XI Congresso Brasileiro de Informática na Educação (CBIE 2022)
Anais do XXXIII Simpósio Brasileiro de Informática na Educação (SBIE 2022)

Pictogram Prediction in Alternative Communication Boards: a Mapping Study

Jayr A. Pereira[1], Sheyla de Medeiros[1], Cleber Zanchettin[1,2], Robson do N. Fidalgo[1]

[1]Centro de Informática,
Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brazil

{jap2, snm, cz, rdnf}@cin.ufpe.br

[2]Department of Chemical and Biological Engineering,
McCormick School of Engineering and Applied Science,
Northwestern University,
Evanston – IL – United States of America

Abstract. *Alternative Communication Boards are high-tech Augmentative and Alternative Communication (AAC) tools that try to compensate for the difficulties faced by people with complex communication needs. Generally, these tools consist of a mobile application in which the user can construct sentences by arranging pictograms (picture+label pair representing a concept) in sequence. This study systematically maps the literature on pictogram prediction in AAC systems. We analyzed eight studies to investigate how computational methods are used for pictogram prediction, how these proposals are evaluated, and what are the studies' outcomes regarding user communication improvement. The main findings indicate the usage of different methods for pictogram prediction and a mixture of automatic and expert evaluation, which lead to inconclusive outcomes regarding user communication improvement.*

1. Introduction

Augmentative and Alternative Communication (AAC) is the area of clinical practice that attempts to compensate for difficulties or disabilities demonstrated (either temporarily or permanently) by individuals with severe disorders of communicative expression [ASHA 2022]. These difficulties and disabilities are due to several factors, such as cerebral palsy, microcephaly, autistic spectrum disorders, stroke sequelae, or apraxia. In general, these people have limitations in gestural, oral, and/or written communication, causing problems in their functional communication and socialization. To overcome these shortcomings, AAC interventions based on the selection of pictograms with captions are widely used by individuals with severe communicative disorders, especially with children, to communicate with others.

Considering the external support that can be used on AAC interventions, non-technological systems are often referred to as *low-tech* (e.g., pictures, objects, and communication books), whereas technological systems are referred to as *high-tech* (e.g., speech-generating devices, or AAC applications installed in smartphones or tablets). The use of AAC systems helps the user to express feelings and opinions, develop understanding, reduce frustration in trying to communicate, and have a greater power of choice

DOI: 10.5753/sbie.2022.225217

Figure 1. Example of Assistive Communication Boards using pictograms.

[Beukelman and Light 2013]. Moreover, AAC can also be used as a support tool for several educational activities. Figure 1 presents an example of *high-tech* AAC system with a content grid (bottom large rectangle), and a sentence area (tiny top rectangle), where pictograms are arranged in sequence.

Some research points out the barriers or difficulties faced by AAC users when using these tools for communication. [Pereira et al. 2019, Donato et al. 2018, Berenguer et al. 2022]. An example of difficulty is the time needed to construct and communicate sentences [Pereira et al. 2019, Berenguer et al. 2022], making communication less fluid and causing frustration in the user and his interlocutor. To overcome this problem, a robust AAC system must support sentence construction by facilitating pictogram searching and selection [Franco et al. 2018]. This requirement can be satisfied by using colors to classify pictograms by their part of speech (e.g., nouns, verbs, and adjectives) or by using prediction techniques to suggest the most appropriate next pictogram [Franco et al. 2018]. Prediction techniques may offer many potential benefits to AAC users [Beukelman and Light 2013]: 1) reduce the number of selections required to construct a sentence, thereby decreasing the effort for individuals; 2) provide spelling support for users who cannot accurately spell words; 3) provide grammatical support; and 4) may increase communication rate. The literature presents a growing number of published studies that use computational resources and techniques to perform pictogram or word prediction in ACBs, driven by the increasing use of AI in AAC [Sennott et al. 2019].

