

A Web-based Multilingual Language Translation

Rosemary L. Buraga, Maria Visitation M. Gumabay, DIT

ABSTRACT--This study was conducted to develop a web-based multilingual language translation which focused on two main points: the development of web-based multilingual translation system for Ibanag, Tagalog and English and the evaluation of system. Specifically, to determine and gather Ibanag words based on the available printed materials and informants, design and develop a website that will translate the words or phrases in Ibanag to Tagalog and English with the related words, an auditory modality and pictures and determine the extent of performance of the developed system. A Greinbach Normal Form (GNF) rule extraction algorithm was utilized to translate the given words and phrase in Ibanag, Tagalog and English with auditory modality and image, while Inference Engine was the method used to execute the source language text / word and analyses the processes based on the rules. Rapid Word Collection (RWC) and Dictionary Development Process (DDP) were the approaches and processes utilized in gathering of Ibanag words. The words gathered were submitted to the language experts for validation where Ethnographic translation was utilized in translating the language.

The system was developed using the Agile Web development process model. To assess the performance of the system, a survey questionnaire was conducted based on the ISO 25010:2011 standard. As assessed by the participants and IT experts, the study revealed that all the prescribed system category (e.g functionality, usability and reliability) had been complied to a "highly acceptable".

Keywords: auditory modality, ethnographic, inference engine, multilingual, translation,

I. INTRODUCTION

Over the years, most ASEAN region had designed their educational system to 12-year Basic Education Program in line with the international standards. The Philippines was the only country in Asia and one of only three countries worldwide with 10-year Basic Education and pre-university cycle. With this, the Philippine government enacted the Enhanced Basic Education Act of 2013 (RA 10533) to gain a competitive edge in the international community and to adhere with the standard of the educational system worldwide. The Department of Education (DepEd) mandated the implementation of the Mother Tongue-Based Multilingual Education (MTB-MLE) as a feature of the Enhanced Basic Education Program which brought about the use of Mother Tongue as a medium of instruction (DepEd order No. 74, series of 2009) and one of which is Ibanag dialect.

The Ibanag is one of the largest ethnolinguistic minorities in the Philippines that lives along the river in the provinces of Cagayan and Isabela, and their language is Ibanag. Ibanag

is also known as "Ibanak" which comes from the words "I"-means "The" and "Bannag" meaning river. The Ibanag language is spoken by 15.3% native speakers or about half a million having a features of doubled consonants which make the language sound "hard" or guttural (e.g. *Mabbabawi*, pronounced *Mab-ba-ba-wi*, meaning *To regret*; *Zozzog*, pronounced *zoz-zog*, meaning *flame*; *Gurammikan* pronounced *gu-ram-mi-kan*, meaning *somewhat sweet*. Also, monophthongization is observable in Ibanag speakers such the word *umay* (to go), *balay* (house) or *aggaw* (day) are sometimes pronounced as *ume*, *bale*, and *aggo*.

One of the purposes of the implementation of mother tongue-based language such Ibanag was to revive its indigenous cultures to ensure the preservation of the history, belief and practice and most specially the language. Language is the bases of one's identity and one of the most important tools for transmitting culture from one generation to another. It is for these reasons that the researchers aimed to develop a multidisciplinary web translator entitled "A Web-based Multilingual Language Translation" that translates the Ibanag to Tagalog (Philippines national language) and English or vice versa. The developed system can serve as one of the tools or repositories for the preservation and survival of the Ibanag language. Also, incorporating rare languages into machine translation is worthwhile, because it could help preserve rare languages and make the internet and the outside world more accessible to speakers of such languages. It is also another rich source of digital tools which can be used in the different sectors including schools, communities and government institutions [6]. The system could be also used by the teachers and students as one of the language learning tools as it enhances interaction between the teacher and the students and among the students themselves [5].

However, the advancement of today's technology has opened up a number of new system developments, particularly language translation. Having different translation methods, there are still issues and challenges being encountered. Researches revealed that despite the vast progress in the development of MT today, systems still fail to recognize which synonym, collocation or word meaning should be used and more than two thirds of all the sentences were translated incorrectly, which means that there is a relatively small possibility that a system will translate sentences correctly and still cannot offer a system that would be able to translate with at least 50% correctness [19].

Revised Manuscript Received on June 10, 2019.

