative communication device use. *Education and Training in Developmental Disabilities*, pages 139–154, 2006.

[6] David R Beukelman, Susan Fager, Laura Ball, and Aimee Dietz. AAC for adults with acquired neurological conditions: A review. *Augmentative and alternative communication*, 23 (3):230–242, 2007.

[7] David R Beukelman, Karen Hux, Aimee Dietz, Miechelle McKelvey, and Kristy Weissling. Using visual scene displays as communication support options for people with chronic, severe aphasia: A summary of AAC research and future research directions. *Augmentative and Alternative Communication*, 31(3):234–245, 2015.

[8] Rolf Black, Annalu Waller, Nava Tintarev, Ehud Reiter, and Joseph Reddington. A mobile phone based personal narrative system. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 171–178, 2011.

[9] Jens Boenisch and Gloria Soto. The oral core vocabulary of typically developing english-speaking school-aged children: Implications for AAC practice. *Augmentative and Alternative Communication*, 31(1):77–84, 2015.

[10] Rosanna Yuen-Yan Chan, Xue Bai, Xi Chen, Shuang Jia, and Xiao-hong Xu. ibeacon and hci in special education: Micro-location based augmentative and alternative communication for children with intellectual disabilities. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1533–1539, 2016.

[11] Lauren Cooper, Susan Balandin, and David Trembath. The loneliness experiences of young adults with cerebral palsy who use alternative and augmentative communication. *Augmentative and alternative communication*, 25(3):154–164, 2009.

[12] John Dattilo, Gus Estrella, Laura J Estrella, Janice Light, David McNaughton, and Meagan Seabury. "i have chosen to live life abundantly": Perceptions of leisure by adults who use augmentative and alternative communication. *Augmentative and alternative communication*, 24(1):16–28, 2008.

[13] Carrie Demmans Epp, Justin Djordjevic, Shimu Wu, Karyn Moffatt, and Ronald M Baecker. Towards providing just-in-time vocabulary support for assistive and augmentative communication. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 33–36, 2012.

[14] L Scott Doss, Peggy Ann Locke, Susan Johnston, Joe Reichle, Jeff Sigafoos, Paul Charpentier, and Dulce Foster. Initial comparison of the efficiency of a variety of AAC systems for ordering meals in fast food restaurants. *Augmentative and Alternative Communication*, 7(4): 256–265, 1991.

[15] Carrie Demmans Epp, Rachelle Campigotto, Alexander Levy, and Ron Baecker. Marcopolo: Context-sensitive mobile communication support. In *Proceedings of the Rehabilitation Engineering and Assistive Technology Society of North America Annual Conference and the 3rd International Conference on Technology and Aging*, 2011.

[16] KAREN Fallon, Janice Light, and Amy Achenbach. The semantic organization patterns of young children: Implications for augmentative and alternative communication. *Augmentative and Alternative Communication*, 19(2):74–85, 2003.

[17] Karen A Fallon, Janice C Light, and Tara Kramer Paige. Enhancing vocabulary selection for preschoolers who require augmentative and alternative communication (AAC). 2001.

[18] Melanie Fried-Oken, David R Beukelman, and Karen Hux. Current and future AAC research considerations for adults with acquired cognitive and communication impairments. *Assistive Technology*, 24(1):56–66, 2012.

[19] Luís Filipe Garcia, Luís Caldas De Oliveira, and David Martins De Matos. Measuring the performance of a location-aware text prediction system. *ACM Transactions on Accessible Computing (TACCESS)*, 7(1):1–29, 2015.

[20] D Jeffery Higginbotham and Jan Bedrosian. Subject selection in AAC research: Decision points. *Augmentative and Alternative Communication*, 11(1):11–13, 1995.

[21] D Jeffery Higginbotham, Howard Shane, Susanne Russell, and Kevin Caves. Access to AAC: Present, past, and future. *Augmentative and alternative communication*, 23(3):243–257, 2007.

[22] D Jeffery Higginbotham, Ann M Bisantz, Michelle Sunm, Kim Adams, and Fen Yik. The effect of context priming and task type on augmentative communication performance. *Augmentative and Alternative Communication*, 25(1):19–31, 2009.

[23] Suzanne Hodge. Why is the potential of augmentative and alternative communication not being realized? exploring the experiences of people who use communication aids. *Disability & Society*, 22(5):457–471, 2007.

[24] Susan S Johnston, Joe Reichle, and Joanna Evans. Supporting augmentative and alternative communication use by beginning communicators with severe disabilities. 2004.

[25] Shaun K Kane, Barbara Linam-Church, Kyle Althoff, and Denise 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*, pages 49–56, 2012.

[26] Gunhee Kim, Jukyung Park, Manchul Han, Sehyung Park, and Sungdo Ha. Context-aware communication support system with pictographic cards. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–2, 2009.

[27] Joanne Lasker and Jan Bedrosian. Promoting acceptance of augmentative and alternative communication by adults with acquired communication disorders. *Augmentative and alternative communication*, 17(3):141–153, 2001.

[28] Janice Light. Interaction involving individuals using augmentative and alternative communication systems: State of the art and future directions. *Augmentative and alternative communication*, 4(2):66–82, 1988.

[29] Janice Light. Toward a definition of communicative competence for individuals using augmentative and alternative communication systems. *Augmentative and alternative communication*, 5(2):137–144, 1989.

[30] Janice Light. Do augmentative and alternative communication interventions really make a difference?: The challenges of efficacy research. *Augmentative and Alternative Communication*, 15(1):13–24, 1999.

[31] Janice Light and David McNaughton. Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication?, 2014.

[32] Janice Light and David Mcnaughton. Designing AAC research and intervention to improve outcomes for individuals with complex communication needs, 2015.

[33] Shelley K Lund and Janice Light. Long-term outcomes for individuals who use augmentative and alternative communication: Part iii–contributing factors. *Augmentative and Alternative Communication*, 23(4):323–335, 2007.

[34] Christine Marvin, David Beukelman, and Denise Bilyeu. Vocabulary-use patterns in preschool children: Effects of context and time sampling. *Augmentative and Alternative Communication*, 10(4):224–236, 1994.

[35] Kathleen F McCoy, Jan Bedrosian, and Linda Hoag. Implications of pragmatic and cognitive theories on the design of utterance-based AAC systems. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 19–27, 2010.

[36] Allison M Meder and Jane R Wegner. ipads, mobile technologies, and communication applications: A survey of family wants, needs, and preferences. *Augmentative and Alternative Communication*, 31(1):27–36, 2015.

[37] Katherine Nelson. *Language in cognitive development: The emergence of the mediated mind*. Cambridge University Press, 1998.

[38] Sonya Nikolova, Marilyn Tremaine, and Perry R 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*, pages 155–162, 2010.

[39] World Health Organization. *International Classification of Functioning, Disability, and Health: Children & Youth Version: ICF-CY*. World Health Organization, 2007.

[40] DongGyu Park, Sejun Song, and DoHoon Lee. Smart phone-based context-aware augmentative and alternative communications system. *Journal of Central South University*, 21(9): 3551–3558, 2014.

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

[42] Ehud Reiter, Ross Turner, Norman Alm, Rolf Black, Martin Dempster, and Annalu Waller. Using NLG to help language-impaired users tell stories and participate in social dialogues. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, pages 1–8, 2009.

[43] Marcia J Scherer and Gerald Craddock. Matching person & technology (mpt) assessment process. *Technology and Disability*, 14(3):125–131, 2002.

[44] Rose Sevcik, Mary Ann Romski, and Lauren Adamson. Measuring AAC interventions for individuals with severe developmental disabilities. *Augmentative and Alternative Communication*, 15(1):38–44, 1999.

[45] Sheela Stuart, Denise Vanderhoof, and David Beukelman. Topic and vocabulary use patterns of elderly women. *Augmentative and Alternative Communication*, 9(2):95–110, 1993.

[46] Sheela Stuart, David Beukelman, and Julia King. Vocabulary use during extended conversations by two cohorts of older adults. *Augmentative and alternative communication*, 13(1): 40–47, 1997.

[47] Jennifer J Thistle and Krista M Wilkinson. Building evidence-based practice in AAC display design for young children: Current practices and future directions. *Augmentative and Alternative Communication*, 31(2):124–136, 2015.

[48] Nava Tintarev, Ehud Reiter, Rolf Black, Annalu Waller, and Joe Reddington. Personal storytelling: Using natural language generation for children with complex communication needs, in the wild.... *International Journal of Human-Computer Studies*, 92:1–16, 2016.

[49] John Todman and Norman Alm. Modelling conversational pragmatics in communication aids. *Journal of Pragmatics*, 35(4):523–538, 2003.

[50] Annalu Waller. Telling tales: Unlocking the potential of AAC technologies. *International journal of language & communication disorders*, 54(2):159–169, 2019.

[51] Annalu Waller, Fiona Dennis, Janet Brodie, and Alistair Y Cairns. Evaluating the use of talksbac, a predictive communication device for nonfluent adults with aphasia. *International Journal of Language & Communication Disorders*, 33(1):45–70, 1998.

[52] Kristin Williams, Karyn Moffatt, Denise McCall, and Leah Findlater. Designing conversation cues on a head-worn display to support persons with aphasia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 231–240, 2015.

[53] Susan Williams White, Kathleen Keonig, and Lawrence Scahill. Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of autism and developmental disorders*, 37(10):1858–1868, 2007.

[54] Bruce Wisenburn and D Jeffery 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, 2008.

[55] Bruce Wisenburn and D Jeffery Higginbotham. Participant evaluations of rate and communication efficacy of an AAC application using natural language processing. *Augmentative and Alternative Communication*, 25(2):78–89, 2009.

Chapter 3

Automated Generation of Storytelling Vocabulary from Photographs for Use in AAC

Mauricio Fontana de Vargas and Karyn Moffatt. 2021. Automated Generation of Storytelling Vocabulary from Photographs for use in AAC. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 1353–1364, Online. Association for Computational Linguistics

Preface

This chapter introduces a new method that generates vocabulary automatically from a user's photographs to support autobiographical storytelling. The objective for creating this method is to enable AAC tools that can provide users with vocabulary relevant for talking about personally relevant topics or events in a timely manner. The evaluation presented how the quality of vocabulary generated by the method behaves under different system's configurations and input photographs, providing meaningful insights for fine tuning the algorithm and enabling the research project to move to the next phase of designing and evaluating, with end users, a novel mobile AAC application (presented in Chapter 4).

Abstract

Research on the application of NLP in symbol-based Augmentative and Alternative Communication (AAC) tools for improving social interaction support is scarce. We contribute a novel method for generating context-related vocabulary from photographs of personally relevant events aimed at supporting people with language impairments in recounting their past experiences. Performance was calculated with information retrieval concepts on the relevance of vocabulary generated for communicating a corpus of 9730 narrative phrases about events depicted in 1946 photographs. In comparison to a baseline generation composed of frequent English words, our method generated vocabulary with a 4.6 gain in mean average precision, regardless of the level of contextual information in the input photographs, and 6.9 for photographs in which contextual information was extracted correctly. We conclude by discussing how our findings provide insights for system optimization and usage.

3.1 Introduction

Augmentative and Alternative Communication (AAC) tools can enhance communication for non-speaking individuals, thus offering improved social interaction and independence. Well established NLP techniques, such as spell check and word prediction, support those with primarily physical barriers to communication (e.g., adults with ALS) to compose complex and nuanced sentences in orthographic-based systems more efficiently. However, those with developmental disabilities (e.g., autism spectrum disorder, ASD) or lexical and semantic processing impairments that limit

their ability to spell out words (e.g., adults with aphasia[1]) must usually rely on less expressive symbol-based systems, for which those techniques offer little support due to unique characteristics of communication with these systems.

Users of symbol-based AAC typically do not construct full, grammatically correct sentences, complete with prepositions and inflections, but rather often only need a few key content words (i.e., nouns, adjectives, verbs)—appearing at any part of the sentence—to supplement other forms of communication, including preserved speech, gestures, or drawings. Such scattered use of vocabulary hinders the typical statistical prediction approach, which relies on patterns learnt from a large training corpus.

Nonetheless, there is much opportunity for improving symbol-based AAC, which is often abandoned because it offers too little communication support relative to the effort required to learn and use [31].

Selecting and organizing vocabularies able to attend user's communication needs in a wide variety of contexts and such that they can find words quickly is one of the major challenges [45, 2]. Alphabetical organizations are not useful, and traditional hierarchical schemes based on abstract categories (e.g., food → apple) are difficult for people with language impairments, making navigation extremely slow for anything but the smallest (least useful) vocabularies. Presenting vocabulary as a flat hierarchy is best [3, 1, 46]; however, only a very limited set of options can be displayed, making communication very reliant on having the desired keywords among the available options.

Providing concise situation-relevant vocabularies currently depends on support from a clini-

[1] a language disorder mostly often caused by a stroke.

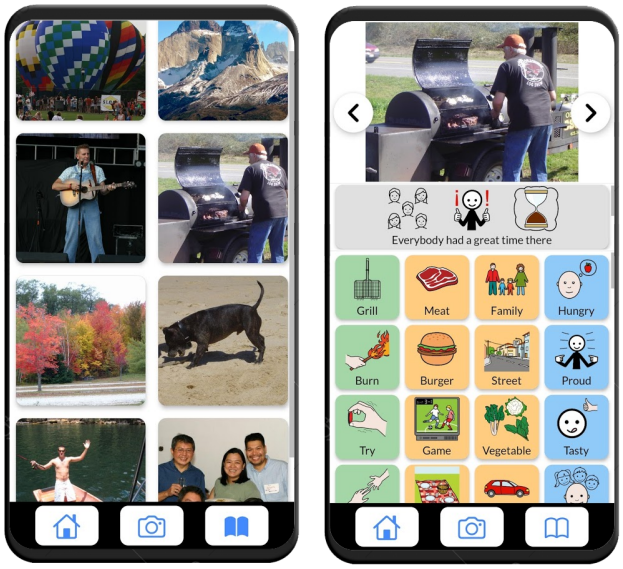


Figure 3.1 An AAC app design demonstrating how context-related vocabulary generated by our method might be presented for use in subsequent conversations. As in many non-orthographic AACs, vocabulary is represented by images that reproduce computer generated speech when selected; however, unlike the status quo, this design eliminates navigation across complicated hierarchies and the need for pre-programming.

cian or caregiver to pre-program the device. But such support is often limited or not available, which consequently limits these devices to supporting generic expressions of wants and needs, i.e., functional communication, and not for social interactions involving spontaneous narratives [47].

Generating vocabulary from user's contextual data through Natural Language Generation (NLG) techniques seems an obvious venue to facilitate social interactions. Although NLG has been successfully applied in the context of task-oriented dialogs [21], question answering [38], text summarization [36], and story generation from photograph sequences [23], it is unclear how these techniques can be adapted to the specific needs of AAC support [40].

In this paper, we call for more research in the NLP community devoted to language generation

for symbol-based AAC systems. We present an overview of the scarce research on the topic and contribute a method that generates vocabulary automatically from a user's photographs to support autobiographical storytelling, demonstrating how it performs under different combination of the system's controllable parameters and a wide range of input photographs.

3.2 Background and Related Work

3.2.1 NLP on Orthographic AAC Systems

NLP research on AAC systems has mainly focused on improving the communication rate of orthographic-based tools, primarily via attempts to reduce keystrokes with letter, word, or message prediction, applying n-grams language models on the user input [39, 18, 16, 44, 42]. Researchers have also explored techniques for improving prediction by including in the language model, some sort of contextual information, such as the topic of conversation [27, 43], the user's location [19], their past utterances [26, 6, 48], or their partner's speech [49]. Virtually all commercial text-based high tech AAC devices employ some form of n-gram prediction [22].

3.2.2 The Need for Symbol-based AACs Able to Support Social Interactions

Many people with developmental (e.g., ASD) or acquired disabilities have difficulty using written language, and therefore need support other than orthographic-based AAC. People with expressive aphasia, for example, present lexical and semantic processing impairments that affect their ability to retrieve the names of objects, combine linguistic elements, and use grammar. Nonetheless,

they usually have good receptive communication skills and intellectual abilities preserved, and typically desire the ability to communicate complex ideas and share social stories spontaneously, such as describing a recent activity or experience [20].[2]

To support this population, researchers from the clinical community [30, 13, 29, 3] have successfully explored the presentation of vocabulary associated with personally relevant and highly contextualized photographs, where people, objects, and activities are depicted in their naturally occurring contexts (also known as visual scene displays, VSDs). Evidence indicates greater conversational turn-taking with fewer instances of frustration and navigational errors [1], and increased lexical retrieval during activity retell [32], for which participants perceived this kind of support as very helpful.

However, the automation of the language production process to support those social narratives is still highly unexplored. For example, Mooney et al.'s system CoChat (2018) generates keywords from human input simulating social network comments. NLP was used only to clean the input and identify nouns and frequent words. In consequence, available commercial tools[3] depend on human effort planning and programming relevant vocabulary, leading to lack of spontaneous and independent communication, and requiring a great amount of time from caregivers [14].

[2]We also witnessed this in interactions observed in conversation groups at a local aphasia institute in which the first author participated for 9 months.

[3]e.g., Tobii Dynavox Snap Scene.

3.2.3 NLG for AAC Systems

Generating language for AAC systems is highly different from typical NLG usage, mainly because the goal of AAC is to provide support for communicating users' thoughts, and not to replace the user by an automatic communicator [40].

The Compansion system [10, 28] was one of the first attempts to apply NLG towards that goal. It was designed to produce grammatically correct sentences from incomplete user input using a small domain model. Although Compansion was dedicated to functional communication, its concept of using domain knowledge served as a stepping stone to Dempster et al.'s system aimed at generating conversational utterances (2010). In their prototype, users populated a personal knowledge base by recording where, when, and with whom they performed an activity shortly after its end. Through a template-driven system, users' knowledge was converted into conversational utterances organized on topics that could be accessed during subsequent conversations. This work showed promising results on how NLG can be able to support social dialogues and increase participation of AAC users. However, their system still required considerable manual linguistic input from users.

Automatic generation of storytelling vocabulary has been successfully explored by researchers [34, 4, 41] to support children with limited memory or with physical and intellectual impairment telling "how was school today" to their parents. In their project, raw sensor data from passive RFID tags relating to locations, objects, and people was aggregated into events, and then transformed to coherent personal narratives using domain knowledge containing the school timetable and the

RFID tags mapping.

To provide just-in-time vocabularies that attend to emergent needs and are not tied to a specific scenario (e.g., school), Demmans Epp et al. [11] explored the use of information retrieval algorithms on internet-accessible corpora such as websites, dictionaries, and Wikipedia pages related to the user's current location or conversation topic. Although this approach was useful for augmenting a base vocabulary with context-specific terms, it is limited to locations (e.g., retail locations) for which internet-accessible corpora are likely to exist.

3.3 Vocabulary Generation Method

Our method generates a rank of key words and short narrative phrases from a single[4] input photo for scaffolding storytelling. It was designed to be used as the back end of interactive AAC systems in which relevant vocabulary is associated with a main photograph, such as Mooney et al.'s CoChat, or as in the example design shown in Fig. 4.1.

We used VIST-TRAIN, a sub-set of the visual storytelling dataset VIST [24] as the main source for vocabulary generation. VIST-TRAIN encompasses 80% of the entire dataset, and is composed of 65,394 photos of personal events, grouped in 16,168 stories. Each photo is annotated with descriptions and narrative phrases that are part of a story, created by Amazon Mechanical Turk workers. We judged VIST to be a good source of vocabulary because i) photos were extracted from personal Flickr albums on a wide range of "storyable" events, related to 69 topics (e.g., graduation, building a house), ii) associated vocabulary is representative of storytelling

[4] to reduce the requirements on users, who may feel discouraged if multiple photos of the event are needed

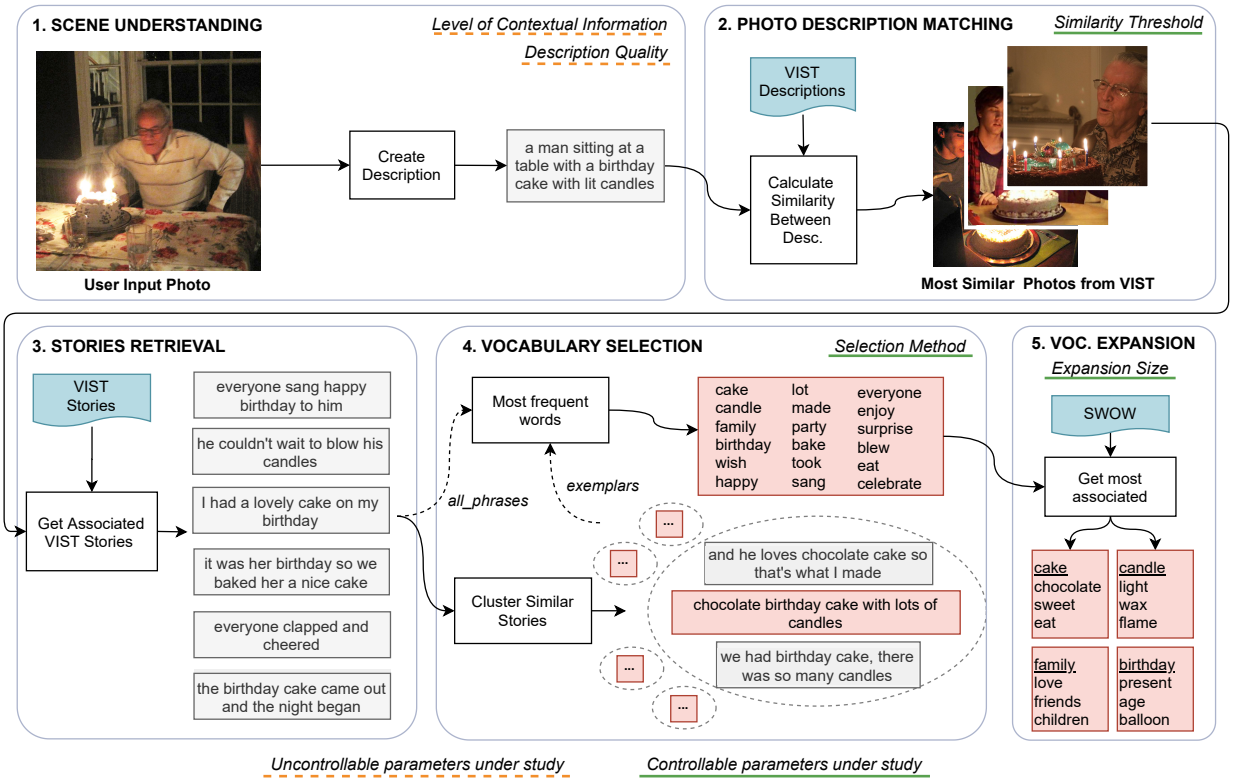


Figure 3.2 Our method. Words and phrases highlighted in red are generated from the input photograph.

and, iii) stories and photo descriptions were constructed by a large number (1907) of workers under a rigorous procedure.

