Grenze International Journal of Engineering and Technology, June Issue

Multilingual Text-to-Pictogram Translation for Augmentative and Alternative Communication

Akashraj Giriraj[1], Eraiyanbu P[2] and Dr.Srinivasan R[3]
[1 -3]Department of Information technology Sri Sivasubramaniya Nadar College of Engineering, Chennai, India
Email: g.akash0704@gmail.com,eraiyanbupandian@gmail.com,srinivasanr@ssn.edu.in

***Abstract*—Augmentative and Alternative Communication (AAC) systems are essential for individuals with language impairments due to genetic conditions, accidents, or strokes. This paper presents two approaches for multilingual text-to-pictogram translation across English, French, and Tamil. The first approach is a hybrid method that combines contextual sentence embeddings extracted using BERT with similarity-based retrieval through an ANN index. The second approach is transformer-based, utilizing contrastive learning for effective mapping of text to AAC pictograms. Both methods aim to enhance communication for cognitively impaired individuals, and their performance is evaluated to determine the most effective solution for AAC applications.**

***Index Terms*— Augmentative and Alternative Communication, Multilingual Text Translation, Transformer Models, Semantic Mapping, Contrastive Learning, Intellectual or Developmental Disabilities (IDD).**

I. INTRODUCTION

Communication is a cornerstone of human interaction that allows individuals to convey information concerning their needs, desires, perceptions and emotional states. Communication relies on language, an organizational system of symbols and rules. But people with profound communication disabilities often find that those traditional methods of language just aren't compatible with their needs. Augmentative and Alternative Communication (AAC) plays an essential role in supporting these individuals in their communication efforts by introducing devices used to facilitate communication, such as manually created signs, symbol-based communication boards, and speech-generating devices.

As a form of AAC, pictograms were created for those who needed more support than a single symbol could provide and were therefore even more effective as a visual communication tool that used simple and recognizable symbols to represent ideas. This is because they provide a useful way to clarify any ambiguous or obscure terminology, enabling users with different degrees of cognitive ability to understand the content. Pictograms allow you to move from simple communication, appropriate for people with low cognitive functioning or in early stages of development, to a more and articulated even at higher levels, but far less robust than written language.

As useful as pictograms are in facilitating communication, there is still a wide gap between AAC users and society at large. This divide is supplemented by a lack of understanding and awareness about AAC methods. As it is, mainstream tech tools for communication today often do not include the voices of individuals with IDD as these tools tend to be complicated to use and often not accessible at all. This gap highlights the urgent need for

Grenze ID: 01.GIJET.11.2.592

digital communication interfaces with pictograms for these individuals to reduce social isolation.
To address the gap, this paper fills this gap by introducing a multilingual text-to-pictogram translation model. For handling English, French, and Tamil respectively, we take reference of lexical simplicity schemes that convert compositional sentences to a succession of pictorials in a clearer format. In addition to this, our goal is to decrease the difficulty someone may have in reading written content through lexical simplification. In addition, it has implications for those who provide everyday communication, with less idiosyncratic language and vocabulary that addresses such inclusiveness while also reducing social exclusion on the part of users with IDD.

