Software Translator From Spanish to Wayuu Language

Ernesto Llerena García

University of Cordoba, Montería, Colombia

Cesar Padilla

Windsor Royal School, Montería, Colombia

This paper aims to understand and preserve the culture and language of Wayuu communities living in Colombia in the Department of La Guajira and in many of the states of Zulia in Venezuela. Because of the contact between languages and cultures, we propose the construction of a software translator from Spanish to simple sentences in Wayuu language supported by a model-based automatic translation based on transfer rules to facilitate communication and the preservation of this language. According to the new *Law for Protection of Indigenous Languages* created by Colombian Congress and other international laws, ethnic cultures have the right for the use of new technologies and communication sistems, and this is a first step of a tool to get this purpose.

Keywords: automatic translator, indigenous, ethnic education, culture, Wayuu language

Introduction

Due to the consequences of globalization processes, many of the world's languages are at risk of extinction or are disappearing altogether. This fact, regardless where in the world it happens, and the quality, tradition, or age of the actors concerned, represents an irreparable loss to the living heritage of humanity, since the linguistic processes have complex implications for identity, communication, social integration, education, and development for the people and they have a strategic importance for communities in the planet. When a language dies, it causes a damage to the world's richest texture of cultural diversity and it also produces loss of opportunities, traditions, memory, unique modes of thought, and expression, all valuable resources for a better future. Cultural diversity is part of the driving force of development, not only with respect to economic growth, but also as means to have an intellectual, emotional, moral, and spiritual life. Cultural diversity is an essential component in reducing poverty and achieving the goal of sustainable development, better if it is added simultaneously innovative use of technology which will lead the dialogue among civilizations and cultures, producing respect and mutual understanding.

Indigenous people are now claiming their right to full access to ICTs (Information and Communication Technologies) that have been limited to their communities, to form virtual networks and new forms of indigenous representation of reality, where they can favor collective senses, virtual languages, and socio-technical framework imposed by the use of new technologies. Nowadays, there are electronic dictionaries of indigenous languages on the Internet, highlighted *freelang.net* website structured by traditional

Ernesto Llerena García, Ph.D. candidate, Linguist Department, University of Cordoba.
Cesar Padilla, bachelor, Computer Department, Windsor Royal School.

dictionaries of Asian languages, such as Japanese and Mandarin; European languages like German, Breton, and Galician, and American indigenous languages like Mapuche, Maya, Quechua, and Guaraní. Another important work is the one made by the National Institute of Anthropology and History of Mexico, which began the preparation of "trilingual visual dictionaries" formed by the ancient languages and dialects of the various indigenous groups living in that country. The texts have indigenous language translations into Spanish and English. They will also have pictures and drawings accompanying the words, which are intended to facilitate their understanding. During this decade, they expect to have the first dictionaries on the languages, such as Chontal (Tabasco), Yoreme (northern Mexico), Zapotec (Juchitán, Oaxaca), Tepehua Popoluca (Veracruz), Potosí and Veracruz Huastec, Nahuatl, Tepehuan (northern Nayarit and southern Sinaloa), Mam (Chiapas), and Chichimeco (Guanajuato).

UNESCO (United Nations Educational, Scientific, and Cultural Organization) proclaimed the *Universal Declaration on Cultural Diversity* (November 2, 2001) for endangered languages. In paragraphs 5, 6, and 10 of the Plan of Action, they make some important recommendations:

> Item 5: Safeguard the linguistic heritage of humanity and offer support to expression, creation and dissemination in the possible number of languages.
> Item 6: Promote linguistic diversity (while respecting the mother tongue) at all levels of education wherever possible and promote the learning of several languages from an early age.
> Item 10: Promote linguistic diversity in cyberspace and encourage universal access through the global network to all information in the public domain. (UNESCO, 2001)

In Colombia, the Ministry of Culture through *Ethnolinguistic Diversity Protection Program* in consultation with the representatives of the people concerned, has created a policy to protect and promote the languages presenting in Colombian territory. Generally this program seeks to: (1) Support existing processes and create new processes regarding the use of languages and its transmission to new generations (promoting the use and revitalization); (2) Implement this use in the areas of modern life (standardization, modernization, and use modern means of communication); (3) Contribute to real languages to be official in "the territories where they speak" as is stated in Article 10 of the Constitution; (4) Advocate for the recognition of national and international society of equity value to humanity of these unique vehicles of culture and memory (public awareness); (5) Collect adequate documentation of these languages with priority attention to those that are endangered (documentation and conservation); and (6) Advance scientific knowledge of languages and promote the training of native speakers to research tasks (research and scientific training).

There are studies on Wayuu language, particularly on issues of phonology and grammar. However, there is not currently Wayuu language translator on the Internet, which would promote the language, as suggested by UNESCO in Item 10, nor linguistic diversity in cyberspace, the standardization, modernization, and use of Wayuu language in modern means of communication as it is stated in the protection program.