Rosemary L. Buraga, Isabela State University, Cabagan Campus (E-mail: Rosemary.l.buraga@isu.edu.ph)

Maria Visitation M. Gumabay, DIT Saint Paul University Philippines (E-mail: gumabayvc@gmail.com)

Therefore, this study focuses on transliteration or literal translation where it emphasizes to word-to word and phrase to phrase; and transcreation or cultural translation where the category of translation was based on the Ethnographic translation. This means that the words and terminologies would fit into cultures either it requires exact equivalents in the Target Language or would have to be put as it is into the Target Text to Source Text reading every word [20] and to explicate the cultural context of the source language and translated language versions. Likewise, the kind of the developed system was done through the Interlingua translation (bilingual or multilingual translation) as this translates into different language such as Ibanag to Tagalog and English [23].

II. OBJECTIVES OF THE STUDY

The general objective of the study is to develop website for the multilingual language translation system.

Specifically, it aimed to:

1. Determine the available printed materials in Ibanag with its translation and gather words for translation from the informants.
2. Design and develop a website that will translate the gathered Ibanag words or phrases to Tagalog and English with the related words, an auditory modality and pictures; displays list of Ibanag words in alphabetically order, the most commonly viewed or search words, and it allows the users to post comments and suggestions
3. Determine the extent of performance of the developed system as assessed by the IT experts and participants using the different software criteria as to Functionality, Usability, Portability and Reliability.

III. METHODOLOGY

Research Design

This study used the mixed method design involving the qualitative approach employing ethnographic designs and evaluative descriptive research designs. The qualitative approach was used to gather the Ibanag terms and descriptive research design was employed with a two-phase process method: The first phase is the development of the website and the second phase is the participant's assessment with respect to the developed web translation system.

This study is based primarily on identifying the printed materials in Ibanag and gathering the Ibanag terms or words where the planning phase begins with the general research questions, the kind of site, and types of participants needed are identified. Non-participant observation, self-reflexive, community immersion and firsthand experience were the methods used in data collection in which the researcher enters a social system [13]. Through the Indigenized *Hiyang Approach* and *Pamamaraan* [9] methods Ibanag terms or words used for everyday conversations were identified and collected. Reference [8], [9] stated that the Indigenized *Hiyang* approach serves as "a pioneer of a research methodology wherein the scientist is both the observer and the experienter". It is further stressed that this methodology

is appropriately designed for the purpose of evolving new development paradigms which reference [9] called *Another Development*, which will be suited (*hiyang*) to specific Third World conditions [8].

The main thrust of this study was to develop a web translation where the Rapid Word Collection [18] and Dictionary Development Process [17] were also the methods used in collecting the Ibanag words from the community members with regards to the frequently and widespread use of Ibanag words (Merriam Webster), prompted by questions related to a selection of semantic domains (families of closely-related words)[17], [18]. These domains were then grouped into different categories or through the Thematic Order [10] such as Part of the House, Animal, transportation, Clothing, Location, People, Job, Society, Food, Parts of the Body, Nature, Materials, Seasons, Activities in Church and Places.

The gathered words were submitted to the language experts for validation as these were also the approaches and processes utilized by Oxford Dictionaries and Merriam Webster. In order to collect more terms an 'on the spot phonetic transcription' was utilized where the researchers travelled from one rural community to another to collect and acquire more information [4], [15]. This method allows the researchers tended to focus on older (age 60 to 80) Ibanag speakers [4], [15] who lived all their lives in the community in which they were born were considered as native language speakers and had the 'purest, most vernacular' speech. Also, elders from this age served as the mirror of the past and as rich resources of information about the indigenous practices and beliefs of the Ibanags that are still being practiced to date [8].

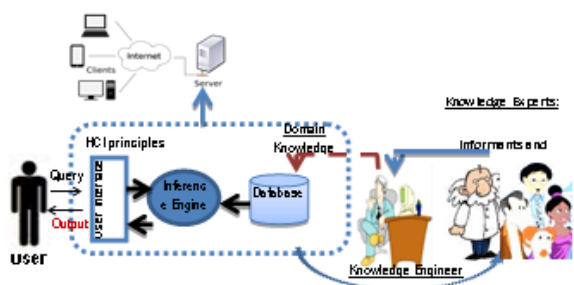
IV. DEVELOPMENT PHASE

One of the software development approaches that was designed and defined for the development process of the software is the Software Development Models. This process model has different life cycle or process that was being followed and used as a guide to ensure the success of the software that was developed. Thus, this study employed the Modern (agile) web development for it is best and suitable for web development. This minimizes overall risk and follows a combination of iterative and incremental approach. Likewise, the development phase of the study includes the design of the system that runs or access through web portal. The system translates a given word/s into the chosen language like Ibanag, Filipino or English or vice versa. The system also provides picture for every translated words with related words and auditory modality that can be used to raise awareness about the issues of correct pronunciation of the Ibanag language thus this preserves the original or proper meaning of words. Also, pictures or images are one way of grabbing users attention and they helped tell meaning of a word and even a story. These can make the translated words easier to understand or immediately learned in a very short period of time [1].



