



XXVII Congreso Nacional y I Internacional de Lingüística, Literatura y Semiótica



Homenaje a
Carlos Patiño Roselli, Rafael Humberto Moreno Durán
y Jairo Aníbal Niño

Software traductor de español a lengua kuna

Ernesto Llerena García
Universidad de Córdoba

En la actualidad y agobiadas por las consecuencias de los procesos de globalización, muchas de las lenguas del mundo están en riesgo de desaparecer o están desapareciendo por completo. Este hecho, independiente del lugar del mundo donde suceda y de la calidad, tradición o antigüedad de los actores afectados, representará una pérdida irremediable al patrimonio vivo de la humanidad, toda vez que los procesos lingüísticos tienen complejas incidencias sobre la identidad, la comunicación, la integración social, la educación y el desarrollo de la gente y revisten una importancia estratégica para las personas y el planeta. Cuando una lengua se extingue, se empobrece el rico tejido de la diversidad cultural y se pierde también posibilidades, tradiciones, memorias, modos singulares de pensamiento y expresión, todos ellos recursos valiosos para lograr un futuro mejor.

La diversidad cultural hace parte de la fuerza motriz del desarrollo, no sólo en lo que respecta al crecimiento económico sino como medio para tener una vida intelectual, afectiva, moral y espiritual más enriquecedora. La diversidad cultural es un componente indispensable para reducir la pobreza y alcanzar la meta del desarrollo sostenible, aún más si simultáneamente se le adiciona la utilización innovadora de los medios y de las TIC en particular, lo que lleva al diálogo entre civilizaciones y culturas, al respeto y a la comprensión mutua.

Los pueblos indígenas reclaman ahora su derecho al total acceso a las TIC que, por el momento, ha sido limitado, para conformar redes virtuales indígenas y nuevas formas de representación de la realidad, en donde se privilegien los sentidos colectivos, el lenguaje

virtual y todos aquellos entramados sociotécnicos que impone la utilización de las nuevas tecnologías. Es así como a nivel internacional encontramos en la actualidad diccionarios electrónicos de lenguas indígenas que aparecen en Internet, en donde se destaca la página web *freelang.ne* en la que aparecen diccionarios de las tradicionales lenguas asiáticas como el japonés y el mandarín; europeas como el alemán, el bretón y el gallego, y americanas de etnias indígenas como el mapuche, maya, quechua y guaraní. Otro de los trabajos destacados es el del *Instituto Nacional de Antropología e Historia* (INAH) de México, el cual comenzó la confección de los “Diccionarios Visuales Trilingües” que reúnen las antiguas lenguas y dialectos de los diversos grupos indígenas que viven en ese país. Los textos tendrán traducciones de la lengua indígena al español y al inglés. Además, tendrán imágenes y dibujos acompañando las palabras, que tienen como fin facilitar su asimilación. Durante esta década esperan tener listos los primeros diccionarios sobre las lenguas chontal (Tabasco), yoreme (norte de México), zapoteco (Juchitán, Oaxaca), popoluca y tepehua (Veracruz), huasteco veracruzano y potosino, náhuatl, el tepehuán (norte de Nayarit y sur de Sinaloa), el mam (Chiapas) y el chichimeco (Guanajuato).

La UNESCO ha proclamado la Declaración Universal sobre Diversidad Cultural (2 de noviembre de 2001) para las lenguas en peligro de extinción. En los puntos 5,6, y 10 del Plan de Acción de la mencionada declaración hace unas recomendaciones muy importantes:

Punto 5: salvaguardar el patrimonio lingüístico de la humanidad y ofrecer apoyo a la expresión, creación y difusión en el mayor número posible de lenguas.

Punto 6: fomentar la diversidad lingüística (siempre que se respete la lengua materna) a todos los niveles de la educación siempre que sea posible y promover el aprendizaje de varias lenguas desde una edad temprana.

Punto 10: promover la diversidad lingüística en el ciberespacio y fomentar el acceso universal a través de la red global hacia toda la información en el ámbito público.

A nivel nacional, el Ministerio de Cultura a través del Programa de Protección a la Diversidad Etnolingüística (PPDE) en concertación con los representantes de los pueblos concernidos, ha

creado una política de protección y fomento a las lenguas de los grupos etnolingüísticos presentes en el territorio de Colombia.