II. Literature Survey

In Augmentative and Alternative Communication (AAC), text to pictogram transformation is an essential function for users who suffer from speech difficulties. Different technologies have been investigated to enhance AAC systems, from independent rule-based approaches to advanced neural networks, empowering users in multiple domains [8].
Natural language processing (NLP) is a newer field and many AAC systems in public use currently take advantage of NLP techniques. Mobile communications applications for medicine, such as *My Symptoms Translator* (Alvarez, 2014; Alvarez and Fortier, 2014) [1] and *MediPicto AP-HP1* serve to facilitate a common means by which physician and patient may communicate through images.
The different approaches to tokenization, especially when it comes to whole texts, with the *Glyph* application focusing more on Natural Language Processing and including preprocessing like sentence splitting, word normalization, synonym matching, etc. It uses terminology and medication databases as well as computer graphics techniques for translation of patient instruction [2] (Zeng-Treitler et al., 2014; Bui et al., 2012). However, this application was not specifically designed for people with disabilities, as was the work of Sevens and of Vandeghinste et al. (2015) [3]. The Text-to-Picto system has had some success. Kultsova et al. (2017) [15] have integrated it into a mobile application which, when used, can help Russian-speaking people with intellectual disabilities in social travel and communication. It has also been adapted for use in Indian languages (Nandy, 2019) [16].
[9] the *AraTraductor* system uses syntactic analysis to disambiguate the translation process as follows, illustrating the need for syntactic parsing to achieve a more comprehensible translation into pictograms. Moreover, a BERT-based approach known as PictoBERT [10] predicts to AAC communication boards sequences of pictograms with word-sense data, taking a step further from traditional n-gram approaches. In line with this, the *PrAACT* methodology modifies large transformer models for AAC, emphasizing user-specific tailoring and adjustability, to support communication for individuals with a wide range of requirements [11]. In medicine, the *BabelDr* system combines the Unified Medical Language System (UMLS) and neural machine translation (NMT) to generate medical text translations that can be converted into pictograms [12].
The task of text-to-pictogram translation in AAC has only been covered in a few works. A system which automatically generates pictorial representations of simple sentences described by Mihalcea and Leong (2008) [4]. They discovered that the comprehension that arises from visual descriptions is equivalent to the translations into target languages that is produced using machine translation. Their method used WordNet (Miller, 1995) [6] as a lexical resource, but ignored WordNet's conceptual relations.
Goldberg et al. [7] (2008) employed a distinctive methodology by structuring pictographs sequences along the semantic roles present in the original sentence, thereby enhancing the understanding of pictograph sequences. Joshi et al. (2006) designed an unsupervised system that enriches text stories with images by finding semantic keywords and searching an annotated image database, although they did not set out to translate entire stories.
More recently, Keskinen et al. For example, [13] proposed a picture-based instant messenger, known as *SymbolChat*. The interaction is controlled through a touch screen, and speech output, specifically oriented for people with intellectual disabilities. Overall, there are many active research pursuits in text-to-pictogram translation for AAC, but most of the work either targets the simple sentence translation of the pictogram use cases or does not directly target the needs of individuals with IDD. Future works in this domain can have an emphasis on enhancement of the Natural Language Processing techniques and better level of tailoring user-specific vocabularies to build up easy to use and adaptive AAC systems with wide range of applications.

III. METHODOLOGY

A. Dataset

The datasets for English, French, and Tamil were structured using a common format as shown in Table I to ensure consistency across languages. Each entry includes three key fields: **src**,**tgt** and **pictos**, ensuring consistency during model training and evaluation.

TABLE I. DATASET CREATION FORMAT

Tags	Definition	Example
src	source sentence of utterance	quel est ton nom
tgt	target sentence of simplified tokens	quelle être ton prénom
pictos	list of pictogram id mapped to each target token (same size as the target output)	[22624, 36480, 12281, 38570]

French Dataset: TCOF Corpus for Pictogram Translation

The French dataset originates from the TCOF corpus (Corpus of Oral French Transcriptions). TCOF consists of various kinds of interactions between children and adults, including scenarios that resemble everyday situations, and also more formal exchanges that resemble debates or medical consultations. The corpus consisted of authentic communication in real life situations between patients with disabilities (in this case, patients who use pictograms to communicate) and caregivers, medical personnel, and family members.

English Dataset: Manually created Text Corpus for Pictogram Translation

The English dataset is built similarly to the French dataset as shown in Fig.1, collecting data from several publicly available repositories like Wikipedia, Project Gutenberg – a digital library with over 60,000 free eBooks, mostly consisting of public domain works -- and Tatoeba, an open, multilingual database of sentences and their translations compiled by a large community of volunteers -- that has a high Flesch readability score, meaning any content can be comprehensible for anyone who needs simple language.

