

TransPic: Using Images as Interlingua for Machine Translations Systems

Mark Vincent B. Melgo ¹⁾, Robert R. Roxas ^{2,*}

^{1,2)}Dept. of Computer Science, College of Science, University of the Philippines
Cebu, Philippines

Abstract. This paper proposes a new approach to machine translation systems that use images as interlingua. The system is just like the usual Text-to-Text systems but with a pictorial output that conveys the image representation of the translated text. The system differs from traditional approaches of machine translation because it generates images together with the translated texts, which would allow the users to better understand the translation, thus avoiding misunderstandings and miscommunication. The system initially contains 945 words together with their corresponding images and are categorized into 29 different categories. Two sets of survey were done on the understandability of images that have been generated, particularly the images for pronouns. The results showed that such images are highly understandable.

Keywords; machine translation; interlingua; TransPic system

1. Introduction

Traveling to foreign places are becoming quite popular today. Although interacting with people with different languages can be hard and complex at times, it can be made possible by the use of different kinds of machine translation systems, e.g. Text-to-Text, Speech-to-Text, and Speech-to-Speech. Currently these different kinds of machine translation systems can translate simple sentences quite well. But a down side with these different kinds of machine translation systems is that they need a lot of work to perfectly translate a complex sentence from one language to another. As what Copper said in [1],

* Corresponding author: robert.roxas@up.edu.ph

Received: 2017.10.20; Accepted: 2018.03.15; Published: 2018.9.30

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

“machine translation right now is pretty much garbage. Everyone knows this. For simple sentences in grammatically similar languages, it works OK. But give it a complex sentence to move between Chinese and English (for example), and the results are typically nonsense.”

There are roughly around 6,909 distinct language in the world [2]. Learning and understanding just one language can considerably require a huge amount of effort and time. To be able to avoid spending such large amount of time and effort, the use of machine translation systems is needed. Machine translation enables the translation of one language to another in a fast manner.

Machine translation can help an individual without prior knowledge of a certain language to communicate comfortably to another individual who is a native speaker of that certain language. The use of machine translator is becoming more increasingly popular as one does not have the time and luxury to learn all the languages. But the current status of today’s machine translation systems is that the translation only works perfectly on simple sentences, and when it comes to translating complex sentences, the translation somehow generates a different meaning leading to miscommunication and misunderstanding [1].

Despite the continuous research and development on how to improve machine translation systems, still the systems are not capable of perfectly translating the source language into the target language. Out of nearly 7,000 languages spoken worldwide, only 15-20 languages are currently taking advantage of the benefits provided by machine translation [3]. But the machine translation systems for those languages still generates imperfect translations because translating a language accurately requires a level of cognitive function that computers simply do not have and are not likely to acquire in the next ten years [1]. So the problem is how we can mitigate the miscommunication and misunderstanding caused by imperfect translation.

The imperfect translation generated by machine translation systems contradicts its goal, which is to provide good communication between two or more parties that have different languages. Researching and developing complex algorithms to be able to support complex machine translation systems would require a huge amount of research and time. So while research and development to create those complex algorithms are still ongoing, one can make use of images to clarify those imperfect translation. The use of images as an Interlingua can be a solution so that an imperfect translation could still be understandable because images can easily be recognized compared to words alone [4].

2. Review of Related Literature

Universal Communication represents one of the long-standing goal of humanity [3] in order to be able to communicate effortlessly without the hindrance of language barrier. This is because through communication, we can learn and share knowledge and information. But language barrier limits our communication skills, thus also limiting our learning and sharing of knowledge and information. That is why finding a way to break this language barrier is a problem that this research will try to solve by using simple images or pictures to denote the meaning of a word together with the translated word.

