Aiding Non-Verbal Communication: A Bidirectional Language Agnostic Framework for Automating Text to AAC Generation

Piyali Karmakar[1], Manjira Sinha [2],
[1] IIT Kharagpur, [2]TCS Research,
Correspondence: piya.karmakar2013@gmail.com, sinha.manjira@tcs.com

Abstract

Persons with severe speech and motor impairments (SSMI), like those with cerebral palsy (CP) experience significant challenges via communication in conventional methods. Many a times they rely on Graphical symbol-based Augmentative and Alternative Communication (AAC) systems to facilitate the communication. Our work aims to support AAC communication by developing specialized datasets for direct translation of Graphical Symbols to Natural Language text. The dataset is enhanced with an automated Text-to-Pictogram generation module. The dataset is enriched with some additive information like tense-based information and subjective information (questionnaires, exclamations). Additionally, we expanded our efforts to include translation into Indian language Bengali, for those individuals with SSMI who are more comfortable communicating in their native language. We aim to develop an end-to-end language agnostic framework for efficient bidirectional communication between non-verbal AAC picture symbols and textual data.

1 Introduction

In today's increasingly connected world, effective communication is vital for personal expression and social interaction. Assistive technology (Cook and Hussey, 2001) plays a crucial role in promoting inclusion and enabling access for individuals with various disabilities and older adults. A significant number of people with various disbailities including severe speech and motor impairments (SSMI) prefers non-verbal picture symbol based communication. This menthod of communication, known as Augmentative and Alternative Communication (AAC) (Mirenda and Erickson, 2000) serves as a system that helps to bridge accessibility gaps for these individuals. Picture Exchange Communication System (PECS) (Sulzer-Azaroff et al., 2009), is a picture-based communication technology that supports writing, reading, and speech through intrinsic methods like symbolic linguistic systems. In the next section, we have described three types of PECS systems.

1.1 Pictographic Communication:

One of the most popular PECS system used for communication is the pictographic communication system (Korpi and Ahonen-Rainio, 2015). A pictogram in a pictographic commutation system, is a schematic symbol that conveys an idea or notion that can represent feelings, a means of reading, comprehending, and eventually visualizing thoughts for persons who struggle with language.

Figure 1: Pictograms

1.2 Bliss Symbolic Language:

Another Graphical symbol-based AAC system is Blissymbols or Blissymbolics [1] which is made up of several hundred fundamental symbols, called composite bliss symbols each of which stands for a different thought and may be combined to create new symbols that represent different ideas. Besides, each character or word has a unique code that determines the word's value, which makes it different from the other major writing systems in use. However, there is a lack of standardization for

[1]https://www.blissymbolics.org

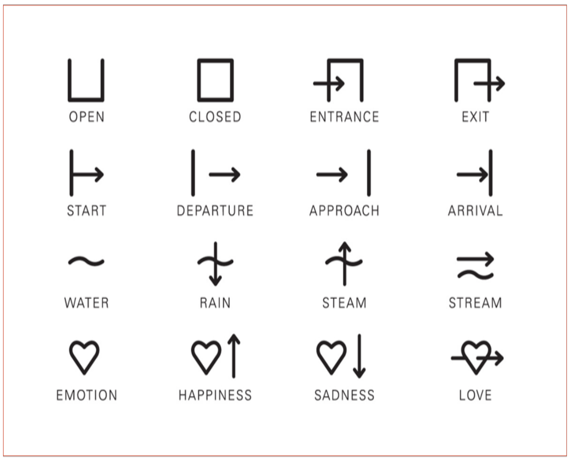

Figure 2: Bliss Symbolic Language

the compositions of Bliss symbols (Muter and Paul, 1986).