V. SYSTEM ARCHITECTURE

Figure 1 show the system architecture of the developed website which explains the process and design utilized in integrating the components Subject Experts, Inference Engine, HCI principles (User Interface) and Greibach Normal Form (GNF) algorithm



Knowledge Expert. The developed system was initiated through the identification of the knowledge or information sources where the terms were derived and obtained from available Ibanag printed materials but most of the data came from the Ibanag indigenous people. The Ibanags words were then used for translation from Ibanag to Tagalog and English, and were submitted to the language experts for the validation.

Knowledge Engineer. In this study the proponent serves as the knowledge engineer who translates the acquired knowledge into a computer usable language that is stored in a database (served as knowledge repository) and designed an inference engine (through the IF-THEN rules) that uses the knowledge appropriately. The researchers also utilize domain knowledge to determine how to integrate the use of certain knowledge in the translation process and what kinds of explanation would be useful to the end user such as if the user inputted words for translation is not found in the database, an error message is displayed otherwise the corresponding translated word will be processed and displayed.

Database. It is a centralized repository of the Ibanag words and the generated Ibanag word with the corresponding term in Tagalog and English with related words, pictures and auditory modality.

Inference Engine. The inference engine is one of the components of expert systems that the proponent utilizes through the logical rules and tries to derive answers from the knowledge based which is represented by the rule base (IF-THEN rules). The IF-THEN rules are typically used in the developed system through the search method and translation process. Sorting algorithm was also employed specifically the quick sort which was used to manipulate the words from the database to searched and retrieved the translated word from Ibanag to Tagalog and English. This algorithm was developed and first used by Tony Hoare in 1959 in his project on machine translation where he needed to sort the words of Russian language prior to looking them up in a Russian-English dictionary that was already sorted in alphabetic order on magnetic tape.

On the other hand, A Greinbach normal form (GNF) rule extraction algorithm was employed which was derived from reference [14] on their study Expressive Hierarchical Rule Extraction for Translation. In the translation process, each line of the specification was read one at a time where the predefined sets of patterns- each were represented in a particular operation and the inputted word is being checked in the available bilingual corpora for an exact match. If the corresponding word is found, the output is retrieved and displayed. However, in case when the input is a pair of words or phrases, a direct translation is employed where the input is broken down into single word, and each word is taken one by one (word-to-word translation) and will run through the algorithm to find for the most probable translated output of the inputted words or phrases.

User Interface. In this phase, Human Computer Interaction Design and Principles in capturing, collecting, modifying of pictures was followed. The selective process of attention and “attention grabbing techniques” were used as bases to select, gather and modify picture for the term. These techniques respond to the appropriate or proper use of pictures and it is also a way to know what should be focus on a picture which is appropriate to the translated term. During this stage, the researcher hired a professional photographer to capture pictures for the Ibanag terms. All the pictures that were captured and downloaded were submitted by the researcher to the professional graphic artist for the editing and enhancing.

VI. PARTICIPANTS OF THE STUDY

The participants of this study were the 160 elders ages 60-80 who are member or resident of Ibanag indigenous group, seven Ibanag Language experts, 10 IT experts consisting of IT practitioners and 500 respondents who also evaluates the system consisting of students, professionals, members of the community and tourist who were purposively selected.

VII. INSTRUMENTATION / DATA GATHERING PROCEDURE

The in-depth interviews and focus groups were the mixed method research strategy used to conduct interviews among the identified elders and other key informants with regard to the Ibanag words/terms, its related words and proper pronunciation. These strategies were more natural environment for interview because participants are influencing and influenced by others- just as they are in real life [7]. A semi-structured interview was prepared as an interview guide with closed and open-ended questions to allow the researcher to follow up points which needed elaboration and to clarify questions that may be misunderstood by the participants [16]. To elicit natural responses for questions, all the participants were interviewed in locations of their own choice and a note-taking was employed to record the participant’s responses.

Usap-Usapan (small talk) and Pagtanung-tanong (asking around) were also the methods employed where the researcher mingled with the people (pakikipagkapwa) to



establish a level of oneness through *usap-usapan* and after which a *pagtanung-tanong* was raised during kwentuhan

(storytelling) with some Ibanag elders. The researcher way of *pagtanung-tanong* started with everyday normal conversation then casually moved to the subject matter [9],[8].