The generation process is composed of five steps, as detailed below and illustrated in Fig 3.2. We explore different implementations for some of the steps, represented by the system's controllable parameters emphasized with bold italic formatting throughout the paper. The different combination of those parameters are evaluated in the next section.

3.3.1 Scene Understanding

The first step extracts contextual information from the photograph in the form of a high-level, human-like description of the scene (i.e., caption) using the computer vision technique from Fang et al. [15]. Captioning was chosen over pure object detection and labelling due to the necessity of communicating more abstract concepts such as the actions being performed and the interactions between the objects, people, and environment during storytelling.

3.3.2 Photo Description Matching

This step finds the subset of VIST-TRAIN photos most similar to the user input by calculating the sentence similarity between the input photo description and all VIST-TRAIN photos descriptions. All photos with description similarity higher than the parameter ***Similarity Threshold*** are selected for processing in the next step, with an upper limit of 30 photos.

Sentence similarity is defined as the soft cosine similarity [37][5] on a bag-of-words representation of the sentences using Word2Vec embeddings, after removing stop words.[6] Soft cosine was chosen as similarity measure due to its ability to capture the semantic relatedness between different words. This strategy was motivated by the fact that soft cosine similarity with Word2Vec was effective for finding similar sentences on question-answering systems, achieving the best performance at the SemEval-2017 Task 3 [5]. Similarity based on entire documents (e.g., Doc2Vec) was not used because it would require a much larger (at present, nonexistent) training corpus to

[5] Gensim library implementation.
[6] as defined by the Natural Language Toolkit (NLTK).

create proper document embeddings, and there are no pre-trained sentence embeddings trained exclusively on photo descriptions.

3.3.3 Stories Retrieval

All narrative sentences associated with the selected photos are retrieved for processing in the next stage. The number of sentences per photo varies from 1 to 5 ($\mu = 3.1, \sigma = 1.4$).

3.3.4 Vocabulary Selection

This step identifies a group of representative sentences and words from the retrieved set by applying the Affinity Propagation[7] clustering [17]—able to generate clusters with less error than other exemplar-based algorithms and not requiring a predetermined number of clusters. The final set of generated phrases is formed by these clusters' exemplars, ranked according to their respective cluster size. By definition, this strategy results in phrases covering the wide range of semantics present in the set of retrieved phrases, while at the same time removing redundant (i.e., very similar) phrases. In case of non-convergence (< 3% in our evaluation), the set of recommended phrases is formed by ranking all phrases according to the sum of their soft cosine similarity against all other phrases retrieved. The generated base vocabulary is formed by a rank of the word frequencies after filtering-out stop words and applying a porter stemmer to merge different variations (e.g., worked, working → work). The parameter ***Selection Method*** determines whether frequencies are calculated considering all retrieved phrases (ALL_PHRASES) or only clusters' exemplars (EXEMPLARS).

[7] damping: 0.5, max. iter: 200, convergence iter.: 15

3.3.5 Vocabulary Expansion

The goal of this step is to diversify the base vocabulary derived from VIST-TRAIN to increase communication flexibility. Thus, to find words that are related to, but distinct from the initial concept (e.g., cake → sweet), our method uses a model of the human mental lexicon as a secondary source of vocabulary. In this model, Swow [9], words are connected with a certain strength representing their relatedness constructed from data of word-association experiments of over 90,000 participants. Therefore, unlike embeddings, SWOW encodes mental representations free from the basic demands of communication.

This strategy was motivated by the fact that word association data was successfully applied in a controlled study to support people with aphasia navigating related words more effectively [33], and that evidence from cognitive science research indicates that the network formed by associations in Swow presents a widespread thematic structure, rather than taxonomic, with words strongly associated often occurring in the same situation (e.g., pick-strawberry; candle-church) [8] . This last step expands the initial set of base vocabulary by adding, for each word, the most strongly associated words in Swow data. The system parameter ***Expansion Size*** determines how many words from Swow are added for each word in the base vocabulary set. Repeated words are not included.

3.4 Evaluation Experiment

The goal of our evaluation is to understand how our design choices, represented by the system *controllable parameters*, along with uncontrollable factors related to the input photograph (i.e., *uncontrollable parameters*), affect the system's performance. Thus, we compared the relevance of vocabulary generated under different combinations of these parameters to investigate the following specific research questions:

1. What combination of controllable system parameters related to the base vocabulary generation optimizes performance?

2. How does the level of contextual information in the input photo affect performance?

3. How does the quality of the contextual description inferred from the input photo affect performance?

4. How does the level of contextual information in the input photo affect the quality of the inferred description?

5. What is the effect of expanding the base generated vocabulary with words from a mental lexicon model on the system's performance?

3.4.1 Performance Metrics

Considering the AAC application usage scenario, the performance of vocabulary generation can be conceptualized by the combination of two factors: i) communication flexibility, i.e., whether

vocabulary needed for composing messages about a specific experience is provided, and ii) communication ease, i.e., the difficulty in finding a particular word among all options generated. These two factors directly map to the information retrieval concepts of precision (P) and recall (R) as a perfect algorithm would provide all words the user needs to communicate the desired message ($R = 1$), and would not contain any irrelevant vocabulary ($P = 1$), thereby minimizing the need for scanning. In contrast, the worst algorithm would provide only irrelevant vocabulary ($P = R = 0$).

Therefore, we tackle the vocabulary generation evaluation as an information retrieval problem, where the input photo is treated as the user query, generated words and phrases are treated as retrieved documents, and crowd sourced narrative sentences about the photograph are the relevant documents, i.e., ground truth (as detailed in Section 3.4.2). For each input photo, difficulty in finding vocabulary and communication flexibility are operationalized as P and R, respectively:

$$P(n) = \frac{|\{rel\_words\} \cap \{G_n\}|}{n}$$

$$R(n) = \frac{|\{rel\_words\} \cap \{G_n\}|}{|\{rel\_words\}|}$$

where n is the number of words displayed to the user, rel_words are the words in the groundtruth sentences, and G_n are the top n words in the generated vocabulary rank. We also calculated the F_1, a common information retrieval measure that captures the trade-off between P and R:

$$F_1(n) = 2 \times \frac{P(n) \times R(n)}{P(n) + R(n)}$$

We calculated these metrics for all $n \in [1, 100]$, and constructed the P-R curves with the arithmetic mean values of P, R, and F_1 across all input photographs under analysis. In contrast to BLEU/METEOR metrics, this analysis allows us to clearly demonstrate trade-offs between the difficulty finding a word among options and communication flexibility, which is important because the number of displayed items will vary for each user.

To obtain a single measure of system performance across this entire interval, considering all input photos, we approximate the area under the P-R curves by calculating the mean average precision:

$$mAP = \sum_{n=1}^{100} P(n)(R(n) - R(n-1))$$

3.4.2 Data

As input photographs and groundtruth sentences, we used VIST-VAL, a sub-set of VIST not employed in our method that contains 8034 photos aligned with crowd sourced stories. We selected all photos from VIST-VAL containing the maximum number of sentences available (5) to act as our input photographs, resulting in 1946 photos. The ground-truth vocabulary for each photograph was formed by joining the five associated narrative phrases (9730 in total), after removing stop words.

3.4.3 Specific Procedures

Controllable Parameters - Base Vocab. (RQ1). We defined four configurations of parameters by crossing two extreme values of ***Similarity Threshold***, i.e., *0* and *best* (highest similarity score

among all VIST-VAL) with the ***Selection Method*** *all_phrases* and *exemplars*, resulting in four configurations: 0_ALL, 0_EXEMPLARS, BEST_ALL, BEST_EXEMPLARS. ***Expansion size*** was set to 0 in all configurations. In the absence of similar AAC generation systems to compare our method to, we created a BASELINE generation formed by a rank of the most frequent words from the Corpus of Contemporary American English (COCA) [7] without stop words. We adopted this baseline because current AAC tools are commonly built on word usage frequency data [35].

The optimal values for the parameters established in this analysis were applied in subsequent analyses.

Contextual Information Level (RQ2, RQ4). To investigate the variability caused by different input photographs, we adopted the concept of context richness from Beukelman et al. [3]. The first author scored each photo from 0–3 based on the number of contextual categories (environment, people/object, activity) it clearly depicts (0 when ambiguous). To validate these annotations, someone unfamiliar with the study also scored a subset of 514 photos (27.8% of the dataset)[8]. Krippendorff's alpha reliability score was 0.82, indicating strong agreement between raters [25].

Context Description Quality (RQ3, RQ4). The first author scored each photo description from 0 to 3 as follows: 0) not generated or completely unrelated; 1) misses most important elements OR contains most of important elements and a few unrelated elements; 2) contains most of important elements OR all important elements and a few unrelated elements; 3) contains all important elements in the photo and does not contain any unrelated elements. As for contextual information level, a subset of 514 were scored by someone unfamiliar with the study. Krippendorff's

[8]all annotations are available at `https://doi.org/10.5683/SP2/NVI701`

alpha reliability score was 0.88, confirming strong agreement.

Effect of Vocabulary Expansion (RQ5). We created 24 pairs of configurations by combining different base vocabulary sizes (5, 10, 15, 20, 25, 30) with the expansion sizes (0, 1, 2, 3). The configuration [5-2], for example, contains five base words plus two expanded words per base word, resulting in a maximum of 15 words (or less if expanded words were already in the base set).

3.4.4 Results

RQ1. To better illustrate the differences in performance, Fig. 3.3 presents the P-R curves, while Table 3.1 shows the *mAP* and maximum *P* and *R* mean values for the pairs of parameter values under investigation, in comparison to the baseline. Overall, 0_ALL results in the best performance, with an *mAP* 4.6 times greater than the baseline, and 1.8 greater than the the worst configuration, BEST_EXEMPLARS.

Configuration	mAP	mAP gain	max P	max R
0_ALL	.058	4.61	.38	.36
0_EXEMP	.039	3.10	.34	.30
BEST_ALL	.042	3.35	.32	.33
BEST_EXEMP	.032	2.52	.27	.28
BASELINE	.013	1.00	.08	.20

Table 3.1 Performance under different configurations.

RQ2. In our input dataset, the proportion of photos according to their context richness score was: 8%(0), 54%(1), 30%(2), 8%(3). A Mann-Whitney U test indicated a significant difference on *P* and *R* only between photos with context richness 0 and the remaining levels ($p < .002$). Table 3.2 shows the mean performance metrics according to level of contextual information.

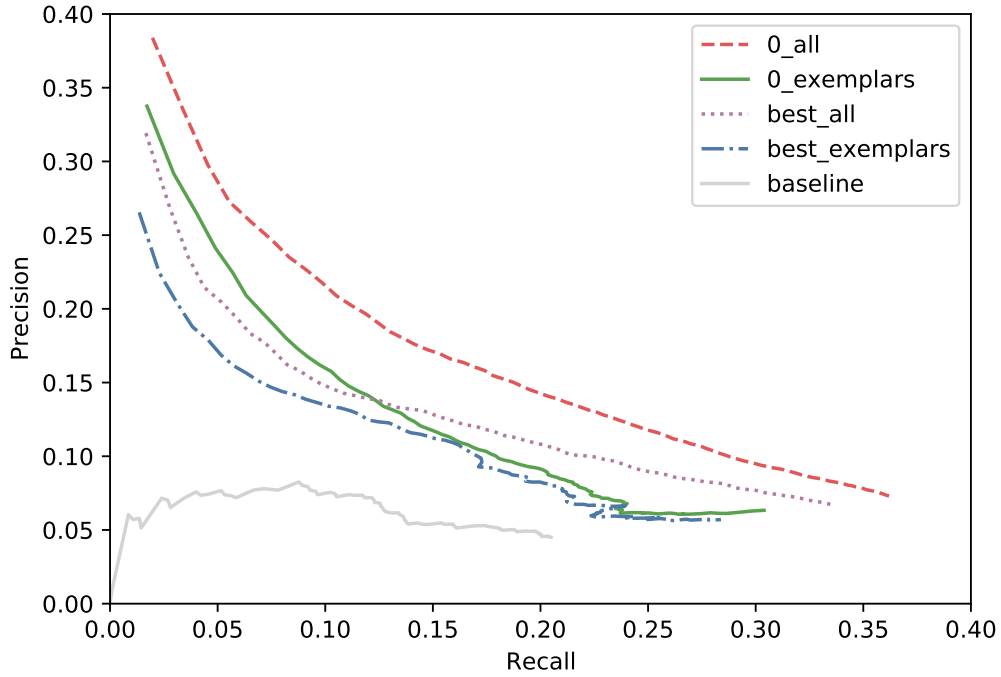


Figure 3.3 P-R curves for different configurations of system's parameters, calculated for all $n \in [1, 100]$.

Context Level	mAP	mAP gain	max P	max R
3	.056	4.44	.43	.37
2	.060	4.72	.38	.36
1	.058	4.57	.38	.36
0	.045	3.54	.29	.23
BASELINE	.013	1.00	.08	.20

Table 3.2 Mean performance according to the level of contextual information in the input photos.

RQ3. The distribution of input photos across context description quality scores was: 16%(0), 16%(1), 30%(2), 38%(3). We plot the P-R curves according to the context description quality scores in Fig. 3.4, and summarize performance metrics in Table 3.3. A Mann-Whitney U test indicated no significant differences between photo quality 1 and 2 ($p > .2$). However, photos with description quality 3 significantly outperformed the other groups ($p < .001$), and quality 0 photos

performed significantly worse than all other groups ($p < .001$).

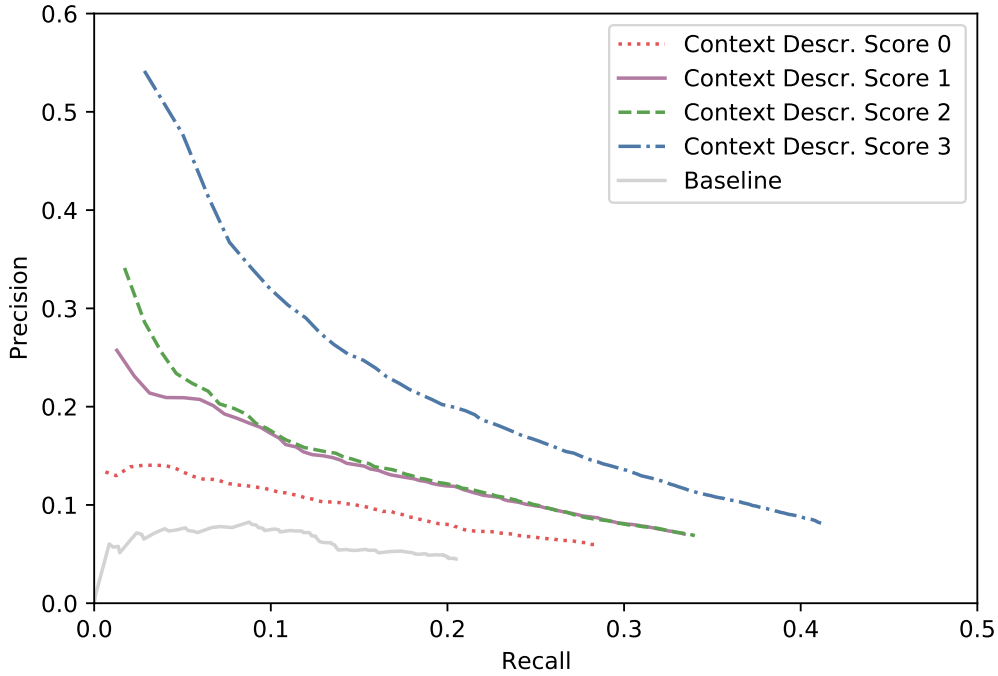


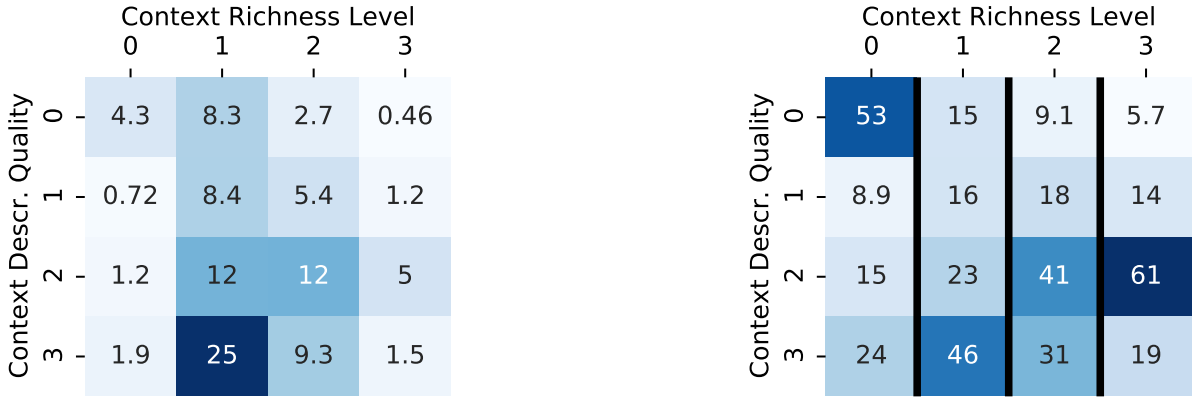
Figure 3.4 Precision-recall curves according to context description quality, under the configuration 0_ALL.

Descr. Quality	mAP	mAP gain	max P	max R
3	.086	6.86	.54	.41
2	.048	3.77	.34	.34
1	.045	3.57	.26	.33
0	.028	2.21	.14	.29
BASELINE	.013	1.00	.08	.20

Table 3.3 Mean performance metrics according to the input photos' description quality.

RQ4. Fig. 3.5 illustrates the relationship between the level of contextual information in the input photos and the quality of the photos descriptions generated using machine-learning.

As expected, photos with ambiguous contextual information (level= 0) most often received bad captions (53%). As context richness increased, the relative proportion of photos with good



(a) Percentages relative to all photos (1946)

(b) Percentages relative to photos with same context richness level

Figure 3.5 Distribution of input photos by contextual richness level and generated description quality

descriptions (scores 2 or 3) also increased (39%, 69%, 72%, 80%), but the relative proportion of perfect descriptions (quality = 3) decreased (46%, 31%, 19%). Photos depicting only one type of contextual information (location, person/object, activity) resulted in the best descriptions: 46% received perfect descriptions, and 66% of all perfect descriptions were given to them. However, when compared to photos with more contextual information, they presented the highest relative proportion of very bad captions (15% vs 9.1% and 5.7%).

RQ5. Fig. 3.6 compares the performance of different combinations of base vocabulary and expansion sizes against base vocabulary only, in function of the number of words displayed n. In general, for a given n, generation without expansion resulted in superior performance. However, on configurations for which a high proportion of expanded words were already in the base vocabulary (e.g., $n = 6, 21, 61$), expansion presented similar or even better F_1 scores than the base vocabulary on its own.

To better understand this phenomenon, we plot the F_1 score, averaged across all photos, in function of the proportion of expansion words not present in the base vocabulary during generation (Fig. 3.7). The mean F_1 for generation without word expansion is also plotted for comparison.

We found that word expansion is able to bring improvement in performance when less than 60% of the expansion words are included in the final generated vocabulary, or in other words, when more than 40% of expansion words is already in the base vocabulary. The tendency is that, the lower the proportion of expansion words not in the base vocabulary, the higher the performance.

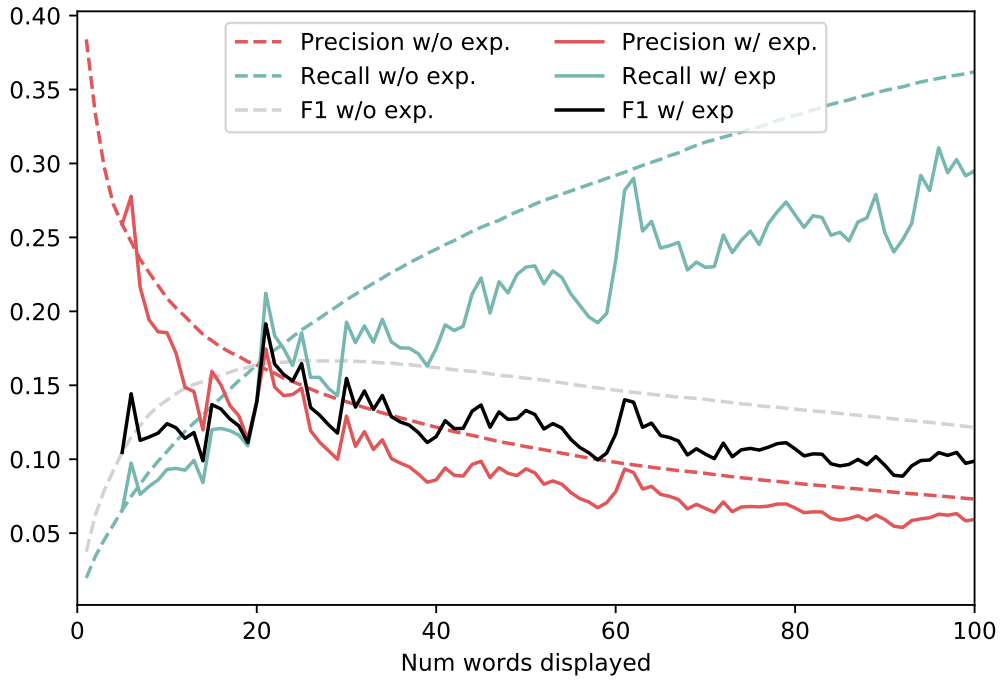


Figure 3.6 Comparison between generation with and without vocabulary expansion.