For individuals with severe speech and motor impairments (SSMI), such as those with cerebral palsy (Haak et al., 2009), graphical symbol-based AAC communication system (Garrod et al., 2007) can effectively close the linguistic gap, facilitating successful daily interactions. To enhance their overall communicative abilities, we developed translation data for graphical symbolic language to natural language text (English) translation. The data is enhanced by applying a Text-to-Pictogram generation module (PicGen) to deal with unknown text. Our system utilizes a newly created graphical symbolic corpus that includes both Bliss symbols and Pictograms. Some additive features like tense-related information and types of statement analysis were added to make the data more expressive. A neural machine translation module was implemented to test our translation data. This translation system is further extended by integrating the Indian Language Bengali to facilitate easier communication in Bengali. By incorporating the Bengali language, the system not only broadens its accessibility but also ensures that native speakers can engage with content in a way that feels natural and intuitive. This enhancement reflects a commitment to magnify the importance of linguistic diversity in technology. Our system aims to improve communication for children with SSMI by providing a more efficient and user-friendly alternative to the existing methodologies.

2 Literature Survey

AAC devices are an essential component of assistive technology that assists people with several speech or language problems communicate. This review of the literature looks at the many AAC systems(Higginbotham et al., 2007) that are used to support persons with SSMI mostly persons with CP as an alternate mode of symbolic communication. In this communication system, some users may use ambiguous symbols for texting, where they must retain and recall all of the meanings associated with each symbolic icon. However, the user needs to have strong linguistic knowledge to obtain the correct output.

Pictographic Communication: A companion system in 1998 (Wiegand and Patel, 2014), was created to expand a collection of uninfected content words into a complete phrase or sentence. Though many users with limited language proficiency may find it challenging to construct messages using the syntactic sequence (Patel et al., 2004). A Web-based translation tool called AraTraductor (Bautista et al., 2017) can convert plain Spanish text into pictograms, primarily matching word-for-word translations. Few researchers have worked on text-to-pictogram translation (Sevens et al., 2016),(Norré et al., 2021), (Mutal et al., 2022), whereas several issues have been mentioned in those systems (Korpi and Ahonen-Rainio, 2015). Even those methodologies cannot express tense-related information, or numeric information in the translated text. These limitations of using pictographic language have led to the use of other graphical symbols in communication.

Bliss Symbolic Communication: Karl Blitz (Crockford, 2003) created bliss symbols to address communication gaps. Researchers have used this Bliss symbolic language for translation (Carlson et al., 1982),(Hunnicutt, 1986) and few researchers have modified the method for the users with several speech and motor impairments (SSMI) especially persons with cerebral palsy (CP) (Olaszy et al., 1994),(Ahani et al., 2014), and also people with several speech and hearing impairments (Sándor et al., 2002).

Bliss symbolic language has a limited number (almost 6000 Bliss words) of data. Thus, most real-world texts lack a description of bliss language. Few researchers have considered pictograms and bliss symbols as graphical icon units for Augmentative Communication (AC) (Mirenda and Erickson, 2000).

Previous models mostly rely on proprietary or authorized datasets for translation, many of which are inaccessible for further research. Currently, there is no online dataset for direct translation between graphical symbols and natural language text. Most available datasets feature either Bliss symbols or pictographic languages, but rarely both. The lack of a structured symbolic language database has stalled progress in this area. In our study, we address this gap by creating a reliable dataset for translating between graphical symbols and natural language (English) text. The dataset is further enhanced by including tense-related and subjective information. A 'Text-to-Pictogram' (PicGen) generation module is applied to manage unknown text. This translation system is also extensible to other languages. We extended our research by embedding the Indian language Bengali into the translation system. Our goal is to provide individuals with SSMI a user-friendly tool to advance symbolic communication.

3 Methodology

In this section, we discussed the experiments conducted during our study. The experiment starts with preparing the translation corpus for the symbolic language-English text translation.

3.1 Translation Corpus creation:

To create the translation dataset that represents the Symbolic-English parallel dataset, we have followed the below steps:

3.1.1 Data preprocessing:

- **English Corpus Preprocessing:** We have collected two datasets: CHILDES Dataset [(a)] and English-dialoge-dataset [b]. We have chosen conversation data as it helps to capture sentiment and emotional undertones, which are vital for context-aware translations as well as for better communication. The data undergoes several preprocessing steps, including the removal of undefined words, duplicates, and extra spaces. Additionally, manual cleaning procedures are applied, such as eliminating junk words and expanding abbreviations to their correct forms. This preprocessing ensures that the English data is well-prepared for translation.

