Image as an Interlingua for Language Translation

By

Mark Vincent B. Melgo

Bachelor of Science in Computer Science

Department of Computer Science

Dr. Robert Roxas

December 14, 2016

CHAPTER 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, e.g. text-text, speech-text and speech-speech. Currently without question that these different kinds of machine translation can translate simple sentences quite good however a down side with these different kind of machine translations is that they need a lot of work to perfectly translate a complex sentence from one language to another.

Background

There are roughly around 6,909 distinct language in the world ("How Many Languages Are There In The World? | Linguistic Society Of America") and learning and understanding just one language can considerably require a huge amount of effort and time. Soo to able to avoid spending such large amount time and effort the use of machine translation is needed. Machine translation enables translation of one language to another in an instant.

Machine translation permits an individual without prior knowledge to a language to communicate comfortably to another individual who is a native speaker to that language. The use of machine translator is becoming more increasingly popular as one doesn't have the time and luxury to learn all the languages. But one great milestone on today’s machine translation 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.

Statement of the Problem

Despite the continuous research and development on how to improve machine translator there are still languages that aren't fully supported to be perfectly translated. Out of 7,000 languages spoken worldwide only 15-20 can currently take advantage of the benefits provided by machine translation (Mihalcea, Leong). Even those languages that are supported to use machine translation still generates imperfect translations. So the problem is how to mitigate the miscommunication and misunderstanding caused by imperfect translation.

Objective

This program will aim 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. 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 (Mihalcea, Leong). The researcher's objective is to make a software or program that acts as a machine translator that generates a simple image together with the translated text. This way misunderstanding and confusion from imperfect translation can be mitigated since text illustration increases understanding by 98% as what Dewan said.

Significance

The imperfect translation generated by machine translation contradicts its goal which is to provide good communication between two or more parties that has different languages. But researching and developing complex algorithms to be able to support complex machine translation requires 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 is one solution so that each imperfect translation could still be understandable since images are effortlessly to be recognized compared to words alone (Dewan).

This software will provide a more understandable translation through the use of images together with the translated words which will be much of use 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 since individuals can search for the word to generate an image and can check if the generated image symbolizes what they actually mean.

Scope and Limitation

The aim of this program is to generate text-image-text translation. The language that will be included in this program for now are English, Tagalog and Cebuano. The program for now will only comprise common English words. The input for Tagalog and Cebuano are restricted in those common English words. But to mitigate these restrictions, selected and verified users can add words and images that are not yet in the database, this includes adding a corresponding word in a different language and adding another image to symbolize a word.

This program will categorize those common English words to ease the users search for the word's corresponding image and translation. Categorization of these words includes people, actions, places, etc.

This program will still be limited to a direct word by word translation and doesn't consider the intonation and pronunciation of a word, thus translating a word that has different meaning but has the same spelling will generate all the different translations of that word. Example a Filipino word “puno” can generate 2 translations, one translation is a picture with a tree and the other translation is a picture that symbolizes full. So to address this issue the program accepts a single word and generates the possible images that corresponds to that word and let the user choose what image they actually mean. Since the program translates in a word by word basis, grammar and context checking are not yet implemented. The program can't determine what kind of word the input is (noun, pronoun, verb), thus there is a possibility that proper nouns will also be translated, an example of this is the proper noun “family park”.

CHAPTER 2
REVIEW OF RELATED LITERATURE

"Universal Communication represents one of the long standing goal of humanity" (Mihalcea, Leong) to be able to communicate effortlessly without the hindrance of language barrier and since 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's why finding a way to break this language barrier is 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.

History

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, pictograms and ideograms for the next 20 centuries and the use of words for the remaining 15 centuries. From that we can say that images and pictures really can be used for communicating. Also Devan 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. To prove this Devan 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, and because of this images is 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 aids recall (Dewan).

Text-to-Image

A text to image converter has been implemented by many researchers. There exist "Easy as ABC?" (Goldberg, et al.), "Text-to-Picture (TTP)" (Zhu, Goldberg, et al.), "Text-to-Picture Synthesis" (Goldberg et al.) that all 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 (Goldberg, et al.) thus translating a text to an image can

still lead to confusions and misunderstandings since the meaning of the word will depend on the viewers perception. All of the aforementioned works makes use of automatic conversion from text to image, thus word sense disambiguation is a problem but has been solved by Leen Sevens, Gilles Jacobs, Vincent Vandeghinste, Ineke Schuurman, Frank Van Eynde 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, about 70%.

Image-to-Text

Two research have implemented image-to-text converter, picoTrans (Finch, et al.) and Pictograph-to-Text (Sevens, et al.) PicoTrans is mainly for travelers to use while Pictograph-to-Text is to be use by people with Intellectual or Developmental Disabilities. Both automatically generates 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 base on the user's input.

Conclusion

Converting text to image only solves half the problem since it will just be like a picture book while image to text doesn't solve any in the goal towards mitigating misunderstanding and confusions made by text to text machine translation since an image can mean many and differently things. So to clarify what the image really means the combination of the two is required, the result should have both the image and the translated text.

Picture paints a thousand words (Goldberg, et al.) and "understanding of pictographs could differ caused by differences of native language and culture, when this happens misinterpretations and misunderstanding occurs"(Munemori, et al.) so to solve this problem on varying user perception both image and the translated text will be generated. And also to mitigate the varying on the 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 has also been implemented, the so called "Pictograph chat communicator III" (Munemori, et al.) which garnered a 91.1% level of understanding.

The use of grouping or categorizing an image will be implemented to facilitate easy and effortless search since as Munemori said "without appropriate grouping, it is difficult to search

for pictographs". But "browsing a large database to find the appropriate image is a long and tedious job" (Sevens, et al.) so to solve this problem the program will implement a search feature to find the desired image.