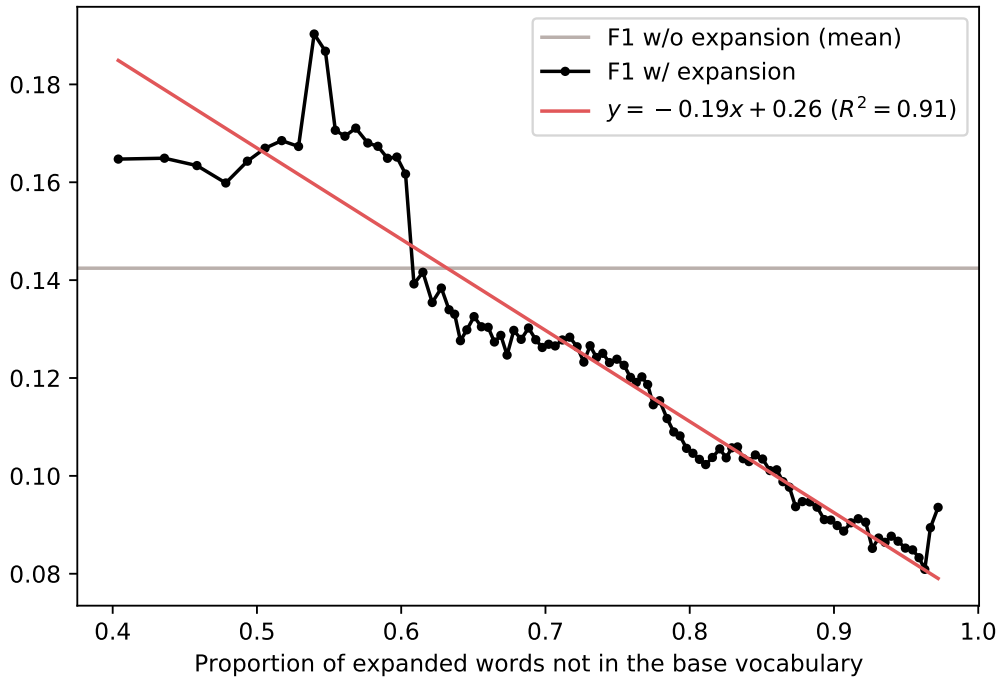


Figure 3.7 Impact of the intersection between base and expanded vocabulary on performance.

3.5 Discussion

The design space for generating AAC storytelling vocabulary directly from photographs is vast and under explored. Design decisions for individual system components will impact other components and ultimately the overall system effectiveness, and therefore cannot be arbitrary. Without a rigorous performance evaluation on different configurations of parameters, users would be at risk of using a flawed or under optimized system, which could lead to user frustration and abandonment, and cause confounds that obscure whether failures are due to the need for algorithmic tuning or mismatch between the intended support and user needs.

The study of controllable parameters (RQ1, 5) demonstrated that **our method is able to pro-**

vide relevant vocabulary, and showed how it can be used to optimize the system and identify areas for further improvement. The exploration of uncontrollable parameters (RQ2, 3, 4) helped illustrate the likely variation in system performance during real world usage (i.e., wide variety of input photos), allowing us to better anticipate potential problems or pitfalls and understand requirements for use.

The similar performance across photos with different levels of contextual information (RQ2) suggests that **our method is robust to variations in the input photograph**. Users will not need to be instructed to take photographs following specific requirements, e.g., "photos should demonstrate an action" or "photos should depict objects only". The similar levels of performance is explained by the pattern observed in the RQ4 analysis; the more elements a photo contains, the better knowledge the machine learning has to infer the central aspect of the photo, but at the same time, the harder it is to capture each and every element. In addition, an element wrongly identified will have less impact on the overall scene understanding since other elements complement the description. An example would be a photo of a birthday party, in which the machine-learning platform is able to infer the central concept (birthday) from the several elements depicted (e.g., cake, candles, balloons), but misses some of the details (e.g. drinks). On the other hand, simplistic photos will rarely lead to elements being cut out, but the computer vision technique will have more variability when performing the inferences, leading to erroneous descriptions more often.

On the other hand, the quality of generated vocabulary was strongly dependent on the computer vision technique employed to extract contextual information about the scene (RQ3) . When a wrong description is generated, the subsequent steps of the algorithm are misled and there-

fore generate vocabulary less relevant for retelling the scene depicted in the photograph. Nonetheless, even in this case, an AAC device using our method would provide vocabulary more relevant than if the most frequent English words were provided. Since photos for which the computer vision technique was able to correctly identify all contextual elements resulted in substantial performance gain, we encourage further exploration of this component. An option would be to use a higher number of raw context labels instead of the single human-like description employed in this work.

Our vocabulary expansion analysis (RQ5) provide valuable insights into how the combination of multiple lexicon sources can generate more relevant vocabulary.

The most promising approach was to combine the visual-to-story dataset with strongly associated words from a mental-lexicon model, but only when there was high intersection between the two vocabularies.

3.5.1 Limitations and Future Work

Although VIST contains a very large range of events, one limitation is that it is unlikely to cover all possible scenarios, and may not accurately reflect AAC communication. However, in the absence of an appropriate AAC-specific corpora (a known issue in the community), we believe the VIST dataset can meaningfully represent the vocabulary needed for scaffolding storytelling. In addition, we do not expect the performance gains observed will directly translate to the same gains in usability. Our goal was to understand fundamental questions necessary for advancing to a usability study, helping fine-tune system components before introducing them to users, avoiding

unnecessary interactions with identifiably poor designs. Our approach also enables larger numbers of parameters to be examined. The low level of social participation commonly observed among people with aphasia, combined with the rate-limited nature of AAC, would require field experiments lasting an impractical amount of time to produce sufficient data to comprehensively explore possible combinations of parameters [26].

As a potential improvement to our method, Sent2Vec trained with BERT may better represent sentence structure and words context for finding similar photo descriptions in step 2 than our use of soft cosine with Word2Vec. Another option would be the use of query expansion to enrich the descriptions. We encourage the exploration of the vast array of strategies for tackling the vocabulary generation process for AAC.

3.6 Conclusion

Developing a photo-to-story vocabulary AAC system presents two challenges; a NLP one in how to generate such vocabularies, and a Human-Computer-Interaction (HCI) one in how to use such vocabulary to offer interactive language support. In this work, we tackle the first challenge.

We demonstrated that our method is able to generate vocabulary with reasonable levels of recall and precision, regardless of the level of contextual information in the input photograph, illustrated the likely variation in system performance during real world usage, and provided meaningful insights for fine tuning the algorithm, enabling us to move to the next phase of designing and evaluating, with AAC users, our mobile interactive application.

Acknowledgments

This research was funded by the Fonds de Recherche du Québec - Nature et Technologies (FRQNT), the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2018-06130], the Canada Research Chairs Program (CRC), and by AGE-WELL NCE, Canada's technology and aging network.

Bibliography

[1] A comparison of visual scene and grid displays for people with chronic aphasia: A pilot study to improve communication using AAC, author = Brock, Kris and Koul, Rajinder and Corwin, Melinda and Schlosser, Ralf, year = 2017, journal = Aphasiology, publisher = Taylor & Francis, volume = 31, number = 11, pages = 1282–1306.

[2] Rita L Bailey, Howard P ParetteJr, Julia B Stoner, Maureen E Angell, and Kathleen Carroll. Family members' perceptions of augmentative and alternative communication device use. *Language, Speech, and Hearing Services in Schools*, 37(1), 2006.

[3] David R Beukelman, Karen Hux, Aimee Dietz, Miechelle McKelvey, and Kristy Weissling. Using visual scene displays as communication support options for people with chronic, severe aphasia: A summary of AAC research and future research directions. *Augmentative and Alternative Communication*, 31(3):234–245, 2015.

[4] Rolf Black, Joseph Reddington, Ehud Reiter, Nava Tintarev, and Annalu Waller. Using NLG and sensors to support personal narrative for children with complex communication needs. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 1–9, 2010.

[5] Delphine Charlet and Geraldine Damnati. Simbow at semeval-2017 task 3: Soft-cosine semantic similarity between questions for community question answering. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 315–319, 2017.

[6] Ann Copestake. Augmented and alternative NLP techniques for augmentative and alternative communication. In *Natural Language Processing for Communication Aids*, 1997.

[7] Mark Davies. The 385+ million word corpus of contemporary american english (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics*, 14(2):159–190, 2009.

[8] Simon De Deyne, Steven Verheyen, Amy Perfors, and Daniel J Navarro. Evidence for widespread thematic structure in the mental lexicon. In *CogSci*, 2015.

[9] Simon De Deyne, Danielle J Navarro, Amy Perfors, Marc Brysbaert, and Gert Storms. The "small world of words" english word association norms for over 12,000 cue words. *Behavior research methods*, 51(3):987–1006, 2019.

[10] Patrick W Demasco and Kathleen F McCoy. Generating text from compressed input: An intelligent interface for people with severe motor impairments. *Communications of the ACM*, 35(5):68–78, 1992.

[11] Carrie Demmans Epp, Justin Djordjevic, Shimu Wu, Karyn Moffatt, and Ronald M Baecker. Towards providing just-in-time vocabulary support for assistive and augmentative communication. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 33–36, 2012.

[12] Martin Dempster, Norman Alm, and Ehud Reiter. Automatic generation of conversational utterances and narrative for augmentative and alternative communication: A prototype system. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 10–18, 2010.

[13] Aimee Dietz, Miechelle McKelvey, and David R Beukelman. Visual scene displays (VSD): New AAC interfaces for persons with aphasia. *Perspectives on Augmentative and Alternative Communication*, 15(1):13–17, 2006.

[14] Kathryn DR Drager, Janice Light, Jessica Currall, Nimisha Muttiah, Vanessa Smith, Danielle Kreis, Alyssa Nilam-Hall, Daniel Parratt, Kaitlin Schuessler, Kaitlin Shermetta, et al. AAC technologies with visual scene displays and "just in time" programming and symbolic communication turns expressed by students with severe disability. *Journal of intellectual & developmental disability*, 44(3):321–336, 2019.

[15] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C Platt, et al. From captions to visual concepts and back. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1473–1482, 2015.

[16] Afsaneh Fazly and Graeme Hirst. Testing the efficacy of part-of-speech information in word completion. In *Proceedings of the 2003 EACL Workshop on Language Modeling for Text Entry Methods*, 2003.

[17] Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.

[18] Nestor Garay-Vitoria and Julio Abascal. Text prediction systems: A survey. *Universal Access in the Information Society*, 4(3):188–203, 2006.

[19] Luís Filipe Garcia, Luís Caldas De Oliveira, and David Martins De Matos. Measuring the performance of a location-aware text prediction system. *ACM Transactions on Accessible Computing (TACCESS)*, 7(1):1–29, 2015.

[20] Kathryn L Garrett. Adults with severe aphasia. In David R Beukelman and Pat Mirenda, editors, *Augmentative and alternative communication for children and adults with complex communication needs*, pages 467–504. Paul H. Brookes, Baltimore, 2005.

[21] He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. *arXiv preprint arXiv:1704.07130*, 2017.

[22] D Jeffery Higginbotham, Gregory W Lesher, Bryan J Moulton, and Brian Roark. The application of natural language processing to augmentative and alternative communication. *Assistive Technology*, 24(1):14–24, 2012.

[23] Chao-Chun Hsu, Zi-Yuan Chen, Chi-Yang Hsu, Chih-Chia Li, Tzu-Yuan Lin, Ting-Hao Huang, and Lun-Wei Ku. Knowledge-enriched visual storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7952–7960, 2020.

[24] Ting-Hao Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, et al. Visual storytelling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1233–1239, 2016.

[25] Klaus Krippendorff. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, 30(3):411–433, 2004.

[26] Per Ola Kristensson, James Lilley, Rolf Black, and Annalu Waller. A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2020.

[27] Gregory W Lesher and Gerard J Rinkus. Domain-specific word prediction for augmentative communication. In *Proceedings of the RESNA 2002 Annual Conference*, 2002.

[28] Kathleen F McCoy, Christopher A Pennington, and Arlene Luberoff Badman. Compansion: From research prototype to practical integration. *Natural Language Engineering*, 4(1):73–95, 1998.

[29] Miechelle L McKelvey, Aimee R Dietz, Karen Hux, Kristy Weissling, and David R Beukelman. Performance of a person with chronic aphasia using personal and contextual pictures in a visual scene display prototype. *Journal of Medical Speech Language Pathology*, 15(3): 305, 2007.

[30] Miechelle L McKelvey, Karen Hux, Aimee Dietz, and David R Beukelman. Impact of personal relevance and contextualization on word-picture matching by people with aphasia. *American Journal of Speech-Language Pathology*, 2010.

[31] Karyn Moffatt, Golnoosh Pourshahid, and Ronald M Baecker. Augmentative and alternative communication devices for aphasia: The emerging role of "smart" mobile devices. *Universal Access in the Information Society*, 16(1):115–128, 2017.

[32] Aimee Mooney, Steven Bedrick, Glory Noethe, Scott Spaulding, and Melanie Fried-Oken. Mobile technology to support lexical retrieval during activity retell in primary progressive aphasia. *Aphasiology*, 32(6):666–692, 2018.

[33] Sonya Nikolova, Marilyn Tremaine, and Perry R 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*, pages 155–162, 2010.

[34] Ehud Reiter. An architecture for data-to-text systems. In *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG 07)*, pages 97–104, 2007.

[35] Kati Renvall, Lyndsey Nickels, and Bronwyn Davidson. Functionally relevant items in the treatment of aphasia (part ii): Further perspectives and specific tools. *Aphasiology*, 27(6): 651–677, 2013.

[36] Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*, 2017.

[37] Grigori Sidorov, Alexander Gelbukh, Helena Gómez-Adorno, and David Pinto. Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas*, 18(3):491–504, 2014.

[38] Yu Su, Huan Sun, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gür, Zenghui Yan, and Xifeng Yan. On generating characteristic-rich question sets for qa evaluation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 562–572, 2016.

[39] Andrew L Swiffin, J Adrian Pickering, John L Arnott, and Alan F Newell. PAL: An effort efficient portable communication aid and keyboard emulator. In *8th Annual Conference on Rehabilitation Technology, Technology-A Bridge to Independence. RESNA'85. Memphis, Tennessee*, pages 197–199. Rehabilitation Engineering Society of North America, 1985.

[40] Nava Tintarev, Ehud Reiter, Rolf Black, and Annalu Waller. Natural language generation for augmentative and assistive technologies. In *Natural Language Generation in Interactive Systems*, pages 252–277. Cambridge University Press, 2014.

[41] Nava Tintarev, Ehud Reiter, Rolf Black, Annalu Waller, and Joe Reddington. Personal storytelling: Using natural language generation for children with complex communication needs, in the wild.... *International Journal of Human-Computer Studies*, 92:1–16, 2016.

[42] Keith Trnka and Kathleen F McCoy. Evaluating word prediction: Framing keystroke savings. In *Proceedings of ACL-08: HLT, Short Papers*, pages 261–264, 2008.

[43] Keith Trnka, Debra Yarrington, Kathleen McCoy, and Christopher Pennington. Topic modeling in fringe word prediction for AAC. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 276–278, 2006.

[44] Keith Trnka, Debra Yarrington, John McCaw, Kathleen F McCoy, and Christopher Pennington. The effects of word prediction on communication rate for AAC. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, pages 173–176, 2007.

[45] Mieke van de Sandt-Koenderman. High-tech AAC and aphasia: Widening horizons? *Aphasiology*, 18(3):245–263, 2004.

[46] Sarah E Wallace and Karen Hux. Effect of two layouts on high technology AAC navigation and content location by people with aphasia. *Disability and Rehabilitation: Assistive Technology*, 9(2):173–182, 2014.

[47] Annalu Waller. Telling tales: Unlocking the potential of AAC technologies. *International journal of language & communication disorders*, 54(2):159–169, 2019.

[48] Tonio Wandmacher, Jean-Yves Antoine, Franck Poirier, and Jean-Paul Départe. Sibylle, an assistive communication system adapting to the context and its user. *ACM Transactions on Accessible Computing (TACCESS)*, 1(1):1–30, 2008.

[49] Bruce Wisenburn and D Jeffery 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, 2008.

Chapter 4

AAC with Automated Vocabulary from Photographs: Insights from School and Speech-Language Therapy Settings

Mauricio Fontana de Vargas, Jiamin Dai, and Karyn Moffatt. 2022. AAC with Automated Vocabulary from Photographs: Insights from School and Speech-Language Therapy Settings . In The 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22), October 23–26, 2022, Athens, Greece. ACM, New York, NY, USA, 18 pages.

Preface

Chapter 3 introduced a novel method for generating context-related vocabulary from photographs of personally relevant events and validated it through quantitative simulations in which the system performance was measured under different system designs (i.e., controllable parameters) and uncontrollable factors (e.g., content of input photograph).

This method and its evaluation lay the technical foundation for research on automated photograph-based AAC, providing insights into the design of automatic vocabulary generation methods. More importantly, Chapter 3 demonstrated that vocabulary generated by the proposed method outperforms a baseline representative of current AAC tools, pointing out the best design choices and discarding system components that did not improve performance. Such evaluation guarantees that end users will not use an under-optimized or flawed generation method that could negatively impact the level of communication support during real life contexts, and that consequently would not properly represent the potential of automatic generation of vocabulary from photographs for AAC.

In this chapter, we move forward in our broad research agenda and design Click AAC, a mobile application that incorporates our generation method. We explore how AAC professionals, such as speech language pathologists, and their clients used our app during their routine activities at school or therapy, and investigate avenues for improving the usability of our design.

Abstract

Traditional symbol-based AAC devices impose meta-linguistic and memory demands on individuals with complex communication needs and hinder conversation partners from stimulating symbolic language in meaningful moments. This work presents a prototype application that generates situation-specific communication boards formed by a combination of descriptive, narrative, and semantic related words and phrases inferred automatically from photographs. Through semi-structured interviews with AAC professionals, we investigate how this prototype was used to support communication and language learning in naturalistic school and therapy settings. We find that the immediacy of vocabulary reduces conversation partners' workload, opens up opportunities for AAC stimulation, and facilitates symbolic understanding and sentence construction. We contribute a nuanced understanding of how vocabularies generated automatically from photographs can support individuals with complex communication needs in using and learning symbolic AAC, offering insights into the design of automatic vocabulary generation methods and interfaces to better support various scenarios of use and goals.

4.1 Introduction

Symbol-based Augmentative and Alternative Communication (AAC) leverages the relative strengths in visual processing of individuals with complex communication needs such as children with autism spectrum disorder. As with other forms of language acquisition, learning symbolic AAC demands a linguistic-rich environment, with frequent opportunities for receiving and producing

language through the symbolic modality. Conversation partners have a crucial role in this process: they need to ensure that the AAC tool is programmed with relevant symbols and then model language use with the tool as conversation opportunities naturally arise [24].

However, the traditional hierarchical organization of symbol-based AAC tools imposes substantial meta-linguistic and memory demands on users searching to find desired words [35, 46], and requires a great amount of time and effort from conversation partners to select and pre-program relevant vocabulary [3]. Consequently, tools are often programmed with only a small set of words that cannot scale to unplanned situations, drastically limiting the opportunities for symbolic language use and acquisition.

One promising approach to alleviate the navigation and pre-programming demands of traditional symbol-based AAC is combining Visual Scene Displays (VSDs) with "just-in-time" (JIT) programming [46, 37, 9, 6]. This approach associates language concepts with a photograph or image of a naturally occurring scene. Conversation partners can program these concepts with the participation of AAC users while the interaction takes place, i.e., "on the fly." For example, while at an amusement park, a family member can take a photograph of the roller coaster and program concepts such as "high," "scared," and "scream" on a page displaying the photograph. This enables the conversation partner to model those concepts quickly, and the individual with complex communication needs to interact with the concepts simultaneously with the relevant real-world referents. While this approach can capitalize on teachable moments [36], and increase symbolic communication turns [21], it still requires effort to manually select and program appropriate vocabularies that is difficult to accomplish in unexpected or emergent situations.

Automated vocabulary generation techniques have been proposed for constructing JIT vocabularies without human assistance, using different types of contextual information as seeds [46]. Researchers have explored the use of geographical locations [19, 44], identification of conversation partners' speech [55, 56], and a combination of different sensor data [45, 8, 53] for generating or retrieving contextually relevant vocabularies. Photographs have also been explored for supporting people with aphasia ordering dinner [43] and retelling past activities [39]. By applying image captioning and optical character recognition (OCR), Obiorah et al.'s prototypes [43] were able to translate photographs of food items and menus of local restaurants into interactive symbols during laboratory experiments. However, their approach is limited to labeling items directly depicted in the photograph and cannot be used to generate additional related concepts. To generate a set of ten words related to a scene photographed, the system by Mooney et al. [39] processed user-generated comments from a fictitious social media. Although their approach showed promising results for supporting people with aphasia in retelling past activities, the approach cannot provide instantaneous support as it is dependent on other people first commenting on the photo.

To date, there has been no research on the creation of AAC tools that automatically generate vocabulary from photographs for use in a broad variety of communication contexts. The design of such tools, both in terms of generation methods and interactive interfaces, and the factors of the dynamics between individuals with complex communication needs, their conversation partners, and automated language support are unexplored. Consequently, the exact kind of support and how such tools could be integrated into real-life settings is unknown.

In this work, we present Click AAC, a prototype tool that generates situation-specific com-

munication boards organized in a VSD-like layout and formed by a combination of descriptive, narrative, and semantic related words and phrases inferred automatically from photographs based on the technique proposed by de Vargas and Moffatt [18]. Through our analysis of semi-structured interviews with AAC professionals, we investigate how these professionals and their clients with complex communication needs used Click AAC during their routine therapy and school activities. We contribute a deep understanding of how vocabularies generated automatically from photographs can support individuals with complex communication needs using and learning symbolic AAC. We offer additional insights into the design of automatic vocabulary generation methods and interactive interfaces to provide adequate support across scenarios of use and goals.

4.2 Background and Related Work

4.2.1 AAC interventions by communicator profiles

Individuals with complex communication needs have an extensive range of expressive communication abilities. Professionals such as speech-language pathologists (SLPs) and assistive technology evaluators are responsible for selecting tools that can adequately attend to the specific communicator's evolving needs. The mapping between available tools and users can be described according to the three broad profiles of communicators, as classified by the speech-hearing community: independent, context-dependent, and emergent communicators [10][1].