Although various mapping studies and systematic reviews investigate the effect of *high-tech* AAC on users' communication [Ascari et al. 2018, Aydin and Diken 2020, Dada et al. 2022], we found no similar research regarding the use of pictogram prediction in AAC systems, and its effect in users' communication. In this paper, we systematically map the literature searching for the strategies used for pictogram prediction in *high-tech*

AAC systems and the methods used to evaluate them. For doing so, we consider the guidelines presented by [Kitchenham 2004] and [Petersen et al. 2015] and assume a systematic mapping study as a particular type of systematic literature review designed to cover and give an overview of a research field by categorizing and counting contributions by predefined categories [Petersen et al. 2015]. We analyzed eight studies selected from a total of 248 and investigated how computational methods are used for pictogram prediction, how these proposals are evaluated, and the studies' outcomes. The main findings indicate the usage of different methods for pictogram prediction, which vary from knowledge databases to neural networks, assessed with a mixture of automatic and expert evaluation. The improvement in user communication is not as evident in the studies, but their results in the used metrics highlight the contribution of each proposal. The results presented in this study can be used as guidelines for AAC developers and researchers when developing AAC tools that perform pictogram prediction.

The remaining of this paper is organized as follows: in Section 2 we present the methods for performing the mapping study; in Section 3, we present the selected studies and discuss the results; and in Section 4, we present our conclusions about the study and give directions for future works.

2. Methods

2.1. Definition of Research Questions

The systematic mapping study presented in this paper aims to analyze the scientific proposals for pictogram prediction in *high-tech* AAC systems concerning the computational techniques and methods used for prediction, the methods used to evaluate the proposals, and their outcomes. Based on this aim, we formulated five research questions, each aimed at a different research facet (cf. Table 1). The facets were designed to help to answer the research questions and obtain a broad view of the current status of research in the field. They serve to classify the articles obtained from the study selection (cf. Section 2.3).

RQ1 (*Prediction method*) aims to identify the study's computational method or technique used for pictogram prediction. This information is essential to understand the field evolution over time regarding the methods employed to attack the task. RQ2 (*Prediction unit*) aims to identify the prediction unit, which is important to understand how the method makes predictions. This question is important because the definition of pictogram may not be the same among the studies. Basically, in AAC, a pictogram is picture+label pair. The label is generally a word or expression that a text-to-speech application will speak. And the picture or photo is the visual support for the user to understand the label's meaning. This question aims to identify what the study uses to perform prediction: the

Table 1. Research Questions

#	Question	Facet
RQ1	What are the computational method/algorithm/artifact used for pictogram prediction?	Prediction method
RQ2	What is the prediction unit?	Prediction unit
RQ3	How the proposal quality is assessed?	Evaluation method
RQ4	What evaluation metric is used?	Evaluation metric
RQ5	What are the study's outcomes?	Outcomes

label, the image, the pair image+label, etc. RQ3 (*Evaluation method*) aims to identify the method used to evaluate the proposal quality. The way the proposal is assessed may indicate the approach maturity. For example, an automatic (intrinsic) evaluation may show that the approach is in an initial stage of development [Jurafsky and Martin 2019]. RQ4 (*Evaluation metric*) investigates what metrics the studies used for evaluation. This information clarifies how the proposal is evaluated. RQ5 (*Outcomes*) aims to investigate the study outcomes. With this question, we want to examine if the results of the studies are positive or not concerning the baseline each study indicated.

2.2. Data Sources and Search Strategy

[Chen et al. 2010] suggest using a search string in scientific databases to combine terms of interest to extract as many related studies as possible and avoid the inclusion of unrelated studies in the results. Figure 2 presents the search string we used. AAC stands for Augmentative and Alternative Communication, which can also be found as Supplementary and Alternative Communication. AAC systems can also be referred to as voice output devices, communication boards, or voice output communication aids (VOCAS). We opted to include all these terms in the string to increase the search range. We opted to include "word prediction" in the string because as stated in Section 2.1, different studies may treat pictograms in different ways. If a study considers that the word in its label better represents a pictogram, so pictogram prediction is word prediction. Besides, pictogram prediction is about supporting sentence construction in AAC, similar to message composition and authoring.

