2017 4th International Conference on Innovation in Economics and Business
IPEDR vol.87 (2017) © (2017) IACSIT Press, Singapore

Mobile Platform for Translation of Brazilian Indigenous Languages of Tocantins

George Lauro Ribeiro de Brito [1 +], Alain Neves Lima [2], Gentil Veloso Barbosa [1] and Antônio Wanderley de Oliveira

[1] Federal University of Tocantins

Adress: Quadra 109 Norte, Avenida NS-15, ALCNO-14, Plano Diretor Norte, CEP:77001-090, Palmas – TO - Brazil

Abstract. The present work highlights the development and use of a platform for translation of indigenous languages from the north of Brazil. This system aims to be a channel of cultural interactivity between indigenous natives and people living in large urban centers that have never had the possibility of establishing contact with indigenous people. The objective is also to reduce the rate of retention and evasion of indigenous students in the schools and universities of the State of Tocantins, ensuring a very efficient support platform made available through the web. In this way, linguistic concepts are explained that were used for the normatization in the writing of this language, linguistic techniques and technological factors. From the development of the platform of translation of the indigenous languages *Xerente* and *Apinajé*, it is possible to affirm that the technologies employed present relevant contributions for the development of educational platforms.

Keywords: Translator, Indigenous Language, Indigenous Education, Educational Informatics.

1. Introduction

Currently, approximately 7 (seven) indigenous tribes live in the state of Tocantins, Brazil, totaling approximately 10,000 (ten thousand) Indians, according to data available from Center for Studies and Indigenous Affairs (NEAI) of the Federal University of Tocantins (UFT), are InyKarajá, InyKarajá / Xambioá, InyJavaé, Xerente, Krahô, Apinajé and Pankararu [1].

In the contemporary world, the exclusion of access to education from disadvantaged social segments, such as the indigenous class, is evident. However, the development in recent years of Brazilian public policies related to social inclusion is noteworthy. This is because, in the field of access to higher education, the National Congress passed the Bill 180/2008 - Law of Quotas - where it was signed the compulsory reservation of vacancies in universities and federal institutions and became Law 12,711 of August 29 of 2012.

The Federal University of Tocantins, Brazil, was already adept with this system reserving 5% (five percent) of the universal vacancies to indigenous people, since 2004. Since, through the Resolution of the Teaching, Research and Extension Council 3a / 2004, the implantation of the system of indigenous quotas was approved, being the pioneer in Brazil to delimit vacancies for the entrance in the education of superior level of such students [13].

Thus, in ten years of this historic conquest, the quota system of these indigenous scholars operates efficiently with respect to their entry into the undergraduate process. However, since then, the university suffers from too much retention of these students, of which a great part extrapolates the time allowed for graduation.

Communication and a lack of intimacy with the Portuguese language are then one of the problems they face. We can show that the difficulties encountered with communication substantially impede the

+ Corresponding author. Tel.: + 55 63 3232-8027; fax: + 55 63 3232-8020.
E-mail address: gbrito@uft.edu.br.

development in the careers of these future trainees. Based on this contradiction, we can assume that in this basic premise of the communication process, can generate in our daily activities not completed, unsuccessful negotiations and inefficient capacities. This setback is also evident in elementary and middle school units, creating an ever-greater immersion in the reality of these students and a real shift from what happens around them.

In this treadmill, a platform was developed that allows the translation of the indigenous Xerente language and the Apinaj é language into Portuguese, which is available via the web to meet the demand of indigenous students and scholars, as well as serving as a Cultural integration for non-indigenous individuals who are interested in knowing the language used in these villages.

1.1. Justifications

The present work is justified due to the considerable educational and social relevance. Therefore, two main topics were taken into account in the preparation of this study: the difficulty of communication with the Portuguese language and the high rate of retention of indigenous students in undergraduate courses at the Federal University of Tocantins.

Thus, this online platform contributes substantially to the dissemination of the indigenous culture in the State of Tocantins, Brazil, allowing the direct contact of non-indigenous people who have never had access to the native language and indigenous tribes. As well, it benefits the indigenous students with regard to a good practice of communication in the Portuguese language.