Independent communicators have literacy skills on par with same-age peers and are able to

[1]This classification is also used in Dynamic AAC Goals Grid-2 (DAGG-2) , a tool for assessment and measurement of an individual's current level of communication popular in the clinical community.

generate completely spontaneous messages about any topics or contexts while interacting with familiar and unfamiliar partners, usually through text-based or *robust* AAC systems—those containing a very large symbol set (e.g., 2000+) organized hierarchically and with consistent arrangement for supporting motor planning, in addition to allowing morphological changes (e.g., past and plural forms), programming of full sentences, and access to keyboard. In contrast, emergent and context-dependent communicators focus on gaining symbolic communication skills, and are the groups most relevant to the focus of our work.

Context-dependent communicators can use symbolic communication reliably but are still limited to certain contexts. Individuals with this profile often use dynamic displays containing larger vocabularies organized hierarchically, but still do not take advantage of all capabilities of a *robust* AAC system (e.g., comprehensive vocabulary and syntax modifiers). They are starting to compose two or more symbol messages, but their interactions are still dependent on familiar partners, who must facilitate communication, selecting and programming words and messages for them, or helping navigating vocabulary [34]. Intervention goals for this group include increasing access to vocabulary, building literacy skills, and expanding the communicator's ability to interact with more partners and contexts.

Emergent communicators use mostly body language, such as gestures, and non-symbolic modalities that are often not easily understandable by unfamiliar partners, and communicate primarily about the current context. AAC interventions for such individuals focus on establishing more reliable communication through symbolic expression and increasing opportunities for communication interactions. To support emergent communicators learning the associations between

real-world objects or actions with their symbolic representation, professionals often pair these individuals with single button communicators or static communication boards composed of a few symbols (e.g., 4–20 on GoTalk series) representing very common words (i.e., *core words*). Recently, VSDs have been proposed as alternative support for this group. VSD tools associate language concepts with photographs taken or uploaded, either as embedded "hot-spots" in the photograph that reveal the concept when selected, or as a dedicated panel attached to the photograph [9, 6].

4.2.2 Challenges learning and using symbolic AAC

Not differently from spoken language acquisition, learning symbolic communication requires regular exposure to a rich linguistic environment and frequent opportunities for language use. SLPs and family members have a crucial role in this process [49]. They must immerse learners in environments rich in AAC language, ensuring the availability of relevant vocabulary and actively performing *aided language stimulation*[2] during meaningful and motivating opportunities [24]. In this technique, a conversation partner models language on the learner's device while they speak. This includes describing their own actions while they engage in parallel play with the learner, describing the learner's actions, providing an example of target production, and repeating the learner's utterances with additional words to create more semantically or syntactically complete sentences [4]. Aided language stimulation has proved effective in increasing learner's semantic understanding of symbols, [14, 15], number of communication turns, and syntax understanding complexity [47], and therefore, researchers recommend that conversation partners should perform it in at least 70%

[2] Also known as *aided language modeling* or *aided language input*.

of interaction opportunities [15].

However, the design of traditional symbol-based AAC devices hinders such frequent exposure to relevant symbolic communication. The main challenge with such tools is the difficulty in organizing a large number of symbols needed for the spontaneous creation of sentences in a manner that allow individuals with complex communication needs and their conversation partners to easily access desired language concepts when needed. Traditional tools display symbols out of context, arranged in grid-based displays that are organized hierarchically following linguistic (e.g., nouns and verbs) or hierarchically-based (i.e., superordinate → ordinate, like "food" → "dessert") categories, imposing significant meta-linguistic and memory demands [35, 46].

To facilitate the availability of relevant vocabulary and reduce navigational demands, conversation partners can create topic-specific communication boards by selecting words related to a topic they deem as useful and grouping them on a single page. Nonetheless, this strategy does not scale to unexpected situations and imposes a heavy workload on conversation partners, who must anticipate learners' vocabulary needs and dedicate time to program that vocabulary into the devices. Consequently, vocabulary availability tends to be restricted to a small set of topics or a series of frequent words that can be used across most contexts (e.g., want, go). Conversation partners are not able to capitalize on naturally occurring opportunities for language learning, further hindering symbolic communication learning and learners independent use of AAC.

4.2.3 Automated JIT support for AAC

The concept of "just-in-time" (JIT) support, as used in the AAC field, refers to the programming and availability of language concepts at the moment they are needed, through technologies that allow the easy creation of VSDs or other AAC content within interactions [46]. This includes either *mentor-generated JITs*, such as the creation of a hotspot on a VSD when a certain activity is happening, or *automated JITs*, which do not require additional human assistance, such as playing a video that demonstrates how to wash hands when a learner enters the bathroom. The benefits of JIT support are hypothesized based on conceptual underpinnings related to working memory demands, situated cognition, and teachable moments [46].

Context-aware computing has demonstrated value as an enabling approach for automated JIT vocabulary support. Through a participatory design involving people with aphasia, researchers [29] explored the concept design of an AAC system that would adapt the vocabulary presented according to the user location or conversation partner to facilitate word finding. In a Wizard of Oz study in a local aphasia center, vocabulary was manually pre-assigned to different contexts (e.g. doctor's office) and presented to participants while they were imagining using the device in that location. Although their study showed the usefulness of providing vocabulary tailored to the user's context, the technical challenges of building algorithms capable of generating those context-related words were not addressed.

In an attempt to generate vocabularies that attend to emergent user needs in various locations, researchers [19] have applied information retrieval and natural language generation (NLG) tech-

niques on internet-accessible corpora such as websites, dictionaries, and Wikipedia pages related to user's current location or conversation topic. Although this approach was useful for augmenting a base vocabulary with context-specific terms during a laboratory experiment simulating two locations and two topics, it is unclear how this approach would behave in naturalistic settings or how to extend it to personal situations (e.g., telling someone about last weekend's trip) for which internet-accessible corpora are unlikely to exist.

Storytelling vocabulary has been successfully generated for supporting children with complex communication needs in recounting "how was school today" to their families [53]. This method clusters unstructured sensor data (e.g., RFID tags determining the user's location within the school) and transforms it into narrative sentences using a knowledge base containing the school's timetable and RFID mapping information.

More recently, researchers have started exploring photographs as the contextual information input. Obiorah et al. [43] designed prototypes aimed at supporting people with aphasia in ordering meals in restaurants by providing automated captioning of scenes using images from the internet and making text-based information and menus interactive through OCR. Mooney et al. [39] utilized comments from a simulated social network to generate context-related words of personally relevant events to support people with primary progressive aphasia in retelling their past events. Although participants in their study demonstrated increased lexical retrieval during controlled experiments, this approach is dependent on the availability of user-comments and thus cannot be used immediately after the photograph is taken.

In this work, we build off of recent research by de Vargas and Moffatt [18], which proposed

a novel method for generating storytelling vocabulary automatically from photographs for use in AAC. Their method generates a rank of key words and short narrative phrases from a single input photo, by matching against the Visual-Storytelling Dataset (VIST) [27] and then expanding this initial word list to include related ideas using the SWOW model of the human lexicon [17]. Their performance evaluation using a subset of VIST as groundtruth vocabulary (1,946 photos and 9,730 narrative sentences) has shown that this method can provide relevant vocabulary for creating narrative sentences. However, it is unclear how well the technique performs when integrated into an interactive application and evaluated by users in real-world contexts.

4.3 Interactive App Design

To explore the usefulness of automatic JIT vocabulary from photographs in supporting symbol-based AAC, we designed Click AAC, an interactive mobile application that integrates different techniques for generating context-related words and phrases. Click AAC runs on Android and Apple smartphones and tablets.

The design of Click AAC is rooted in evidence-based recommendations from HCI and AAC literature, including the design of well-established AAC tools. Throughout the design process, the first author volunteered for eight months in a local aphasia center, and integrated lessons from first-hand communications with SLPs and people with aphasia into the design of Click AAC. Before the launch of this user study, hundreds of AAC professionals working directly with individuals with complex communication needs informally checked the design and overall concept of the prototype

through a post on specialized social media groups. They confirmed that the prototype design was suitable to be tested with end users during therapy and school activities.

We detail the design rationale and important facets of Click AAC's vocabulary generation and user interface below.

4.3.1 Vocabulary Generation

AAC tools must support users in a variety of communication functions across different contexts, such as commenting, describing, asking and answering questions, and engaging socially [33, 23]. Therefore, Click AAC employs a combination of three generation methods (descriptive, related, narrative) that provide vocabulary spanning the main parts of speech for symbolic AAC (i.e., pronouns, nouns, verbs, and adjectives).

The first step for all methods consists of creating a set of candidate description tags and a human-like description sentence (i.e., caption) for the input photograph using the computer vision technique from Fang et al. [22][3], as done in the work from de Vargas and Moffatt [18]. By applying captioning rather than pure object detection and labelling, Click AAC obtains abstract concepts representing the interactions between the objects, people, and environment depicted in the photograph (e.g., "playing", "angry").

This initial vocabulary is then used with distinct goals in each method:

1. **Descriptive:** Simple description of the scene. It includes lemmas of all description tags, as well as the description phrase.

[3]Microsoft Azure API implementation.

2. **Related (Expanded):** Words semantically related to the elements in the scene. It includes lemmas of all description tags plus lemmas of the three words most strongly connected in SWOW—a model of the human mental lexicon constructed from word-association experiment data—for each description word[4].

3. **Narrative:** Words and phrases used for creating narratives about the scene photographed, obtained through the technique proposed by de Vargas and Moffatt [18]. This technique selects vocabulary associated with similar photographs (i.e., having semantically similar captions) from the visual storytelling dataset VIST [27], which contains 16,168 stories about 65,394 photos created by 1,907 mechanical Turk workers.

By default, the final set of vocabulary presented to the user is the combination of all methods, limited to 20 verbs, 20 nouns, 15 adjectives, and 6 phrases, and fixed pronouns (I, you, he, she, they, and we). We chose this combination of values to maximize the number of vocabulary items displayed while keeping symbols size similar to current tools, and minimizing scrolling. Vocabularies from the Descriptive method have the highest priority, followed by the Related (Expanded).

The app categorizes words by their parts of speech applying the NLTK library's tagger. Users can enable and disable each method in the settings menu. Finally, symbols representing the vocabulary are retrieved from ARASAAC, a repository containing more than 11,000 AAC symbols[5]. If the language set in the application is different from English, generated words and phrases are translated to the target language through the Google Translate API.

[4]The words human, person, man, men, woman and women are not expanded with SWOW vocabulary.

[5]ARASAAC is maintained by the Department of Culture, Sports and Education of the Government of Aragon (Spain): https://arasaac.org/

4.3.2 Interface

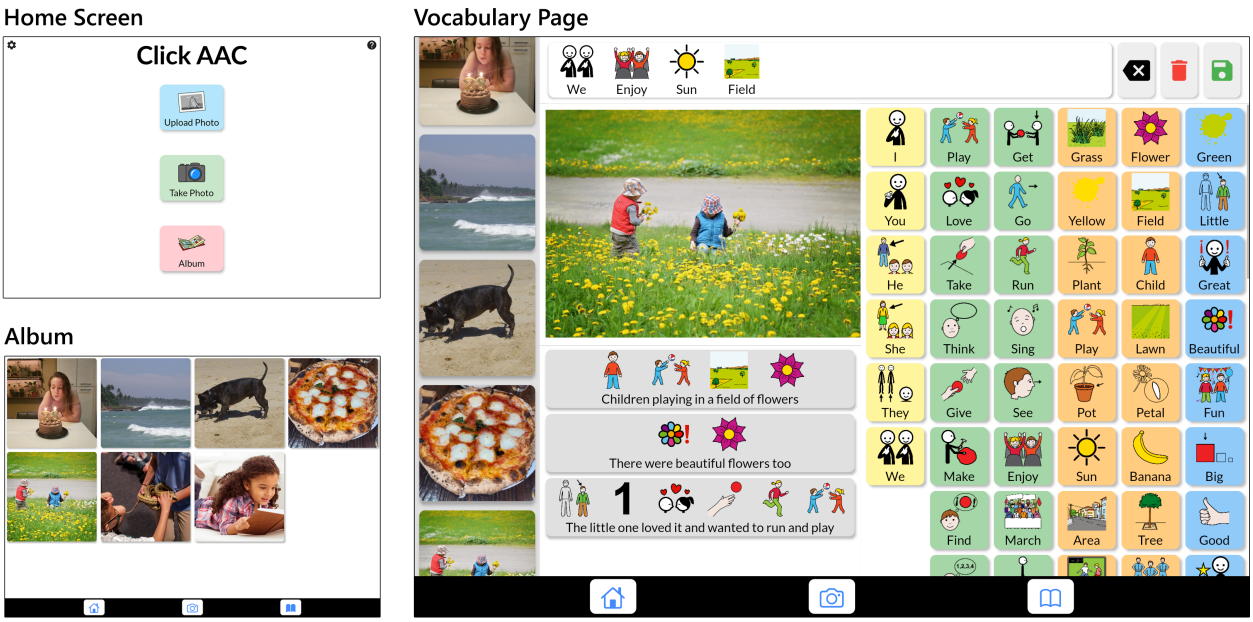


Figure 4.1 Click AAC's Home Screen, Album, and Vocabulary Page containing words and phrases generated automatically from a photograph. Within a Vocabulary page, users can navigate to other photos through the vertical panel on the left, or interact with symbols. Taping on a symbol activates text-to-speech and adds the concept to the message bar on the top. Users can reorder, remove, edit the symbol associated with a word, and add new words and sentences. The size of all elements and the number of vocabulary items generated are customizable.

Our mobile application is composed of three main screens designed to provide direct access to its main features, as shown in Fig. 4.1: (1) a home screen from which the user can import existing photos from the device's gallery, take a new photo, or view their album, (2) an album screen from which the user can navigate through all their previously imported photos and open associated communication boards, and (3) the vocabulary page screen that presents the vocabulary generated for an individual photo. The smartphone version consists of the same three screens, with minor differences in the Vocabulary Page, as detailed next.

Vocabulary Page

Click AAC borrows the overall layout concept and key features from VSDs, a state-of-art AAC support for early symbolic communicators and individuals with cognitive and linguistic limitations [37, 9, 36, 6, 1]: vocabulary is organized in communication boards around a center topic represented by the main photograph (e.g., "eating quesadillas"), rather than in hierarchical categories representing abstract concepts (e.g., "actions" or "foods").

We follow evidence-based guidelines for the design of VSDs and grid displays for children with developmental disabilities and adults with acquired conditions given by Light et al. [37].

First, the set of words is displayed in a grid layout with symbols grouped and colored according to their part of speech, following the Modified Fitzgerald Key [38] color coding, as in popular AAC tools. We chose this configuration over embedding vocabulary in the photograph itself through "hotspots" to allow a larger number of symbols to be displayed without navigation to other pages, and to facilitate the transition between Click AAC and other popular, grid-based tools. Each generated sentence is displayed as a single button containing the symbols of its content words. Users can trigger synthesized audio output by tapping on the vocabulary buttons. Scrolling up or down possibly reveals items hidden due to lack of available space on the screen.

Second, tablet users can navigate to other vocabulary pages by selecting thumbnails of the signature photos, available via the navigation bar on the left of the communication board currently open, a strategy demonstrated beneficial by clinical researchers [37, 54, 6]. Selected words are displayed in a message bar on top of the screen, allowing users to compose sentences combining

individual symbols, as in typical AAC devices. On smartphones, due to the restricted screen size, the message bar and navigation bar are not displayed. The main photograph is displayed on the top of the screen, with the associated vocabulary in the bottom. Thus, smartphone users must swipe left or right, or tap on arrows located on the sides of the photo to navigate to other vocabulary pages.

Editing Vocabulary Generated

To support user's agency during communication [52], Click AAC allows the editing of the initial vocabulary set generated automatically to correct errors and enter missing items. To edit vocabulary generated, users must enter the edit mode by tapping on the main photograph and holding it for at least 500 ms. In line with the design of other AAC apps[6], we chose a non-obvious interaction for editing to prevent unintentional activation. The editing mode displays a new menu bar next to the photo with options for reordering and removing words and phrases, adding new words, and editing symbols associated with the words. To perform one of those actions (e.g., remove a word), users must select the option (remove) and then select the item that will receive the action for at least 500 ms.

4.3.3 Personalization settings

To maximally support a range of different user profiles, Click AAC allows personalization of several aspects of vocabulary generation and interface: type of vocabulary (generation method), max-

[6]Popular tools employ three-finger swipe or a small button in the corners of screen

imum number of words generated for each part of speech, number of phrases generated, language, number of columns used for each part of speech (automatically adjusting the vocabulary buttons to fit in the available space), and size of interface's components (main photo in the vocabulary page, phrases panel, words grid, and menu bar). Users can also modify the font size (or remove fonts completely, as suggested by Light et al. [37]), the spacing between vocabulary buttons, colours attributed to each part of speech, number of columns in the photo album, voice type, rate, and pitch. The interface in Fig. 4.1 shows the default configuration.

4.4 Methods

We conducted a user study involving AAC professionals and their clients with complex communication needs who used Click AAC in their routine practices of therapy sessions or school activities. Through questionnaires and semi-structured one-on-one online interviews with AAC professionals, we investigated an overarching question:

> **How can situation-specific vocabularies automatically generated from photographs support communication and language learning for individuals with complex communication needs?**

We explore this research question in terms of professionals' reflections on their experiences with our prototype, as well as broader factors and concepts envisioned through their experiences. This approach allowed us to understand the broad application of automatic generation of vocabularies from photographs, without limiting use scenarios or introducing artificial ones. As a long-

established practice, AAC interventions must consider not just the needs of individuals that require AAC, but also those of their conversation partners [34, 4, 36]. These professionals regularly try novel AAC technologies and combine multiple tools to accommodate emerging needs dependent on the situation and client profile, in addition to practicing symbolic communication with clients and instructing family members on how to support AAC at home. Therefore, their expertise can provide unique higher-level perspectives than individual users. This broad perspective was particularly pertinent to this stage of our longer term research. HCI researchers working in similar contexts, e.g., designing technologies for dementia care [20], have also noted the value of working with clinicians.

In this exploratory study, our goal was not to specifically evaluate Click AAC, but rather to understand the use and expand the design space regarding automatic JIT vocabulary from photographs in AAC. Engaging directly with users and observing them using the application on defined tasks might bring valuable insights for designing an application but would offer little support for such exploration. Nonetheless, during all interviews, care was taken to ensure that the participants not only shared their perspectives but also relayed the experiences of their clients. The virtual format of the interviews also enabled us to reach a broad set of use cases across learning, therapy, and cultural contexts, without geographic constraints.

As a secondary investigation, we looked into user experiences with Click AAC to understand its overall usability in naturalistic settings. This investigation was not intended to obtain performance metrics of Click AAC in comparison to existing approaches through controlled experiments, but rather to ensure that the app had a reasonable usability and to provide evidence to help us interpret

findings that are directly influenced by our particular implementation. This analysis could also shed light on how to improve the interactive vocabulary support to inform future designs of such tools.

4.4.1 Participants

We made Click AAC publicly available through mainstream app store platforms, and recruited AAC professionals through a message displayed in its initial screen. This message prompted SLPs, who were trying or expecting to try Click AAC with one or more individuals with complex communication needs, as well as AAC consultants or evaluators, who assessed the app independently based on their professional expertise, to enter their contact information if they were interested in participating in the study. Eighty-four (84) individuals agreed to participate, and 53 answered a preliminary questionnaire regarding their experience with AAC, the professional setting of use, and their expected timeline for trying the app with individuals with complex communication needs.

Within this group, 14 SLPs used Click AAC with their clients on private therapy sessions or with their students in special education for at least four weeks, and additional 6 consultants/evaluators tested the app by themselves. This study includes the data from these 20 professionals. Through them, we reached a variety of settings and user profiles (detailed in Tables 4.1 and 4.2). We refer to these clients and students as *AAC learners* throughout the paper.

ID	Profession	Years Exp.	Current Caseload (AAC Users)	Setting	AAC Users in this Study	User Profiles
P1	SLP	6	250	PT	4	non-verbal and min-verbal children; 3 with ASD, 1 with cerebral palsy
P2	SLP	7	20	SES	5	non-verbal and min-verbal children; severe sensory dysregulation
P3	SLP	20	1	PT	1	5 yo with ASD and apraxia ; non-verbal
P4	SLP	6	many	PT	1	child with down syndrome; literate; dysarthirc speech
P5	SLP	43	25	PT	3	2 teenagers and 1 adult, all non-verbal
P6	SLP	10	4	SES	1	non-verbal child; fine-motor skill difficulties
P7	SLP	30	50	EC	20–25	3–22 yo non-verbals and min-verbals; intellectual disabilities; some with fine-motor skill difficulties
P8	SLP	5	39	SES	3	non-verbal child with ASD; 2 young adults min-verbal
P9	SLP	20	2	SES	2	6 yo with ASD, min-verbal, sensory needs; teen min-verbal, apraxia
P11	SLP, AAC consultant	20	3	PT	2	8 and 16 yo with significant cognitive and social issues, dependent on conversation partner
P13[b]	SLP	16	50	PT, SES	2	min-verbal child; literate min-verbal 17 yo
P17[ac]	SLP	20	16	PT	16	verbal and non-verbal children with intellectual disabilities
P18[ac]	SLP	15	22	PT	6	verbal and non-verbal children with ASD
P19[ae]	SLP, AAC specialist	25	180	PHI	1	non-verbal child with ASD

SES: Special ed. school PT: Private therapy PS: Public school CATI: Center for AT innovation PHI: Public health institute

[a] Email interview [b] Hebrew app [c] Spanish app [d] French app [e] Italian app

Table 4.1 Participants in our user study

ID	Profession	Years Exp.	Current Caseload (AAC Users)	Setting	AAC Users in this Study	User Profiles
P10	SLP, AAC specialist	11	80–150	PS	0	diverse intellectual disabilities
P12	SLP, AT consultant	20	40	PT	0	diverse disabilities
P14	AT specialist	28	300 class-rooms	PS	0	diverse disabilities
P15	SLP, AAC researcher	25	3	PT	0	illiterate individuals with language impairment or intellectual disabilities
P16[ad]	SLP, AAC advisor	20	2800 sub-scribed	CATI	0	emerging communicators and AAC experts; aphasia, diverse cognitive impairments
P20[ac]	SLP, AAC advisor	24	anyone in the country	PHI	0	people with difficulty naming objects and navigating vocabulary

SES: Special ed. school PT: Private therapy PS: Public school CATI: Center for AT innovation PHI: Public health institute
[a] Email interview [b] Hebrew app [c] Spanish app [d] French app [e] Italian app

Table 4.2 Participants in our user study (continued)

4.4.2 Procedure

Since this study aimed to understand the use in naturalistic settings, participants were not instructed on how or where to use the app, but rather asked to use or continue to use it in their routine practices in the ways they judged to be most appropriate. Accordingly, professionals used their own expertise and judgment in selecting which clients to try the application with.