- **Bliss Corpus Preprocessing:** Bliss Corpus has been collected from the Bliss symbolic organization [(c)]. Each bliss symbol comes with one or more corresponding words as multiple words in English describe one bliss symbol. In such cases, English words are separated and kept in separate fields with corresponding bliss words (see Table 1).

BCI reference number	Gloss words
8521	a
12321	an
12321	any
25520	ability
12322	capability
12322	capacity
23401	aboard
23401	onboard

Table 1: Bliss Dataset after Data Preprocessing

- **Pictographic Corpus Preprocessing:**

  We collected pictographic corpora from different sources. A few pictogram corpuses were also collected from the Indian Institute of Cerebral Palsy (IICP) [1] to enhance the contextual features of our corpus. Pictographic Corpus contains pictographic images along with corresponding texts. We assigned numeric IDs to each of the images preceded by the English letter 'P' to make them unique in the corresponding corpus.

3.1.2 Symbolic Corpus Creation:

The Bliss corpus is not expandable in nature because of the limited number of Bliss symbols thus Pictograms were added to the corpus that addressed the limitation. The text whose corresponding bliss symbols are not present in the bliss corpus will be mapped with pictograms. Pictograms with no assigned IDs were annotated manually with unique IDs prefixed by the English letter 'P' to identify them uniquely in the whole corpus.

3.1.3 Special Features Addition:

- **Addition of Tense Information:**

  There are specific Bliss symbols for each of the tense indicators. Thus, we added bliss tense indicators to make the text more contextual and expressive.

  Figure 2 shows the output data with Parts-of-Speech (POS) tags for each English sentence and its symbolic representations. However, many words in the English sentences do not have corresponding symbolic data in the limited symbolic corpus (e.g., the word 'woof')

[1]https://www.iicpindia.org/

English Text	Symbolic Sentence with IDs
what is that present	18231 12639 16482 23547
it's a woof present	14960 12369 8521 -9999 23547
is that a hat present	12639 16482 8521 14682 23547
oh look at that present	16747 12591 1648 23547
go for a ride present	14449 14382 8521 21836 12547
look at this present	16747 12591 17720 23547

Table 2: English-Symbolic data with Tense Information

and are marked with the unique ID -9999 for unknown words.

- **Implementation of Text-to-Pictogram Generation (PicGen) Feature:** The limited number of graphical symbols makes the created graphical corpus inefficient in expressing real-world natural language text. To deal with unknown text, we applied a text-to-picture generation module renamed as 'PicGen'. The unknown English text will be passed through the module, which includes a web-based application under a Creative Commons license that accepts the textual data and returns a probable pictographic representation of the data. The generated pictograms are then assigned a unique ID prefixed by the English letter 'I' (e.g. 'woof' = "I1004") to make them unique to the whole symbolic corpus. Finally, we added these new pictograms to our created corpus. The user interface of the output dataset is displayed in Figure 3 after applying the PicGen feature, where the unknown word 'woof' is represented by the respective generated image. The final corpus is named as 'symbolic corpus', where each symbol is mapped with corresponding English words. This corpus is useful for understanding unusual message patterns in target users' communication. The full process of creating the final symbolic corpus is presented in Algorithm 1,

3.1.4 Tokenization and Mapping:

First, we lowercased the words in the English text and applied the tokenization technique. Finally, we mapped each of the tokens with the corresponding graphical symbols (symbolic IDs) from the symbolic data file. Our symbolic corpus is also associated with special indicators like question_Mark "?", and exclamatory_mark "!" with their corresponding IDs. Theadditionn of these indicators made our data more contextual.