The aforementioned works shared similar features which will also be implemented in this research. The image for a noun can be used in one or multiple aspect of that noun. The image for a verb are deprived of tenses. Also articles, possessive nouns, inflection, abstract texts (politics, regeneration, etc.), singularity and plurality will not be implemented. And to depict approval a green colored checkmark image will be shown, while to depict disapproval a red colored cross image will be shown.

CHAPTER 3
METHODOLOGY

Requirements and Tools

Before developing the program the researcher must have these as a requirements and tools:

(1) To be able to make this software a list of the most common words in English must be present which will be provided by the 5000 common words list by the site wordfrequency. ("Word Frequency: Based On 450 Million Word COCA Corpus")

(2) The images for the words will either be from the website sclera.be, for black and white pictures, and/or from image-net.org, for colored pictures, since those sites where the sources of images of the previous researches, but the researcher can or will make the image himself if there are no fit image to be used.

(3) In terms of the language to be used in developing the program the researcher will user JavaScript as the core language to be used since it can provide the program to be deployed through the internet, in hopes that the program will grow its number of supported words. To be specific it will be developed using meteor platform to handle server and client activities together with ReactJs as the front-end for the program to be able to support real-time rendering. MeteorJs is a good choice in order to support future works if the researcher will develop a mobile version of the program since MeteorJs is multiplatform, which means that the same code can be deployed in mobile, both android and iOS.

(4) The data will be stored in the database and the database to be used is Mongodb. But the images will either be converted into base64 using image to base64 converter and then store the base64 string to the database. Another way is to store the image locally or through a cloud service and just store the relative path or link of the image to the database.

Database Design

The program for now requires just a simple database schema to hold the image/s, the words in each different languages. The different language will be sorted alphabetically but for now only English, Filipino, and Cebuano languages will have a word in it, other languages will be implemented in the future. The image attribute in the database will be a list or an array to accommodate many images in a single word, this way the user can understand better through the use of many image examples. But a word may contain 1 or more images. Categorization will also be implemented and a different table for categories will be done. It has also the same

attributes as the translation table except the “category” attribute. See table 1 for the list of the languages that will serve as an attribute of the database. See table 2 in appendix for the design of the database schema for translation table. See table 3 for the category schema. See table 4 for the contributor schema.

User Interface Design

User interface is the one of the key aspects in the usability of a programs. “The design of a system's user interface is critical to software success. No matter what expectations decision-makers hold for technology, goals won't be reached if users are uncomfortable or hampered by confusing tools.” (Gillert). The researcher aims to implement a user-friendly interface for the ease and comfortability of the user using the system.

Client’s User Interface

The main look of the user interface will just be as simple as possible. The home page will be the page where translation occurs which is comprise of the logo of the program, a top navigation bar which contains the “translate” and “log-in” button.

The translation page will have the top navigation bar. It has two types, the translation page with the search for the word, and the translation page with the search for the categories. It has a search bar that has a radio button which let the user choose what to search, either “category” or “word”.

The translation page with the word search will still have the top navigation bar. The upper portion of the translate page will be for the generated image and translated text. While directly below it is a search bar that the user can use to search for images. The search bar will have two dropdown buttons to let the user choose what languages the conversion will take place, both to and from. A result box that contains the result, both the image/s and the translated text, of the searched query will be shown below the search bar. And if multiple results happen they will also be shown. Below each item, both image/s and text, is a button that adds the item to the translation result. While in the translation result box each item has an “x” button in its upper right to delete or remove that item from the translation. To clarify the translate page with the word search interface see Figure 1 in appendix.

The translation page with the category search will just like be the translation page with the word search that has the upper portion of the translation page will be for the generated image/s and translated text and a search bar for categories that has a dropdown button to what the language of the category name will be. And instead of the result as an individual item with the specific word, the category of that image/s with the specific word is shown depending of the language selected. Instead of an “add” button it has a “view” button which user can click

which will lead to showing the specific item that are present in that category. Each individual item still contains the button that adds that item to the translation result. And if the user has nothing in the search bar then a paginated alphabetical list of categories will be shown. See Figure 2 in appendix for details.

Contributor's User Interface

The log in page will still have the top navigation bar and below is a form that has the field's username and password in it. Take note that there is no sign up page since only the researcher and trusted persons can be able to sign up to the program to make the data reliable and consistent. But for those user who are interested to contribute to the program a "contact us" button will be provided directly below the log-in form. This way the researcher can validate the user and can give him/her a log in credentials. The log in page is shown in Figure 3 in appendix.

Once the user has successfully logged in the contributor's page will be shown. A top navigation bar will have the button "log-out". Just like the translation page the contributor's page will have two types, for category and word search. The contributor's page will consist of a search bar and a result page.

In the contributors page for word search it will be just like translation page with word search which contributors can search for a word in a given language and then sees the results, both image/s and text. But this time instead of an "add" button, a "view" button is present and once clicked it will open a modal containing the details of that item. See Figure 5 in appendix to see the view modal. Also in the upper right of the each item exist a dropdown menu that has "delete" and "edit" as an option. Once the contributor clicks the delete button a prompt telling the contributor that this will be deleted will be displayed and once the contributor clicks "okay" that word and image/s will be deleted from the database, which is why only trusted people are given log in credentials. The contributor can also click the "edit" button which will render another page, the edit page. It has also a button below the result panel to add a new word to the database which when click opens the add word page. See Figure 4 in appendix for contributor page for word search.

In the contributors page for category search it will be just like translation page with word search which contributors can search for a category in a given language. The result item has also the "view" button which shows the modal (Figure 8) and the dropdown menu that has "delete" and "edit" as an option which has the same functionality mentioned above. And if the user has nothing in the search bar then a paginated list of categories will be shown. It has also a button below the result panel to add a new category to the database which when click opens the category word page. See contributor page for category search in Figure 7 in appendix.

