A Blissymbolics Translation System

Usman Sohail
USC Institute for Creative Technologies
12015 Waterfront Dr. Playa Vista,
CA 900094, USA
msohail@usc.edu

David Traum
USC Institute for Creative Technologies
12015 Waterfront Dr. Playa Vista,
CA 900094, USA
traum@ict.usc.edu

Abstract

Blissymbolics (Bliss) is a pictographic writing system that is used by people with communication disorders. Bliss attempts to create a writing system that makes words easier to distinguish by using pictographic symbols that encapsulate meaning rather than sound, as the English alphabet does for example. Users of Bliss rely on human interpreters to use Bliss. We created a translation system from Bliss to natural English with the hopes of decreasing the reliance on human interpreters by the Bliss community. We first discuss the basic rules of Blissymbolics. Then we point out some of the challenges associated with developing computer assisted tools for Blissymbolics. Next we talk about our ongoing work in developing a translation system, including current limitations, and future work. We conclude with a set of examples showing the current capabilities of our translation system.

1 Background

An estimated 7.7% of children aged 3-17 have had a communication disorder, 44.8% of which receive no intervention services (Black et al., 2015). Blissymbolics was created to provide a tool for cognitive, and speech related communication disorders. Blissymbolics(Bliss, 1965), uses pictographic symbols to represent language as opposed to existing alphabetic writing systems in order to provide an alternate that may be easier to learn for people with low literacy.

In 1985, Muter and Johns conducted three experiments to see if ideographic symbols made it easier to extract meaning from words compared to alphabetic symbols. Their experiments showed shorter reaction times for extracting meaning from symbols of Blissymbolics than for words spelled in an unfamiliar language (Muter and Johns, 1985). Therefore Blissymbolics may be easier to learn for people with low literacy. In addition, Blissymbolics can be used without any speech, which may be useful for people with speech related communication disorders.

Although many people use Blissymbolics, they still have to rely on an interpreter to communicate with the general population. In this paper, we discuss a prototype system we developed that translates Blissymbolics utterances to English. We also discuss the future work we think is necessary for this to become feasible for mainstream use.

Blissymbolics is composed of graphic Bliss characters that form the smallest unit of meaning. There are four categories of reasoning for creating a glyph for Bliss characters illustrated in figure 1.
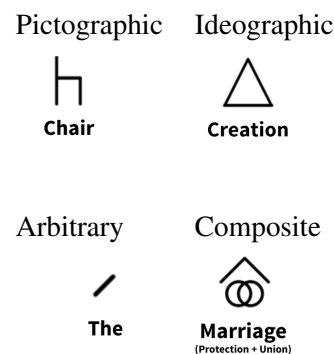

Figure 1

Bliss characters can be combined to form Bliss words with new meanings similar to the way English words can be composed of one or more letters. However, individual symbols in Bliss correspond to a morpheme, or smallest unit of meaning, unlike the phonetic correspondence of written English. In figure 2, the symbol for *house* combined with the symbol for *medical* form the word *hospital, clinic*.

Figure 2

2 Challenges of Blissymbolics

2.1 Encoding

Currently there is no agreed upon encoding for Bliss characters, making it difficult to develop computer assisted tools. The official Blissymbolics dictionary contains a unique 4-5 digit code associated with each Bliss character or Bliss word. This encoding scheme does not differentiate between Bliss characters and words. This is the only encoding we were able to find.

2.2 Computer Assisted Tools

Currently, users of Blissymbolics are restricted by the need for an interpreter. Although the internet has provided many tools for Blissymbolics, there has yet to be a satisfactory translation tool from Blissymbolics to natural language. Most online tools are focused on creating customized Bliss charts. For example, the chart in figure 3 about food was created using *blissonline* [1].

Figure 3

Bliss charts help users communicate with non-Blissymbolics users since the symbols are annotated with their translation. However, users are restricted to the number of symbols that fit on one chart and the expressiveness of Blissymbolics is reduced. Previous work has addressed the large number of symbols by dynamically changing the chart as symbols are input so that only valid options are presented to the user at each step (Netzer and Elhadad, 2006).

Attempts have been made to create a translation system from Blissymbolics to natural language. Several systems have a digital bliss chart that synthesizes speech for a given bliss word that is selected [2]. The digital nature of such devices helps increase the number of symbols that a user can access. Still, users are not able to build words up from the characters that compose them.

At the University of Dundee (Waller and Jack, 2002), a predictive translation system prototype was built using a trigram language model. The system took Blissymbols as input and output English sentences. The gloss of each Blissymbol contains one or more words of the target language. The system consulted the trigram model to find the most probable word from a given gloss. The system also looked for words that probably belonged between any two words, such as articles (which are often implied in Blissymbolics). Although the results of the system were not good enough for mainstream use, the study paved the way for Natural Language Processing techniques to be applied to Blissymbolics, and highlights some shortcomings that need to be addressed for our work.