Algorithm 1 Symbolic Corpus Creation

```
Input English Word
if Word not in Bliss Corpus then
    search Word into Pictographic corpus
    if Word in Pictographic Corpus then
        Add the corresponding pictogram into Bliss Corpus
        if Pictogram has its ID then
            Assign the pictogram with the ID prefixed by letter 'P'
        else if Pictogram has no ID then
            Assign the icon with a unique incremental ID prefixed by 'P'
            Add all pictograms into Bliss corpus and renamed as symbolic corpus
        else if Word not in symbolic corpus then
            Generate new pictogram of the Word using word2pictogram generation module
            Assign unique ID to the pictogram prefixed by letter 'I'
            Add new pictograms to symbolic corpus
        end if
    end if
end if
```

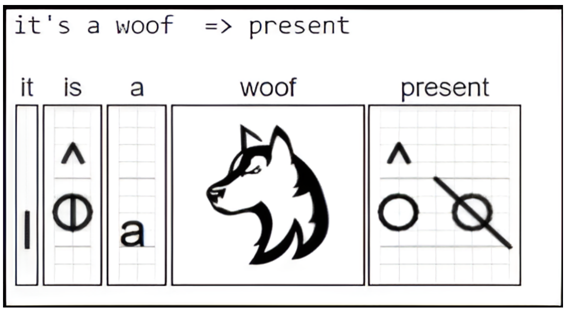

Figure 3: Outputs of Text-to-Pictogram Generation Feature

The symbolic dataset consists of words, assigned IDs, and image sources. Unique IDs enable direct capture of corresponding Bliss symbols and pictograms. The symbolic images will be displayed in front of users when they choose the assigned text. Users will learn the symbols and communicate by those displayed images of bliss and pictograms. The display of the output is shown in Figures 4.

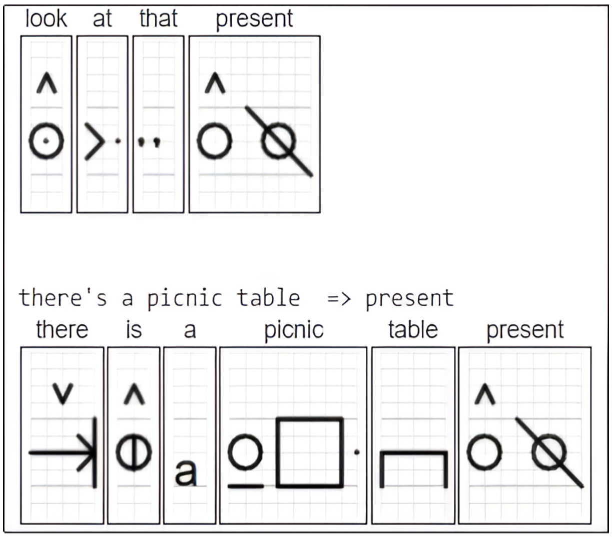

Figure 4: Display of English-Symbolic Image Data (User Interface)

3.2 Graphical Symbolic context to English text translation

We created a bidirectional Symbolic-English machine translation model that employs the state-of-the-art Neural Machine Translation (NLP) technique.

Modelling of bidirectional Symbolic-English NMT:

Our bidirectional Symbolic-English machine translation model is built on an encoder-decoder architecture, incorporating self-attention mechanisms and feed-forward neural networks. The model architecture of this sequence-to-sequence translation model was illustrated in Figure 5.

The input to the translation model consists of a sequence of tokens, such as word tokens or sub-word tokens, which are generated by the tokenizer from the input English sentence. Each token is represented as an embedding vector within a continuous vector space.

We constructed a self-attention-based transformer model with 6 encoder layers and 6 decoder layers. Each encoder consists of two sub-layers: a multi-head self-attention layer with 6 heads and a fully connected feed-forward network. Layer normalization is applied after each sub-layer to stabilize training.

The embedded tokens of the input English sentence are processed through the self-attention layer of the encoder. The attention mechanism then calculates the weights for each token in the sequence relative to a given token, producing a weighted sum for each token. These weighted sums were then passed through the feed-forward neural network. Layer normalization was applied after each sub-layer to ensure training stability. The output of one encoder serves as the input for the subsequent encoder. Finally, the output from the top-most encoder is converted into a set of attention vectors, Q, K and V which were then fed into the decoder stack.