Images have been widely used throughout history to record experiences and events. Images have been long used for communication before humans made use of words or text to communicate. As Dewan said that pictures have been used for about 250 centuries, pictographs and ideograms for the next 20 centuries, and the use of words for the remaining 15 centuries [4]. From such statement, we can say that images and pictures can really be used for communicating. Dewan also said that humans are neurologically wired with an astounding visual sensory ability, making the use of images and pictures to communicate not hard to do [4]. To prove this, Dewan mentioned a researcher named Pavio that said, "Pictures enter the long-term memory with two codes: one visual and the other is verbal, while words or text just enter in a single code." The dual-coding of images in a human brain allows humans to access the information in two independent ways: visual or verbal, because of those images are easier to remember and can be recognized effortlessly more than words. Pictures facilitate learning by providing clarifying examples, extra-lingual information, contexts for interpretation, and redundancy, which aid recall [4].

Images has been used in machine translation systems. One kind of machine translation that makes use of images is the text-to-image machine translation system. Examples of text-to-image machine translation systems include Easy as ABC? [5], Text-to-Picture (TTP) [6], and Text-to-Picture Synthesis [7]. All of them make use of an input text to generate an image that can denote the meaning of the text. The aforementioned works all aim to convert text into meaningful pictures for communicating mainly to aid and assist people with limited literacy. But a picture paints a thousand words [5]. Thus, translating a text to an image can still lead to confusions and misunderstandings since the meaning of the word will depend on the viewer's perception. All of the aforementioned works made use of automatic conversion from text to image. Word sense disambiguation is a problem but has been partially solved by Sevens, et. al. in their work "Improving Text-to-Pictograph Translation through Word Sense Disambiguation," but only led to a 14 out of 20 ambiguous words that was correctly translated, that is, about 70% accuracy rate only [8].

Another kind of machine translation systems that makes use of images is the image-to-text machine translation systems. Examples of this approach are picoTrans [9] and Pictograph-to-Text [10]. PicoTrans is mainly for travelers to use, while Pictograph-to-Text is to be used by people with Intellectual or Developmental Disabilities. Both systems automatically generate sentences from the given image, but misleading and confusing sentences were also generated since only 74% of the generated sentences were semantically correct. PicoTrans made use of a friendly user interface to facilitate image search by placing images in a category. While Pictograph-to-Text made use of a search bar that automatically suggest images based on the user's input.

Picture paints a thousand words [5], and understanding of pictographs could differ because of the differences in native language and culture. When this happens misinterpretations and misunderstanding occur [11]. To solve this problem on varying user perception, both image and the translated text will be generated. Moreover, to mitigate the different viewer's perception of the image, a simpler image will be used. The use of simple images to communicate more precisely than using complex images has been proved by a chat application that makes use of only simple images in chatting. It has also implemented the so called "Pictograph chat communicator III" [11], which garnered a 91.1% level of accuracy.

To be able to generate the images to be used for the machine translation systems, a simple Text-to-Image-and-Text system will be implemented. The simple Text-to-Image-and-Text system will be called TransPic. This system enables one to easily add, remove, and update the database system that will be used for the machine translation systems. To control the unwanted or incorrect data to be added to the database, only trusted and selected people are allowed to use the administrator role of TransPic.

3. The TransPic System

This section presents the machine translation systems called TransPic. It is given that name because it is a system of Translating through Pictures. Once TransPic has been built, it will serve as a mode of the user to interact with the database. TransPic will facilitate the creation, deletion, search, and update of the data, both data from the *word* table and *category* table. TransPic at the same time will give the user of the system ease and comfort by having a user-friendly interface.

TransPic aims to solve the miscommunication and misunderstanding caused by imperfect translation by the use of simple images to go with the translated text. Images will serve as an Interlingua that will help both parties to understand the output of the machine translation systems. The images will also clarify the translated text since

humans, regardless of the language they speak, share almost the same ability to understand the content of an image [3]. TransPic will act as a machine translator that generates a simple image together with the translated text. Generating both image and the translation of the word will mitigate the difficulty in misunderstanding and the confusion caused by imperfect translation since text with illustration increases understanding by 98% as what Dewan's study found out [4].