First, the input to the prototype translation system is full Bliss words, not necessarily the characters that compose them. In the current official dictionary, there are 404 unique Bliss characters, and 4,626 unique Bliss words that are composed of one or more characters. If the system had the ability to build up words from the characters that compose them, then users would only need access to 404 unique symbols, as opposed to all symbols (characters and words).

Second, the translation system does not allow the creation of new words. The official dictionary contains words that are agreed upon by Blissymbolics International, but does not contain all possible words, or even all conjugations of those words. The Blissymbolics Fundamental Rules includes a section on building new vocabulary words, acknowledging that not all words will necessarily be built the same way by all users, and that users may want to express words that are not in the official lexicon. The rules provide an example of a word being built in a different way than the official dictionary.

For example, in Figure 4, the official spelling of teacher is composed of the characters for *person*

---

[1] www.blissonline.org

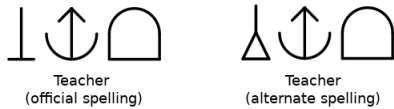[2] www.tobiidynavox.com, www.minspeak.com

Figure 4

(non-gendered) + *giving* + *knowledge*. The fundamental rules concede that the same word should be able to be built with the symbol *female* replacing *person (non-gendered)*. Additionally, there is no official spelling for the word *cried*, although there is a word for *cry* and there is a *past action* indicator character that is made to be used as in figure 5.
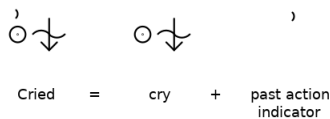

Figure 5

A translation system that could handle alternate spellings and unseen conjugations would help relax the strict spelling requirements of the official dictionary as intended by the Blissymbolics community, and increase the expressibility of the system.

3 Our Translation System

We built a translation system in Python available on github[3]. We made a few assumptions about how Blissymbolics would be used. First, we assumed that any input sequence would have a word separating token, as the Fundamental Rules of Blissymbolics dictate. Second, we used the encoding scheme found in The Official Blissymbolics Dictionary, where each Bliss symbol, word or character, is given a unique 4-5 digit numeric ID. Our translation system only accepts these IDs as input. We will need to create a graphical user interface that allows users to select the Bliss characters to input in order to make this system usable.

The work associated with building our translation system focuses on Morphological Realization, and a Language Model.

3.1 Morphological Realization

We wanted users to have the ability to express words or conjugations of existing words that are

not in the Official Blissymbolics Dictionary. For now, we only applied morphological realization on recognized Bliss words, meaning only officially recognized words can be conjugated in new ways.

We used the SimpleNLG realizer (Gatt and Reiter, 2009) to conjugate Bliss words. For example, if an input Bliss character sequence had a Bliss past tense indicator, we applied the past tense realization to it. So a user could input the spelling for the bliss word translating to *cry, weep* and append the past tense indicator to the end, and get the resulting words *cried, wept*.

Currently the system is limited to morphological realization supported by SimpleNLG. There are over 40 morphological relationships included in Blissymbolics. Each relationship needs its own realizing mechanism, not all of which can be found in SimpleNLG. For example, there is a Bliss character *combine* meant to combine two concepts found in Bliss words. This is not a morphological relationship and cannot be done using SimpleNLG.

3.2 Language Model

We needed a language model to help choose the best word from a given set of translation gloss, and to decide when to add articles. The system first builds all words using the machine readable dictionary, and the morphological realizer outputting a list of sets, where each set contains the possible English words that the given Bliss word may translate to. The system looks at each set to determine if it contains nouns using wordnet (Miller, 1995). If a noun is found, then a set of articles *a, the,* or a blank is inserted before the set of nouns. From here, the language model needs to decide the most probable gloss words from each set, and also which article, if any, is most probable.

We created an N-gram model trained on the Gutenberg, brown, conll2000, and nps-chat corpora using NLTK (Bird and Loper, 2004). We used interpolation smoothing as in equation 1.

$$P(w_1, w_2, w_3) = c_1 P(w_3) + c_2 P(w_3|w_2) + ...$$
$$... + C_n P(w_3|w_1, w_2)$$
$$s.t. \sum c_i = 1 c_1 < c_2 < c_n$$

(1)

---

[3]www.github.com/usmansohail/Nighat

4 Test Set

We created a test set composed of 15 bliss utterances from children's books (Bruna, 1978; Andy and Mann, 1979; Chait, 1992; Cocking, 1979), 6 of which are shown in figure 6 for discussion.