Each decoder layer contains the same two sub-layers as the encoder: a multi-head self-attention layer with 6 heads and a fully connected feed-forward network. In addition, the decoder has a third sub-layer that performs multi-head attention over the encoder stack's output. This attention mechanism allowed the decoder to focus on relevant parts of the input sequence during decoding. As in the encoder, the output of one decoder was passed to the next decoder, where the decoding results are accumulated. Positional embeddings are added to the decoder inputs to capture the order of the tokens.

The final linear layer, followed by a Softmax layer, generates the predicted output sequence corresponding to the translated version of the input sentence.

Model Hyperparameters:

We used the following hyperparameters: **Adam** optimizer, **512** max_length, **4000** learning_rate_warmup_steps, learning_rate = **1e-4**, Dropout rate = **0.5**, batch_size = **64**, Loss = **sparse cross-entropy**. We ran our model for **10 epochs** and observed that by increasing the number of epochs, accuracy increased.

3.3 Details of Resultant Symbolic-English Translation Dataset:

Finally, we created a dataset that includes the English sentences and their corresponding symbols, here, symbols are represented in terms of their unique numeric form. Table 4 shows the snapshot of our created translation data.

After reversing the data, the final data includes the symbolic sentences and the corresponding English sentences. Symbolic sentences appear in numeric forms (unique symbolic IDs). This Symbolic-English dataset is displayed in Table 5.

3.4 Graphical Symbolic context to Bengali text translation

We extended our work by implementing a translation system in Bengali. We translated the Bengali

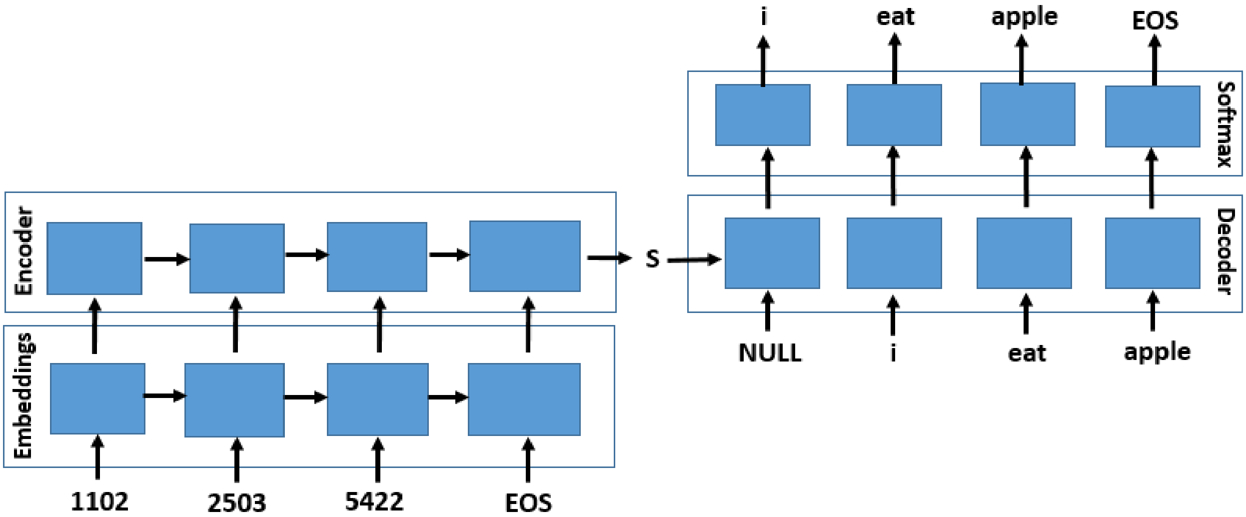

Figure 5: The Sequence-to-Sequence Model Architecture for Graphical Symbol to English Translation

sentences into corresponding English sentences using a pre-trained Bengali NMT model and then translated English sentences were passed through the process of mapping to generate respective graphical symbolic sentences. The user interface of the Bengali to Graphical symbolic Translation output data is displayed in Figure 6. Reversing the Bengali-Symbolic translation data, we structured the final Symbolic-Bengali Text translation data (see Table 3).