TransPic will provide a more understandable machine translation system through the use of images together with the translated texts, which will be very useful for individuals who will be communicating to other individuals with a different language, especially travelers and tourist. Individuals will be assured of the correctness of the information they are communicating because the system also generates images, and they can check whether or not the generated images symbolize what they actually meant.

A. The Database Design

The program for now requires just a simple database schema to hold the images, and the words in different languages. The different languages present in the database have been retrieved from the list found in *aboutworldlanguages.com* website [12] and have been sorted alphabetically. But for now only English, Filipino, and Cebuano languages have entries. Other languages will be implemented in the future. The *word* table is composed of the languages, the suitable image, and the category it belongs to.

One major concern is that storing images in the database is very heavy to the database, thus the images are first encoded to their Base64 equivalent before they will be added to the database. And in terms of retrieving an image from the database, the Base64 string will be decoded to its image equivalent, then the decoded image will then be presented. Figure 1 shows how the image is added into the database, and how it is retrieved from the database.

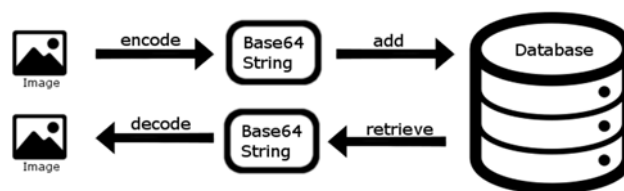


Fig 1. Adding and retrieving of image to/from the database.

The image attribute of the *word* table in the database will be a list or an array of Base64 strings. An array was created and not just a plain Base64 string, so that in the future, it can accommodate many images in a single word. In this way, the user of the system can understand better through the use of many image examples. But initially, it can only facilitate one image per word.

Categorization will also be implemented in order to facilitate easy word search. Categories includes persons, things, places, etc. The *category* table will have the same schema as that of the *word* table, which will have the image attribute and the attribute for each language.

B. The User Interface and System Design

User interface is one of the key aspects in the usability of a system. The design of a system's user interface is critical to software success. No matter what expectations the decision-makers have for technology, goals will not be reached if users are uncomfortable or hampered by confusing tools [13]. Because of this, TransPic is designed to implement a user-friendly interface for the ease and comfort of the user using the system.

A storyboard of the user/contributor interactions is given below:

- 1) *The user opens his/her account.*
- 2) *The user then is taken to the home page where the user can choose from a radio button what he/she wants to update, delete, search, or add and either a category or a word.*
- 3) *The user is then given some options like to search, add, or view all the categories or words, depending on what he/she has previously chosen.*
 - a) *If the user would choose Search or View all the categories or words, then the results would be generated and presented, then he/she could view, update, or delete a certain word or category.*
 - i) *If the user would choose View, a modal would automatically open. The modal contains the information about the category or word being viewed.*
 - ii) *If the user would choose Update, then he/she would be taken to another page in which he/she could update the category or word.*
 - iii) *If the user would choose Delete, then the category or word would be deleted.*
 - b) *If the user would choose Add a category or word, he/she then would be taken to another page in which he/she could add new category or word.*

TransPic's search is achieved through the use of Regular Expressions, thus if the user would search for "a," then all words containing the searched parameter (in this case the letter "a") would be retrieved from the database. But retrieving such large number of results in the database will lead to the decrease of the program's speed thus limiting the user's interactions with the program. So to solve this issue, TransPic limits the results only to 8 outcomes and to be able to retrieve more results, a "load more" button has been implemented.

Even though the results have been limited to 8 outcomes, the user can still be confused so the results have also to be sorted alphabetically. Sorting the results alphabetically can also make easy and convenient searching.

C. Word and Image Generation

Artificial intelligence or automation has not been used for the creation of the image and the translation. The words that are included in the database have been chosen from the *wordfrequency.info* website [14], where the website provides the list of English words, which are sorted by the frequency of its usage. But the translation of the words is not included in the website. Thus the translation of the words has been encoded manually. Another reason why the translations were encoded manually because as of now the system only supports *word for word* translations.