The sentences chosen represent everyday English, simplified using the **Flesch** reading ease score, with all sentences scoring **above 80** to ensure linguistic simplicity.


```
{
    "src": "She is learning how to ride a bike in the neighborhood.",
    "tgt": "she learn ride_a_bike neighborhood",
    "pictos": [7028,8029,36945,6241]
}
```


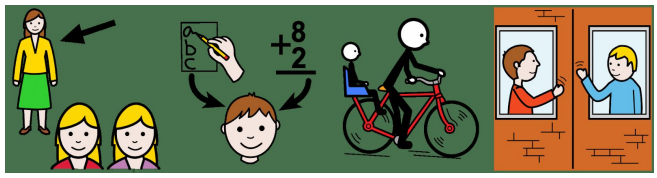

Figure 1. Example text to pictogram sequence mapping for English

Tamil Dataset: Manually Created Text Corpus for Pictogram Translation

The Tamil dataset involved scraping Wikipedia along with Project Madurai, a digital library of Tamil literary works. However, the lack of simplified Tamil text resources limited the number of sentences to only 340. These sentences were manually annotated to indicate correct pictogram mapping. No Flesch readability score was they applied unlike the English and French datasets, as there is currently no standardized scoring system for Tamil. These challenges arise from both the limited size of the dataset available and the complexity of Tamil's morphology, making the development of a useful Tamil text-to-pictogram translation system difficult.

The mappings of the target keywords (Table II) to their respective pictogram id's are used to create a pictogram vocabulary json file. The pictogram vocabulary is based on the **ARASAAC** collection consisting of more than 25.000 pictograms which have been released under Creative Commons license.

The dataset comprises tailored corpora for the task, containing sentence-pictogram mappings across three languages to support AAC applications, as outlined in Table III.

TABLE II. PICTOGRAM VOCABULARY JSON

Tags	Definition	Example
keyword	contains word/token target key-value (picto_id) pair	"pavement": [2247]

TABLE III. COMPREHENSIVE OVERVIEW OF CURATED DATASET

Language	Sentence Count	Pictogram Vocabulary
French	21,348	3,439
English	984	741
Tamil	341	692

B. Proposed Methods

Hybrid Approach for French and English

We propose a hybrid method that first extracts contextual sentence embeddings using BERT, followed by similarity-based retrieval using an Approximate Nearest Neighbor (ANN) index.

BERT-Based Embedding Extraction

Transformers are models deployed with self-attention layers to generate in-context vectors for words. BERT's WordPiece tokenizer processes the input sequence by splitting words into their subword constituents while adding special tokens [CLS] and [SEP]. Next, the tokenized sequence is passed into a pre-trained BERT model, and a set of contextual embeddings of all the tokens is returned. Finally, a mean-pooled vector over all tokens is taken to get a sentence-level embedding. These are high-dimensional embeddings that retain both semantic and syntactic relationships on which downstream classification and retrieval tasks will be based.

ANN Retrieval with Cosine Similarity

The generated embeddings are indexed with an ANN algorithm to allow efficient similarity search. At inference, the cosine similarity between the query embedding and indexed vectors are employed to retrieve the most semantically similar pictogram sequences.

Transformer based approach

This part describes a transformer-based approach of creating pictogram sequences given an input sentence in French, English, and Tamil. Sequence-to-sequence (seq2seq) models such as MarianMT, mBART-50, and Google T5 to give numerical representations of text input to convert textual input to a number that represents some semantic meaning. These embeddings are then used to generate the correct sequence of pictograms.

The proposed model is trained for the translation task of French and English Language as described in architecture Fig. 2.

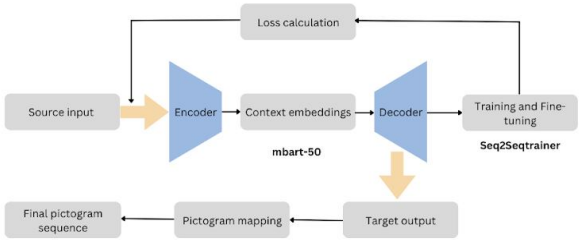

Figure 2. Architecture of proposed mbart-50 model

Encoder-Decoder Framework

It uses an encoder-decoder structure, where the input sentence is tokenized and runs through the encoder to create contextualized representations in the latent space. The decoder subsequently uses these representations to generate the corresponding sequence of pictograms while respecting the needed syntactic and semantic constraints. The encoder learns to represent relationships and structures at the level of words and sentences, while decoder generates information based on most relevant encoded data.