We applied the search string in Figure 2 to six scientific databases: Scopus, Science Direct, ACM Digital Library, Taylor & Francis Online, PubMed, and Springer[1]. This study was conducted from May to June 2022, considering studies published from 2015 to 2022. The search yielded 248 studies, which we organized using the StArt tool [Fabbri et al. 2016]. Figure 3 shows the study distribution along sources. Note that the bases related to the health areas have more significant articles (PubMed and Taylor & Francis Online). This is because AAC is a clinical practice field, with studies generally conducted by speech therapists or other health professionals. However, the usage of high-tech AAC is becoming common, especially the use of AI [Sennott et al. 2019]. Therefore, studies are also found on sources more related to technologies (e.g. ACM Digital Library). We identified and removed 18 duplicated studies by using StArt duplicates classification.

[1]The string may suffer some modifications depending on the database search format.

Figure 2. Search string

(OR "alternative communication" OR "AAC" OR "voice output devices" OR "communication boards" OR "voice output communication aids" OR "VOCAS") AND ("sentence construction" OR "pictogram prediction" OR "pictogram suggestion" OR "predictive composition" OR "word prediction" OR "message composition" OR "message authoring")

Table 2. Selection Criteria

#	Criteria
E1	The study is written in a language other than English;
E2	The study is not a primary study;
E3	The study is not of Augmentative Alternative Communication field;
E4	The study focuses on AAC but does not use any strategies for pictogram suggestion;
E5	The study focus on word prediction with no pictogram;

2.3. Study Selection

To assess the relevance of the studies to be included in the final results, we applied the criteria presented in Table 2. Notice that these are exclusion criteria, meaning that the mapping results exclude the studies that fall on at least one of them. We opted to include only primary studies as they may fit better the research questions. This criterion avoids including editorials, keynotes, biographies, opinions, tutorials, workshop summary reports, progress reports, posters, thesis, dissertations, book chapters, panels, or literature mappings or reviews. These studies may not propose new approaches for pictogram prediction or are not pair revised. Some studies are in the AAC field but focus on word-based systems. We excluded these studies because they may present word prediction techniques that cannot perform pictogram prediction.

The procedure for applying the criteria consisted of screening the studies' title, keywords, and abstract. In some cases, accessing the study's full text was necessary as insufficient information is provided in the abstract to decide. It is required when studies are about AAC and mention prediction but does not specify if it is about words or pictograms in the abstracts. Two researchers performed the screening procedure to avoid individual biases. Uncertainties are solved through a researcher's meeting.

2.4. Data Extraction

For data extraction, we applied the keywording technique, as proposed by [Petersen et al. 2008]. The method assigns labels or keywords to concepts found in the study's text. Some open codes would be obtained, which have to be put into an overall structure. In the process, the codes representing the categories may be merged or renamed [Petersen et al. 2015]. According to [Petersen et al. 2015], the process may only be applied to the

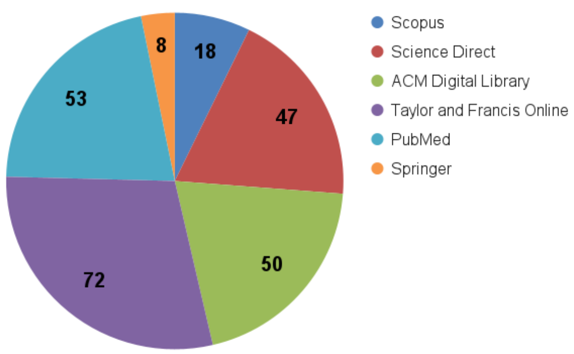

Figure 3. Studies by sources.