Another factor to foster this work is the considerable scientific relevance. The proposed theme is based on Tocantins indigenous tribes, where some of these ethnic groups present symbols in their native alphabet, not common of the Roman alphabet.

Given the above, the process to work in this segment is reused for other areas of knowledge in the field of computer science. The use of techniques for web programming and research on databases that best behaves with the cause, contributing positively to future work in information technology academies.

2. Methodology

This work, in view of the objectives to which it was proposed to fulfill, employed the typology of methodological and applied research, since the development of the platform proposed here is an instrument for capturing reality, that is, a survey of words Indigenous people to be used in the application, motivated by the need to make available via the web a translator for the indigenous students of the State of Tocantins, as a form of aid in the learning mechanism of the academic environment [7].

In addition, this work was elaborated in four stages. First, a study was carried out on the indigenous languages of the state of Tocantins, Brazil, in order to find out which one had the greatest amount of published data, books and dictionaries.

In the third step, the development of the application, the implementation and the configuration of the platform were evidenced. Here, project management techniques were used to ensure that all phases of the schedule met the planning deadlines.

Finally, the final stage was summarized in the experimental phase, testing and improvement adjustments. The improvement of the application consisted in performing software tests focused on the use in order to find defects, errors and failures (IEEE 610, 1990).

3. Theoretical Foundation

In this part will be presented some theoretical points about this project. Initially a study on linguistics will be exposed. Thus, a fraction will be used to expose the vocabulary of the indigenous languages to be used, the Xerente and the Apinaj é.

Next, the most commonly used translation mechanisms will be presented in the internet, as well as some technologies and characteristics that many of them use.

3.1. Spelling and articulation of the language Xerente and Apinajé

In the Xerente dictionary developed by Guenther Carlos Krieger (1994), the practical alphabet consists of 12 (twelve) consonants, 9 (nine) oral vowels and 5 (five) nasalized vowels. Already the sounds of the Apinajé language that have a certain degree of similarity with the sounds of the Portuguese language, are represented by the same letters in both languages. The alphabet of the Apinajé is composed of 12 (twelve) consonants, 10 (ten) oral vowels and 6 (six) nasal vowels [5].

a	ã	â
e	ẽ	ê
i	ĩ	
o	õ	ô
u	ũ	û

Fig. 1: Alphabetical Grouping-Vowel Xerente.

Vowel Oral	A	À	E	Ê	I	O	Ô	U	Û	ỳ
Vowel Nasal	Ã	ẽ	ĩ	Õ	ũ	ỹ				

Fig. 2: Alphabetical Grouping-Vowel Apinajé.

3.2. Automatic translators

The process of automatic online translation has improved satisfactorily with the advancement of recent years; however, there is much to go forward yet. One of the primary factors relates to human translation, since it deals with infinite possibilities of words, taking into account cultural elements and peculiarities of each language.

As much as the search engines of these translators bring infinite information to a search, these in turn cannot organize all this data succinctly, where you must use now of choosing the ideal word for the translation. Thus, the algorithm responsible for sentence semantics does not operate efficiently.

One of the most widely used algorithms today uses the Machine Statistical Translation (MET) methodology. This procedure consists of a machine of translation of paradigms, where the translations are carried out based on the statistical models. The differential is based on the way of analyzing the translated texts that are already in the network, because, instead of inserting several grammatical rules and uncharged words, he realizes this capture of translations that are already on the web [8].

Since the search engines of automatic translators find a defined translation standard, it uses the same standard to translate other texts that may emerge in the future. It is an artificial intelligence procedure where the machine learns that the best expression to use, will be the one in which it was most often requested.