Self-Attention Mechanism

This mechanism enables the model to dynamically focus on different parts of the input sequence, allowing it to capture long-range dependencies and contextual relationships between tokens. The self-attention mechanism in the encoder helps contextualize each word based on surrounding words, while in the decoder, it assists in generating tokens by attending to both previous outputs and the encoded input sequence.

Contrastive Learning

Constraining the learned embeddings through contrastive learning hones the learned representations by delineating the embeddings space. Positive and negative pairs are created, where for example, similar sentences

would be pulled closer together while dissimilar sentences would be pulled further apart. A triplet loss function is applied to facilitate this process, further refining the model's capacity to differentiate semantically relevant inputs. This change leads to better retrieval performance, as well as embeddings that more closely reflect the meaning of input sentences. The entire procedure, including the construction of the pairs, the loss based on them, and the update of the model, is displayed in Fig.3.

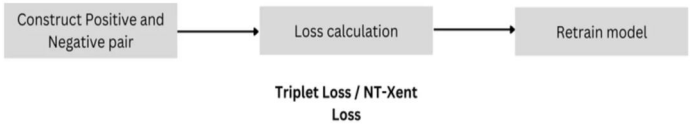

Figure 3. Contrastive learning framework for embedding refinement

C. Evaluation Metrics

BLEU (Bilingual Evaluation Understudy):
BLEU is a precision-based evaluation metric commonly used in machine translation tasks. It measures how many words or n-grams in the generated text (c) overlap with the reference text (r). The score ranges from 0 to 1, with higher scores indicating better translation quality. BLEU is computed by comparing the n-grams (N) of the generated translation with those of the reference , and a brevity penalty (BP) is applied to discourage very short translations. While efficient and widely adopted, BLEU does not account for synonyms or semantic similarity, focusing solely on exact matches.

METEOR (Metric for Evaluation of Translation with Explicit ORdering):
METEOR is another recent machine translation evaluation method which was designed to overcome some of the limitations of BLEU by taking synonymy and stemming into account. It compares the generated and reference translations by matching based on exact matches, synonyms and word stems. METEOR is based on a weighted harmonic mean of precision (P) and recall(R) and fragmentation penalty (Pf) for misaligned words, so it is more sensitive to the meaning of words than exact matches. This metric is especially useful for judging translations that preserve semantics even when the used words differ from the reference.

IV. Results And Discussions

A. Training Process and Performance Analysis

All sequence-to-sequence model training also used a **batch size** of **8** with a **learning rate** of **5e-5** and was performed for a total of **8 epochs**. We employed the AdamW optimizer to demonstrate improved convergence and the early stopping to ensure our model is not overfit by tracking the validation loss. For the **BERT**-based approach, the AutoTokenizer and AutoModel were used to produce contextual embeddings. This process iterated many times until a final list of neighbors was found relative to target tokens from the training set, using the ANN algorithm with Cosine similarity as the distance metric. By employing this method, semantically-relevant tokens closely matching the intended pictogram outputs could be retrieved.
As seen in the training graphs (see Fig.4), the models exhibit consistent convergence, with the training loss continuously decreasing and stabilizing around epoch 5. Importantly, the gradient norm steadily reduced showing increased model stability and the learning rate decay effectively retained updates to the weight prompting additional convergence. These observations were corroborated by the evaluation metrics, demonstrating that effective optimization was performed during the entire training phase.