A huge number of words will be added to the database, thus the use of groups or categorization will also be implemented to facilitate easy and effortless search because without appropriate grouping, it is difficult to search for pictographs [11], and browsing a large database to find the appropriate image is a long and tedious job [8]. So to solve this problem, the program implemented a search feature to find the desired image or word.

The images used in the system have either been manually made or have been taken from the *sclera.be* image website database [15]. By the help of TransPic, the translation of a word and its corresponding image is added to the database.

Images are not just simply created without any constraints. The images are created in a manner that they will be simple and can easily be understood by the majority. Images can be used in more than one word as long as the image suits the word. But both images and words as of now are deprived of tenses. Also articles, possessive pronouns, inflection, and abstract words such as politics, regeneration, and the like are not yet included because these words are hard to depict in images but will be taken later.

Singularity and plurality of nouns are not yet implemented. But pronouns that can both denote singular and plural in some language but distinct in other languages are included and two different images were made to differentiate the singular from the plural equivalent. Example of such word is “you,” which can denote both singular and plural. There are also words that can denote different membership. Examples of these words include the word “we,” which has exclusive or inclusive meaning. Table I shows the images for plural “you” and singular “you” and the images for inclusive “we” and the exclusive “we.”

TABLE I. SEMANTICS OF IMAGES

			
Singular: You	Plural: You	Exclusive: We	Inclusive: We

4. Results and Discussion

One major concern of this system is to know whether or not people can really understand the images that would be shown by the machine translation system. Therefore, two sets of survey were conducted to determine whether or not people can successfully link images to its corresponding word and vice versa. Only pronouns have been included in the survey since pronouns are the ones that are harder to understand and harder to depict into images as compared to nouns, which can be easily understood and depicted into images.

The results of the survey would serve two purposes. The first purpose is for the system user. If the images are easy to understand, then the people using the system can easily decipher the meaning of the translation in another language. The second purpose is for people who are interested to add their own language into the system. They can easily give the corresponding text in their language without exerting stressful mental activity.

The first survey was about linking an image to its corresponding word. This survey gives the respondents an image and they were asked to select the corresponding word from the given choices of words. This survey garnered 88 respondents. Based on the results, questions regarding pronouns that represent countable and non-countable objects garnered less correct answers compared to other questions. Respondents had a hard time figuring out the difference on these pronouns, like figuring out the difference between the words “several” and “many” as an example. Because of this, we can conclude that the correct association of the images to their corresponding word depends on the knowledge and vocabulary of a person. Table II shows the overall percentage of the respondents' correct answers, which is calculated by adding all their total correct answers, which is 3,515 correct answers, divided by the number of respondents, which is 88. The result shows that out of 49 questions, the respondents got the average of 39.94 questions, which is about 82% of the questions.

TABLE II. SURVEY RESULTS

Average Score	Incorrect	Correct
Set A: Linking Images to Words	18%	82%
Set B: Linking Words to Images	15%	85%

The last survey was about linking a word to its corresponding image. This survey gives the respondents a word and they were asked to select the suitable image from the given choices of images. While the first survey garnered 88 respondents, this survey only garnered 73 respondents. The result was just like the previous survey, questions regarding pronouns that represent countable and non-countable objects garnered less correct percentage compared to other questions. This result also supports the observation mentioned above that the correct association of the images to their corresponding word depends on the knowledge and vocabulary of a person. Table II shows the overall percentage of the respondents' correct answers, which is calculated by adding all their total correct answers, which is 3,050 correct answers, divided by the number of respondents, which is 73. It is shown that out of 49 questions, the respondents got the average of 41.78 questions correct, which is about 85% of the questions.