For participants who tested with AAC learners, frequency of use ranged from a few sessions spread over four weeks to continuous use during approximately two months. The time using the app within their routines also greatly varied because most of the usage occurred as the need and opportunities arose, rather than during time slots dedicated to testing the app. Consultants/evaluators

who tested the app by themselves used the app less extensively, given that their evaluation mostly consisted of uploading several photographs and investigating vocabulary generated without engaging in specific activities with AAC learners. Most participants used only the tablet version, with the exception of P20 and P7 who also used the smartphone one.

We interviewed the professionals through online video meetings once they deemed their evaluation was complete and they were ready to provide feedback. Each interview took approximately 20–50 minutes. The semi-structured interview was guided by eight questions, covering scenarios of use, profile of users, comparison against current AAC tools, adequacy of the tool in professionals and learners routines, and strengths and weaknesses of the current prototype.

Professionals who used the application with their clients or students also responded to a 5-point Likert scale questionnaire containing 16 questions about interaction, vocabulary, and usage factors (Fig. 4.2). Four participants (P16–P19) were uncomfortable with communicating in English and were instead interviewed in their preferred language (i.e., French (1), Spanish (2), Italian (1)) by email. Interview questions and participant answers were translated with third-party services and checked by the first author who has basic knowledge of those languages. Each participant received a 10$ honorarium.

4.4.3 Data Analysis

We conducted a reflexive thematic analysis [11, 12] on the interview transcripts within MAXQDA2022[7]. The first author performed inductive open coding, guided by our overarching research question.

[7]https://www.maxqda.com/new-maxqda-2022

The open codes were iteratively developed into themes and sub-themes through axial coding, followed by selective coding. All authors discussed the inductive codes as they were evolving, until reaching an agreement on the themes and their interpretations.

4.5 Overall Usability

To help with the interpretation of the thematic analysis, we first present the results regarding the overall usability. These illustrate how users perceived the quality of different elements in the app, such as the vocabulary generated and the overall interaction style. It also revealed necessary improvements, new features, and possible avenues for improved interactive language support that can help inform the design of future applications with automatic vocabulary from photographs.

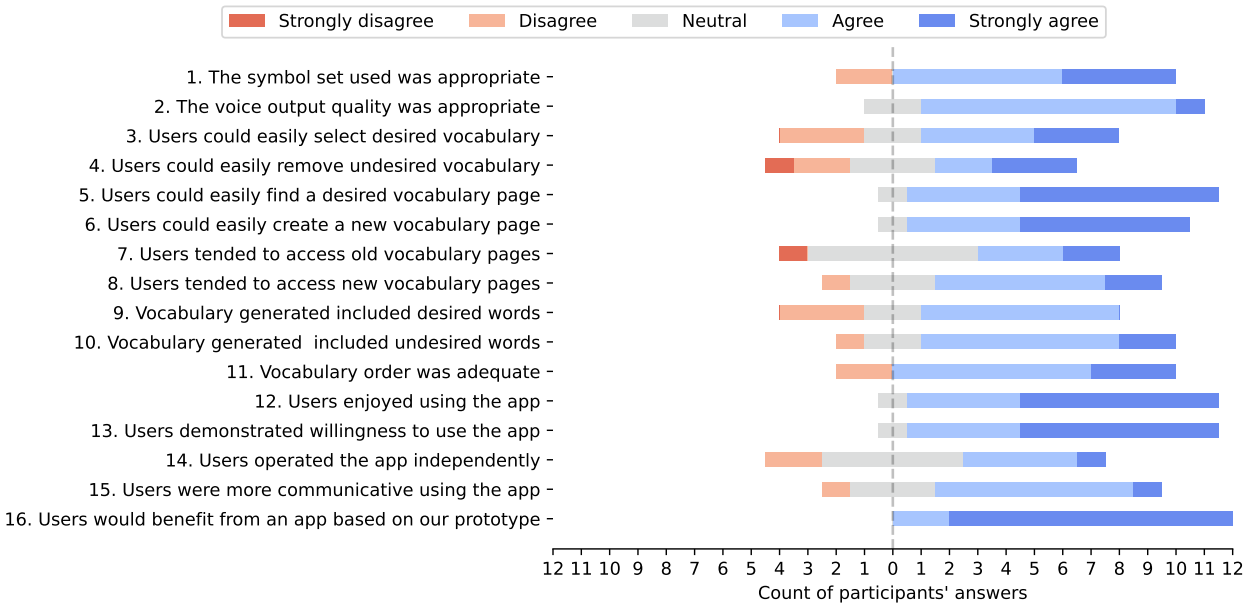


Figure 4.2 Post-questionnaire scores for 12 professionals who used Click AAC with AAC learners during their practices. Original questions are provided as supplementary material.

Fig. 4.2 shows the post-questionnaire answers as a diverging stacked bar chart, with the count of participants' answers[8] represented in the x-axis. Horizontal bars are aligned by the center of the "Neutral" category. In general, professionals were satisfied with the design of Click AAC and the support it provided. Most users could operate the application without major issues (Q1–8). Some participants commented that some learners' motor abilities required additional methods of access, such as external switches or scanning options that were not supported by Click AAC, and thus could not easily select desired items on a vocabulary page (Q3). Three participants also disagreed that removing undesired vocabulary was easy to perform. During the interviews, they explained that they had not figured out how to access the editing options, but once they were instructed, it became straightforward to remove undesired items, highlighting the importance of more evident instructions.

Regarding the perceived quality of vocabulary generated (Q9–11), seven out of the nine participants who used the app in English agreed that generated vocabulary included desired words. However, most participants also agreed that the vocabulary set included non-relevant words. Participants who had more issues with the quality of vocabulary generated commented during interviews that Click AAC was not recognizing the photographs they wanted to talk about, and thus vocabulary generated was mostly irrelevant. Participants who used translated vocabulary were least satisfied with its quality, with two disagreeing that generated options were relevant and one being neutral. This was not unexpected given the simplistic way translation was handled.

The answers covering the app usage (Q12–16) indicate that, although independent use of app

[8] P5 and P13 did not answer the post-questionnaire.

occurred, professionals mostly operated the app together with the learner or by themselves to accommodate learners' needs (e.g., physical access, level of proficiency with symbolic communication). Also, participants strongly agreed that learners demonstrated willingness and enjoyed using the app, and that learners would benefit if there was a complete, commercially ready application based on our prototype.

Importance of personalization

Besides vocabulary editing, participants often had to perform other personalization actions to accommodate learner's needs (e.g., "I made some reduced choice boards for students that couldn't handle as many choices" (P14)).

The importance of limiting the number of vocabulary items for each part of speech category and modifying the layout to increase symbol buttons or the main photograph were highlighted among participants. P3 attributed the observed improvement on her learner's communication to the ability of personalizing the tool to match the learner's profile:

> P3: I think your board was what was able to start her on more modalities of communication, in more areas of her day, then what I had been previously doing with her . . . because I was able to make it more specific for her needs and for her level of communication . . . and I like how easy it is to add and take off a lot of the icons, so that it's not so overwhelming

Improving interactive language support

Participants identified two main improvements required for providing better interactive language support. First, they noted that, although the app was highly customizable, they also need to be able to add familiar people as pronouns symbols and that automatic identification of the names of people photographed "would be an amazing feature" (P5), speeding up the process. The second relates to the availability of frequent words, i.e. "core" vocabulary, for all photographs. Participants highlighted that would be "extremely useful" (P3) to have a personalized core vocabulary present on all pages, and displayed in the same location on the screen to leverage motor planning.

4.6 Findings

Our thematic analysis revealed three main themes that together answer our overarching research question. The first theme describes the situations and ways in which Click AAC was incorporated in school and therapy activities. It details the kinds of support provided for different learner profiles, in addition to presenting envisioned use cases. The second theme interprets how people benefited from the immediacy of vocabulary provided by Click AAC during those activities. The third theme explores the dynamics between AI and users, weighting the benefits and issues introduced by automation and revealing the importance of keeping humans in the loop.

4.6.1 Click AAC offers a flexible, complementary AAC tool for a wide range of user profiles

A wide range of learner profiles can benefit

Professionals selected learners for trying Click AAC through the feature-matching process [48, 25], the gold standard for AAC evaluations, in which they consider the "learner profile, the environment they are in, and the tasks they need to do" (P14) to select appropriate tools from their "toolbox" for communication support or language instruction.

Overall, professionals felt that a wide range range of learner profiles may benefit from technology similar to Click AAC. Selected learners were, in majority, emerging or context-dependent communicators that were non- or minimally-verbal children with diverse developmental disabilities (e.g., ASD, cerebral palsy). Professionals also used Click AAC with a smaller number of children and young adults with functional communication and some literacy skills. Professionals described how a wide range of user profiles could benefit from auto JIT from photographs depending on how the professional incorporates it in their practice. For example, P2 explained their optimism about Click AAC potentially benefiting more profiles as it extended the expressive capabilities for kids along the verbal and nonverbal spectrum in terms of visual cues and structures, working in tandem with existing AAC devices.

> P2: I don't know if there's one profile that [Click AAC is] better for. I think it's pretty open to whatever profile you're working with and tweaking ... based on what the kid's goals are ... [For kids who] are verbal ..., this really helps them visually ... build those sentences ... to expand their expressive language and hopefully generalize out-

> side of just the app. [For] completely nonverbal kids . . . this is really awesome because . . . you're limited to the amount . . . you can talk about so it gives them more structure. So, hopefully then [they'll] be able to build their language in whatever communication app that they're using, that has the 3000+ words . . . not just whatever picture they've taken.

Professionals also envisioned the use of such technology for other populations such as adults with aphasia and dementia. P2, for example, described how people with Broca's aphasia[9], who have "telegraphic speech" and often "want to be talking about their favorite things", such as a visit to the "museum or zoo over the weekend", could "take pictures of" such an event and "bring it to like, a family party and talk about it". P15, who has experience with people with aphasia also envisioned such use cases, but warned that the user would need to have "pretty good reading comprehension" to discern whether the vocabulary generated was adequate or not. Otherwise, the app would get the user "into trouble because [users] would be selecting messages that were maybe not appropriate or relevant to the photo and not realize it."

A complementary tool to talk about past and present contexts, "giving them a voice" and "facilitating language development"

Professionals viewed Click AAC as a complementary tool to facilitate language development and enable communication about specific topics. They reported a variety of main goals for using the

[9]Individuals with Broca's aphasia have trouble speaking fluently but their comprehension can be relatively preserved. This type of aphasia is also known as non-fluent or expressive aphasia. The National Aphasia Association. https://www.aphasia.org/aphasia-resources/brocas-aphasia/

app, including to "expand expressive language" (P2), "augment and facilitate speech language development" (P4), and reduce prompt dependency (P8), "build[ing] up towards using it as an alternative form of communication" (P4).

They did not see the tool as a substitute to existing *robust* systems because of the uncertainty of vocabulary that will be available and because it does not give access to all language concepts the user may need at all times—which are limitations inherent to the concept of automatic JIT vocabulary and topic-specific communication boards, respectively. In addition to those limitations, some aspects of our specific design further hindered the adoption of the tool as substitute AAC device, such as the the lack of a fixed core vocabulary set across all pages and a variable arrangement of symbols that does not promote motor planning.

On the other hand, the ability to easily access language in situations where they "want[ed] to talk about something that's unique or [to] tell a story" (P12) supported learners in different ways towards the intervention goals. While learners used the app independently in some instances (as reported by P2, P9, and P17), on most occasions professionals operated the app in conjunction with learners due to the learner's cognitive and motor difficulties, in line with their regular practice involving other AAC tools. They used the app for: **i)** taking photos of the current context (e.g., "desk space", toys, "table top activity", and doing horticulture); **ii)** uploading generic photos obtained from the internet about topics personally relevant to learners (e.g., favourite toys and cartoon characters); **iii)** uploading photos of past activities or events (e.g., "past vacation", "last weekend, and "cooking").

Then, professionals used the vocabulary generated to work on different activities that encour-

aged symbolic AAC, such as performing aided language stimulation (i.e., modeling language) with emerging communicators while describing, asking questions, or making comments about the scene photographed. With context-dependent communicators, professionals extended the activity by instructing learners to construct their own sentences. For example, P20 explained how she guided her student to compose sentences by using the part of speech categorization, and P8 commented on how Click AAC enabled her to model language faster and supported one learner in constructing his own sentences about an activity previously performed in the school, as he was telling a story:

> P20: [I] used it for sentence construction. At first it was me who used it to teach my students. Then they learned how to do it and they are the ones who choose the picture and the pictures to compose the sentence. I guided them to put the subject first, then the verb and finally the complements.

> P8: The speaker is able to [model] quicker. It's like an easier application to be able to have [language] modeled and then have them either replicate it or have them generate their own sentence from it. [later] everyone was cooking . . . for the week, . . . making quesadillas. And so, . . . we just generated sentences based on that and . . . one of them . . . could structure it to where it was almost like he was telling a story. Like, he could say: first, I added the cheese, and then we used the cheese, . . . kind of a sequencing story

P11 further illustrated how she used Click AAC in therapy as a bridge between the auditory-verbal realm and the symbolic concepts for emerging communicators, and discussed the potential

of auto JIT vocabularies from photographs for enabling communication about past events to family after therapy:

> P11: It can be used flexibly either way. ... So these are not individuals with cognitive ability that you can just hand [Click AAC] to them, and they can start talking and modify it ... They're dependent on me to present something that's relevant to them, and they may not have the physical capabilities to access it ... [Click AAC] is a great bridge for me to use as a therapy tool so that I'm not just existing in the auditory-verbal realm with them, because then I can become over narrating everything, and I'm not anchoring them to any concepts. So when I have something that's concrete, I can use that as my anchor to bridge what I'm saying, and ... the concepts I'm wanting to teach, and they're learning. ... If you have somebody that can ... take a picture of ... this event you're seeing, and when you go home, you have the picture, and you can have some way of communicating that one time event you saw, I see that as being very, very powerful.

In fact, P6 and her learner used the app for a scenario similar to the one envisioned by P11. They relied on the generated vocabulary to start talking about a past visit to the dentist when the learner's main AAC device could not provide support:

> P6: So I used [Click AAC] to download a photo of the a doctor or dentist, and I asked [a learner] where she had been ... and when we pulled up the dentist picture, we got a few more words about it. So, we used it as a supplement to an AAC that she's already

using . . . it did help us kind of open up a conversation about the dentist which we didn't have easy access to use in her standard device.

Theme summary

This theme revealed that a range of user profiles can benefit from automated generation of vocabulary from photographs, and that Click AAC was used as a complementary tool to facilitate symbolic language learning and to enable communication about specific things, addressing some previously unmet needs. These findings signal potential directions to expand the existing design to better attend users needs across the naturalistic activities within therapy and school settings. The next theme explains how the immediacy of vocabulary benefited and can benefit users during those activities.

4.6.2 Immediacy of vocabulary facilitates communication and language learning "on the spot" with reduced workload

Our findings in this theme reveals how learners and professionals benefited and may benefit from immediate availability of situation-specific vocabulary from photographs across the different activities and intervention goals described in the first theme.

Reduced workload opening up opportunities for AAC stimulation

Professionals stressed the importance of selecting and programming appropriate fringe vocabulary to support learners in the various situations encountered in their routines. They further described

how this is typically an arduous task, but that vocabulary generated automatically from photographs can alleviate it. Not surprisingly, conversation partners were not only overloaded by the need of selecting and programming vocabulary on current tools, but also unable to plan and perform these tasks for all situations encountered by learners. The instant availability of relevant vocabulary allowed participants to increase the frequency of moments in which they could model language or engage learners with AAC in general (which is fundamental for successful AAC interventions, as introduced in Section 4.2.2). For example, P7 pointed out the challenges of helping multiple students with different tasks concurrently in teaching routines. She then commented how she was currently able to provide only core vocabulary throughout the school day, and how auto JIT from photographs encouraged communication on the spot by providing easy access to relevant fringe vocabulary:

> P7: One of the big drawbacks in teaching in that kind of environment . . . is you don't know what everybody's doing . . . I could be outside doing pruning and snipping and lopping, or I could be in working with somebody on hand washing, or somebody with feeding . . . [Later] You always want a child to have the ability to communicate, but the time for teachers to do that is very limited. . . . I just I can't keep up with fringe [vocabulary]. With something like [Click AAC], a teacher could take a picture and could encourage that communication and they could do it quickly and they could do it easily . . . So this is brilliant.

Continuing, P6 commented that Click AAC offered "more specific words" to talk about what

was happening in their environment (school room) than the learner's main AAC device, and P7 described one episode in which the instant generation of vocabulary supported an unexpected situation for which she did not have vocabulary material prepared in advance, enabling her to engage the learner with AAC:

> P7: The other day in horticulture . . . I had some fringe vocabulary, but I had not taken pictures of pruners, and loppers and snipers [for low-tech AAC] . . . So, I was able to take a picture of those three tools [with Click AAC] [and] start talking about those tools.

They also postulated that auto JIT from photographs might particularly benefit families, who are less experienced with AAC technologies and vocabulary selection than professionals, and therefore face challenges in creating adequate on-topic communication boards.

> P1: [It] takes me probably like a few minutes to be able to create a new page in someone's communication application, but that's because I do that . . . five days a week [for] seven years . . . but imagine a family who either is not tech savvy [or] just trying to keep up with their child who has special needs. It takes them like hours . . . they just don't have that time. . . . the convenience and quickness of being able to program information into it is the most impressive thing that I've seen.

The significantly reduced programming workload required by Click AAC could "engage more families" to implement AAC at home, as well as teachers in classrooms. P14, who works on

selecting and recommending assistive technologies, commented that she shared the app to families and teachers in her county as an attempt to encourage them to use AAC more often at home and in classroom:

> P14: We have a weekly that we share to families with free apps on it where we give information and [Click AAC is] one of the tools for a student that's starting with communication where we want to ... get the family engaged with, this would be nice because it's not a lot of heavy programming ... We have a lot of families that are not super comfortable with technology. So, if it's easier, then we're more likely to see implementation ...

In addition, P7 discussed how the easiness of having immediate relevant vocabulary can help parents offering opportunities for immersing learners in AAC, and thus keeping learners engaged during a variety of scenarios where is not be possible to have other AAC tools to support communication:

> P7: So, for the parent who wants to communicate with their child, but they're not going to carry around 15 core boards with the possibility of what might be there, this is terrific! I mean, you could be at Sea world and take a picture of Shamu and you're going to get great stuff to talk to your kid about. You could be at at a restaurant and take a picture of the food and be able to talk about what you're doing or while you're standing there waiting in line. Endlessly ... to go to the Harry Potter word, you could take different pictures of things ... keep them engaged .

Facilitated symbolic understanding and sentence construction

Professionals discussed how Click AAC benefited learners and themselves during aided language stimulation and sentence constructions activities. Having "something visual [to] anchor some concepts" was deemed particularly important for modelling language for emerging communicators because a picture taken was "live" and "connecting [the symbols with] something very physical" (P13). Professionals explained how the immediate creation of symbols from real world concepts support teaching learners on "how to use symbolic communication to communicate something more specific" (P11). Because learners could see and use the symbols at the same time as they engaged with the associated object or concept, it was easier to "understand that a referential symbol replaced the presence of that object" when composing messages on the AAC tool, "such as words [do] in oral language," as P20 described:

> P20: The person can have a pictographic representation of an object or scene immediately, and therefore, have at his disposal the symbolic representation that he or she must learn to use communicatively to refer to similar objects and scenes.

Participants noticed that having a concise set of vocabulary displayed next to a photograph setting the communication context in "real-time" supported learners in engaging in the formulation of spontaneous sentences. P13 commented that this strategy gives "situational context cues", "making it easier to compose sentences and to make it more of a conversation" with minimum navigation. As P14 mentioned, "sometimes the navigation between boards, as you're learning to build sentences, it's like a lot together". P8 described how the layout particularly supported users

with limited attention spans:

> P8: [It's] an easier application to ... have them either replicate [sentences] or have them generate their own sentence from it. It's already right there as opposed to needing to return back to the picture or go down to the choices. Everything is all right there, which keeps them focused ... if they're limited by their attention span [or] cognition ... Everything is all right there in front of them.

Spontaneous sentence construction was also supported by the automatic classification of vocabulary into part-of-speech columns. Participants found the organization of vocabulary into columns following the "modified Fitzgerald coloring system" helpful because it was "easier to read" (P9), and "one of the most common ones, so the kids [were] familiar with that color coding" (P2), "match[ing with] another couple of AAC" (P6). P3 further illustrated the ease of use and the consistency of this organization:

> P3: It was very easy ... with the categories where you have them lines up, like, the pronouns are on the left ... then the verbs ... it makes it easier to follow through when you're trying to formulate sentences or questions. I think it's an easier flow, and it was consistent across all the boards.

Support for communicating personal interests

Professionals noted that instant situation-related vocabulary from photographs enabled children who relied on nonverbal communication and had major difficulties navigating traditional AAC

systems to initiate communication about personally relevant topics, allowing professionals to "expand and build on whatever modality [learners were] using" (P2). P6 explains how nonverbal learners often want to initiate communication about topics interesting to them but are hindered due to the lack of easy access to relevant vocabulary, comparing how auto JIT from photographs provided better support in relation to existing tools:

> P6: My main purpose for it would be that "on the go," when I have a student that needs to talk about something that is just too frustrating to find the words for on their device. . . . right now it's snowing, we could take a picture of the snow and the playground, and then the language that comes up about that is concise and related. And, the Proloquo, the Snap Scene . . . we have to dig and dig and dig and find "go", back to the page that has "playground", and go forward to the page that has "weather". And I go back to the page that has "clothes" . . .

The ability to choose a communication topic by selecting a photograph and having related vocabulary instantly available was also deemed important because it allows users to talk about things that were popular among other children, as P5 detailed:

> P5: What my young people are screaming at me about . . . is that [they] can't say what [they] want to say, and talk like the other kids that are out there (their peers without disabilities) And this application gives them the ability to do that, if the vocabulary that's being generated is right, 'cause they can pull up the things that are popular, the things that are of interest to them.