B. Model Performance Analysis

The performance comparison across models as reported in Tables IV & V indicated that for the same problem there are distinct strengths and weaknesses across models in terms of their architecture for text to pictogram translation. ANN with BERT, though deeply semantic, scored low with 3.41% BLEU in French and 2.31% in English. This output underlines BERT's insufficiency in sequence-to-sequence settings due to its encoder-only architecture, where an autoregressive decoder is not available to produce coherently. Despite successfully capturing finer semantics at the word level, the inability to generate fluent sequences using BERT makes it difficult to apply it directly to AAC tasks that require structured pictograms to be aligned. On the other hand, MarianMT, which specializes in translation, produced much better scores of 58.19% for French translation and 62.54% for English translation. In contrast, MarianMT performed very well in direct text translation but

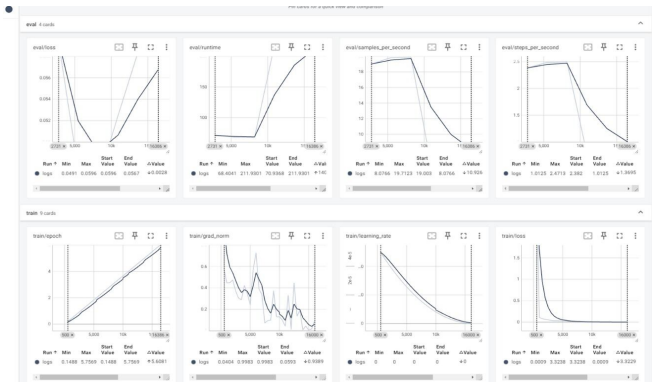

Figure 4. Convergence Analysis of Training and Evaluation Metrics Across Epochs

struggled when generating simplified phrases needed to accurately map to pictograms, and frequently generated verbosity that was misaligned with the target pictograms.

We achieved **BLEU** scores of **74.44%** and **79.64%** in **French** and **English**, respectively, with **METEOR** scores above **90%** with **mBART**, which demonstrated the most reliable and good performance among the models we tested. mBART's denoising autoencoder training and simple to complex approach using a regressive decoder that reconstructs simplified outputs are credited with this success (preserving semantics). Using mBART, which captures contextual dependencies while generating coherent sequences, played a pivotal role in aligning text to the associated pictograms. By contrast, T5 performed middlingly: 50.32% in French, 34.70% in English (BLEU). T5's encoder-decoder architecture facilitates structured text generation, but performance varied widely with data scaling, affirming the model's reliance on larger datasets to approximate fluency and correctness. T5 also achieved better performance than MarianMT in larger datasets, indicating better adaptability for extensive data availability.

TABLE IV. COMPARATIVE MODEL METRIC ANALYSIS FOR FRENCH

Model	BLEU(%)	METEOR(%)
BERT + ANN	3.41	14.35
MarianMT	58.19	78.61
mBART	74.44	92.08
T5	50.32	75.24

TABLE V. COMPARATIVE MODEL METRIC ANALYSIS FOR ENGLISH

Model	BLEU(%)	METEOR(%)
BERT + ANN	2.31	7.52
MarianMT	62.54	91.89
mBART	79.64	96.99
T5	34.70	86.86

C. Dataset Size Impact Analysis

In exploring this with measures of performance impacted via dataset size, when applied through T5, an overall positive correlation between data volume and translation quality was found. This is reflected in Table 6 where by increasing the dataset size we see that BLEU and METEOR scores get slightly improved and this shows the importance of diversity of data in the enhancement of the performance of the AAC systems.

D. Tamil Dataset and Ongoing Development

Tamil morphology is also a complex one, and due to limited available data, the translation systems for Tamil are still maturing. Although a Tamil pictogram vocabulary with 672 entries has been created, only 340 manually annotated sentences have been prepared for training the model. Continued work is underway to enhance translation precision using more sophisticated linguistic techniques and broader dataset inclusivity.

V. CONCLUSION

The results suggest significant advances in this respect based on the generated outputs of the state-of-the-art multilingual text-to-pictogram translation system presented here for both English and French text to pictogram

TABLE VI. DATASET VS PERFORMANCE ANALYSIS FOR T5

Sentence Count	BLEU(%)	METEOR(%)
50	3.40	29.91
100	19.04	46.64
500	25.22	68.01
1000	33.23	80.62

generation. And consistently mBART managed to have the highest scores of BLEU and METEOR, highlighting that this model has an excellent semantic understanding of the context and generate the output in a very structured form. How the combination of denoising training techniques, regressive decoding and semantic analysis were pivotal in improving over subsequent challenges of accuracy and maintaining context coherence with pictogram sequences as input.