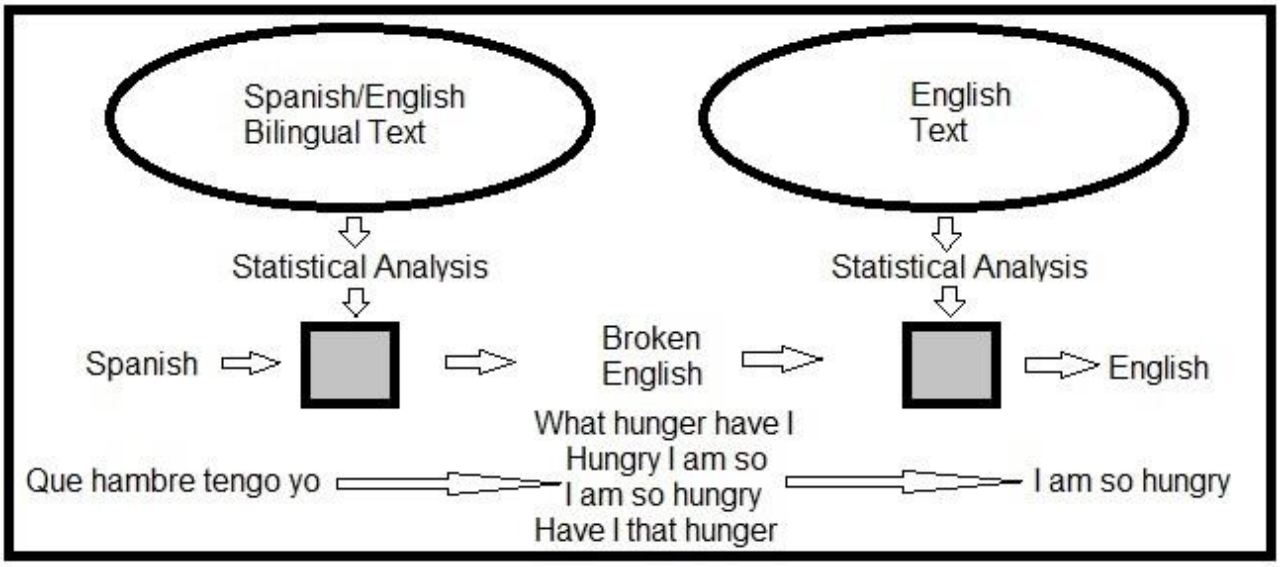

Fig. 3: Statistical MT Systems.

4. Methods Used

In the initial stage of the construction of the indigenous translator, a bibliographical research was carried out to verify which indigenous language had the largest collection of materials published by scholars, as this would facilitate the loading of the application database in the development stage. As it is the insertion of information where they were analyzed and studied by third parties, during this period of research the descriptive method was also present, where the authors of the present work do not manipulate the phenomena obtained.

After this decomposition, a common name was reached in the state of Tocantins, Guenther Carlos Krieger [5]. With numerous published works related to the Xerente culture and language, Krieger made viable through the Portuguese-Xerente / Xerente-Portuguese Indigenous Dictionary the collection of words necessary for the creation of the Online Translator.

The designated time period for this phase of the system was the largest of the schedule, as the author made available a printed dictionary and a scanned version in .DOC file extension. In order for the files to support the creation of the application, they had to undergo a process of adaptation, since in the way the words were distributed in the file (.DOC) they did not present the necessary conditions for automatic insertion In the database using Structured Query Language (SQL) technology.

Numerous removals of information that would not be used were carried out from the original file. To do this, the file was inserted into a .XLSX extension spreadsheet, giving the possibility to configure filters for the data that would serve as the basis for the database, at that stage the collection quantity was 17,148 lines, since the rule departed Of the principle that for each id of a word the nomenclature would be: *(word in Xerente) / n (its syntax) / n (translation into Portuguese) / n / n (word in Xerente) / n (its syntax) / N (translation into Portuguese)*.

With the file slightly appropriate, it was necessary to leave it adapted so that the SQL insert command could be used, that is, to insert the corrected information into our database. This adequacy procedure reduced the number of lines abruptly, leaving the insertion file with 3,692 lines in the nomenclature*: (its syntax) (n.) / N / n (its word in Xerente) (its syntax) (English translation)*.

The intention is that at the end of this step, the .TXT extension file containing the words, could receive the SQL parameters *into tb_word (pal_xerente, syntax, pal_portugues) values ("X", "Y", "Z")*; So this step was successfully completed.

Another important variable to be emphasized, refers to 3 (three) symbols characteristic of the Xerente language. In Portuguese, there are 5 vowels in the alphabet, 'a', 'e', 'i', 'o' and 'u'. Already in the Xerente alphabet, three vowels are added that are not characteristic of the Latin symbols, 'ẽ', 'ĩ' and 'ũ'. When executing the necessary truncations in the table of the word with the function *SQL insert into (table name)*

((field name, field name, field name)) values ("abtuwawẽ", "subst", "bee- Large "), for example, MySQL allowed to save this information, but when displaying words that contained these special characters did not return a satisfactory result, because the database collation was not suitable for these symbols.