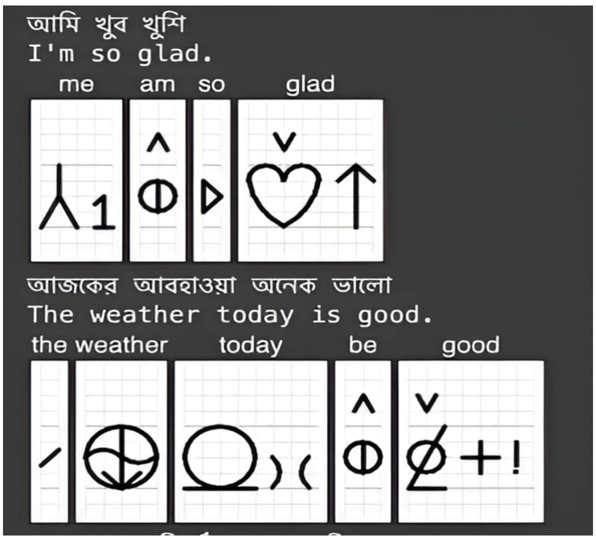

Figure 6: Bengali to AAC Picture Symbols

Symbolic Sentences	English Text	Bengali Text
24017 12639 8521 17493	there is a table	একটি টেবিল আছে

Table 3: Output Example of Symbolic Sentence to Bengali Translation

4 Conclusion

The study aimed to address the communication needs of individuals with several speech impairments. A graphic symbolic communication system provides an alternative communication tool that makes the environment more accessible to persons with Speech and Motor impairments (SSMI) especially persons with cerebral palsy (CP). With the help of creating graphical symbolic corpora, the translation datasets were generated to develop a Graphical symbolic-Natural Language Text translation system using an emerging NLP-AAC domain. The dataset is freely accessible for direct use in graphical language translation and reverse translation to enhance communication. The proposed dataset includes the usefulness of symbolic data like Bliss symbols and pictograms. For the persons with SSMI, a unique user-system translation method was designed to overcome practical difficulties. We are aiming to provide those users with an intuitive and user-friendly interface, enhanced by an additive feature such as Text-to-Pictogram Generation (PicGen) module to deal with unknown inputs. Our dataset is also enriched with contextual information like tense-related information to make the data more reliable. Our system is also extensible to other languages, such as Bengali. Translation into the Indian language Bengali makes communication easier for individuals who are comfortable with this Bengali language. Our objective is to complete the project with additional features and resources to promote greater independence for those individuals.

English Sentence	Symbolic Sentences with Respective IDs
go for a ride present	14449 14382 8521 21836 23547
here the dog's gonna go up the ladder present	14708 17700 12380 12639 14449 17739 14449 17983 17700 15166 23547
look at this present	16747 12591 17720 23547
there's a table and there's some chairs present	24017 12639 8521 17493 12374 24017 12639 17207 13148 23547
it's a woof present	14960 12639 8521 I1004 23547
shall we go eat? future	24261 18212 14449 13906 18229 25568
let's go eat present	24732 12639 14449 13906 23547
look at this little boy present	16747 12591 17720 14171 12888 23547

Table 4: Final English-Symbolic Data

English Sentence	Symbolic Sentences with Respective IDs
14708 12639 8521 13148 23547	here is a chair present
24834 17739 16440 24011 16161 14932 17700 13148 23547	want to put another person in the chair present
15671 13416 23547	very good present
15736 24732 12639 16747 18231 -9999 18212 14435 14688 12670	now let is see what else < unk > got her past
12360 P378 23547	all right present
24264 18212 13906 12591 17700 17493 18229 25568	should we eat at the table? future
13114 18465 14435 14960 17983 24017 12374 13360 17700 15166 25568	can you get it up there and climb the ladder future
12639 16482 8521 15662 18229 23547	is that a mommy ? present

Table 5: Symbolic-English Data (Reverse Data)

References

Asieh Ahani, Karl Wiegand, Umut Orhan, Murat Akcakaya, Mohammad Moghadamfalahi, Hooman Nezamfar, Rupal Patel, and Deniz Erdogmus. 2014. RSVP IconMessenger: icon-based brain-interfaced alternative and augmentative communication. *Brain Comput. Interfaces (Abingdon)*, 1(3-4):192–203.