The edit page will also have two types, one for the word and one for category. Both will have the top navigation bar. It will also consist of input fields for each string attribute and a “cancel” and “update” button which when clicked will either cancel the editing or saves the edit to the database.

The edit page for the word will show the image/s and each image has an “x” button that will delete the image, but it has also a button that adds a new image to the word. All languages will be shown in the edit page with their corresponding value which will be placed in an input field. Also it has a checkboxes to select the category for that word and those categories will be automatically checked if that word belongs to it. See figure 6 in appendix.

The edit page for the category will be just like the edit page for the word but without the checkboxes to select the category. See figure 9 in appendix.

The add word and add category page will be just like the edit page but instead of an update button, an add button will be placed. See figure 10 and 11 in appendix.

APPENDIX

Afrikaans	**Cebuano**	Hausa	Kurdish	Pashto	**Tagalog**	Yucatec
Albanian	Chechen	Hawaiian	Kyrgyz	Persian	Tajiki	Zapotec
Amharic	Cherokee	Hawaiian Creole	Lao	Polish	Tamil	Zulu
Arabic	Croatian	Hebrew	Latin	Portuguese	Tatar	
Aramaniac	Czech	Hiligaynon	Latvian	Punjabi	Telugu	
Armenian	Dakota	Hindi	Lingala	Quechua	Thai	
Assamese	Danish	Hungarian	Lithuanian	Romani	Tibetic	
Aymara	Dari	Icelandic	Macedonian	Romanian	Tigrigna	
Azerbaijani	Dholuo	Igbo	Maithili	Russian	Tok Pisin	
Balochi	Dutch	Ilocano	Malagasy	Rwanda	Turkish	
Bamanankan	**English**	Indonesian	Malay	Samoan	Turkmen	
Bashkort	Esperanto	Inuit	Malayalam	Sanskrit	Ukrainian	
Basque	Estonian	Irish Gaelic	Mandarin	Serbian	Urdu	
Belarusan	Éwé	Italian	Marathi	Shona	Uyghur	
Bengali	Finnish	Japanese	Mende	Sindhi	Uzbek	
Bhojpuri	French	K'iche'	Mongolian	Sinhala	Vietnamese	
Bislama	Georgian	Kabyle	Nahuatl	Slovak	Warlpiri	
Bosnian	German	Kannada	Navajo	Slovene	Welsh	
Brahui	Gikuyu	Kashmiri	Nepali	Somali	Wolof	
Bulgarian	Greek	Kazakh	Norwegian	Spanish	Xhosa	
Burmese	Guarani	Kmer	Ojibwa	Swahili	Yakut	
Cantonese	Gujarati	Khoekhoe	Oriya	Swedish	Yiddish	
Catalan	Haitian Creole	Korean	Oromo	Tachelhit	Yoruba	

Table 1. List of to be Supported Languages. *Note: The one in bold are the currently supported languages*

_id: String	**Primary Key**
Image: [String]	
Category: [category_id]	**Foriegn Key**
English: String	
Cebuano: String	
Tagalog: String	
...	
Zulu: String	

Table 2. Database Schema for translation table

_id: String	**Primary Key**
Image: [String]	
English: String	
Cebuano: String	
Tagalog: String	
...	
Zulu: String	

Table 3. Database Schema for category table

_id: String
Name: String
Occupation: String
Organization: String
Username: String
Password: String

Table 4. Database Schema for contributors table

Note: Below figures are just a sketch and not the actual look of the user interface. It will be replaced with the actual interface once the program has been completed.