De manera general este programa busca:

1. Apoyar procesos existentes y crear procesos relativos a:
 - Favorecer el uso de las lenguas y su transmisión a las nuevas generaciones (**fomento del uso y revitalización**).
 - Implementar este uso en las esferas de la vida moderna (**normalización, modernización y uso en los medios modernos de comunicación**).
 - Contribuir a una real **oficialización de las lenguas** en “los territorios donde se hablan” tal y como reza el artículo 10 de la Constitución.
 - Propender la toma de conciencia por parte de los usuarios de estas lenguas del inmenso valor y ventaja que supone su conocimiento y su uso (**concientización y trabajo sobre actitudes y valores**).
 - Abogar por el reconocimiento por parte de la sociedad nacional e internacional del valor patrimonial para la humanidad de estos vehículos únicos de cultura y de memoria (**sensibilización pública**).
 - Recoger una adecuada documentación de estas lenguas con una atención prioritaria a las que están en peligro de extinción (**documentación y conservación**).
 - Avanzar en el conocimiento científico de las lenguas, favoreciendo la capacitación de hablantes nativos a las tareas investigativas (**investigación y formación científica**).
2. Recomendar mecanismos de fortalecimiento institucional para ayudar el Estado a desarrollar una política consistente y duradera en el área, en concertación con los representantes de los pueblos que hablan lenguas propias (**institucionalización de instrumentos**).

En Colombia y Panamá encontramos el grupo étnico kuna. En Colombia se encuentran en Caimán Nuevo (Antioquia) y en Arquía (Chocó); también en Panamá en las islas y costas del archipiélago de San Blas y a lo largo de los ríos Bayano, Tuira, Chucunaque y Paya. Los kunas se autodenominan *tule* que significa “persona”. Los plátanos, cocos, y el pescado forman parte de la dieta kuna, suplementado con comida importada y animales domésticos.

Existen estudios sobre la lengua kuna, en particular sobre temas de fonología y de gramática, como los del investigador Rito Llerena Villalobos en su libro *Relación y determinación en el predicado de la lengua kuna*. Sin embargo, no existe actualmente un traductor de la lengua kuna en Internet, el cual promovería, como lo plantea la UNESCO en el Punto 10, la diversidad lingüística en el ciberespacio; asimismo, se fomentaría la normalización, modernización y uso de la lengua kuna en los medios modernos de comunicación como lo plantea el PPDE.

Objetivos

Objetivo general

Desarrollar e implementar un software de traducción automática basado en reglas por transferencia, que permita traducir oraciones simples de español a la lengua kuna, como un elemento de comunicación y preservación de la lengua kuna.

Objetivos específicos

- Caracterizar los modelos de traducción automática más utilizados para la creación de software traductor.
- Analizar la estructura de las oraciones simples en español y kuna, estableciendo los requerimientos técnicos que se deben contemplar al momento de desarrollar un software traductor automático.
- Desarrollar un software de traducción automática basado en reglas por transferencia que permita la traducción de oraciones simples de español a kuna.
- Implementar el software traductor en una página en Internet.

Antecedentes

De manera general existen diccionarios de las lenguas indígenas de diversos países americanos o europeos en ediciones en papel y muchas otras ya se encuentran en páginas de

internet. A nivel internacional se encuentra en Internet el *Software Traductor de Español a Náhuatl*. Para el caso colombiano encontramos en la dirección de internet *Language and Cultures of Colombia* varios diccionario en línea como el *Diccionario bilingüe Koreguaje-Español*, *Español-Koreguaje* y el *Diccionario básico ilustrado Wayuunaiki-Español*, *Español-Wayuunaiki*. Sin embargo no se encuentra traductores de oraciones simples o complejas para lenguas indígenas o criollas a nivel mundial.

Marco Teórico

En materia de traducción automática son muchas las posibilidades de orientar la manera en que se puede llevar a cabo el desarrollo de la aplicación. Según López (2002) La traducción automática (MT) es el proceso mediante el cual se utiliza software de computadora para traducir un texto de un lenguaje natural (como el inglés) a otro (como el español). Para el autor al realizar cualquier tipo de traducción ya sea por medio de la intervención humana o un computador, el objetivo de la traducción es convertir la información en una lengua de origen, restaurada en una lengua de destino. Afirma que existen dos tipos de traducción: la traducción automática estadística y la traducción automática basada en reglas. La primera es una traducción basada en formulas estadísticas, cuyo fin es la formación de un corpus de palabras bilingües (conjunto de frases) para hacer la traducción lo más exacta posible, sin tener en cuenta la gramática. La segunda consiste en realizar una traducción literal del texto de la lengua origen, después se le aplica un conjunto de reglas, para realizar la alineación correspondiente a la estructura gramatical de la lengua destino.

La traducción automática basada en reglas permite más calidad en la coherencia para adecuar la traducción al contexto; además, utiliza las reglas gramaticales para la alineación de las palabras, aunque es difícil manejar con la traducción automática basada en reglas las excepciones gramaticales de la lengua origen y destino.

Debido a que la lengua kuna tiene una estructura gramatical compleja, desarrollar un traductor automático estadístico no sería viable para un traductor español-kuna debido a que este modelo no tiene en cuenta la gramática para realizar la traducción; este modelo suele utilizarse más para lenguas cuya estructura gramatical son similares.