Table 3. Included studies

Title	Author and Year	Venue
A semantic grammar for beginning communicators	[Martínez-Santiago et al. 2015]	Knowledge-Based Systems
Context-aware communicator for all	[García et al. 2015]	International Conference on Universal Access in Human-Computer Interaction
An augmentative and alternative communication tool for children and adolescents with cerebral palsy	[Saturno et al. 2015]	Behaviour & Information Technology
Evaluating pictogram prediction in a location-aware augmentative and alternative communication system	[Garcia et al. 2016]	Assistive Technology
Compositional Language Modeling for Icon-Based Augmentative and Alternative Communication.	[Dudy and Bedrick 2018]	Association for Computational Linguistics Meeting
Predictive composition of pictogram messages for users with autism	[Hervás et al. 2020]	Journal of Ambient Intelligence and Humanized Computing
A semantic grammar for augmentative and alternative communication systems	[Pereira et al. 2020]	International Conference on Text, Speech, and Dialogue
PictoBERT: Transformers for next pictogram prediction	[Pereira et al. 2022]	Expert Systems with Applications

papers' abstract. However, if the abstracts are unclear, the method may consider the paper introduction, conclusion, or other parts. We applied keywording to the papers' full text to fit the research questions better. This way, the labels we code while reading the papers help to answer the research questions presented in Table 1.

3. Results

In this study, we analyzed 248 studies retrieved using the search strings presented in Figure 2. However, applying the criteria shown in Table 2, only eight studies were included in the final results. In Table 3, we present the included studies, their references, and publishing venue. Notice that we got three studies published in conferences [García et al. 2015, Dudy and Bedrick 2018, Pereira et al. 2020], and five published in journals. Besides, most of the venues are from the Computer Science field, except for [Garcia et al. 2016], published in a multidisciplinary journal. AAC is a multidisciplinary field [Beukelman and Light 2013], and the participation of the Computer Science community in this field is due to the need to improve AAC interventions to maximize communication and participation outcomes for individuals with complex communication needs [Light and McNaughton 2012] by using mobile applications. Besides, word or pictogram prediction may involve natural language processing techniques, which rely on machine learning and statistical analysis [Sennott et al. 2019], fields generally populated by computer scientists.

Table 4 presents the results of applying the keywording technique (cf. Section 2.4), which generated 22 keywords along the five studied facets. Next, we discuss these results regarding each research facet.

Prediction Method: We identified five methods used to perform pictogram prediction in the studies. We can say that the most common methods are those based on

Table 4. Studies keywording

	Facets and Keywords				
Study	**Prediction Method**	**Prediction Unit**	**Evaluation Method**	**Evaluation Metric**	**Outcomes**
[Martínez-Santiago et al. 2015]	Semantic Grammar	Pictogram sense	Automatic	None	No baseline
[García et al. 2015]	concept network	Pictogram label	None	None	not reported
[Saturno et al. 2015]	Direct graph	Pictogram label	Quasi experiment	Number of Pictograms, Time	Positive
[Garcia et al. 2016]	n-gram	Pictogram label	Automatic	Keystroke saving	Positive
[Dudy and Bedrick 2018]	Deep learning	Pictogram related words	Automatic	MRR, Top-n Accuracy	No baseline
[Hervás et al. 2020]	n-gram	Pictogram label, Pictogram POS	Quasi experiment	Time, Number of Pictograms, Top-n Accuracy	Positive
[Pereira et al. 2020]	Semantic Grammar	Pictogram sense	Automatic	Precision	No baseline
[Pereira et al. 2022]	Deep learning	Pictogram sense	Automatic	Perplexity, Top-n accuracy	Positive