Potential impact on motivation and confidence

Although our user study did not focus on measuring language outcomes, our findings provided preliminary evidence that such technology may improve motivation and confidence for some learners, particularly for those who had been least successful with current tools. Some professionals commented that learners were receptive to the technology and motivated to used it. P9 also observed improved communication in a particular moment for a student who was more confident to speak words thanks to the support provided by Click AAC.

> P9: I've seen the most improvement with the one who's minimally verbal ... This actually happened just yesterday ... We had used the app the whole session with a puzzle, ... so I would push the thing to say *my turn* and take a turn and then I would prompt him to do the same and he started doing that. ... But then, after just about a couple of minutes, he decided to be verbal, so I would say *my turn* and then he would just verbally say *my turn* instead of pushing the button. So I think that kind of gave him like a little bit of confidence ... and an understanding of ... I see what I need to do ... I'm okay with being verbal with with this part. So it was it was a really cool moment I thought for him to take that app in and start with it, ... and he's shown some emergence of that kind of skills a little bit through therapy.

P13 commented that her learner, who was working on literacy skill and had no interest in engaging in language learning activities with other tools, got motivated by trying Click AAC:

> P13: [For] example, [a boy] had no interest... he [tried Click AAC] and it was very,

> very emotional 'cause ... Once we started it till now, he's like, I can't read, I don't know, I'll never know ... but when I started talking to him about the app ... [he] started saying, yeah, I'm going to learn to read and learn to write. It ... got him ... motivated to even try, which is very new for him.

Theme summary

This theme detailed the impact of having vocabularies generated instantly from photographs, revealing four main benefits in terms of how such vocabularies: (1) reduced the workload for selecting and programming situation-specific vocabulary for professionals, which led to increased opportunities for AAC practice , (2) facilitated the immersion of learners in symbolic communication during language modeling and sentence construction activities, (3) supported the communication of personal interests, and (4) impacted on motivation and confidence engaging with symbolic AAC.

4.6.3 Biases introduced did not compromise support but highlighted the importance of AI-human cooperation

Automation of vocabulary selection proved helpful and led to positive outcomes, but participants' experiences highlighted the importance of keeping humans in the loop and revealed new aspects and challenges intrinsic to human-AI cooperation for AAC. This theme first demonstrates how participants' perceptions of the vocabulary quality was related to the type of photographs they used as input and the context of use. Then, it shows common biases and errors caused by the

algorithms powering Click AAC and revealed how participants cooperated with the AI not only to overcome those issues, but also to achieve improved support that would not be possible by the AI or themselves alone.

Quality of vocabulary was directly related to the photograph's content, failing for some relevant situations

Our analysis revealed common patterns in the quality of vocabulary generated across different input photographs, signaling the system's high dependency on the input photograph's content.

Overall, professionals judged individual words generated to be mostly relevant and requiring only a few modifications, when the scene photographed was correctly identified (which users could verify through the descriptive phrase displayed at the top option). Participants positively noted that words were "not limited," "not too predictable," and included not only the names of objects depicted in the photo but also a broader set of words related to the scene, "expand[ing] language." However, they noticed that some words were unrelated and deemed the quality of the narrative phrases as inappropriate to support communication in the naturalist settings they experimented in, as the instances described by P14 below:

> P14: We took a picture of rainbow fish, and that did pull up a lot of really good vocabulary about stuffed animals that it . . . had all the colors, . . . things like draw. And things . . . that were good to go with the book. But then there were things that showed up that we were like, I'm not sure how this fits in. [Later] the sentences often didn't go with the activity that I was putting together.

When the scene was not properly identified, vocabulary was generally not useful and the application was left aside. Some participants got frustrated due to the misidentification, but then, by experimenting, learned the kinds of pictures the app was able to properly identify.

> P14: [There was] this little learning period where I was . . . a little frustrated with the AI, because it was misreading the pictures, but once we got kind of figuring it out, I was very happy with it.

In general, participants reported that Click AAC was able to correctly identify the scene in the majority of photographs ("most of the time it picks up what you're doing" (P7)). However, although identification on "cluttered pictures", or with very specific elements or details occurred in some instances, such as specific gardening tools (e.g., pruners and loppers (P7)), TV characters (White, Rue, and Bea from Golden girls (P5)), facial expressions ("straight face" (P10), and age-related attributes ("historic" (P7)), Click AAC often misinterpreted photographs relevant for learner's common activities.

Participants who encountered the most difficulties cited input photographs containing "two-dimensional" "images that are not real", such as "cartoon's characters", "specific toys", a "door knob", a "smiley" face (to talk about emotions), "super heroes", "ax throwing place", "play-doh", "bubbles", "holiday tree Tu BiShvat", "body parts for Mr. Potato head", and "Peppa Pig". P2, for example, discussed how learners wanted to take photos of cartoon characters or games, but the incorrect identification of the specific toy resulted in unusable vocabulary:

> P2: A lot of the pictures they're wanting to take are flat pictures . . . of Spiderman or,

> of Dora the Explorer . . . one of the car mats, so it's like the carpets and it just has the cars that you can drive on it . . . But then the vocabulary that came up, I couldn't really talk about it at all cause [Click AAC] thought it was a box of cards . . . so those are the kind of pictures that we're ending up taking. It's not of your normal three dimensional objects

Besides the lack of specificity in the identification of "uncommon" scenes, participants noticed some items being constantly identified as other similar, but totally unrelated objects. For example, P10 discussed how "random cylinder objects" were being recognized as soda containers, and P7 experienced "laundry soaps . . . and some softeners . . . " being 'identified as food' ". Some participants also acknowledge how tricky it is to correctly identify some photos, given potential similarities. For example, P15 described an instance where Click AAC identified a goat as a dog, expressing "but to be fair, he does kind of look like a dog in this picture".

Part of the vocabulary that participants judged inappropriate were gender or language style-related biases introduced by the datasets used when training the machine learning models adopted in our generation methods. Identifying people with long hair as woman was a common comment among participants. Another common issue reported was that the language was not adequately matched to user's age-group. For example, P6 noted: "I would want to make sure that when the water bottle came up, [Click AAC] doesn't show me 'wine' " Professionals—especially those using Click AAC in languages other than English—also noticed differences in the vocabulary style of our generated options and their community norms, describing these as "not adapted to

our country" (P18). P10 further expressed the need for recognizing subtle differences between apparent synonyms, especially in language and cultural contexts:

> P10: I don't think that my community [would] really use the word "filth" in that way. We would use the word "dirty" So, just being able to tweak that would be helpful.

Errors and biases introduced did not compromise support, and effort correcting and complementing the AI "was worth it"

Despite the aforementioned errors and biases introduced by AI, participants noted that the automation still facilitated performing language stimulation with the learners during meaningful activities. In addition, the great majority of participants found that the effort filtering, complementing, and correcting the AI was worthwhile in comparison with the amount of work needed for programming the current tools ("It takes less time to create a few boxes than to recreate a complete page" (P6)).

Participants found it easy to edit and add individual words once they had learned how to perform those actions, either through the app's embedded tutorial or by asking for researcher instructions: "It was easy for me to move it around, take off what I didn't want and add what I did want" (P3). They highlighted the importance of being able to quickly edit vocabulary for professionals and learners, as P5 elaborated below:

> P5: When you're busy, when you're programming . . . and also for users who are doing their own programming, [editing] needs to be streamlined to be as easy as they can go in and do it . . . So, the least amount of effort is the best thing.

In instances where professionals were mostly working on aided language stimulation, professionals just mentally ignored irrelevant symbols and focused on relevant ones to maximize the immediacy of the symbolic representation, as P9 discussed: "I don't delete [anything]. I can . . . go through and determine which ones I like best".

In most situations though, participants edited the vocabulary prior to engaging with learners, as a preparation for a specific activity, or in conjunction with the learner when the communication was taking place. P9 described how she and her learners worked together to add missing words:

> P9: I think it's really easy to add words, . . . I can do it in real time during the session [when] we really need this word. So, "let me put it in real quick." . . . and my other student who can already program the words on his own . . . if he comes up with a word, I'll say, "oh, that's a great idea. Why don't you add that in there?"

Cooperation led to extended support

In most cases, once Click AAC displayed a new vocabulary page, participants checked if the overall scene identification was correct, and scanned (with learners in some instances) through the items to remove undesired items and/or add missing words. Professionals reported that during this scanning process, the initial set of words generated by the AI often "served the role of a prime," stimulating them to think of new relevant words that they would have not thought if they were selecting the vocabulary by themselves, as P15 discussed:

> P15: I might see something that was generated by the app that makes me think: "Oh,

> that's a good idea." ... this would also be appropriate and I might not have thought of that before. ... when it comes to ... vocabulary development, it's kind of the difference between a blank slate, where you're thinking, okay where do I start? What do I? How do I come up with something that's relevant that ... and having the app generate some stuff for you, based on a relevant picture, and then that triggers more ideas. So then you might think of other things that you would try programming to see if that would work for the client.

P11 discussed a similar effect, highlighting "the endless potential" of having vocabulary that is not strictly related to the input photograph, using "imaginative" sentences as a starting point for stimulating conversation through other forms of AAC:

> P11: [When I tried with a photo of] my dog, it said "a dog standing on a wood floor," and then it came up with something imaginative like "He decided to dress up his dog." So, then I could take that and run with [the learner]. I wouldn't have thought of that myself, and I would be "What a great idea!" I could go to the markup tool and on the iPhone or the iPad and start dressing my dog up in different things. So, it could be a springboard to something ... I see that as being endless potential.

P7 also illustrated that the mutual collaboration between users and AI led to novel levels of support. She discussed how she adapted her communication to incorporate words offered by Click AAC and expanded the interactions with the learner:

> P7: [After uploading a photo of a dog,] if a child is not really scanning, but they touch "mammal", I can go ahead and talk about that and I can say: yeah, she's a mammal, let's think of some other mammals. Let's see ... animals that have "fur" (points to the vocab button). ... You can really expand just with a handful of vocabulary like that, that you would go. ... Why would I want the word "mammal" on a fringe board? ... that's exactly why! So, you can go ahead and expand on language so that no matter what they touch, I can go further with them ...

Theme summary

This theme demonstrated how the perceived quality of vocabulary was directly related to the photograph content, informing future selection of machine learning models and training dataset for improved scene recognition. It also explained how participants cooperated with the AI to overcome the errors and biases introduced, providing insights into how this cooperation can be leveraged to reach improved support.

4.7 Discussion

Our findings revealed insights into the potential for automatic generation of context-related vocabulary from photographs to support AAC, as well as on aspects specific to our implementation in school and speech-language therapy settings. We now discuss how the observed and envisioned benefits of such technology relate to the conceptual underpinning of JIT support introduced

by Schlosser et al. [46], moving to the implications for the design of such tools for the variety of user profiles and contexts of use identified in our analysis. We conclude with reflections on the study design employed and directions for future research.

4.7.1 Conceptual underpinning for the benefits from immediacy of vocabulary

The benefits of being able to immediately generate vocabulary as needs arise, as revealed in the second theme, included reduced workload leading to increased opportunities for AAC stimulation, facilitated symbolic understanding and sentence construction, support for communicating personal interests, and potential impact on motivation and confidence. These benefits are all tightly related to the conceptual foundations of the JIT support: working memory demands, situated cognition, and teachable moments [46].

When communicating with the aid of a traditional dynamic grid display, learners must keep the desired concept in mind while simultaneously remembering the page where that symbol is located, how to navigate to that page, and the location of the desired symbol on the target page, while avoiding distractions that may arise during this process. When forming sentences, users must go through this process several times [51]. With the combination of automated generation of vocabulary from photographs and VSD-like interface, users do not need to hold in memory the symbols previously selected nor to remember how to navigate to a desired symbol while constructing sentences, reducing **memory demands**. Our participants emphasized how this was particularly useful for constructing sentences to model language because learners can focus on the language concepts rather than being burdened with the navigation task.

Our approach enabled users to have symbols representing the real world concepts they were engaging with readily available, which can not only alleviate working memory demands but also facilitate **situated cognition**. Cognition and learning are inherently dependent on the social and cultural contexts in which they occur, and this is no different for language learning and comprehension [13]. Associating language elements with perceived referents while a situation takes place is crucial for learners to comprehend and use language. The immediacy of symbolic representation helps to clarify the relation between objects, symbols, events, and agents participating in that situation [2]. By providing related vocabulary instantly without requiring users to anticipate the situation, our approach can increase the frequency of moments for which the learning of symbolic representation through aided language stimulation is possible.

This relates to the third conceptual underpinning of JIT support, **teachable moments**. According to the education literature [28], teachable moments are those opportunities that emerge when students are excited, engaged, and primed to learn. Adults must provide activities to children according to their level of development, allowing them to "learn what they want and when they are ready to learn" [28]. The provision of automatic JIT vocabulary can support conversation partners capitalizing on those teachable moments by being able to adapt the offered support to emerging and unforeseen situations quickly, and to engage in topics of interest of the learner, which can activate background knowledge about those contexts, consequently promoting comprehension [26]. Our findings indeed demonstrated how the auto generation of vocabulary in Click AAC enabled or facilitated communication in those *teachable moments*, even when the generated vocabulary had missing or irrelevant words. Participants explained how the app could provide relevant vocabulary

during unplanned, very specific activities (e.g., horticulture), or when finding the words on the main device was too frustrating for the learner (e.g., visit to dentist).

Bringing these three concepts together we can see that auto JIT vocabulary from photographs not only reduced the workload of AAC professionals, but also enabled them to take advantage of teachable moments that arose during school or therapy activities, facilitating the use of situated cognition in stimulating symbolic AAC. These benefits also speak to the need of considering modeling of AAC use and the creation of AAC user-friendly environments when designing new technologies, as highlighted by Bircanin et al. [7], as well as some of the key facilitators (increased availability of technical solutions, motivated community of stakeholders) and barriers (complex technologies, resource restrictions) for AAC adoption and support in special education setting revealed on the ethnographic study by Norrie et al. [42]. Looking beyond our results, the impact of reduced workload on families, as hypothesized by our participants, could be even more remarkable. Learners naturally spend most of their time with family, but family members often lack the expertise and time for selecting and programming vocabulary needed to perform aided language stimulation sufficiently throughout their daily routines [36]. To underscore the magnitude of this workload, note that it's been recommended that at least 70% of interaction opportunities should be modeled through the aid tool [15]. With automatic vocabulary from photographs, families have the opportunity to greatly increase the frequency of *teachable moments* without adding workload to their routines. Future research can investigate the use of such tools with family members, potentially revealing new facets of the cooperation between AI and users, and new naturalist use cases unexplored in therapy or school settings.

4.7.2 Expanding auto-generation of vocabulary from photographs for other populations

The fact that auto JIT vocabulary from photographs can support a wide range of profiles opens up opportunities for broader user groups. As envisioned by our participants, for example, people with aphasia or dementia could use such an app independently as an alternative mode of communication, which would likely result in very different dynamics of cooperation between AI and users.

First, independent use of the tool would require individuals with complex communication needs themselves to judge the relevance of generated vocabulary. In our study, professionals acted as a filter, either by editing the vocabulary or ignoring unrelated words. That dynamic would be impacted during independent use of the tool, and arguably would be very different across individuals with autism, aphasia, and dementia, or other disabilities. In those cases, a generation method with much higher precision may be required to minimize the frustrations and confusion that the provision of unrelated or not-so-related vocabulary may cause. For example, as emerged from our findings, some narrative sentences generated were very imaginative, not exactly fitting the scene, but nonetheless were useful as a springboard to initiate communication. Such serendipitous prompts can potentially better support agency for people with aphasia in creative activities [40, 50] and people with dementia in art therapy [32] and social sharing [31, 16].

Second, although professionals judged the editing interactions as easy to perform once they had learned how, and the smartphone form-factor facilitated one-hand operation, it is unclear how our design would support individuals with complex communication needs operating the app to edit vocabulary and access it when needed. More research is needed to investigate the design of

interaction gestures and layout configurations for supporting independent editing of vocabulary. For example, people with aphasia often have motor deficits that would dictate different layout or interaction capabilities. Another interesting venue for investigation is the trade-off between vocabulary precision and editing effort across individuals. Some people may find it easier to have a larger set of vocabulary generated automatically, and then scan through it deleting undesired items; others may find this task too difficult and require a smaller, but more precise set of words.

Third, it is unclear exactly how other populations would make use of the vocabulary generated. AAC use in aphasia ranges from prompting to compensation [5]; for example, in the study by Obiorah et al. [43], involving the use of AI for helping people with aphasia ordering dinner, a participant used the support provided by a prototype to rehearse what he wanted to say rather than having the system speak out loud or automatically place the restaurant order for him. Literate users proficient of AAC, who do not need the symbolic representation of vocabulary, may use the generated words as a supplement for the next-word prediction mechanism running in their current devices, potentially increasing the communication rate. Future research is needed to understand the impact of removing the conversation partner and increasing the participation of the individual with complex communication needs in the cooperation with the AI in tools that provide automatically generated vocabulary from photographs.

4.7.3 Designing for specific use cases

Our findings from the first theme provide insights into the scenarios in which automatic JIT vocabulary from photographs can provide support, as well as how people used the support offered across

these situations. This enables future research to narrow down the design of tools such as Click AAC. Since our study was exploratory in nature, we designed Click AAC as a generic tool aimed at supporting a wide set of contexts. Future research can now explore how to leverage the capabilities of automatic JIT vocabulary from photographs to facilitate the specific activities identified, including language modeling, sentence construction, language expansion, and past event recount.

For example, researchers can explore different interfaces for facilitating single word modeling for emergent communicators, such as providing only the symbol of the main object identified in the scene, maximized in the display. Continuing, future work can look into the design of tools that facilitate the practice of sentence construction using language concepts extracted from photographs. This may include exercises for filling gaps in sentences related to the identified scene, in which sentences and available options are generated automatically. For example, taking a photograph with a boy playing soccer as input, the application could automatically generate the sentence "the boy is playing", and ask the user to complete it from a option list including baseball, tennis, and soccer.

To support language expansion, future directions include probing new interactive interfaces and organization strategies that allow easy exploration of semantically related words. For example, words semantically related to the concepts appearing in the photographs, generated by the related-expanded method, could be displayed in a secondary level that would appear only when the user selects the main concepts in the photograph. Finally, we propose studying how to generate more meaningful sentences to retell a past event, in addition to facilitating the presentation and editing of such phrases for maximum personalization. The exploration how to combine multiple photos of

the same event for providing support is another possibility, given that people often capture different moments and angles of personally relevant events.

Another avenue for future research is to study how to create a robust AAC that integrates automatic vocabulary photographs. Our findings pointed to some design opportunities, such as the use of a customizable core vocabulary board across all pages, consistent spacial arrangement of items to support motor planning, access to a keyboard, and possibility to do morphological inflections (e.g., plural and past tense).

The understanding of the usefulness of such tool on school settings also raises questions on the other form factors that such application could be created for. The use of tabletop displays and smart boards, for example, may provide new opportunities for providing a "shared communication space" among the entire classroom, potentially increasing the participation of peers in the interaction.

4.7.4 Improving quality of vocabulary generated

Our study did not focus on evaluating the quality of vocabulary generated through controlled experiments. Nonetheless, our findings are able to provide insights into some common, general patterns in the quality of vocabulary generated in relation to the photograph content, in addition to the use cases for such technology, informing i) the future selection of machine learning models and training dataset for improved scene recognition, ii) context-related vocabulary generation methods, and iii) the selection of adequate datasets for evaluating generation methods during early stages of system design.

Future research can integrate existing techniques for identifying cartoon's characters [58, 41]

and person re-identification [57], for example, and study whether these models are able to attend the needs of AAC professionals and learners during their routine activities. Another thread of research can study forms of cooperation between AAC users, professionals, and AI to achieve enhanced support. This includes, for example, new techniques that incorporate corrections on the image descriptions and vocabulary set generated made by all users for retraining or reinforcing the image identification model and/or vocabulary generation method over time, aiming at improving their overall accuracy and precision.

The findings that emerged in the third theme also inform how novels methods for expanding the image description into a set of contextually related terms following user's own style are needed. The narrative method used by Click AAC used corpora from adults in the USA. This was insufficient, leading to mismatch between users language styles and support offered. Future research should investigate generation methods for AAC that accommodate regional styles, and more importantly, that provide children and teenagers with language that sounds like their peers'. One possible avenue is to reproduce the user language style by applying the lexicon terms manually associated with a certain photograph to new photographs containing similar elements (as judged by the AI) during the generation process. Other strategy could be to reinforce the generation method with vocabulary selected during communication.

The necessity of running performance evaluations of AAC systems on datasets has been discussed in the field [30, 18]. Obtaining quantitative findings that are statistically significant and can inform the fine-tuning of internal components for optimizing the system, and anticipating flaws before testing the system with end users are the main reasons. In the initial evaluation of the sto-

rytelling generation method by de Vargas and Moffatt [18], authors found that the method was robust to variations in the input photograph. However our findings revealed that the technique for identifying the scene failed for several AAC use cases, leading to unrelated vocabulary and lack of support. Our findings on what kind of photographs professionals and learners want to use the technology with inform the construction of new datasets for this first stage of system evaluation that better represents AAC use. A possible next step would be to extend the VIST dataset with photos and vocabulary for cartoon characters, popular people, school objects, and toys.

4.7.5 Limitations

Our approach to recruiting professionals interested about the concept uncovered perspectives essential for exploring the broad use of automated vocabulary from photographs for AAC. Professionals expertise allowed us to understand the unique needs of users when learning symbolic AAC, and how auto JIT vocabulary can be integrated into their existing practices to support the learning process. However, it hindered direct investigation into how AAC learners interact with the tool.

Future work could perform on-site observations during therapy and school sessions to better understand the interactions between professionals, learners, and the tool, as well as assessing the level of language support provided for different situations through more controlled experiments. A possible approach is to employ a single-subject treatment design to measure the difference in the individuals lexical retrieval skills when using Click AAC and other tools to support the person retelling past activities, such as in the study by Mooney et al. [39].