Por tal razón el modelo de traducción automática basada en reglas permite realizar una mejor traducción, puesto que se basa en la creación de un conjunto de reglas gramaticales sin importar la similitud entre las lenguas, como es el ejemplo, entre la lengua kuna y el español.

Metodología

Esta investigación es de un enfoque de tecnológica aplicada según Andrea Burg (1982), porque se basa en la utilización de los conocimientos en la práctica, para aplicarlos, en la mayoría de los casos, en provecho de la educación, creando mecanismos que facilitan la mediación de las experiencias, con el fin de mejorarlo y hacerlo más eficiente; así se llegó a la conclusión de que han sido muchos los trabajado a nivel nacional con tecnología aplicadas a las comunidades indígenas, tales como páginas web, podcats, diccionarios electrónicos, entre otros, pero aún no se han desarrollado los primeros avances en traductores automáticos para comunidades indígenas que permitan el entendimiento de su lengua a la vez que su preservación.

Resultados

Caracterización de los modelos de traducción automática más utilizados para la creación de software traductor.

En materia de traducción automatizada también existen varias formas o modelos de programar los procesos que se deben llevar a cabo para realizar tal fin; en esta investigación fueron estudiados algunos de ellos para determinar el más apropiado para traducir oraciones simples de español a kuna.

Uno de ellos es el modelo de traducción automática estadística, en el cual son programadas unas formulas estadísticas por medio de las cuales se toma el input o texto de entrada y al ser procesado por estas fórmulas estadísticas, se va tomando la palabra en la lengua destino que tenga más aproximación o que ha sido empleada en la mayoría de los casos por los usuarios, ya que estos sistemas cuentan con un módulo de corrección de errores por parte de los usuarios; entonces, a medida que los usuarios de todo el mundo realizan correcciones el sistema es capaz de tomar esas sugerencias y realizar los cambios pertinentes en su código para no cometer el mismo error. Estos sistemas son empleados en la mayoría de los casos en lenguas que tiene una estructura gramática similar como es el caso del español y el inglés, entre otras.

Otro es el modelo de traducción automática basados en reglas, el cual se divide en dos: el basado en reglas por transferencia y el basado en reglas interlingua; el primero realiza un análisis de lo que se quiere traducir y reemplaza palabra por palabra por su equivalente más apropiado; el segundo utiliza una tercera lengua como punto de referencia que tenga una estructura similar tanto para la lengua origen como para la lengua destino, de esta forma las palabras u oraciones van pasando de una lengua origen a una lengua intermedia y finalmente llegan a la lengua destino.

Análisis de la estructura de las oraciones simples en español y kuna, estableciendo los requerimientos técnicos que se deben contemplar al momento de desarrollar un software traductor automático.

Como todo programa informático, se deben dar una serie de órdenes al computador en un lenguaje de programación para que este ejecute lo que el usuario o programador desea que éste haga. El principio en el cual fue basado este traductor fue trasladar el proceso que la mente realiza para traducir, a condiciones que el computador puede interpretar; para este caso fue utilizado el lenguaje de programación ActionScrip 3.0.

La primera de las reglas consiste en tomar las palabras que se digitan en la caja de texto y separarlas por medio de un método que detecta cada vez que se hace un espacio, entonces toma cada palabra separada por espacio y lo almacena en un arreglo; es así como se tiene lo siguiente: “mañana tengo partido de fútbol” = “mañana”, “tengo”, “partido”, “de”, “fútbol”, de esta forma el programa comienza a detectar cuáles fueron las palabras que se han escrito para traducir. Además de esto, previamente se ha condicionado al usuario, a que en la caja de texto no puede escribir mayúsculas, comas y puntos, todo esto porque pueden presentarse errores, debido a que aún el software no ha alcanzado su desarrollo total.

Posterior a esto, el software elimina las palabras que no tienen ninguna traducción a la lengua kuna, sin que estas puedan afectar la traducción, y traslada solamente las palabras que son necesarias en otro arreglo nuevo.

Construcción de la base de datos

Un factor que determina el éxito de un traductor es la confiabilidad de su base de datos y la efectividad del programa para buscar rápidamente las palabras en ella. Por esta razón, la base de datos de este software fue construida en XML, puesto que permite una fácil construcción y acceso a sus nodos; cada nodo de la base de datos cuenta con cuatro atributos importantes, la palabra en español, su traducción en kuna, el tipo de palabra, y elementos gramaticales.

El programa cuenta con más de 5000 nodos que tienen la misma estructura. De esta forma al acceder a un nodo se está accediendo a toda la información necesaria para su procesamiento posterior.