knowledge bases: semantic grammars (2 studies), concept network (1 study), and direct graph (1 study). However, the mentioned bases have different characteristics. For example, a semantic grammar has a component that limits the sentence constructions to pre-defined grammatical structures. The direct graph used in [Saturno et al. 2015] has a component that indicates the probability of the connections between its nodes, while the other knowledge-based approaches do not. For this reason, we opted to maintain their label classification separated instead of grouping them as knowledge bases. Two studies using statistical language models based on n-grams [Hervás et al. 2020, Garcia et al. 2016]. These studies trained bi-gram language models by using pre-defined text corpora. Another characteristic they have in common is that they enrich the models' knowledge with the user's actual usage. Two other approaches employed deep learning models [Pereira et al. 2022, Dudy and Bedrick 2018]. They used neural networks trained with synthetic text corpora generated from natural language samples of text. The literature suggests that neural networks based language models may perform better than statistical models [Goldberg and Hirst 2017]. Besides, [Pereira et al. 2022] compared their model with knowledge-based approaches and demonstrated improvements. Their models outperformed the semantic grammar on predicting the correct pictogram to complete a sentence. However, neural networks may require more computational resources than statistical models or knowledge bases, making their deployment difficult in production.

Prediction unit: The analyzed studies used four types of prediction units: the pictogram label, part-of-speech (POS), set of related words, and the word-sense. As discussed in Section 1, in *high-tech* AAC systems, each pictogram has an associated label or caption, which can be a word or a multi-word expression. Some of the analyzed studies consider this label enough for making pictogram prediction [García et al. 2015, Saturno et al. 2015, Garcia et al. 2016, Hervás et al. 2020]. This way, they perform a word prediction and do not take care of polysemic words. For example, the English word "bat" can have many meanings (e.g., "nocturnal mouselike mammal" or "a club used for hitting a ball") and, similarly, many related pictograms in a given vocabulary. In addition to the label, [Hervás et al. 2020] opted to use its POS tag (e.g., verb, noun, or adjective) as a prediction unit. They trained a bi-gram language model using the sequence of POS tags as training data. The aim is to suggest to the user the pictograms labeled with the predicted POS tag. The authors compared the two approaches and noticed that the prediction improvement based on POS sequencing is not as clear. [Dudy and Bedrick 2018] treated a pictogram as a set of synonyms. For a given pictogram they look for the labels used in the Symbolstix database[2] and generate a real-valued vector using pre-trained word embeddings vectors. For example, if a pictogram has four associated words, the authors get the words' vectorial representation in the embeddings matrix and average them. The result is used as the pictogram vectorial representation. The authors used these vectors as input to their recurrent neural network. Other studies followed an approach similar to [Schwab et al. 2020], which consider that a pictogram is better represented by a concept from a dictionary (e.g., person: a human being) [Martínez-Santiago et al. 2015, Pereira et al. 2020, Pereira et al. 2022]. This approach assumes that the concept is a link between the pictogram label and its figure. [Martínez-Santiago et al. 2015] used concepts from the FrameNet database, [Pereira et al. 2020] used WordNet synsets (a set of synonyms with a glossary definition, e.g., a person is a human being), and [Pereira et al. 2022] used WordNet word-senses (a link between a word and a synset). For more details about the differences between a synset and a word sense, refer to WordNet documentation[3]. [Pereira et al. 2022] encodes each word-sense to a real-valued vector using the embeddings constructed by [Scarlini et al. 2020]. Approaches based on concepts (synsets, word-senses) may fit better polysemic words. However, it may require a prepossessing step in the prediction pipeline. An example is [Pereira et al. 2022], which parsed the text corpus for word-sense disambiguation, and [Dudy and Bedrick 2018], which requires the preexistence of a list of words for each pictogram. Approaches that use labels may not need a preprocessing step, but it does not treat polysemy.