Objectives

General Purpose

Developing and implementing a software-based machine translator based on transfer rules of simple sentences to translate from Spanish to Wayuu language as an element of communication and preservation.

Specific Objectives

(1) Characterize software models most used for creating translator;

(2) Analyze the structure of simple sentences in Spanish and Wayuu, establishing technical requirements that must be considered when developing a software translator;

(3) Develop a software-based translator based on transfer rules that allows the translation of simple sentences from Spanish to Wayuu language;

And (4) implement a software translator website.

Review

There are, in some countries in America or Europe, dictionaries of indigenous languages in paper editions and many others are already on websites such as *Software Translator From Spanish to Nahuatl*. In Colombia on Internet, we can find *Languages and Cultures of Colombia* in which there are some on line dictionaries as *Bilingual Dictionary Koreguaje-Spanish*, *Spanish-Koreguaje*; and the illustrated basic dictionary *Wayuunaiki-Spanish*, *Spanish-Wayuunaiki*. However, there are no translators of simple or complex sentences for indigenous or native languages worldwide.

Theoretical Framework

In machine translations, there are many possibilities to guide the way that you can carry out the development of the application. According to López (2002), MT (Machine Translation) is the process by which it is used computer software to translate texts from one natural language (like English) to another (like Spanish). He pointed out that there are two types of translation: automatic statistical translation and automatic translation based on rules. The first is a translation based on statistical formulas, aimed at forming a bilingual word corpus (set of sentences) to make the translation as accurate as possible, regardless of grammar. The second makes a literal translation of the text of the source language, and then applies a set of rules for the alignment corresponding to the grammatical structure of the target language.

The rule-based machine translation allows more consistency quality to adapt the translation to context and it also uses the grammatical rules for the alignment of words. For this reason, the model of rule-based machine translation enables a better translation, since it is based on the creation of a set of grammatical rules regardless the similarity between languages as is seen between Wayuu and Spanish languages.

Methodology

This research belongs to the applied technological approach stated by Burg (1982), because it relies on the use of knowledge applied in practice, in most cases, for the benefit of education, creating mechanisms facilitate mediation of experiences in order to improve and make it more efficient. Thus, we concluded that there have been many national and international works with technology applied to indigenous communities, such as webpages, podcasts, and electronic dictionaries among others, but in Colombia, nobody has developed automatic translators for indigenous communities that allow understanding of their language, and in the world are a few (on Internet no more than four).

Results

Characterization of the Machine Translation Models Used to Create Software Translator

In the field of automatic translation, there are several ways to program or construct models for the processes of translation. In this research, we studied some of them to determine the most appropriate to translate simple sentences from Spanish to Wayuu.

One is statistical translation model, in which are programmed statistical formulas using the input or input text and when it is processed by these statistical formulas, the software seeks the correct words in the target language, similar words or words that have been used in most cases by users; as these systems have an error correction module for users, then, all users around the world make corrections and the system is able to take these tips and make the appropriate changes in the code to not make the same mistake. These systems are used in most languages where grammar is similarly structured as the case of Spanish and English among others.

Another model is automatic rule-based translation model, which is divided in rule-based transfer model and rule-based interlingua model; the first makes analysis of what you want to translate word by word and replaces the more appropriate equivalents; the second uses a third language as a benchmark that has a similar structure for both the source language and the target language, so the words or sentences go from a source language into an intermediate language and finally reach the target language.

Analysis of the Structure of Simple Sentences in Spanish and Wayuu, Establishing Technical Requirements When Developing a Software Machine Translator

Like any software, you must give a series of commands to the computer in a programming language to make this run as the user or developer wants them to do. The principle on which was based the translator was to move the process that takes the mind to translate, to conditions that the computer can interpret; for this case, the programming language used was Action Scrip 3.0.
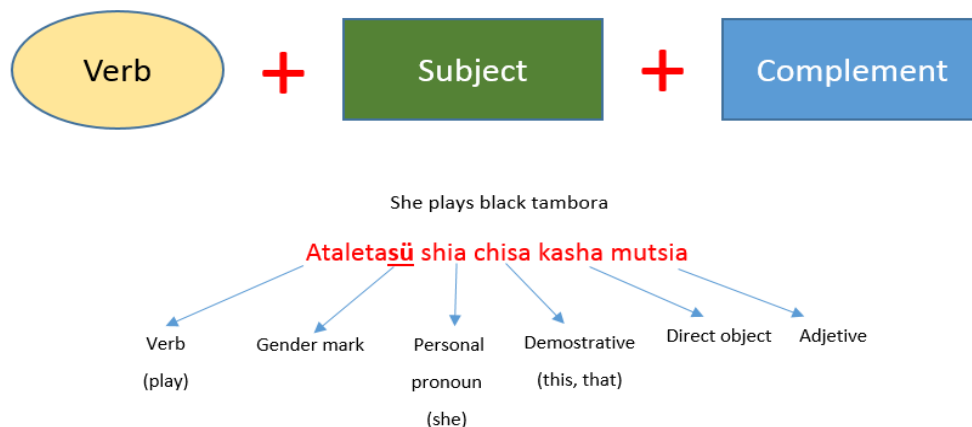

Figure 1. Basic simple sentence in Wayuu language.