4.8 Conclusion

The immense potential of the "iPad and mobile technology revolution" for benefiting AAC users has been discussed for more than a decade, but current symbol-based tools still have not realized the advantages brought by recent advancements in artificial intelligence and context-aware computing. In this work, we integrated computer vision and machine learning techniques proposed by de Vargas and Moffatt [18] to create Click AAC—a mobile application that generates situation specific communication boards automatically from photographs. We conducted a user study with AAC professionals and their clients with complex communication needs who used the application in their routine practices for therapy sessions or school activities. We contribute a nuanced understanding of how situation-specific vocabularies automatically generated from photographs can support communication and language learning for individuals with complex communication needs, offering new insights into the design of automatic vocabulary generation methods and interactive interface to provide adequate support across naturalistic scenarios of use and goals.

4.9 Acknowledgements

We would like to thank all participants, who generously shared their time and expertise in the development of this work. This research was funded by the Fonds de Recherche du Québec - Nature et Technologies (FRQNT), the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2018-06130], the Canada Research Chairs Program (CRC), and by AGE-WELL NCE, Canada's technology and aging network.

Bibliography

[1] A comparison of visual scene and grid displays for people with chronic aphasia: A pilot study to improve communication using AAC, author = Brock, Kris and Koul, Rajinder and Corwin, Melinda and Schlosser, Ralf, year = 2017, journal = Aphasiology, publisher = Taylor & Francis, volume = 31, number = 11, pages = 1282–1306.

[2] Lawrence W Barsalou. Language comprehension: Archival memory or preparation for situated action? 1999.

[3] David Beukelman, Jackie McGinnis, and Deanna Morrow. Vocabulary selection in augmentative and alternative communication. *Augmentative and alternative communication*, 7(3): 171–185, 1991.

[4] David R Beukelman, Pat Mirenda, et al. *Augmentative and alternative communication*. Paul H. Brookes Baltimore, 1998.

[5] David R Beukelman, Susan Fager, Laura Ball, and Aimee Dietz. AAC for adults with acquired neurological conditions: A review. *Augmentative and alternative communication*, 23 (3):230–242, 2007.

[6] David R Beukelman, Karen Hux, Aimee Dietz, Miechelle McKelvey, and Kristy Weissling. Using visual scene displays as communication support options for people with chronic, severe aphasia: A summary of AAC research and future research directions. *Augmentative and Alternative Communication*, 31(3):234–245, 2015.

[7] Filip Bircanin, Bernd Ploderer, Laurianne Sitbon, Andrew A Bayor, and Margot Brereton. Challenges and opportunities in using augmentative and alternative communication (AAC) technologies: Design considerations for adults with severe disabilities. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, pages 184–196, 2019.

[8] Rolf Black, Joseph Reddington, Ehud Reiter, Nava Tintarev, and Annalu Waller. Using NLG and sensors to support personal narrative for children with complex communication needs. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 1–9, 2010.

[9] Sarah Blackstone, J Light, D Beukelman, and H Shane. Visual scene displays. *Augmentative Communication News*, 16(2):1–16, 2004.

[10] Nancy C Brady, Kandace Fleming, Kathy Thiemann-Bourque, Lesley Olswang, Patricia Dowden, Muriel D Saunders, and Janet Marquis. Development of the communication complexity scale. 2012.

[11] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[12] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4):589–597, 2019.

[13] John Seely Brown, Allan Collins, and Paul Duguid. Situated cognition and the culture of learning. *Educational researcher*, 18(1):32–42, 1989.

[14] Joan Bruno and David Trembath. Use of aided language stimulation to improve syntactic performance during a weeklong intervention program. *Augmentative and Alternative Communication*, 22(4):300–313, 2006.

[15] Shakila Dada and Erna Alant. The effect of aided language stimulation on vocabulary acquisition in children with little or no functional speech. 2009.

[16] Jiamin Dai and Karyn Moffatt. Making space for social sharing: Insights from a community-based social group for people with dementia. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–13. ACM, 2020.

[17] Simon De Deyne, Steven Verheyen, Amy Perfors, and Daniel J Navarro. Evidence for widespread thematic structure in the mental lexicon. In *CogSci*, 2015.

[18] Mauricio Fontana de Vargas and Karyn Moffatt. Automated generation of storytelling vocabulary from photographs for use in AAC. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1353–1364, 2021.

[19] Carrie Demmans Epp, Justin Djordjevic, Shimu Wu, Karyn Moffatt, and Ronald M Baecker. Towards providing just-in-time vocabulary support for assistive and augmentative communication. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 33–36, 2012.

[20] Emma Dixon and Amanda Lazar. Approach matters: Linking practitioner approaches to technology design for people with dementia. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–15. ACM, 2020.

[21] Kathryn DR Drager, Janice Light, Jessica Currall, Nimisha Muttiah, Vanessa Smith, Danielle Kreis, Alyssa Nilam-Hall, Daniel Parratt, Kaitlin Schuessler, Kaitlin Shermetta, et al. AAC technologies with visual scene displays and "just in time" programming and symbolic communication turns expressed by students with severe disability. *Journal of intellectual & developmental disability*, 44(3):321–336, 2019.

[22] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C Platt, et al. From captions to visual concepts and back. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1473–1482, 2015.

[23] Brenda Fossett and Pat Mirenda. Augmentative and alternative communication. *Handbook on developmental disabilities*, pages 330–348, 2007.

[24] Carol Goossens'. Aided communication intervention before assessment: A case study of a child with cerebral palsy. *Augmentative and Alternative Communication*, 5(1):14–26, 1989.

[25] Jessica Gosnell, John Costello, and Howard Shane. Using a clinical approach to answer "what communication apps should we use?". *Perspectives on augmentative and alternative communication*, 20(3):87–96, 2011.

[26] Audrey L Holland. Language therapy for children: Some thoughts on context and content. *Journal of Speech and Hearing Disorders*, 40(4):514–523, 1975.

[27] Ting-Hao Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, et al. Visual storytelling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1233–1239, 2016.

[28] Eunsook Hyun and J Dan Marshall. Teachable-moment-oriented curriculum practice in early childhood education. *Journal of Curriculum Studies*, 35(1):111–127, 2003.

[29] Shaun K Kane, Barbara Linam-Church, Kyle Althoff, and Denise 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*, pages 49–56, 2012.

[30] Per Ola Kristensson, James Lilley, Rolf Black, and Annalu Waller. A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2020.

[31] Amanda Lazar, Caroline Edasis, and Anne Marie Piper. Supporting people with dementia in digital social sharing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2149–2162, 2017.

[32] Amanda Lazar, Jessica L. Feuston, Caroline Edasis, and Anne Marie Piper. Making as expression: Informing design with people with complex communication needs through art therapy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–16. ACM, 2018.

[33] Janice Light. Interaction involving individuals using augmentative and alternative communication systems: State of the art and future directions. *Augmentative and alternative communication*, 4(2):66–82, 1988.

[34] Janice Light. Toward a definition of communicative competence for individuals using augmentative and alternative communication systems. *Augmentative and alternative communication*, 5(2):137–144, 1989.

[35] Janice Light, Kathryn Drager, John McCarthy, Suzanne Mellott, Diane Millar, Craig Parrish, Arielle Parsons, Stacy Rhoads, Maricka Ward, and Michelle Welliver. Performance of typically developing four-and five-year-old children with AAC systems using different language organization techniques. *Augmentative and Alternative Communication*, 20(2):63–88, 2004.

[36] Janice Light, David McNaughton, and Jessica Caron. New and emerging AAC technology supports for children with complex communication needs and their communication partners: State of the science and future research directions. *Augmentative and Alternative Communication*, 35(1):26–41, 2019.

[37] Janice Light, Krista M Wilkinson, Amber Thiessen, David R Beukelman, and Susan Koch Fager. Designing effective AAC displays for individuals with developmental or acquired disabilities: State of the science and future research directions. *Augmentative and Alternative Communication*, 35(1):42–55, 2019.

[38] Eugene T McDonald and Adeline R Schultz. Communication boards for cerebral-palsied children. *Journal of Speech and Hearing Disorders*, 38(1):73–88, 1973.

[39] Aimee Mooney, Steven Bedrick, Glory Noethe, Scott Spaulding, and Melanie Fried-Oken. Mobile technology to support lexical retrieval during activity retell in primary progressive aphasia. *Aphasiology*, 32(6):666–692, 2018.

[40] Timothy Neate, Abi Roper, Stephanie Wilson, and Jane Marshall. Empowering expression for users with aphasia through constrained creativity. pages 1–12. Association for Computing Machinery, 5 2019.

[41] Nhu-Van Nguyen, Christophe Rigaud, and Jean-Christophe Burie. Comic characters detection using deep learning. In *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, volume 3, pages 41–46. IEEE, 2017.

[42] Christopher S Norrie, Annalu Waller, and Elizabeth FS Hannah. Establishing context: AAC device adoption and support in a special-education setting. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 28(2):1–30, 2021.

[43] Mmachi God'sglory Obiorah, Anne Marie Marie Piper, and Michael Horn. Designing AACs for people with aphasia dining in restaurants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

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

[45] Ehud Reiter. An architecture for data-to-text systems. In *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG 07)*, pages 97–104, 2007.

[46] Ralf W Schlosser, Howard C Shane, Anna A Allen, Jennifer Abramson, Emily Laubscher, and Katherine Dimery. Just-in-time supports in augmentative and alternative communication. *Journal of Developmental and Physical Disabilities*, 28(1):177–193, 2016.

[47] Samuel C Sennott, Janice C Light, and David McNaughton. AAC modeling intervention research review. *Research and Practice for Persons with Severe Disabilities*, 41(2):101–115, 2016.

[48] H Shane and J Costello. Augmentative communication assessment and the feature matching process. In *Mini-seminar presented at the annual convention of the American Speech-Language-Hearing Association, New Orleans, LA*, 1994.

[49] Martine M Smith. Language development of individuals who require aided communication: Reflections on state of the science and future research directions. *Augmentative and Alternative Communication*, 31(3):215–233, 2015.

[50] Carla Tamburro, Timothy Neate, Abi Roper, and Stephanie Wilson. Accessible creativity with a comic spin. pages 1–11. Association for Computing Machinery, Inc, 10 2020.

[51] Jennifer J Thistle and Krista M Wilkinson. Working memory demands of aided augmentative and alternative communication for individuals with developmental disabilities. *Augmentative and Alternative Communication*, 29(3):235–245, 2013.

[52] Nava Tintarev, Ehud Reiter, Rolf Black, and Annalu Waller. Natural language generation for augmentative and assistive technologies. In *Natural Language Generation in Interactive Systems*, pages 252–277. Cambridge University Press, 2014.

[53] Nava Tintarev, Ehud Reiter, Rolf Black, Annalu Waller, and Joe Reddington. Personal storytelling: Using natural language generation for children with complex communication needs, in the wild.... *International Journal of Human-Computer Studies*, 92:1–16, 2016.

[54] Sarah E Wallace and Karen Hux. Effect of two layouts on high technology AAC navigation and content location by people with aphasia. *Disability and Rehabilitation: Assistive Technology*, 9(2):173–182, 2014.

[55] Bruce Wisenburn and D Jeffery 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, 2008.

[56] Bruce Wisenburn and D Jeffery Higginbotham. Participant evaluations of rate and communication efficacy of an AAC application using natural language processing. *Augmentative and Alternative Communication*, 25(2):78–89, 2009.

[57] Di Wu, Si-Jia Zheng, Xiao-Ping Zhang, Chang-An Yuan, Fei Cheng, Yang Zhao, Yong-Jun Lin, Zhong-Qiu Zhao, Yong-Li Jiang, and De-Shuang Huang. Deep learning-based methods for person re-identification: A comprehensive review. *Neurocomputing*, 337:354–371, 2019.

[58] Yi Zheng, Yifan Zhao, Mengyuan Ren, He Yan, Xiangju Lu, Junhui Liu, and Jia Li. Cartoon face recognition: A benchmark dataset. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2264–2272, 2020.

Chapter 5

Discussion

This thesis explores the design of the first AAC tool able to generate vocabulary symbols automatically from input photographs with the goal of supporting language learning and use for people with communication disabilities. By proposing a novel natural language generation method that creates a set of words and phrases related to input photographs, and integrating such method into an interactive mobile application, the project enables the investigation of how this technology can support users learning and using symbolic language. The methods and systems introduced in chapters 3 and 4 lay the technical foundation for research on automated photograph-based AAC, providing insights into the design of automatic vocabulary generation methods and interactive interfaces for adequate communication support on naturalistic school and therapy settings.

In addition to its core contributions, this thesis applies a creative and ambitious approach for assessing novel AAC tools in naturalistic settings and in the context of pandemic restrictions. To avoid redundancy with the previous chapters' discussions, this section elaborates on the process

through which these contributions were obtained, focusing on the challenges in creating and validating novel context-adaptive AAC systems, and studying them in naturalistic settings. We also provide practical recommendations for future researchers exploring the topic.

5.1 Designing and validating context-adaptive AAC vocabularies

The Natural Language Processing field offers several techniques that enable the understanding and generation of natural language for various applications, such as text translation, dialogue systems, and automatic summarization [6, 10]. Independently of the technique, the quality of vocabulary generation is always highly dependent on the dataset used for training the language models [3]. Not only is a huge amount of data needed, but the data must be representative of the envisioned use cases. These two requirements are extremely difficult to obtain in the AAC realm [16, 15, 12, 14]. Previous research in the field has circumvented these challenges by constructing crowd-sourced corpora of fictional AAC messages [16, 15] or selecting existing datasets [14] that approximate the envisioned contexts of use. The present thesis tackles the challenge in a similar way, applying the largest narrative dataset linked with photographs of real events available at this time (VIST [8]), both as a core component of the generation method, and as ground-truth corpora for the performance exploration phase.

The difficulties of NLG for AAC are not limited to the design and implementation stages. Validating and fine-tuning intelligent systems that infer or predict users' needs is even more challenging due to the impossibility of applying user-centred design practices established in the HCI

field. While HCI researchers have explored AAC systems through various user studies, from field studies of prototype systems [1, 2, 18] to controlled performance evaluations [5, 13] and qualitative explorations of design concepts [4, 9, 19], a user study designed to investigate the trade-offs between different system designs (i.e., system controllable parameters) or to understand the impact of factors external to the system (e.g., input photographs) would encounter major practical and ethical barriers. First, the low level of social participation commonly observed among people with complex communication needs, combined with the rate-limited nature of AAC, would require field experiments lasting months, if not years, to produce sufficient data to comprehensively explore all possible combinations of parameters and enable statistical comparisons of performance metrics. Second, the heterogeneity of AAC users would introduce external factors that could lead to erroneous interpretations of the effect of the different system parameters under investigation. For example, determining which generated words are relevant based on users' vocabulary choices during conversation would be subject to the individual's visual acuity and symbol processing abilities, and not only the system's provision (or lack) of desired vocabulary. Third, providing participants with a tool that has been deliberately configured for the sole purpose of covering a possible point in the design space, and exposing them to real-world situations to probe the quality of support would be ethically questionable.

The need to analyze the theoretical performance of context-aware AAC systems before undertaking research with users has recently been raised and debated within the HCI community. In the best paper of CHI 2020, Kristensson et al. [12] proposed a novel approach adapted from design engineering. According to their approach, researchers should determine a minimal functional design

in which key functions and necessary sub-functions of the system are identified, without concerning the actual implementation of these functions. Then, the functional design is parameterized into controllable and uncontrollable factors affecting the system performance. Finally, research should evaluate different conceptual models by visualizing the theoretical keystroke-saving rates under different assumptions of the controllable and uncontrollable system parameters. This technique has been successfully applied to investigate sentence prediction in AAC [12], predictive text on mobile phones [11], and a multi-turn dialogue system for AAC [14].

Although our exact methodological approach in Chapter 3 departs from that proposed by Kristensson et al. [12], their goals and motivations guided our own evaluation design. Instead of studying a conceptual model based on assumptions of underlying components (e.g., the accuracy of word prediction and context tagging, the level of sentence re-use), we conducted a quantitative exploration of an actual implementation of our method, observing the system performance under different controllable parameters while monitoring uncontrollable factors (e.g. content of input photograph). This approach allowed us to validate the method by demonstrating that generated vocabulary outperforms a baseline representative of current AAC tools, to find the best design choices for our method, and to discard system components that did not improve performance, before moving to the second stage of the project involving human participants using our prototype in naturalistic settings (Chapter 4). Without such evaluation presented in Chapter 3, participants of our user study would have used the system with a sub-optimal design that could have impacted the level of communication support during school and therapy activities, possibly leading to different qualitative findings that do not represent the potential of automatic generation of vocabulary

support from photographs.

5.2 Studying novel AAC technologies in naturalistic settings

The quantitative exploration approach for validating the vocabulary generation method is of great importance in the initial stages of a novel context-adaptive AAC system. However, once validated and fine-tuned through computer simulation, the generation method should be ultimately integrated into an actual AAC system for the assessment or exploration of its communication support with end users. However, this assessment is not trivial because of the unpredictability of real-world situations and the technical challenges in creating a complete prototype and deploying it ”in the wild” . As a result, most research in the field to date has focused on experiments in settings contrived by researchers to study isolated factors, such as the appropriateness of a certain interaction technique or layout for selecting desired vocabulary [19, 17], or the communication support in specific activities [7]. While providing valuable evidence of certain aspects of the support provided, these approaches lack the spontaneity of use across necessary and desirable moments and assume that the imposed experimental conditions are representative of real-life events.

This thesis presents an approach to studying broad questions related to the use of a novel AAC technology in naturalistic settings, including target populations, scenarios of uses, and improvements needed in the prototype AAC tool implemented. This approach evolved from another research design because of the impossibility of conducting research in close contact with human participants created by the Covid-19 pandemic. In this initial research design, our prototype would

be employed as an autobiographical storytelling tool for supporting people with aphasia retelling personally relevant events and activities. The user study would take place in a local conversation group for people with aphasia. We would provide each participant with a mobile device running the prototype app, and they would use it to capture personally relevant moments in their daily lives. Researchers and participants would meet once a week to have a conversation about things they photographed with the app, and language-production metrics (e.g., wpm, number of target words communicated) would be collected and analyzed through a single-subject experimental design. However, pandemic-related public health measures halting all research involving human participants prevented us from interacting with that population and collecting data.

To tackle these broad questions in the context of a global pandemic, the approach adopted in Chapter 4, at its core, relies on the public distribution of the app through mainstream platforms (i.e., Apple and Android app stores), and on recruiting AAC professionals for testing the app in their usual work routines with their clients with disabilities. Although moving from the evaluation of communication production levels in a laboratory setting to a broad investigation of app usage in naturalistic settings allowed us to reach stronger research implications, two major technical and practical challenges hindered the execution of the study design: i) reaching a large number of interested professionals who meet the selection criteria, and ii) developing and providing them with a software that can be installed on their own devices, and used "in the wild" with no direct support from researchers.

The next sections detail and discuss those challenges and how they were approached within the scope of this thesis, providing meaningful advice to future researchers willing to apply a similar

approach.

5.2.1 Reaching out participants

Local entities and organizations, such as community centres and support groups are valuable resources for connecting researchers to stakeholders during in-person research. Similarly, web communities in social networks can be of great value for remote AAC studies such as the one conducted in this thesis. For example, on Facebook, several groups[1] provide support to AAC users, their families, and professionals. Members have formed a community on which they often rely for learning about best practices, new tools, scientific findings, and new research projects.

Being a member of these groups was fundamental for the development of this thesis for two main reasons. First, it enabled me to understand the interest in automated photo-to-vocabulary technologies and the feasibility of conducting the intended user study. In the initial stages of the research project, discussions between professionals provided me the means for forming a solid understanding of current clinical and pedagogical practices involving AAC. While scientific literature in the field provides a reliable summary of best practices and novel directions, professionals' and families' discussions in web communities can facilitate the understanding of the main needs and struggles, and how they are usually approached by the involved parties. Member discussions also serve to highlight literature that has being widely accepted and thus incorporated in the practical field. In the following stages, before submitting the study design to the slow REB review process, being a member of these groups enabled me to gauge the initial impression of professionals

[1] "AAC Through Motivate, Model, Move Out Of The Way", "Ask Me, I'm an AAC user!", "AAC for the SLP"

and family members and to plan the study logistics. Through a post in the group "AAC Through Motivate, Model, Move Out Of The Way" that briefly explained the potential study design and the app concept (illustrated by a high-fidelity wireframe), 287 members expressed the necessity of such technology and their excitement to participate in the study, generating 353 comments in a few hours. In addition, many members from non-English speaking countries (e.g., Spain, Israel, Denmark, France, Brazil) questioned the possibility of having the app translated and highlighted the lack of AAC tools for their native languages. Through this initial interaction, we decided to narrow down the participant pool to only AAC professionals, given that several speech language pathologists and AAC evaluators were interested in the research, and their expertise evaluating AAC solutions with a range of individual profiles would be more beneficial than that of family members in answering the research questions at this stage. Second, after the app development and REB approval, it was possible to advertise the app to numerous potential participants (37 thousand members in the group "AAC for the SLP") from widely diverse social and cultural backgrounds. The advertisement of the app in a social network group also facilitated word-of-mouth: people would tag other members whom they thought would be an appropriate fit for the app, and would start discussing the possibilities of the technology, drawing the attention of other community members.

Although our approach yielded approximately 1500 downloads and attracted 180 prospective participants (i.e., who entered their email in the app to receive the consent form) in the first month after advertising in social network groups, data collection was not smooth. To reach a total of 13 participants who tested the app with their AAC learners, it was necessary to continue recruiting for

roughly 6 months and 5000 app downloads. The delay in recruiting participants and collecting data was due to several factors. The initial screening questionnaire revealed that the great majority of the interested participants were not eligible, mostly because they were not currently working with people that could potentially benefit from the technology. In other instances, the professionals would experience caseload changes during the study and stop working with potential app users. The possibility of using the app freely without participating in the research, in combination with the professionals' heavy workload, especially in the context of a world pandemic, may also have discouraged several potential participants from participating. Therefore, future research adopting a similar approach should account for the extra time needed to recruit participants and letting them use the app amid their varied life circumstances.

5.2.2 Developing and maintaining an app for participants' devices

Laboratory studies mostly involve participants operating equipment from the research team. This allows researchers to have total control in the selection of hardware and software used in the experiments, thus minimizing the odds of incompatibilities between the system's components. In addition, researchers have the freedom to choose platforms and tools for developing prototypes in accordance with their expertise and the project's requirements. On the other hand, our approach precludes researchers from choosing the platforms/systems and instead obliges the development of prototypes that meet the requirements of participants' devices.