Evaluation method: The analyzed studies performed two types of experiments for evaluating their proposals: automatic evaluation and quasi-experiments. One of the papers only presents the proposal and some usage examples but did not carry out an assessment [García et al. 2015]. First, we describe the studies that performed an automatic evaluation. [Martínez-Santiago et al. 2015] evaluated the semantic grammar at each step of its construction. They tested how well the controlled language (i.e., set of sentences) fits into the semantic grammar. [Garcia et al. 2016] ran several software simulations to measure the performance of the different pictogram prediction approaches they proposed. They evaluated the models over a set of sentences indicated by specialists as adequate for the AAC domain. [Dudy and Bedrick 2018] used the synthetic text corpus they created to evaluate their models. They divided the corpus into a five-fold split and computed the model performance in each fold. [Pereira et al. 2020] assessed the quality of the predictions made by their semantic grammar by using it to reconstruct subject+verb+object sentences extracted from the CHILDES database [MacWhinney 2014]. All the experiments performed by [Pereira et al. 2022] were in an automatic setting. They used part of the synthetic text corpus they built to assess the quality of the proposal on predicting

[2] https://www.n2y.com/symbolstix-prime/
[3] https://wordnet.princeton.edu/documentation/wngloss7wn

pictograms to complete the sentences. Besides, they asked practitioners to inform examples of sentences usually constructed in AAC systems and evaluated the models' ability to complete them. Two studies performed a quasi-experiment involving humans. [Saturno et al. 2015] analyze a student's performance through a dialogue with and without using the proposed AAC system. They observed the efficiency and satisfaction of using the system with predictions. The student is a child with complex communication needs. In [Hervás et al. 2020], a teacher working with autistic children with complex communication needs participated in the experiments, which involved reproducing the children's conversations in the class over five weeks in the AAC tool. This way, most of the studies used an automatic evaluation and assessed their proposal quality without the participation of actual AAC users. This situation can be explained by the difficulties of accessing people with complex communication needs, but it also indicates that the field is more exploratory than experimental.

Evaluation metric: Two studies did not report the used evaluation metrics [Martínez-Santiago et al. 2015, García et al. 2015]. [Saturno et al. 2015] assessed the number of pictograms used by the experiment participant to construct the proposed sentences and the time spent. [Hervás et al. 2020], which also performed a quasi-experiment, used the same metrics and a top-n accuracy, which indicates whether the pictogram the participant used was on top-n predicted by the model. Top-n accuracy was also used by [Dudy and Bedrick 2018] and [Pereira et al. 2022], the two approaches based on deep learning. These approaches used other most common metrics in the natural language processing field. [Dudy and Bedrick 2018] used Mean Reciprocal Rank (MRR), generally used to assess information retrieval systems quality, where is wanted to the best item to be in a higher position in the ranking. [Pereira et al. 2022] used a metric called Perplexity, which indicates how surprised a language model is when exposed to a new distribution of text. [Pereira et al. 2020] evaluated their proposal's precision for reconstructing the sentences from a corpus. And finally, [Garcia et al. 2016] assessed the system's quality on saving keystrokes. This way, there is not a consensus on what metric is most adequate to the task. However, top-n accuracy is the most used metric among the analyzed studies. As mentioned in Section 1, AAC systems use to present pictograms in a grid. This way, we can say top-n accuracy measures how accurate the system is in predicting the pictograms that will be shown in a grid of size n.

Outcomes: Regarding the outcomes of the presented studies, we labeled the results of the experiments as positive, neutral, negative, or no baseline. We labeled as positive the studies that presented some improvement in the used metrics when compared with preview proposals [Saturno et al. 2015, Garcia et al. 2016, Pereira et al. 2022]. No study was labeled as negative or neutral. We marked those studies that did not compare their proposal to any other as having no baseline [Martínez-Santiago et al. 2015, Dudy and Bedrick 2018, Pereira et al. 2020]. [García et al. 2015] performed no experiment, and there is no what to compare. Overall, the results presented by the papers are positive. However, it can be noticed an absence of a baseline in some studies. They are not compared to anything, so we can not say how good or worse they are on the task considering other approaches.

4. Conclusions

This paper presents a mapping study aimed to analyze the scientific proposals for pictogram prediction in *high-tech* AAC systems concerning the computation techniques and methods used for prediction, the methods used to evaluate the proposals, and their outcomes. We searched in scientific databases for papers matching this aim and found a total of 248, of which we selected eight based on pre-established criteria. We read the selected papers to extract the information needed to accomplish the aim of this study and answer the research questions.