Búsqueda de las palabras de la lengua destino

Lo importante de la traducción automática es asegurarse de que el computador realice los pasos necesarios para que al momento de hacer el cambio de palabras sea lo más certero posible. Es por esto por lo que la base de datos que fue creada para este traductor automático kuna cuenta con unas 5.000 palabras aproximadamente, pero al tener una base de datos tan

grande, puede ocasionar que el programa tarde mucho tiempo en buscar las palabras y reemplazarlas.

Organización de las palabras en la estructura kuna

En vista de que realizar una organización teniendo en cuenta la palabra que va entrando a la función encargada de esto puede hacer que el programa tarde mucho más de lo que dura haciendo la búsqueda, se pensó en seguir la estructura de las oraciones simples en kuna que previamente se había estudiado: sustantivo –adjetivo – sustantivo – adverbio – verbo. Teniendo en cuenta esta estructura se creó un arreglo con el número de campos necesarios para alojar la cantidad de palabras requeridas y conformar esta estructura; entonces, al momento de que las palabras van llegando a la función encargada de organizar, ésta busca el tipo de palabra que es y la ubica en su posición.

De esta manera es que el software realiza la organización de las palabras, siguiendo la estructura planteada, recibiendo una palabra a la vez e inmediatamente organizándola en su posición, de acuerdo al tipo de palabra que le indique la base de datos. Al final, el programa arroja el resultado. Las demás palabras son eliminadas por el software debido a que en kuna no tienen traducción alguna de ellas. Sin embargo, el sentido de la oración simple es el mismo en español y en kuna.

Desarrollo del software traductor

Al finalizar el proceso de organización lo que queda por hacer es mostrar el resultado en otra caja de texto para que el usuario pueda ver el resultado de la traducción; este es una función sencilla que solamente se encarga de recorrer el arreglo donde ha sido almacenada la traducción e ir mostrando palabra por palabra.

Validación del software de traducción por expertos en la lengua, para la verificación de su confiabilidad en su uso.

En general, los traductores de lenguas tradicionales como el inglés o el francés, se encuentran de manera virtual (on line) o en procesadores de un menor tamaño. Estos traductores presentan cierta confiabilidad, pero hasta el momento no se tiene uno que sea cien por ciento confiable, ya que se presentan muchos problemas con la semántica y la sintaxis de las lenguas al realizar las traducciones. Para el caso de lenguas indígenas, son escasos los traductores de estas lenguas ya que muchas no tienen sistemas de escritura, o si lo tiene varían dependiendo de la comunidad o de las propuestas de los investigadores. Por eso este software se encuentra en la página del grupo LEXICÓN <http://www.lexicondecordoba.org/> para aquellas personas que deseen hacer aportes al traductor para su mejora.

Conclusiones

A partir de las necesidades de protección de lengua nativas promulgadas por la UNESCO, la Ley de Protección de Lenguas Indígenas de Colombia, y muchas otras leyes se plantea la necesidad de utilizar las TIC como un mecanismo para la conservación y preservación de las lenguas y culturas del mundo. La creación de traductores automáticos de lenguas indígenas y criollas permitirá hacer un aporte importante en este sentido, al acercarnos al mundo de estas lenguas y culturas a partir del uso de instrumentos informáticos para este propósito.

Referencias

Burg, Andrea (1982). *Ciencia y tecnología en el mundo*. p. 224. CONACYT.

CAPTAIN, David M. Compiler (2005). *Diccionario básico ilustrado; Wayuunaiki-Español, Español-wayuunaiki*. Recuperado de <http://www.sil.org/americas/colombia/pubs/abstract.asp?id=928474518949>

Diccionario bilingüe Koreguaje-Español Español Koreguaje. Recuperado de http://www.sil.org/americas/COLOMBIA/show_subject.asp?pubs=biblio&code=Dictionaries&Lang=spael

Freelang. Net . Recuperado de <http://es.freelang.net/enlinea/index.php#lista>

Instituto Nacional de Antropología e Historia (INAH). Recuperado de http://www.cronica.com.mx/nota.php?id_nota=172098

Ley de Protección de Lengua indígenas. Recuperado en <http://www.lenguasdecolombia.gov.co/content/ley-de-lenguas-nativas>

Llerena Villalobos, Rito (1987). *Relación y determinación en el predicado de la lengua kuna*. Lengua Aborígenes de Colombia. Descripciones número 1. Centro Colombiano de Estudios de Lenguas Aborígenes. Bogotá: Talleres del Centro de Publicaciones de la Universidad de los Andes.

López, V. (2002). Posibilidades y Realidades de la Traducción Automática. *La Linterna del Traductor* (ISSN1579-5314), 3.

MINCULTURA. (09 de Diciembre de 2009). Proyecto de Ley de Protección de Lenguas Indígenas. *Ley de Lenguas Indígenas* , 4. Bogotá: Colombia.

UNESCO. *Declaración Universal sobre Diversidad Cultural*. Recuperado en <http://unesdoc.unesco.org/images/0012/001271/127162s.pdf>