In other Database Management Systems (DBMSs) such as SQL Server, one solution to this problem would be to use the *nvarchar* or *nchar* storage data types in the fields corresponding to these inserts. However, MySQL does not support these types of data. To meet this need, a bibliographic search was used to know which languages were used of special characters that had accentuations in vowels different from Latin.

The Swedish alphabet has 3 (three) plus letters, besides the 26 letters used in the Portuguese alphabet, they are: å, ä and ö. With this approximate information, the collation was used in MySQL *latin1_swedish_ci*, so it was possible to continue the system development process by inserting the information successfully.

With all the structure of the database and the information inserted in its corresponding fields, we started to the elaboration phase of the screens and forms. Here, we made use of prototyping techniques, using applications to perform this activity.

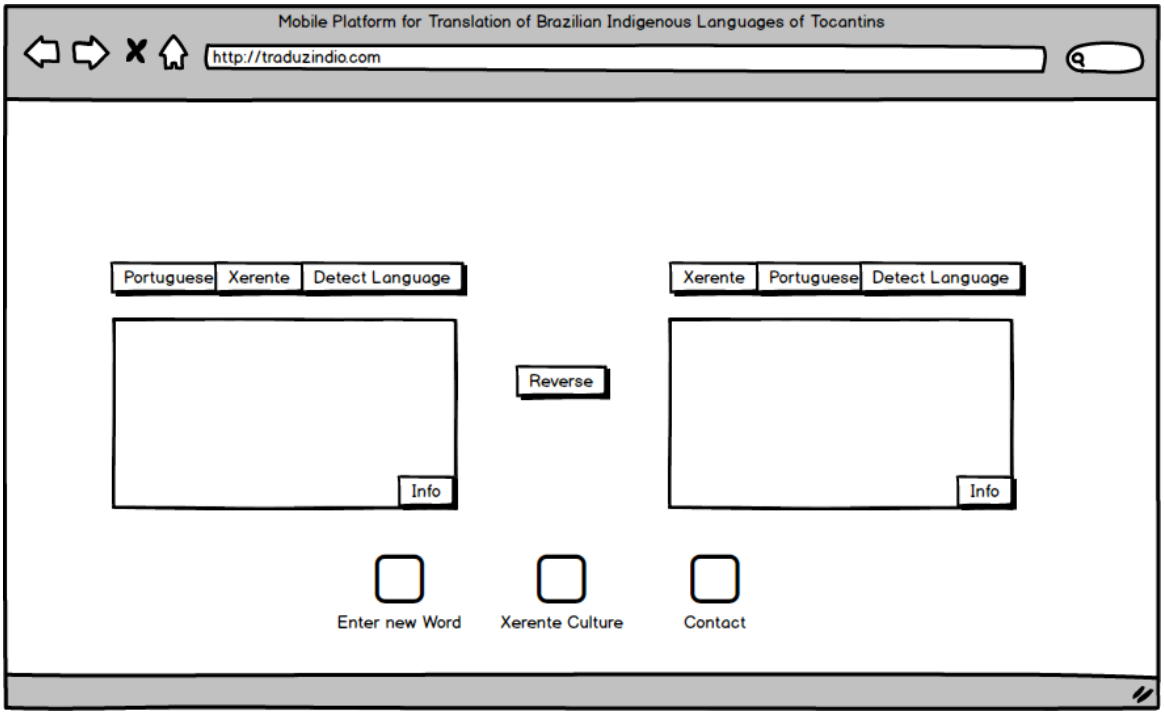

Fig. 4: Prototype of the translation platform [13].

Then, with all the elaborate prototypes, we enter into the development part of the application. To unify all the linguistic content researched and studied, we built an application on the Android platform and made available free on the web.

This is available for download on Google Play at ***https://play.google.com/store/apps/details?id=br.com.traduzindio***.