The results indicate that the studies used different pictogram prediction methods, varying from knowledge databases to neural networks. Regarding the evaluation, some studies performed automatic evaluations to demonstrate the quality of their proposals. Other studies executed quasi-experiments with actual users and AAC experts. The improvement in user communication is not as evident in the analyzed studies, but the contribution of each proposal is highlighted by their results in the metric they used.

It can be noticed that the method for accomplishing the task (pictogram prediction) is changing over time. First, the studies used knowledge bases. Then, some studies used statistical models based on n-gram. And recently, studies employed more complex models based on neural networks. However, using only automatic evaluations and quasi-experiments demonstrate that this field is yet exploratory. The analyzed papers present initial versions of their approaches, which are not yet adequately tested by the actual users. This way, one can not conclude whether pictogram prediction affects AAC user communication. This fact opens a horizon of possible future works involving the prediction of pictograms. Some of the analyzed studies point to the difficulties of accessing AAC users, given their cognitive limitations. However, this problem must be faced by someone one day to help the field to progress.

We can consider each researcher's subjective understanding of applying exclusion criteria as a threat to the validity of this study. However, to ensure greater consistency in the process and the extracted data, all accepted papers were reviewed by two researchers. Besides, the used scientific databases may have some limitations on applying the search string. Therefore, relevant studies may not have been included in the study. Similarly, studies that predict pictograms but do not clarify this in the title, abstract, or keywords may not have been included simply because search engines did not find them.

We consider that this study may serve as guidelines for AAC developers and researchers to embassies the decisions regarding pictogram prediction in AAC systems. In future work, we intend to implement an AAC tool with pictogram prediction for Brazilian Portuguese and assess its ability to facilitate user communication in real scenarios.

Acknowledgement

This research was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES). Grant code: [88887.481522/2020-00]. The pictographic symbols used in Figure 1 are the property of the Government of Aragón and have been created by Sergio Palao for ARASAAC (http://www.arasaac.org), that distributes them under Creative Commons License BY-NC-SA.

References

Ascari, R. E. d. O. S., Pereira, R., and Silva, L. (2018). Mobile interaction for augmentative and alternative communication: a systematic mapping. *Journal on Interactive Systems*, 9(2).

ASHA (2022). Augmentative and alternative communication. `https://www.asha.org/practice-portal/professional-issues/augmentative-and-alternative-communication/`. June, 2022.

Aydin, O. and Diken, I. H. (2020). Studies comparing augmentative and alternative communication systems (aac) applications for individuals with autism spectrum disorder: A systematic review and meta-analysis. *Education and training in autism and developmental disabilities*, 55(2):119–141.

Berenguer, C., Martínez, E. R., De Stasio, S., and Baixauli, I. (2022). Parentsrsquo; perceptions and experiences with their childrenrsquo;s use of augmentative/alternative communication: A systematic review and qualitative meta-synthesis. *International Journal of Environmental Research and Public Health*, 19(13).

Beukelman, D. R. and Light, J. C. (2013). *Augmentative & Alternative Communication: Supporting Children and Adults with Complex Communication Needs*. Paul H. Brookes Baltimore.

Chen, L., Babar, M. A., and Zhang, H. (2010). Towards an evidence-based understanding of electronic data sources. In *14th International conference on evaluation and assessment in software engineering (EASE)*, pages 1–4.

Dada, S., van der Walt, C., May, A. A., and Murray, J. (2022). Intelligent assistive technology devices for persons with dementia: A scoping review. *Assistive Technology*, 0(0):1–14. PMID: 34644248.

Donato, C., Spencer, E., and Arthur-Kelly, M. (2018). A critical synthesis of barriers and facilitators to the use of aac by children with autism spectrum disorder and their communication partners. *Augmentative and Alternative Communication*, 34(3):242–253.