5 Results

The translation system received a BLEU score of 34.53 when evaluated on the 15 utterance test set. Some sentences preserve the general meaning, whereas others do not. Some of the errors are related to the language model, while others are related to Blissymbolics. In Figure 6, examples 1, and 2 have errors that are related to the language model. Sentence 1 is missing an *a*. Sentence 3 is an example of a sentence that preserves the meaning, although it incorrectly translates *fat* to *thick*. The system translates *Julius* from sentence 2 and 1 to *a boy*. This translation relies on context to be interpreted correctly. Sentence 4 translates *other* to *you*. This error is caused by the fact that *you* and *other* are spelled the same way with a minor difference. The word *you* is spelled with *person + 2*, while *other* is spelled with *person [modified] + 2*. The current encoding scheme assigns each symbol with a unique ID, however modified symbols do not have a unique ID. Therefore, *person* and *person [modified]* both have the same ID. Figure 7 shows the difference between the two words.
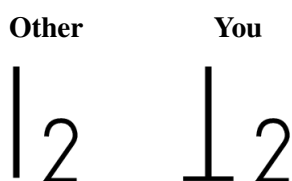

Figure 7: Other vs You

6 Future Work

In order to make a usable system, we think it is necessary to address the following topics:

1. **Encoding scheme**
   As the examples from figure 6 show, the current encoding scheme is not able to capture all of the capabilities of Bliss. In order to make a usable system, an encoding scheme needs to be chosen that can work with computer systems, and also preserves the capabilities of Bliss, such as modification of symbols.

2. **Language model**
   The language model that we used was implemented to show a proof of concept. In order to make the system more applicable, we think that the training corpora used should be composed primarily of dialogue utterances, since this is the way that the translation system is intended to be used.

3. **User Interface**
   If users are ever to use the system, there needs to exist a way for them to easily input. In order to build this, the encoding scheme needs to be chosen first. A critical component of a UI is a text to speech component so that users can be independent of a human interpreter.

4. **Context**
   A system that is able to exploit the context of a dialogue would decrease the reliance on a human interpreter. The way Bliss is used typically involves a human interpreter who can infer context, such as the name *Julius* from figure 6.

7 Conclusion

Our translation system adds some new features to Blissymbolics translation systems, namely the ability to create new words based on existing words. We also address some topics that need to be addressed for mainstream use. We believe our morphological perspective is useful for Blissymbolics, however more work is necessary to assess it's impact on translation. We hope to work with the Blissymbolics community for future work.
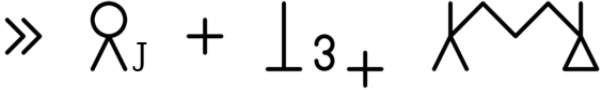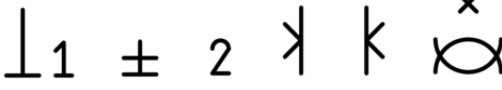
8 Acknowledgments

Reference Translation/Bliss	Result Translation
1 For Julius and his family	for **boy** and his family
2 Julius has many friends	**a boy the gain much persons**
3 I have two fat fish	I have two **thick** fish
4 the other fat fish is called **Bottom**	the **you the** fat fish is called bottom

Figure 6: Each row contains an utterance written in Bliss annotated with it's reference translation. Adjacent to that is the corresponding result translation using our system. Any written English inside of the Bliss utterance is taken as is.

References

Project gutenberg (n.d.). www.project.gutenberg.org. Accessed: 2018-7-29.

Andy and Gwen Mann. 1979. *Autumn*. University of Toronto Press.

Steven Bird and Edward Loper. 2004. Nltk: the natural language toolkit. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, page 31. Association for Computational Linguistics.

Lindsey I Black, Anjel Vahratian, and Howard J Hoffman. 2015. Communication disorders and use of intervention services among children aged 3-17 years: United states, 2012. nchs data brief. number 205. *Centers for Disease Control and Prevention*.

Charles Kasiel Bliss. 1965. Semantography (blissymbolics): a simple system of 100 logical pictorial symbols, which can be operated and read like 1+ 2.

Dick Bruna. 1978. *I Can Dress Myself*. Methuen Publications.

Thelma Chait. 1992. *Julius and his friend the computer*. Oxford University Press Southern Africa.

Althea Cocking. 1979. *The Blissymbol Opposite Series: Book 1*.

Albert Gatt and Ehud Reiter. 2009. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93. Association for Computational Linguistics.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Paul Muter and Elizabeth E Johns. 1985. Learning logographies and alphabetic codes. *Human Learning*, 4:105–125.

Yael Netzer and Michael Elhadad. 2006. Using semantic authoring for blissymbols communication boards. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, NAACL-Short '06, pages 105–108, Stroudsburg, PA, USA. Association for Computational Linguistics.

Annalu Waller and Kris Jack. 2002. A predictive blissymbolic to english translation system. In *Proceedings of the fifth international ACM conference on Assistive technologies*, pages 186–191. ACM.