Future work can focus on improving Tamil text-to-pictogram translation, which remains under development due to linguistic challenges and limited resources. Expanding the pictogram vocabulary for Tamil and enhancing model training with specialized datasets tailored for Tamil syntax and semantics could improve performance. Additionally, integrating reinforcement learning techniques and prompt-based fine-tuning strategies may further enhance model adaptability and precision. Extending the system to support real-time communication interfaces and evaluating the model's usability with AAC device users would enhance its practical impact and scalability.

REFERENCES

[1] J. Alvarez, "Visual design: A step towards multicultural health care," *Arch. Argent. Pediatr.*, vol. 112, no. 1, pp. 33–40, 2014.

[2] D. D. A. Bui, C. Nakamura, B. E. Bray, and Q. Zeng-Treitler, "Automated illustration of patients instructions," *AMIA Annu. Symp. Proc.*, vol. 2012, p. 1158, 2012.

[3] M. Norré and Vandeghinste, "Can pictograph translation technologies facilitate communication and integration in migration settings?" in *Proc. Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, 2017, pp. 1–4.

[4] R. Mihalcea and C. W. Leong, "Toward communicating simple sentences using pictorial representations," *Mach. Transl.*, vol. 22, no. 3, pp. 153–173, 2008.

[5] M. Norré and Vandeghinste, "Investigating the medical coverage of a translation system into pictographs for patients with an intellectual disability," in *Proc. Annu. Meeting Assoc. Comput. Linguistics*, pp. 311–318, 2017.

[6] G. Miller, "WordNet: A lexical database," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[7] A. Goldberg, X. Zhu, C. Dyer, N. Eldawy, and L. Heng, "Facilitating pictorial communication via semantically enhanced layout," in *Proc. Conf. Comput. Nat. Lang. Learn.*, 2008, pp. 119–126.

[8] J.Cataix-Nègre, *Communiquer Autrement: Accompagner les Personnes avec des Troubles de la Parole ou du Langage : Les Communications Alternatives*, De Boeck Supérieur, 2017.

[9] S. Bautista, R. Hervás, A. Hernández-Gil, C. Martínez-Díaz, S. Pascua, and P. Gervás, "Aratraductor: Text to pictogram translation using natural language processing techniques," in *Proc. Int. Conf. Hum. Comput. Interact.*, 2017, pp. 1–4.

[10] J. A. Pereira, D. Macêdo, C. Zanchettin, A. L. I. de Oliveira, and R. N. Fidalgo, "Pictobert: Transformers for next pictogram prediction," *Expert Syst. Appl.*, vol. 202, p. 117231, 2022.

[11] J. A. Pereira, C. Zanchettin, and R. N. Fidalgo, "PrAACT: Predictive augmentative and alternative communication with transformers," *Expert Syst. Appl.*, vol. 240, p. 122417, 2024.

[12] J. Mutal, P. Bouillon, M. Norré, J. Gerlach, and L. O. Grijalba, "A neural machine translation approach to translate text to pictographs in a medical speech translation system – The BabelDr use case," in *Proc. Biennial Conf. Assoc. Mach. Transl. Amer.*, 2022, pp. 252–263.

[13] T. Keskinen, T. Heimonen, M. Turunen, J. Rajaniemi, and S. Kauppinen, "SymbolChat: A flexible picture-based communication platform for users with intellectual disabilities," *Interact. Comput.*, vol. 24, no. 5, pp. 374–386, 2012.

[14] V. Vandeghinste, I. Schuurman, L. Sevens, and F. Van Eynde, "Translating text into pictographs," *Nat. Lang. Eng.*, Accepted, 2017.

[15] M. Kultsova, D. Matyushechkin, A. Usov, S. Karpova, and R. Romanenko, "Generation of pictograph sequences from Russian text for assistive mobile applications," in *Proc. Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, 2017, pp. 1–4.

[16] A. Nandy, "Beyond words: Pictograms for Indian languages," *Int. J. Res. Sci. Technol.*, vol. 9, no. 1, pp. 19–25, 2019.