In Figure 5 we can verify that the application has 3 (three) compartments defined as: i) Translator; (Ii) Culture, and; (Iii) About. Where "Translator" is the space that can translate the languages Xerente and Apinajé into Portuguese and vice versa; In "Culture" one can verify a little of the history of the peoples Xerente and Apinajé, and; And in "About" you can check the details of the developers and the contacts to them.

Fig. 5: Illustration of applicative *Traduzindio*.

Fig. 6: Illustration of applicative *TraduzIndio*.

5. Conclusion

The present work sought to highlight the technological possibilities for the area of linguistics and education, materializing the work of historians and researchers in a translation platform made available via the web and a mobile application. The application obtained a very satisfactory result, because at the end of its elaboration it was evident that such a system assists the native community of the villages Xerentes and Apinajés for the training and training of Portuguese. In addition, the system facilitated people who never had the opportunity to have a direct contact with the native culture of the native tribes.

Today with the indigenous translator, these individuals can access this application online and gain knowledge of a population that yesterday was unknown to many and which today can be known and understood without losing its millenarian culture. This work also highlighted the good moment of online systems due to the expansion of these resources on the Internet and its importance for the reach of isolated populations and with great difficulty of access to more traditional technologies for both education and cultural development.

In the next steps of this work, we will include two more indigenous languages of the state of Tocantins, the Krahô and Karajá, which will include three (3) indigenous peoples of the state and who have already shown interest in this work. In addition, we will include a timeline of the indigenous peoples of Tocantins with greater interactivity for non-Indians to know more about these cultures of the original peoples and the possibility of translation into languages other than Portuguese, for example English and French.

6. References

[1] BARROSO, Lídia Soraya Liberato, Indigenous of Tocantins. In: *The Indigenous Peoples of Tocantins*. Available: <http://www.palmas.org/tocantinsindios.htm>. Accessed in: Aug 18 2014.

[2] BRANDÃO, Alessandro. Online translators: Evaluation of the most famous services. Available: <http://www.englishexperts.com.br/2007/04/02/tradutores-on-line-avaliacao-dos-servicos-mais-famosos/>. Accessed in: 10 Jul. 2014.

[3] IFES. Schaeffe. Available in: https://www.youtube.com/watch?v=SWVRVuMtQnY. Accessed in: 07 de set. 2014.

[4] KOEHN, Philipp. Statistical machine translation. Available:<http://www.statmt.org>.Accessed in: 14 set. 2014.

[5] KRIEGER, Wanda Braidotti; KRIEGER, Guenther Carlos. School dictionary: Xerente - Portuguese; English - Xerente. Rio de Janeiro: National Missions Council of the Brazilian Baptist Conversion, 1994.

[6] LEE, Seung Hwa. Introduction to Language Studies I. Phonemic. Available in: <http://relin.letras.ufmg.br/shlee/Apostila_Fonemica.pdf>. Accessed in: 07 sept. 2014.

[7] MARCONI, Marina de Andrade. Fundamentals of Scientific Methodology. 7 ed. São Paulo: Atlas S.A, 2010.

[8] NIELSEN, Jakob; MOLICH, Rolf. Heuristic evaluation of user interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1990, pp. 249-256.

[9] NIELSEN, Jakob. 10 usability heuristics for user interface design. Available: <http://www.nngroup.com/articles/ten-usability-heuristics/>. Accessed in: 24 set. 2014.

[10] NIELSEN, Michael. Imprementing statiscal machine translation using mapreduce. Available: <http://michaelnielsen.org/blog/implementing-statistical-machine-translation-using-mapreduce/>. Accessed in: 13 set. 2014.

[11] SMAAL, Beatriz. Google translate: How does it work?. Available: <http://www.tecmundo.com.br/google/4883-tradutor-do-google-as-funciona-.htm>. Accessed in: 10 Jul. 2014.

[12] THE nist metrics for machine translation 2010 Challenge (MetricsMaTr10). Available: <http://www.nist.gov/itl/iad/mig/upload/NISTMetricsMaTr10EvalPlan.pdf>. Accessed in: 14 set. 2014.

[13] LIMA, Alain Neves. Development of an online translation platform for the *Xerente* indigenous language for the Portuguese language. 2014. Course Completion Work. (Undergraduate Degree in Computer Science) - Federal University of Tocantins.