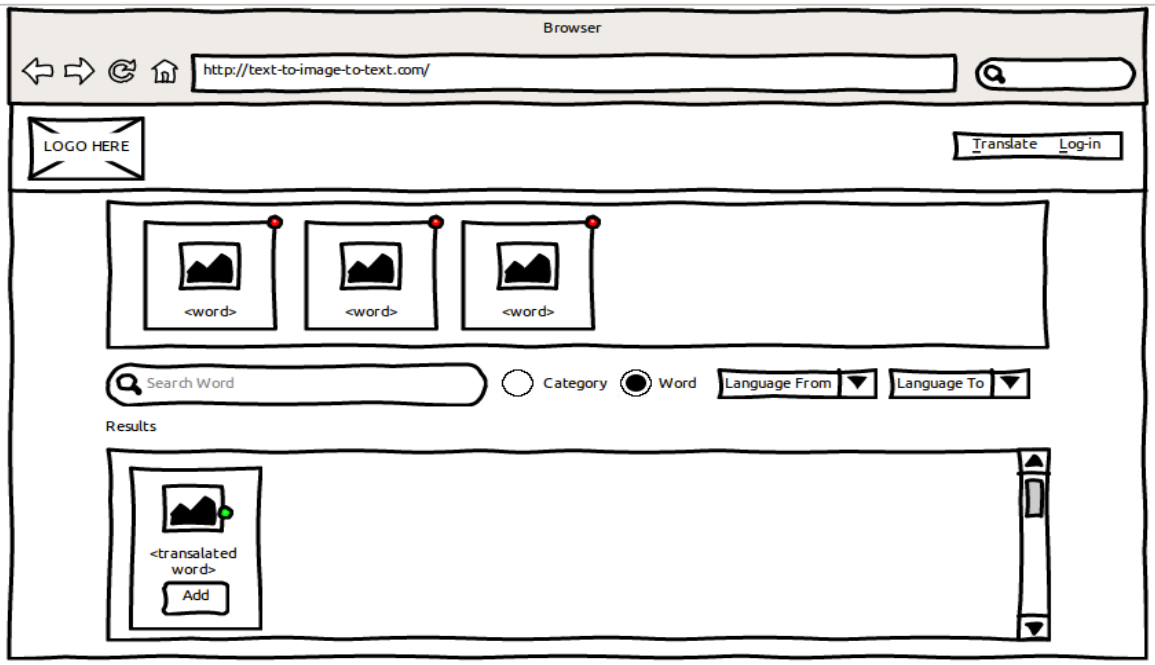

Figure 1. Word Search – Translation Page

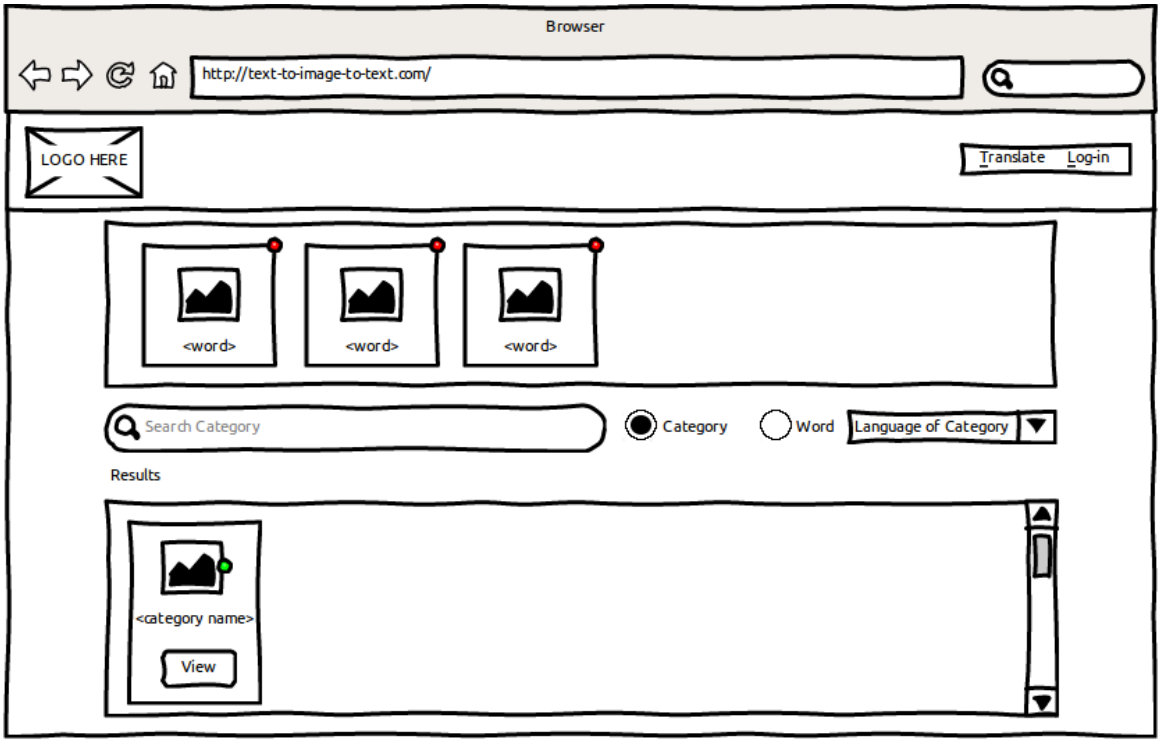

Figure 2. Category Search – Translation Page

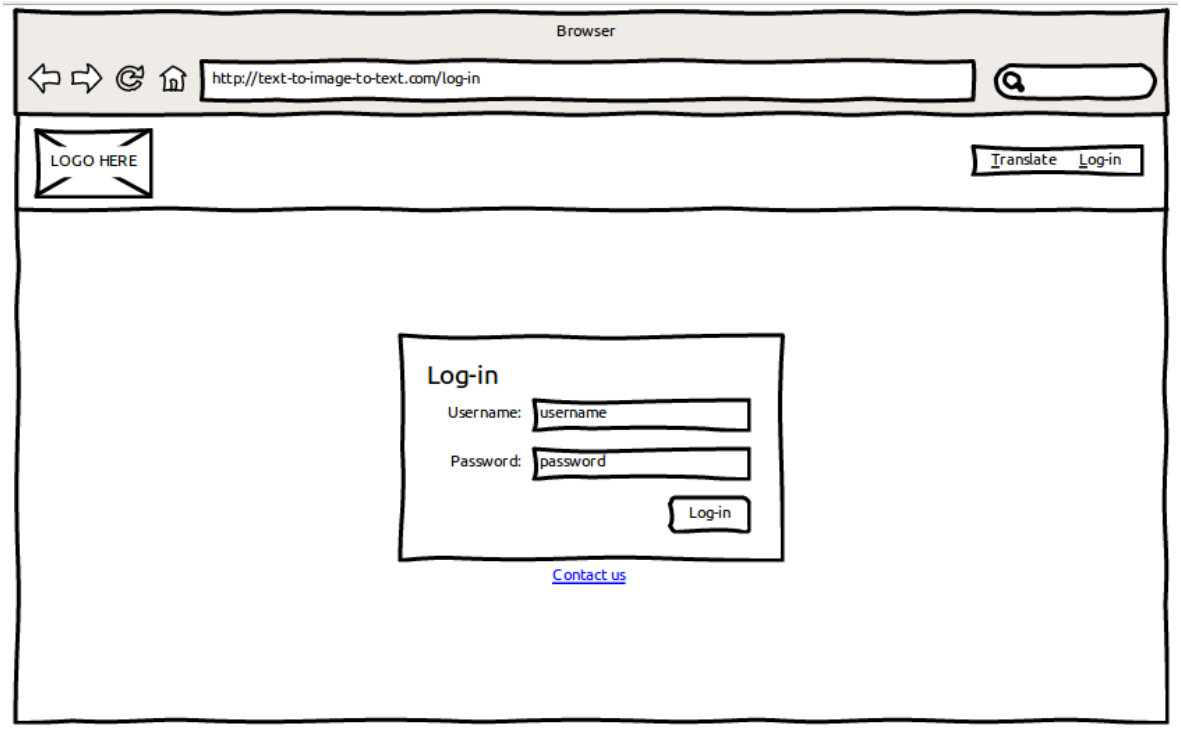

Figure 3. Login Page

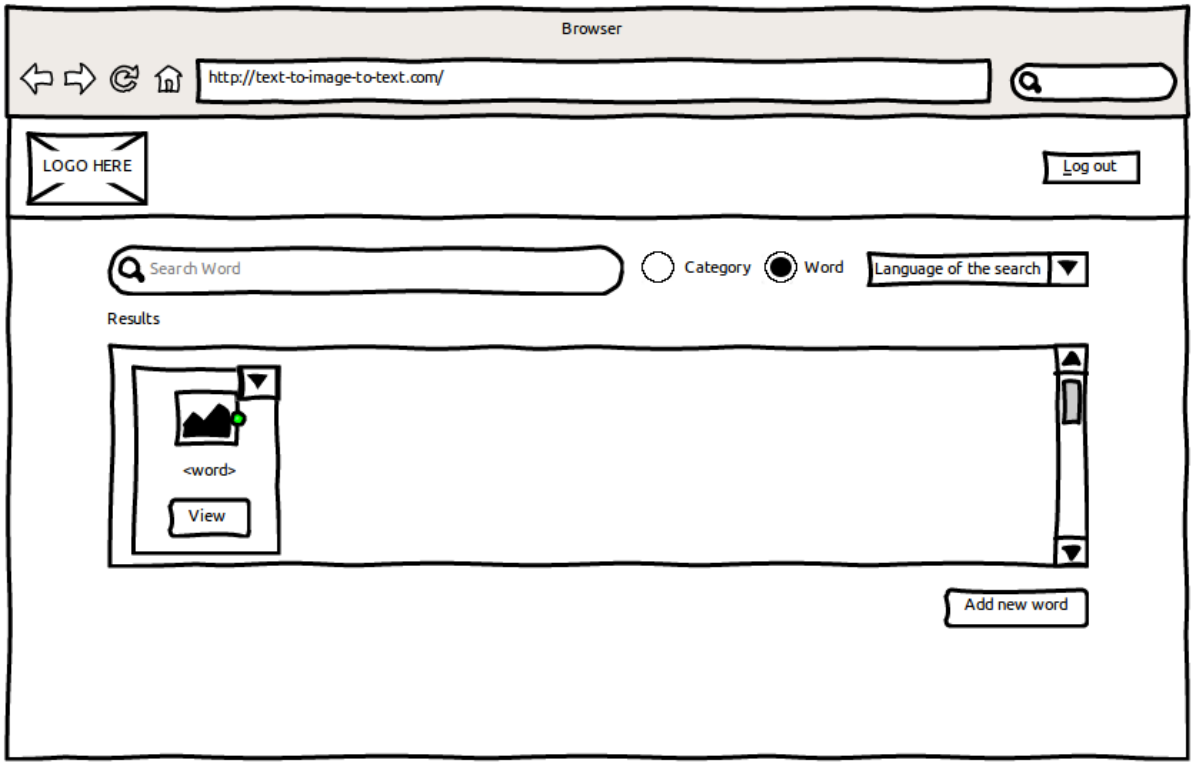

Figure 4. Word Search – Contributor's Page

Figure 5. Word View Modal – Contributor's Page

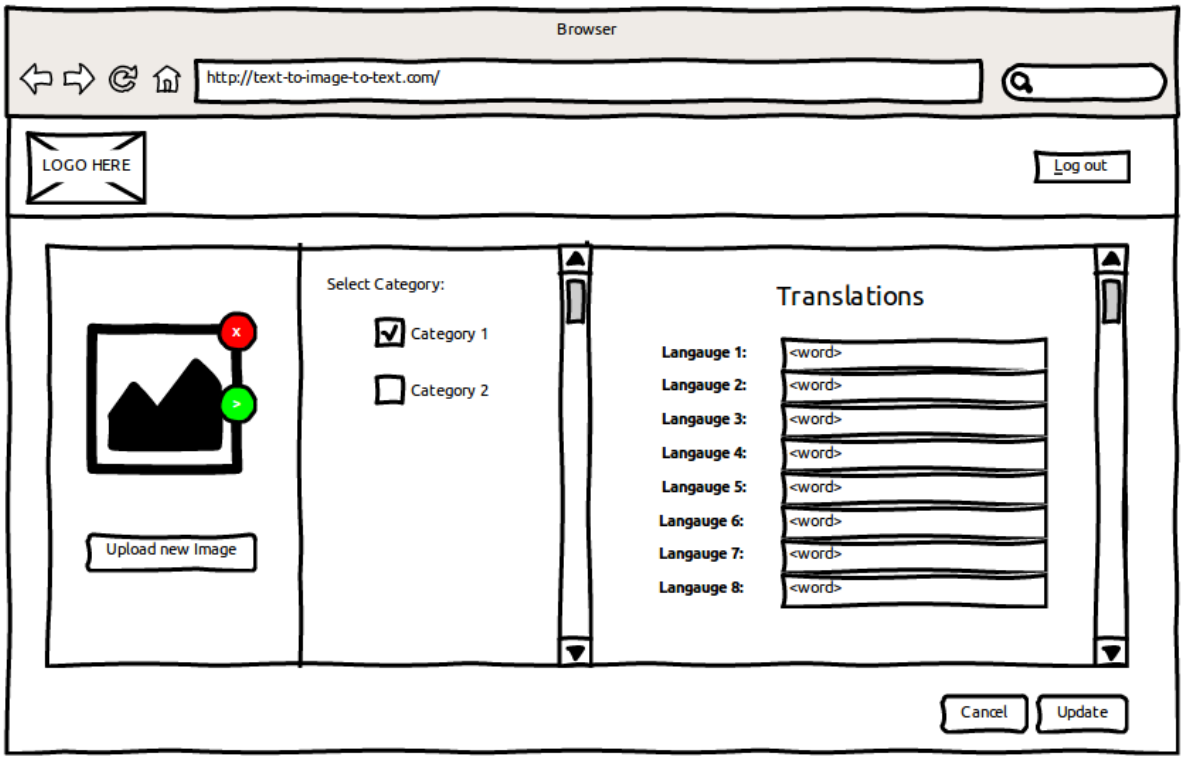

Figure 6. Edit Word Page

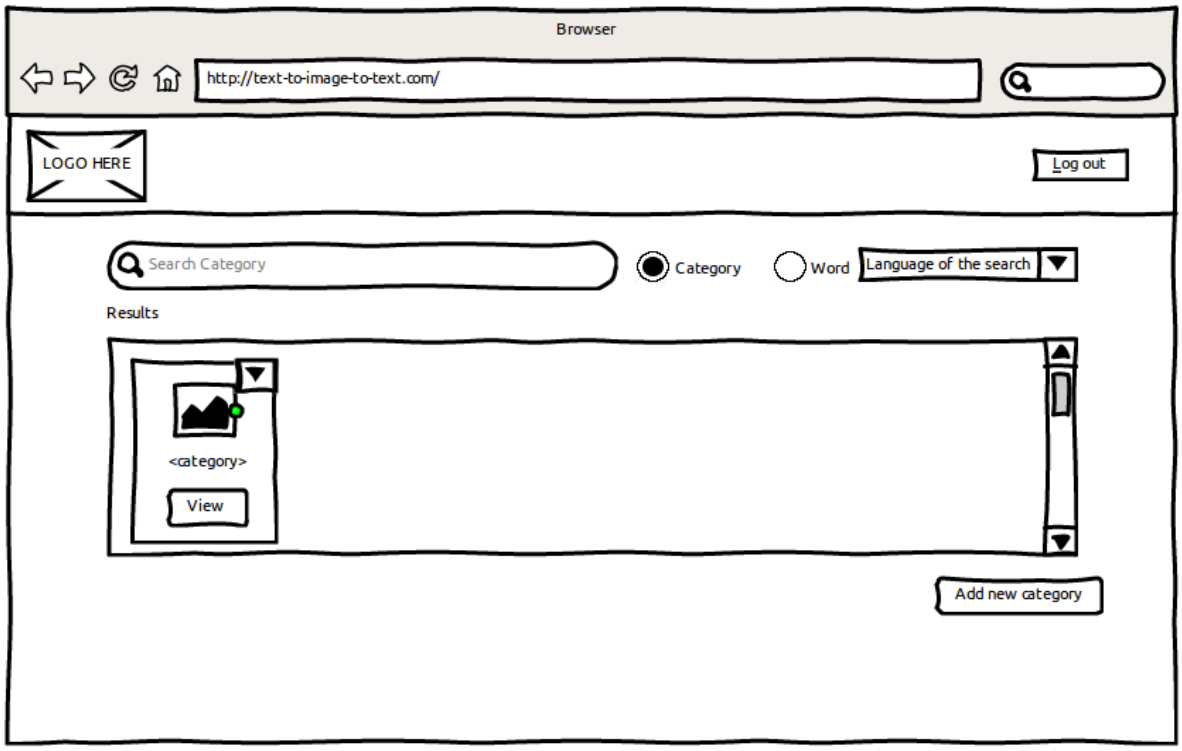

Figure 7. Category Search – Contributor's Page

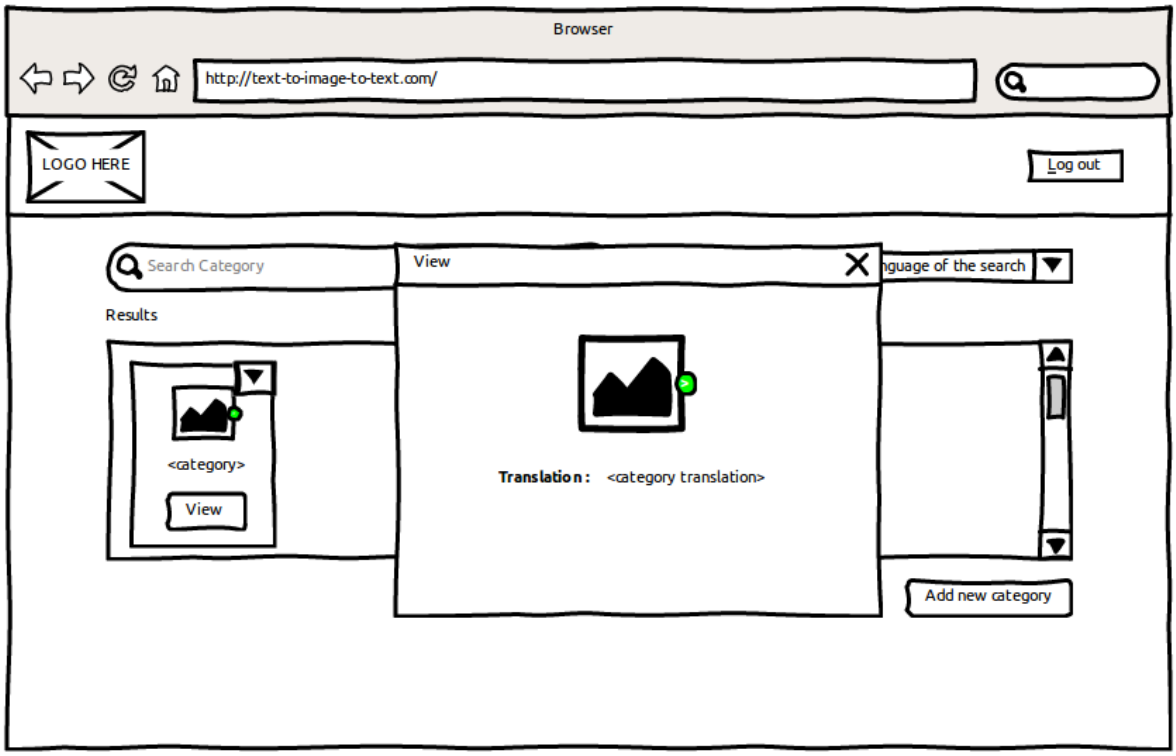

Figure 8. Category View Modal – Contributor's Page

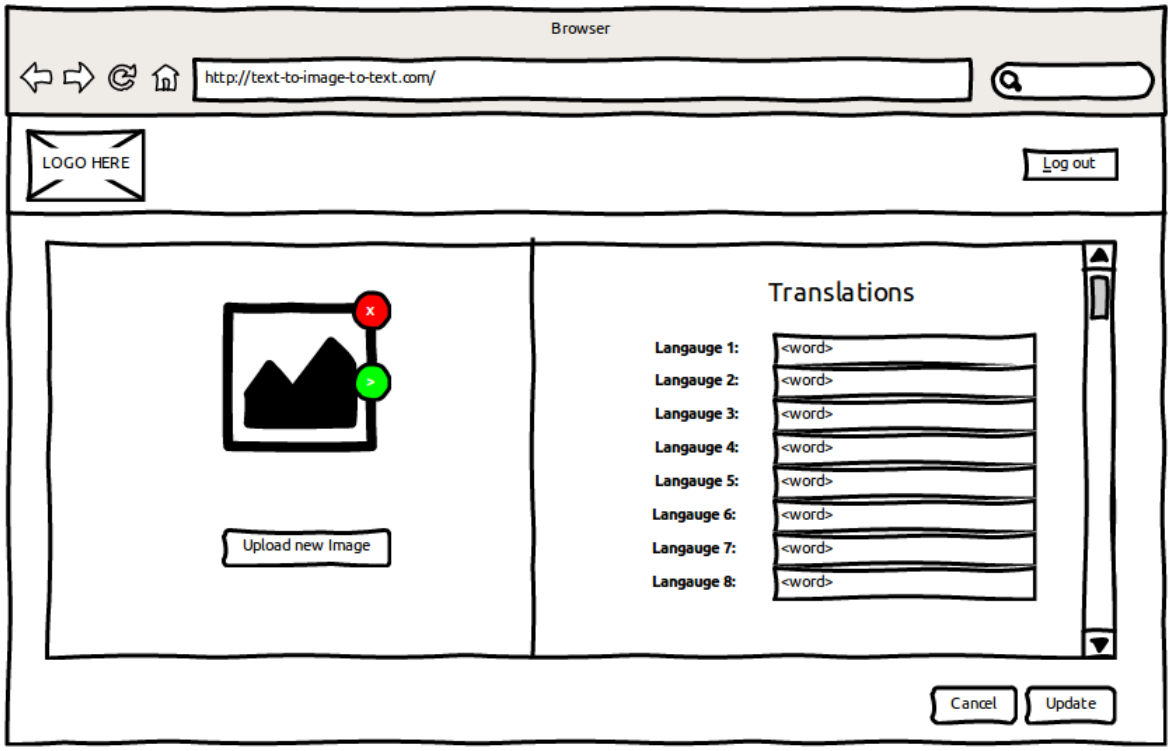

Figure 9. Edit Category Page