Susana Bautista, Raquel Hervás, Agustín Hernández-Gil, Carlos Martínez-Díaz, Sergio Pascua, and Pablo Gervás. 2017. Aratraductor: Text to pictogram translation using natural language processing techniques. In *Proceedings of the XVIII International Conference on Human Computer Interaction*, Interacción '17, New York, NY, USA. Association for Computing Machinery.

Rolf Carlson, B Granstrom, and Sheri Hunnicutt. 1982. Bliss communication with speech or text output. In *ICASSP'82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 7, pages 747–750. IEEE.

AM Cook and SM Hussey. 2001. Assistive technologies: Principles and practice, volume 2a ed. *Mosby-Year Book*.

Douglas Crockford. 2003. Lesson one: An introduction to blissymbols. *Uppsala: Blissym Language Institute*.

Simon Garrod, Nicolas Fay, John Lee, Jon Oberlander, and Tracy MacLeod. 2007. Foundations of representation: where might graphical symbol systems come from? *Cognitive science*, 31(6):961–987.

Peterson Haak, Madeleine Lenski, Mary Jo Cooley Hidecker, Min Li, and Nigel Paneth. 2009. Cerebral palsy and aging. *Developmental Medicine & Child Neurology*, 51:16–23.

D. Jeffery Higginbotham, Howard Shane, Susanne Russell, and Kevin Caves. 2007. Access to aac: Present, past, and future. *Augmentative and Alternative Communication*, 23(3):243–257. PMID: 17701743.

Sheri Hunnicutt. 1986. Bliss symbol-to-speech conversion:'blisstalk'. *Journal of the American Voice I/O Society*, 3(June).

Jari Korpi and Paula Ahonen-Rainio. 2015. Design guidelines for pictographic symbols: Evidence from symbols designed by students.

Pat Mirenda and Karen A Erickson. 2000. Augmentative communication and literacy.

Jonathan Mutal, Pierrette Bouillon, Magali Norré, Johanna Gerlach, and Lucia Ormaechea-Grijalba. 2022. A neural machine translation approach to translate text to pictographs in a medical speech translation system-the babeldr use case. In *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 252–263.

Muter and Paul. 1986. Blissymbolics, cognition, and the handicapped. In Erland Hjelmquist and Lars-Göran

Nilsson, editors, *Advances in Psychology*, volume 34 of *Advances in Psychology*, pages 233–251. North-Holland.

Magali Norré, Vincent Vandeghinste, Pierrette Bouillon, and Thomas François. 2021. Extending a text-to-pictograph system to French and to arasaac. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1050–1059, Held Online. INCOMA Ltd.

Gábor Olaszy, SL Kálmán, and Péter Olaszi. 1994. Blissvox—voice output communication system for teaching, rehabilitation, and communication. In *Computers for Handicapped Persons: 4th International Conference, ICCHP'94 Vienna, Austria, September 14–16, 1994 Proceedings 4*, pages 421–428. Springer Berlin Heidelberg.

Rupal Patel, Sam Pilato, and Deb Roy. 2004. Beyond linear syntax: An image-oriented communication aid. *Assistive Technology Outcomes and Benefits*, 1(1):57–66.

Leen Sevens, Gilles Jacobs, Vincent Vandeghinste, Ineke Schuurman, and Frank Van Eynde. 2016. Improving text-to-pictograph translation through word sense disambiguation. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 131–135, Berlin, Germany. Association for Computational Linguistics.

Beth Sulzer-Azaroff, Anne O Hoffman, Catherine B Horton, Andrew Bondy, and Lori Frost. 2009. The picture exchange communication system (pecs) what do the data say? *Focus on autism and other developmental disabilities*, 24(2):89–103.

N Sándor, Géza Várady, Attila Szalmas, and Cecilia Sik-Lányi. 2002. Electronic usage of bliss symbols.

Karl Wiegand and Rupal Patel. 2014. Rsvp-iconchat: A single-switch, icon-based aac interface. In *Journal on Technology & Persons with Disabilities, 29th Annual International Technology and Persons with Disabilities Conference*.