Dudy, S. and Bedrick, S. (2018). Compositional Language Modeling for Icon-Based Augmentative and Alternative Communication. *Proceedings of the conference. Association for Computational Linguistics. Meeting*, 2018:25–32.

Fabbri, S., Silva, C., Hernandes, E., Octaviano, F., Di Thommazo, A., and Belgamo, A. (2016). Improvements in the start tool to better support the systematic review process. In *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*, EASE '16, New York, NY, USA. Association for Computing Machinery.

Franco, N., Silva, E., Lima, R., and Fidalgo, R. (2018). Towards a reference architecture for augmentative and alternative communication systems. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 29, page 1073.

Garcia, L. F., de Oliveira, L. C., and de Matos, D. M. (2016). Evaluating pictogram prediction in a location-aware augmentative and alternative communication system. *Assistive Technology*, 28(2):83–92.

García, P., Lleida, E., Castán, D., Marcos, J. M., and Romero, D. (2015). Context-Aware Communicator for All. In Antona, M. and Stephanidis, C., editors, *Universal Access in Human-Computer Interaction. Access to Today's Technologies*, pages 426–437, Cham. Springer International Publishing.

Goldberg, Y. and Hirst, G. (2017). *Neural Network Methods in Natural Language Processing*. Morgan & Claypool Publishers.

Hervás, R., Bautista, S., Méndez, G., Galván, P., and Gervás, P. (2020). Predictive composition of pictogram messages for users with autism. *Journal of Ambient Intelligence and Humanized Computing*, 11(11):5649–5664.

Jurafsky, D. and Martin, J. (2019). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3 edition.

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews.

Light, J. and McNaughton, D. (2012). The changing face of augmentative and alternative communication: Past, present, and future challenges. *Augmentative and Alternative Communication*, 28(4):197–204. PMID: 23256853.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.

Martínez-Santiago, F., Díaz-Galiano, M. C., Ureña-López, L. A., and Mitkov, R. (2015). A semantic grammar for beginning communicators. *Knowledge-Based Systems*, 86:158–172.

Pereira, J., Franco, N., and Fidalgo, R. (2020). A Semantic Grammar for Augmentative and Alternative Communication Systems. In Sojka, P., Kopeček, I., Pala, K., and Horák, A., editors, *Text, Speech, and Dialogue*, pages 257–264, Cham. Springer International Publishing.

Pereira, J., Pena, C., de Melo, M., Cartaxo, B., Soares, S., and Fidalgo, R. (2019). Facilitators and barriers to using alternative and augmentative communication systems by aphasic: Therapists perceptions. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 349–354.

Pereira, J. A., Macêdo, D., Zanchettin, C., de Oliveira, A. L. I., and do Nascimento Fidalgo, R. (2022). Pictobert: Transformers for next pictogram prediction. *Expert Systems with Applications*, 202:117231.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*, pages 1–10.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18.

Saturno, C. E., Ramirez, A. R. G., Conte, M. J., Farhat, M., and Piucco, E. C. (2015). An augmentative and alternative communication tool for children and adolescents with cerebral palsy. *Behaviour & Information Technology*, 34(6):632–645.

Scarlini, B., Pasini, T., and Navigli, R. (2020). With More Contexts Comes Better Performance: Contextualized Sense Embeddings for All-Round Word Sense Disambiguation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages "3528 – 3539". Association for Computational Linguistics.

Schwab, D., Trial, P., Vaschalde, C., Vial, L., Esperança-Rodier, E., and Lecouteux, B. (2020). Providing semantic knowledge to a set of pictograms for people with disabilities: a set of links between wordnet and arasaac: Arasaac-wn. In *LREC*, pages 166–171.

Sennott, S. C., Akagi, L., Lee, M., and Rhodes, A. (2019). AAC and Artificial Intelligence (AI). *Topics in language disorders*, 39(4):389–403.