In the present thesis, this shift implicated the need for an app able to run in both iOs and Android ecosystems. The major preference for Apple devices in the AAC field—in part because most

AAC tools to date have been developed for that platform—necessitates the development of the app for iPads and iPhones, which is often more complex than developing for Android because of the open-source and community-based nature of the latter. In addition, Apple devices are often cost-inaccessible in developing countries. Thus, if the prototype is not available for other platforms such as Android, the pool of potential participants can be severely limited in terms of their cultural backgrounds. To minimize the constraints and provide our prototype to a more diverse population, increasing the prospective participant pool, research on the use of novel AAC tools in naturalistic settings should consider the need for implementation on a range of mobile platforms and its inherent challenges.

First, researchers either need the skill set and time resources to develop and maintain at least two different code-bases (one for each platform), or they must deal with frameworks able to generate apps for different platforms from a unique code source, such as Flutter (adopted in this thesis) or React Native. With all heavy load computing depending on more advanced libraries (e.g., for computer vision and natural language processing tasks) being performed in the cloud, it is also important to adopt a client-server architecture and to minimize the number of external libraries embedded in the mobile application to guarantee compatibility across various versions of the operational systems.

Furthermore, given the extensive range of mobile devices, and the plentiful combination of their of screen dimensions, operational system versions, and hardware, our approach hinders researchers from being certain about the app behaviour in the device used, including the interface's exact appearance, which may lead to elements being too small or awkwardly misplaced. In this

thesis, these challenges were minimized by enforcing a responsive design in all screens of the app, followed by extensive testing to cover the most common screen resolutions and pixel densities. In addition, the app allows users to change the dimensions of any element in the user interface to accommodate eventual devices that require different proportions between the interface components and were not covered in pilot tests.

Finally, in addition to the requirements for the software architecture, deploying the application to the app stores imposes extra steps that are usually needed only in the commercialization phase of a new technology, and not during the research phase. While both Google and Apple have facilitated processes for beta testing (e.g., TestFlight) that circumvent the comprehensive review process involved in the app submission to the app stores, the current requirements for using them significantly reduces participation to users with the most recent operational system (i.e., launched 2 years prior to the study). So as not to limit participation, it is thus optimal to deploy the app to the actual app stores, which necessitates adhering to specific AppStores' rules. When these rules conflict with the needs of the assistive technology, developers must spend additional time addressing the reviewers' concerns, dealing with bureaucracy such as creating and paying their accounts, creating images showing the application under different screen sizes, writing the legal terms of use and creating a website to host them, and offering support, among others.

5.2.3 Ethical concerns of our approach

We have explicitly called Click AAC a prototype and announced that it was under development as a scientific research project. However, the app's availability in app stores may induce users to rely on

app support for the medium-long term, raising ethical concerns about the sudden discontinuation of this tool. In the case of successful adoption of the technology, researchers are required to maintain the app's functionality as long as possible.

Another point to be considered is the ethics of collecting and using data through the app. A concern often raised by participants involved the actual use of the photographs taken or uploaded in the app. Furthermore, participants working in public schools and other institutions often required detailed explanations about data use, which they had to submit for administrative approval. To minimize these concerns in the present thesis, we opted to collect all data exclusively outside of the app. We also clearly communicated to participants that no data was being obtained through the app, and that data used for generating vocabulary (i.e., photographs) was being used solely for this purpose. Given that we took advantage of third-party APIs for implementing some of the system's components (i.e., Microsoft Azure Vision), we also carefully examined how the third party would use the data, guaranteeing that photographs were not used for any purpose other than processing the image, and not being kept after the processing ended.

Bibliography

[1] Abdullah Al Mahmud, Rikkert Gerits, and Jean-Bernard Martens. XTag: Designing an experience capturing and sharing tool for persons with aphasia. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 325–334, 2010.

[2] Meghan Allen, Joanna McGrenere, and Barbara Purves. The field evaluation of a mobile digital image communication application designed for people with aphasia. *ACM Transactions on Accessible Computing (TACCESS)*, 1(1):1–26, 2008.

[3] KM Annervaz, Somnath Basu Roy Chowdhury, and Ambedkar Dukkipati. Learning beyond datasets: Knowledge graph augmented neural networks for natural language processing. *arXiv preprint arXiv:1802.05930*, 2018.

[4] Elke Daemen, Pavan Dadlani, Jia Du, Ying Li, Pinar Erik-Paker, Jean-Bernard Martens, and Boris De Ruyter. Designing a free style, indirect, and interactive storytelling application for people with aphasia. In *IFIP Conference on Human-Computer Interaction*, pages 221–234. Springer, 2007.

[5] Aimee Dietz, Kristy Weissling, Julie Griffith, Miechelle McKelvey, and Devan Macke. The impact of interface design during an initial high-technology AAC experience: A collective case study of people with aphasia. *Augmentative and Alternative Communication*, 30(4): 314–328, 2014.

[6] Albert Gatt and Emiel Krahmer. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170, 2018.

[7] D Jeffery Higginbotham, Ann M Bisantz, Michelle Sunm, Kim Adams, and Fen Yik. The effect of context priming and task type on augmentative communication performance. *Augmentative and Alternative Communication*, 25(1):19–31, 2009.

[8] Ting-Hao Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, et al. Visual storytelling. In *Proceedings of the 2016 Conference of the North American Chapter of the*

Association for Computational Linguistics: Human Language Technologies, pages 1233–1239, 2016.

[9] Shaun K Kane, Barbara Linam-Church, Kyle Althoff, and Denise 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*, pages 49–56, 2012.

[10] Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, pages 1–32, 2022.

[11] Per Ola Kristensson and Thomas Müllners. Design and analysis of intelligent text entry systems with function structure models and envelope analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2021.

[12] Per Ola Kristensson, James Lilley, Rolf Black, and Annalu Waller. A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2020.

[13] Sonya Nikolova, Marilyn Tremaine, and Perry R 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*, pages 155–162, 2010.

[14] Junxiao Shen, Boyin Yang, John J Dudley, and Per Ola Kristensson. Kwickchat: A multi-turn dialogue system for AAC using context-aware sentence generation by bag-of-keywords. In *27th International Conference on Intelligent User Interfaces*, pages 853–867, 2022.

[15] Keith Vertanen. A collection of conversational AAC-like communications. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–2, 2013.

[16] Keith Vertanen and Per Ola 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*, pages 700–711. Association for Computational Linguistics, 2011.

[17] Sarah E Wallace and Karen Hux. Effect of two layouts on high technology AAC navigation and content location by people with aphasia. *Disability and Rehabilitation: Assistive Technology*, 9(2):173–182, 2014.

[18] Annalu Waller, Fiona Dennis, Janet Brodie, and Alistair Y Cairns. Evaluating the use of talksbac, a predictive communication device for nonfluent adults with aphasia. *International Journal of Language & Communication Disorders*, 33(1):45–70, 1998.

[19] Kristin Williams, Karyn Moffatt, Denise McCall, and Leah Findlater. Designing conversation cues on a head-worn display to support persons with aphasia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 231–240, 2015.

Chapter 6

Conclusion

The advent of mobile devices such as smartphones and tablets has impacted the daily lives of individuals with communication disorders. First because they are now able to use mainstream technology as support for communication, instead of being restricted to proprietary devices that are substantially more expensive and bring additional stigma to users due to the general population's unfamiliarity with the device's form factor. But more importantly, because researchers and developers are now able to design innovative AAC tools taking advantage of recent advancements in artificial intelligence and context-aware computing to facilitate users' communication, such as through the prediction of relevant vocabulary from photographs. However, no research to date has dealt with the technical challenges in building such tools, and consequently there is no evidence on the different kinds of support and how they can be used in real-life settings. The factors of the dynamics between individuals with complex communication needs, their conversation partners, and automated language support were also unknown. This thesis explored, for the first time, the

creation of a prototype tool able to generate situation-specific communication boards formed by a combination of descriptive, narrative, and semantic related words and phrases inferred automatically from photographs.

Through the presentation of novel techniques, findings, and methodologies, this thesis successfully fulfilled the objectives set in Chapter 1 and supported the central hypothesis that photographs are a rich and sufficient source of information that can be used to generate contextually relevant vocabulary in an automated manner, without restricting scenarios of use or demanding actions alien to everyday life.

Our novel language generation method and the demonstration that vocabulary generated by it is more relevant than the most frequent English words for talking about events photographed enabled us to design a novel AAC tool, and investigate how AAC professionals and their clients with complex communication needs used the tool during their usual routines.

Besides the technical contributions brought by the vocabulary generation algorithm and interactive application design, our analysis of the interviews with AAC professionals who evaluated the application with their clients with disabilities contribute a deep understanding of how vocabularies generated automatically from photographs can support individuals with complex communication needs using and learning symbolic AAC. Our findings revealed that a range of user profiles can benefit from automated generation of vocabulary from photographs, and that Click AAC was used as a complementary tool to facilitate symbolic language learning and to enable communication about specific things, addressing some previously unmet needs. We also detailed the impact of having vocabularies generated instantly from photographs, revealing four main benefits in terms of

how such vocabularies: (1) reduced the workload for selecting and programming situation-specific vocabulary for professionals, which led to increased opportunities for AAC practice , (2) facilitated the immersion of learners in symbolic communication during language modeling and sentence construction activities, (3) supported the communication of personal interests, and (4) impacted on motivation and confidence engaging with symbolic AAC. Finally, this thesis also demonstrated how the perceived quality of vocabulary was directly related to the photograph content, informing future selection of machine learning models and training dataset for improved scene recognition. It also explained how participants cooperated with the AI to overcome the errors and biases introduced, providing insights into how this cooperation can be leveraged to reach improved support.

It is important to note, though, that while chapters 3 and 4 focus on particular implementations of vocabulary generation methods and interactive interfaces for AAC, the design space for both vocabulary generation methods and interactive interfaces for AAC is large and underexplored. Design Engineering offers tools to study alternatives in the architectural design at a functional level, without direct assignment of *function carriers*, i.e., particular solutions, and to explore candidate function carriers for each critical function of the system (e.g., image captioning, vocabulary expansion). I encourage future research to apply such tools in the context of vocabulary generation from photographs for symbolic AAC, as previously done for ortographic-based AAC.

While this thesis introduces an innovative system that enables studies to analyze several aspects of automated generation of vocabulary from photographs, I acknowledge that there are many other possible research directions and end-user applications that the application of artificial intelligence and context-aware computing to the field of AAC enables. Based on our findings, we encourage

future research to investigate the following questions:

- What's the optimal user interface design to support language modeling and sentence construction activities powered by automatic vocabularies from photographs?
- How can computer vision and NLG methods be created and/or applied to generate relevant vocabulary for AAC-specific use cases for which the current approach failed (e.g., related to cartoon's characters, toys, personal objects, familiar people)
- How can literate AAC users benefit from vocabularies automatically generated from photographs?
- What's the impact of using Click AAC at home on family and users?
- What's the impact of using Click AAC on communication rate?

Appendix A

Annotation Data - Contextual Information Level and Context Description Quality (Chapter 3)

All manual annotations of Contextual Information Level and Context Description Quality of VIST-VAL dataset employed in the study presented in Chapter 3 are available at `https://doi.org/10.5683/SP2/NVI701`. They are not directly included in this appendix due to their extensive length.

A.1 Files

The annotations are divided in two files:

1. **all-annotations.json**: include annotations of all 1946 photos from VIST-VAL that have 5 narrative sentences associated. Annotations were made by the first author, Mauricio Fontana de Vargas.

2. **external-annotations.json**: annotations of 514 photos made by a person unfamiliar with the study, used to calculate the interrater reliability score.

A.2 Dataset entries

Each key in the json files corresponds to one photo of the VIST-VAL dataset. Each photo entry has the following attributes:

1. **photo_id:** the original photo id in VIST-VAL.

2. **azureCaption:** the caption generated automatically by the machine learning technique adopted.

3. **photo_quality:** a score between 0 and 3 based on the number of contextual categories (environment, people/object, activity) it clearly depicts (0 when ambiguous). It is the sum of "photo_quality_location", "photo_quality_subject", and "photo_quality_activity".

4. **photo_quality_location:** a 0/1 score indicating whether the location of the scene photographed in clearly depicted.

5. **photo_quality_subject:** a 0/1 score indicating whether the subject (person or object) of the scene photographed in clearly depicted.

6. **photo_quality_activity:** a 0/1 score indicating whether the activity present in the scene photographed in clearly depicted

7. **azureCaption_quality:** a score between 0 and 3 given to the azureCaption generated, according to these rules : 0) not generated or completely unrelated; 1) misses most important elements OR contains most of important elements and a few unrelated elements; 2) contains most of important elements OR all important elements and a few unrelated elements; 3) contains all important elements in the photo and does not contain any unrelated elements

8. **groundTruthSIS:** a set of five narrative sentences from VIST-VAL associated with the photo_id

Appendix B

Preliminary Questionnaire (Chapter 4)

Background

1. What is your age ?
2. What is your gender?
3. What is your educational background?
4. What is your profession?

Experience with AAC

5. How long have you been working with AAC users for?
6. How many AAC users you currently work with?
7. What is the profile of AAC users you currently work with (e.g., communication and motor abilities)?

8. What kind of activities do you generally perform with AAC users?

Expectations about our application

9. Are you already using our application with AAC users in your practice? If not, when do you expect to try it?

10. How many of your AAC users do you think our application could support to communicate?

11. Could you describe their profiles in terms of communication and motor abilities?

12. What kind of AAC tools and other assistive technologies do they currently use?

13. How do you expect them to use our application for supporting communication?

14. In what communication scenarios do you expect our application to be best suited for?

15. Do you expect to perform any specific activity involving our application? Could you describe them?

16. What features do you think an application such ours (that generates vocabulary automatically from photos) must have?

Appendix C

AAC Professional's Feedback Interview Guide (Chapter 4)

1. Could you describe how you tried the app and how many users tried it?
2. Was it you or your students who mostly operated it?
3. Could you describe the users' communication profile?
4. Did you perform any specific activity involving our application? Could you describe them?
5. What did your students/clients/family members mostly use our application for?
6. In what scenarios (communication contexts) was the application best suited for?
7. What profile of users mostly benefited from the application? (In case of multiple AAC users)
8. Was the application used as the main AAC tool or rather as a complementary tool?

9. What are the improvements of our application in comparison to existing AAC tools?

10. Could you talk about the quality of vocabulary generated? What happened when the generation was too bad or completely unrelated? Was deleting vocabulary you did not want more work than creating the page from scratch?

11. Did you play with different generation options? Did you find one better than other?

12. Could you elaborate on the strengths and weakness of the current prototype?

13. What specific changes should be made in the application for its successful adoption (e.g., interface elements, vocabulary items, new features)

14. Could you comment if the app met the expectations you had when you first learned about the app?

15. Would you like to be contacted for testing new versions of the prototype (refined based on all participants' feedback)?

Appendix D

AAC Professional's Feedback Questionnaire (Chapter 4)

Based on your experience using our application with your clients/family members, please indicate to what extent you agree or disagree with the following statements: (each question is followed by 5 options: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree)

Interaction

1. The symbol set used was appropriate
2. The voice output quality was appropriate
3. Users could easily select a desired vocabulary item within a page
4. Users could easily remove undesired vocabulary

5. Users could easily navigate through existing pages to find a desired photo and the associated page
6. Users could easily create a new page with a new photo
7. Users tended to access/use vocabulary from previously created pages (e.g., previous days)
8. Users tended to access/use vocabulary from newly created pages (e.g., instants or minutes after creating)

Vocabulary quality

9. The generated vocabulary included words users wanted to use
10. The generated vocabulary included words users did not want to use
11. The order the vocabulary was presented was adequate

Usage

12. Users enjoyed using the application
13. Users demonstrated willingness to use the application
14. Users operated the application independently
15. Users were more communicative using the application than they usually are using other AAC tools

16. Users would benefit it there was a complete, commercially ready application based on our prototype/beta-version

Appendix E

Coding Scheme (Chapter 4)

Figures E.1, E.2, E.3, and E.4 present the coding scheme developed in Chapter 4 when analyzing data from participants' interviews. Numbers in the right column indicate the number of segments in the transcribed interviews assigned to a specific code.

Figure E.1 Coding Scheme developed in Chapter 4: Theme 1 - Flexible, complementary AAC tool for a wide range of user profiles

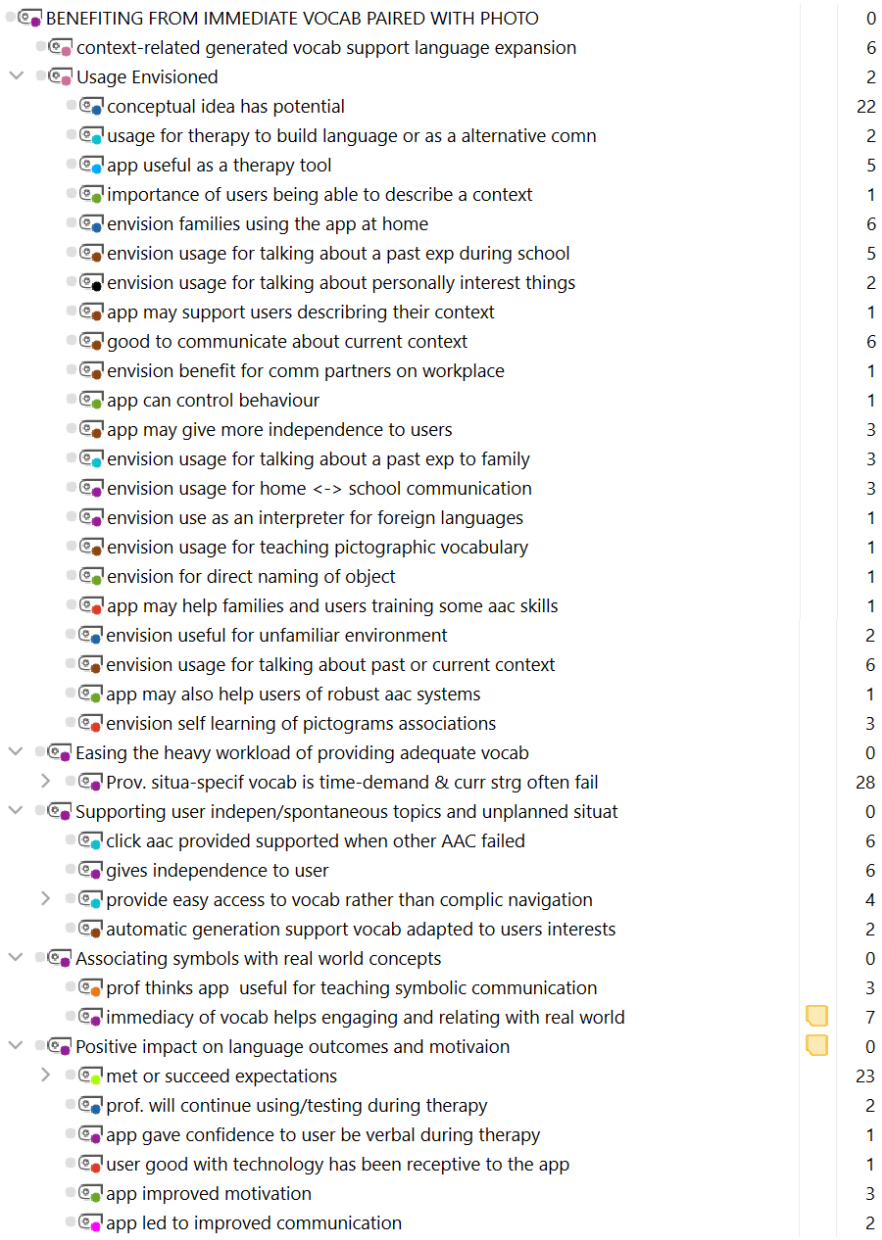


Figure E.2 Coding Scheme developed in Chapter 4: Theme 2 - Benefiting from immediacy of vocabulary

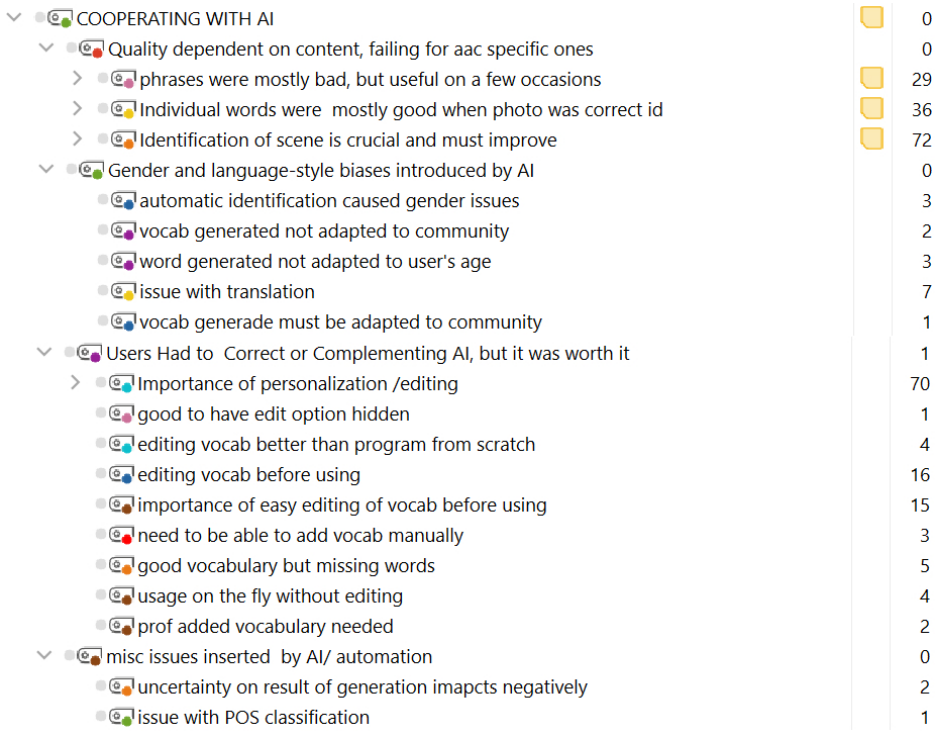


Figure E.3 Coding Scheme developed in Chapter 4: Theme 3 - Cooperating with AI

Figure E.4 Coding Scheme developed in Chapter 4: Theme 4 - Improving interactive support and new features (presented in the usability section)