The first rule is to take the words that are entered in the textbox and separate them by a method that detects each time you make a space, then take each word separately by spaces and stores it in an array, as the following: "Tomorrow I have football match" = "tomorrow", "I", "have", "football", and "match", so the program starts to detect which were the words that have been written to translate. In addition, previously it is explained in the textbox that the user cannot write capital letters, commas, and periods, all because errors may occur, because the

software has not yet reached its full development. After this, the software removes the words that have not been translated in Wayuu, and translates only the words that are needed again in another arrangement. The software allows creating new rules as many as it is needed. The basic simple sentence in Wayuu language is in Figure 1.

Construction of the Database

One factor that determines the success of a translator is the reliability of your database and the effectiveness of the program to quickly search for words in it. For this reason, this database was built on XML (eXtensible Markup Language) software, as it allows easy construction and its nodes; each node of the database has four important attributes: the Spanish word, its translation in Wayuu, type of word, and grammatical elements.

The program has more than 5,000 nodes that have the same structure. Thus, accessing a node is accessing to all the information required for further processing.

Organization of the Words in Wayuu Structure

Since making an organization considering the word upon entering the function responsible for this can make the programs low in doing the search, it was thought to follow the simple sentence structure in Wayuu previously been studied: verb-noun-noun-pronoun. Given this structure, it created an array with the number of fields needed to accommodate the required number of words and form this structure; then, when the words are coming to the function responsible for organizing, it searches for the word type and places it in position.

Thus, the software does the organization of words, according to the proposed structure, receiving a word at a time and immediately organizing it in place, according to the type of word that tells the database. Finally, the program gives the result. Other words are eliminated by the software, because Wayuu have no translation for some of them. However, the meaning of the simple sentence is the same in Spanish and Wayuu.

Software Development Translator

At the end of the process of organizing what remains is to show in another textbox the result of the translation; this is a simple function that only handles arranging the translation that has been made and shows it in the textbox.

Software Validation Expert Translation in the Language, to Verify Its Reliability in Use

In general, traditional language translators like English or French, are on line or in smaller processors. These translators have reliability, but so far no one is 100% reliable, since there are many problems with the semantics and syntax of the language to make the translations. In the case of indigenous languages, there are few translators of these kind of languages, even more in Colombia where there is any. So this software is in the research group page Lexicon—http://www.lexicondecordoba.org/—for those who wish to contribute to the translator for improvement and want it to be used for communication and preservation.

Conclusions

From the need of protecting native languages enacted by UNESCO, the *Law on Protection of Native Languages* in Colombia, and many other laws raises the necessity to use ICT as a mechanism for the conservation and preservation of languages and cultures in the world. Creating automatic translators for native indigenous languages will make an important contribution in this regard, as we approach the world of these languages and cultures through the use of software tools for this purpose.

References

Bilingual Dictionary Koreguaje-Spanish Spanish-Koreguaje. (2005). Retrieved from http://www.sil.org/americas/COLOMBIA/show_subject.asp?pubs=biblio&code=Dictionaries&Lang=spael

Burg, A. (1982). *Science and technology in the world*. México: CONACYT (Consejo Nacional de Ciencia y Tecnología).

Cabero Almenara, J. (2002). *Impact of new technologies of information and communication in educative institutions*. Granada: Grupo Editorial Universitario.

INAH (Instituto Nacional de Antropología e Historia). (2013). *Visual and trilingual dictionaries for indigenous languages*. Retrieved from http://www.cronica.com.mx/nota.php?id_nota=172098

Llerena Villalobos, R., & Llerena Garcia, E. (2004). *Etnolinguistic dictionary of embera language from Alto Sinú*. Montería: Talleres Gráficos Augusto Yepes.

López, V. (2002). *Possibilities and realities of machine translation*. La Linterna del Traductor. Madrid: ASETRAD.

Ministry of Culture. (2010). *Native language protection law*. Retrieved from http://www.lenguasdecolombia.gov.co/content/ley-de-lenguas-nativas

Oliver, A., Moré, J., & Climent, S. (2007). Translation and technologies. *FUTUROS, II*(6).

Oscar, D. Á. (2004*).* Technological hope: Internet for indigenous people of Latin America. *FUTUROS, II*(6).

Ramones, Q. S., & Madueño, L. (2006). *Süchiki Walekerü: An example of the use of technology in indigenous schools, case Wayuu language*. Mérida: Universidad de los Andes.

Ruíz, M., & Follossá, J. (2005). *Better techniques for translation based on phrases*. Barcelona: Universidad Politécnica de Cataluña.

UNESCO (United Nations Educacional, Scientific, and Cultural Organization). (2001). *Universal declaration of cultural diversity, parts 5, 6, 10*. Retrieved from http://unesdoc.unesco.org/images/0012/001271/127162s.pdf