The Ibanag language also has terms that have similar to spelling, but may differ in meaning and pronunciation. Thus, the auditory modality of the system was also a great help not just for the correct pronunciation but also for the proper use of words. As to the auditory modality feature, the researcher sought the approval from the FM Radio Station Manager as to the recording of Ibanag words with the proper guidance from one of the language experts and the sound engineer and radio technician as to the proper blend and editing of sounds. Furthermore, an audition was conducted to select and identify a recording artist for the auditory modality feature of the system.

VIII. DATA ANALYSIS AND STATISTICAL TOOLS

The Ibanag word/s or terms that were collected and organized were submitted to the identified language experts for a series of validation and verification. This was performed to ensure the reliability and accuracy of the terms, the translated words, the appropriate pronunciation of words and the proper used of pictures to represent the Ibanag term. The validated words carefully reviewed by the researchers to know the word that needs to be updated. The translation process was done through the Ethnographic translation where the purpose is to explicate based on the cultural context of the source language to target language [2]. Every file that was obtained from the language experts were later on cross checked from the work of experts to another. The corresponding corrections from the experts were rolled together with the updated word/s or term/s.

The questionnaire was submitted to the IT experts for a pre-test to ensure the validity of the test items and reliability as regards to ISO25010:2011. An evaluation was conducted after an actual demonstration of the system. The survey questionnaire was used to determine the extent of performance of the system as to functionality, usability, portability and reliability. All the data gathered were tabulated, analyzed and interpreted using weighted mean.

IX. RESULTS AND DISCUSSION

The results of the study revealed that the presence of available printed and learning materials in Ibanag is not evidence except the Ibanag bible. Moreover, even after the mandated of RA 10533, the availability of printed or reading materials that display or present the mother tongue of the indigenous group particularly Ibanag are scarce. Amidst the dramatic advancement of technology, the availability of the technological tools such as website or application or system to meet the demands of the revival; restoration and the preservation of the language of the Ibanag is not evidence.

On the evaluation part the assessment of the IT experts and participants on the performance of the system with

respect to Functionality was “highly acceptable”. They generally agreed that “*The system provides correct translation of word/s and the related words*” and “*The system provides pictures to serve as reminder of the words and has audio that provides proper pronunciation of the word.*”. The results is supported by reference [1] that pictures are one way of grabbing user/learner attention which this can make the translated words easier to understand or learned in a very short period of time and this can help educate the users and can help tell a meaning of a term/word and even a story.

The assessment of the IT experts and participants on the performance of the system with respect to Usability was “highly acceptable”. The IT and participants agreed that the “system is easy to use, learn and its concept is easy to understand”. This could be indicated that the HCI principles that was utilized makes the learners or users easily learn and understand the concepts of the system. This also conforms to reference [24] that the implementation of HCI aims to make an interactive product that was easy and enjoyable to use. However, some of the respondents were not agreed that the “*System can navigate without difficulty*”. This could be presumed that some of the participants particularly the elders were not knowledgeable enough in browsing and navigating a website, while some IT experts were not convinced that “*The system search buttons give a response immediately after the click is made and has no broken links*”. The slow internet connection or weak signal could be factors why the buttons did not response immediately after the click is made.

Also, the assessment of the IT experts and participants on the performance of the system with respect to Portability was “highly acceptable”. Both the IT and participants generally agreed on the item that “*The system can perform the translation of words using different platforms.*” This could be attributed to the compatibility testing that was performed during the testing phase and this has been a factor why this item rated as highest. On contrary, the item that was rated lowest is on items “*The system can be accessed using android devices that have a web browsing features*”. This could be presumed that the minimum hardware and software requirements for the android devices to view the website were not met which resulted to inaccessibility of the system. Moreover, the assessment of the IT and participants on the performance of the system with respect to Reliability was “highly acceptable”. The highest mean rating was given on “*The system informs users when the word/s entered are not found in the records*”. On the other hand, the item that was rated lowest by the respondents is “*The system handles errors systematically*”. This could be attributed to the wrong spelling of the words inputted by the users, wherein the system was not able to inform or provides users an option showing the correct spelling of words.

X. CONCLUSION

The developed multilingual language translation system implementing the concepts of HCI has paid special attention to knowledge acquisition from the experts (native Ibanag members and language experts), storage and retrieval



of knowledge. The generated and validated of 3000 Ibanag words from the indigenous Ibanag members holds most of the language tools to start having basic or functional

beginner for everyday conversations. Most of the world's languages, 500 words will be more than enough to get an individual through any tourist situations and everyday introductions [11]; while 1000 words or 1000- 3000, a tourist would be able to ask people how they're doing, tell them about their day and navigate everyday life situations like shopping and public transit [3] ; and 1000 words allows a tourist to understand about 80% of the language which surrounds him, as long as it is not too specialized [22].