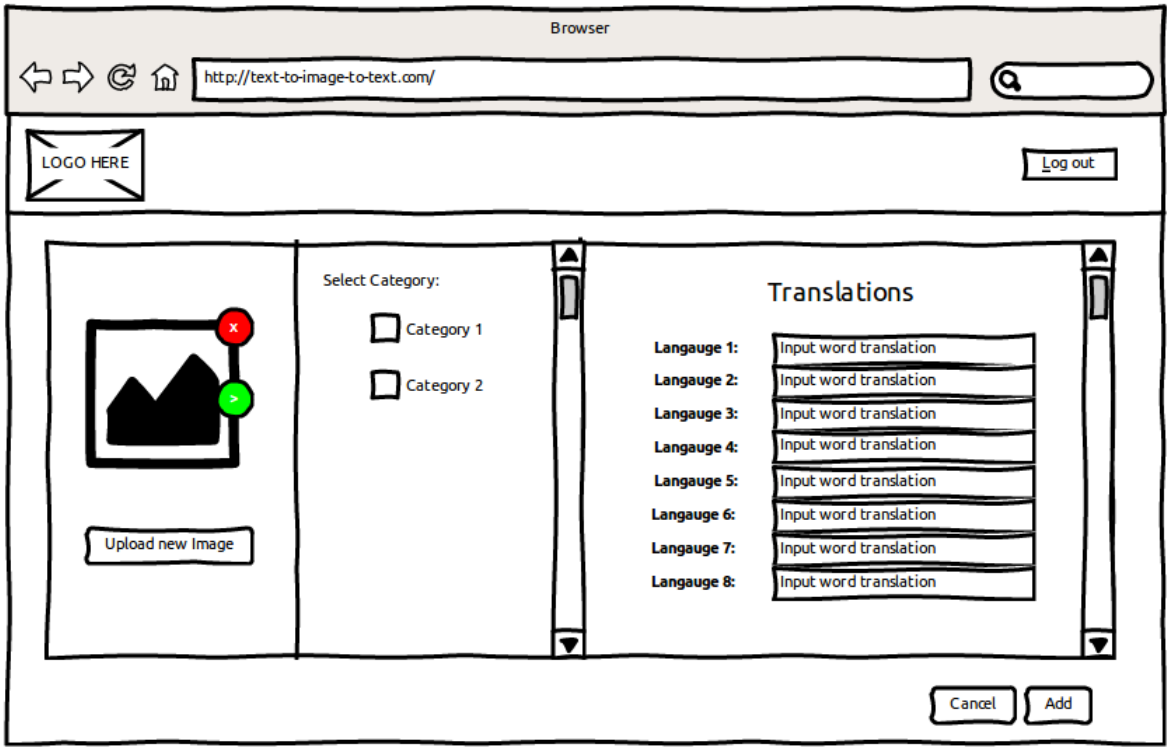

Figure 10. Add Word Page

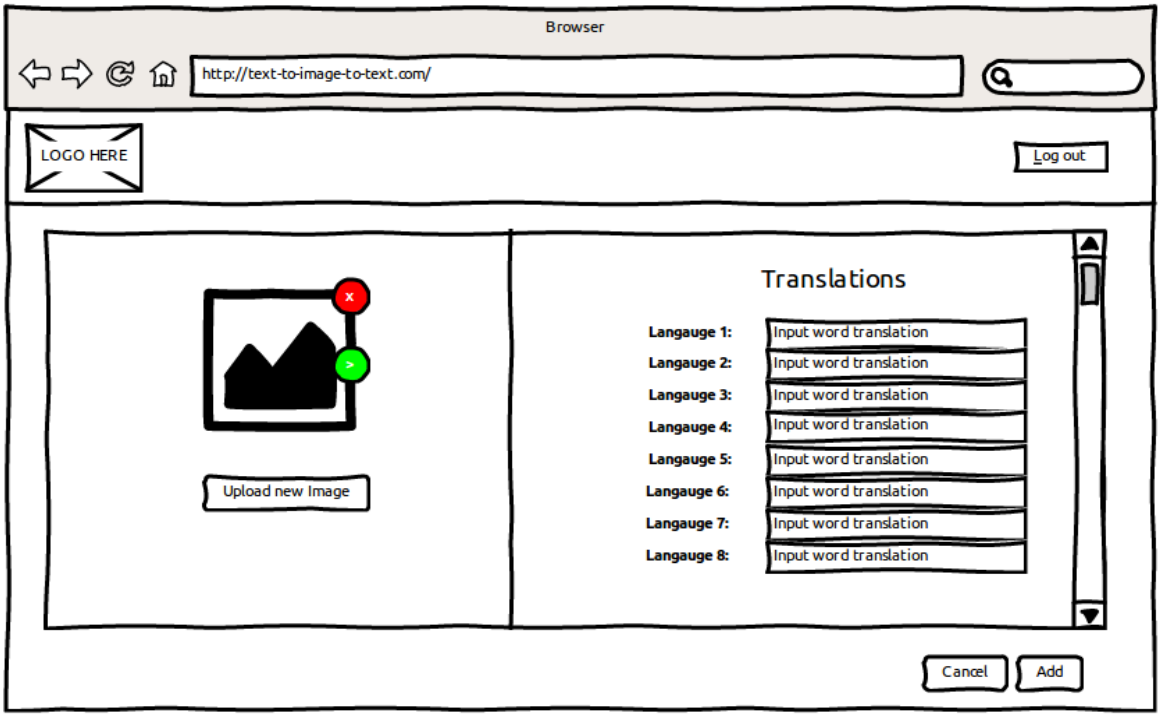

Figure 11. Add Category Page

BIBLIOGRAPHY

Journals

Dewan, Pauline. "Words Versus Pictures: Leveraging the Research on Visual Communication." *Partnership: The Canadian Journal of Library and Information Practice and Research* 10.1 (2015): 1.

Finch, Andrew, et al. "picoTrans: using pictures as input for machine translation on mobile devices." *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*. Vol. 22. No. 3. 2011.

Goldberg, Andrew B., et al. "Easy as ABC?: facilitating pictorial communication via semantically enhanced layout." *Proceedings of the Twelfth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2008.