4. Conclusion and Future Works

This paper presented an idea of using images as an Interlingua in machine translation system, which would be displayed regardless of what target language is used. To verify the understandability of the images for pronouns, two sets of survey were conducted, that is, one for image-to-text and another for text-to-image association. The result of survey showed that 82% of images presented to the respondents was correctly linked to their corresponding pronouns, and 85% of the pronouns presented to the respondents was correctly linked to their corresponding images. These surveys show that using these images would make TransPic a very promising machine translation system in the future.

For our future works, we would like to incorporate the grammar and semantic features of the languages that will be included in the system. We would also prepare the system to allow other nationalities to encode the corresponding text of the stored images in their language so that the images would truly become the Interlingua that can be used in machine translation systems.

References

- [1] C. Copper, "Tech in Asia - connecting Asia's startup ecosystem." Retrieved 12 December 2016, from <https://www.techinasia.com/dsiconnect>
- [2] S. Anderson, "How many languages are there in the world?," in Linguistic Society of America. Retrieved 12 May 2017, from <https://www.linguisticsociety.org/content/how-many-languages-are-there-world>
- [3] R. Mihalcea, and C. W. Leong, "Toward communicating simple sentences using pictorial representations." in *Machine Translation*, 22(3), 2008, pp. 153-173.
- [4] P. Dewan, "Words Versus Pictures: Leveraging the Research on Visual Communication," in *Partnership: The Canadian Journal Of Library And Information Practice And Research*, 10(1), 2015. <http://dx.doi.org/10.21083/partnership.v10i1.3137>
- [5] A.B. Goldberg, X. Zhu, C.R. Dyer, M. Eldawy, and L. Heng, "Easy as ABC?: facilitating pictorial communication via semantically enhanced layout," in *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pp. 119-126, 2008.
- [6] X. Zhu, A.B. Goldberg, M. Eldawy, C.R. Dyer, and B. Srtock, "A text-to-picture synthesis system for augmenting communication," in *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, pp. 1590-1595, 2007.
- [7] A. B. Goldberg, J. Rosin, X. Zhu, and C. R. Dyer, "Toward text-to-picture synthesis," in *Proc. NIPS Mini-Symp. Assistive Mach. Learn. People Disabilities*, 2009, pp. 1-3.
- [8] L. Sevens, G. Jacobs, V. Vandeghinste, I. Schuurman, and F. Van Eynde, "Improving text-to-pictograph translation through word sense disambiguation," in *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 131-135, 2016.
- [9] W. Song, A. Finch, K. Tanaka-Ishii, K. Yasuda, and E. Sumita, "picoTrans: an intelligent icon-driven interface for cross-lingual communication," *ACM Transactions On Interactive Intelligent Systems*, 3(1), 1-31, April 2013.
- [10] L. Sevens, V. Vandeghinste, I. Schuurman, and F. Van Eynde, "Natural language generation from pictographs," in *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, pp. 71-75, 2015.
- [11] J. Munemori, T. Fukuda, M.B. Mohd Yatid, T. Nishide, and J. Itou, "Pictograph chat communicator III: a chat system that embodies cross-cultural communication," in Setchi R., Jordanov I., Howlett R.J., Jain L.C. (eds) *Knowledge-Based and Intelligent Information and Engineering Systems. KES 2010. Lecture Notes in Computer Science*, vol 6278. Springer, Berlin, Heidelberg, 2010, pp. 473-482.
- [12] Languages A-Z | About World Languages. [Aboutworldlanguages.com](http://aboutworldlanguages.com/languages-a-z). Retrieved 8 March 2017, from <http://aboutworldlanguages.com/languages-a-z>.
- [13] S. Gillert, "The importance of user interface design," Retrieved 6 December 2016, from <http://www.jobscience.com/productstechnology/the-importance-of-user-interface-design/>
- [14] Word frequency: based on 450 million word COCA corpus. Retrieved 27 November 2017, from <http://www.wordfrequency.info/>
- [15] Sclera vzw. Retrieved 27 November 2016, from <http://sclera.be/nl/vzw/home>