As regards the extent of performance of the developed system with the standard set by the International Organization for Standardization (ISO 25010:2011) as assessed by the participants and IT experts, the study revealed that all the prescribed system category (e.g functionality, usability, portability and reliability) had been complied to a "highly acceptable". The IT experts from the IT industry during the evaluation of the system have shown interest with the study and appreciate the usefulness of the translation system to the language learners. The participants have also shown interests and enthusiasm in using the system as it can translate the Ibanag words to Tagalog and English with related words, pictures and auditory modality. The students are particularly more interested in the system as this can serve as a tool for learning a language. Reference [12] elaborates that aside from the good effects of translation such an aid to learning, it also helps create and maintain good relations between teacher and student for it facilitates classroom management and allows students to maintain their own sense of first language identity, while also building a new bilingual identity.

REFERENCES

1. J. Balm, "The power of pictures. How we can use images to promote and communicate science". 2014. Retrieved from: <http://blogs.biomedcentral.com/bmcblog/2014/08/11/the-power-of-pictures-how-we-can-use-images>
2. Brislin in Cholliludin, "Purposes of Translation", 2007: 26-30.
3. FluentU. "Foreign Language Immersion Online". Retrieved from <https://www.fluentu.com/blog>
4. J.K. Chambers, and P. Trudgill, "Dialectology". Cambridge: Cambridge University Press. pp. 18-23, 1980. ISBN 10: 0521294738 ISBN 13: 9780521294737
5. A. Duff, "Translation", Oxford: Oxford University Press. p. 55, 1989.
6. J. Carbonell, D. Farwell, R. Frederking, S.Helmreich, E. Hovy, K. Knight, L. Levin, and S. Nirenburg "The Pangloss MT System", Proceedings of the First Conference of the Association for Machine Translation in the Americas, Columbia, Maryland, USA, pp.240-241, 1994.
7. R. A. Krueger and M. Casey, "Focus Groups. A Practical Guide for Applied Research", Thousand Oaks, CA: Sage Publications, 3rd Edition, 2000. Retrieved from <http://www.qualitative-research.net/index.php/fqs/article>
8. M. C. Cureg, "Pannonono: Ibanag Cultural Narratives and Discourses in a Changing Communication Environment", 2004.
9. T. B. Obusan, and A. R. Enriquez, "Pamamaraan : Indigenous Knowledge and Evolving research paradigms", 1994. Retrieved from <https://catalogue.nla.gov.au/Record/2193964>
10. G. Wyner, "Your First 625 (in Thematic Order, with notes)", 2017. Fluent-Forever.com
11. J. Gibbons, "The Numbers Game: How Many Words Do I Need to Know to Be Fluent in a Foreign Language?", 2018.
12. G. Cook, "Review Translation in Language Teaching: An Argument for Reassessment", Oxford University Press, 177, 2010. ISBN 978-0-19-442475-2
13. F. Liu, & S. Maitlis, "Nonparticipant Observation", 2010.
14. M. Siahbani and A. Sarkar, "Two Improvements to Left-to-Right Decoding for Hierarchical Phrase-based Machine Translation", 2014.
15. L. Milroy, "Observing and analyzing natural language: A critical account of sociolinguistic method", Oxford: Basil Blackwell, p. 10., 1987.
16. G. T. Ndamba, "Mother tongue usage in learning: An examination of language preferences in Zimbabwe," Great Zimbabwe University, 2008.
17. R. Moe, "Dictionary Development Program" - SIL International. Retrieved from sil-philippines-languages.org/ical/papers/vandenberg-mas
18. SIL. "Language Technology. Rapid Word Collection: An updated approach to dictionary creation", 2018. Retrieved from <https://www.sil.org/enterprise-news/rapid-word>
19. G. Stankevičiūtė, R. Kasperavičienė and J. Horbačiauskienė, "Issues in Machine Translation", 2017, ISSN 2453-7101
20. UKessays, "Translation", 2013. Retrieved from <https://www.ukessays.com/essays/education/translation>
21. Oxford Dictionaries, "How our dictionaries are created", 2018. Retrieved from <https://www.oxforddictionaries.com/our-story/>
22. Universe of Memory, 2015
23. Jacobson in Leonardi, "Jakobson and the concept of equivalence in difference", 2000
24. K.-Y. Huang, "Challenges in Human-Computer Interaction", 2009.