Goldberg, Andrew B., et al. "Toward text-to-picture synthesis." *NIPS 2009 Mini-Symposia on Assistive Machine Learning for People with Disabilities*. 2009.

Mihalcea, Rada, and Chee Wee Leong. "Toward communicating simple sentences using pictorial representations." *Machine Translation* 22.3 (2008): 153-173.

Munemori, Jun, et al. "Pictograph chat communicator III: a chat system that embodies cross-cultural communication." *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer Berlin Heidelberg, 2010.

Sevens, Leen, et al. "Extending a Dutch Text-to-Pictograph Converter to English and Spanish." *6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*. 2015.

Sevens, Leen, et al. "Improving Text-to-Pictograph Translation Through Word Sense Disambiguation." *cc2016 The* SEM 2016 Organizing Committee. All papers cc2016 their respective authors. This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons. org/licenses/by/4.0* (2016): 131.

Sevens, Leen, et al. "Natural Language Generation from Pictographs." *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*. Association for Computational Linguistics, 2015.

Vandeghinste, Vincent, and Ineke Schuurman. "Linking Pictographs to Synsets: Sclera2Cornetto." *LREC*. 2014.

Zhu, X., Goldberg, A. B., Eldawy, M., Dyer, C. R., & Strock, B. (2007, July). A text-to-picture synthesis system for augmenting communication. In *AAAI* (Vol. 7, pp. 1590-1595).

Websites

Gillert, Sabine. "The Importance Of User Interface Design". Jobscience. N.p., 2016. Web. 27 Nov. 2016.

"How Many Languages Are There In The World? | Linguistic Society Of America". Linguisticsociety.org. N.p., 2016. Web. 22 Oct. 2016.

"Imagenet". Image-net.org. N.p., 2016. Web. 27 Nov. 2016.

"Languages A-Z | About World Languages". Aboutworldlanguages.com. N.p., 2016. Web. 27 Nov. 2016.

"Sclera Symbols". Sclera.be. N.p., 2016. Web. 27 Nov. 2016.

"Word Frequency: Based On 450 Million Word COCA Corpus". Wordfrequency.info. N.p., 2016. Web. 27 Nov. 2016.