

TRADUCTOR AUTOMÁTICO EN LINEA DEL ESPAÑOL A QUECHUA, BASADO EN LA PLATAFORMA LIBRE Y CÓDIGO ABIERTO APERTIUM

AUTOMATIC TRANSLATOR IN LINE SPANISH A QUECHUA, BASED ON FREE AND OPEN SOURCE PLATFORM APERTIUM

Hugo David Calderón¹, Vilca César David Mamani Calderón², Flor Cagniy Cárdenas Mariño³ & Edwin Fredy Mamani Calderón⁴

¹Universidad Nacional Micaela Bastidas de Apurímac. C.E. hdcalderon@gmail.com

²Escuela Superior de Formación Artística de Puno. C.E. cesardavid23@gmail.com

³Universidad Nacional Micaela Bastidas de Apurímac. C.E. C.E.clavelyfcm@gmail.com

⁴Universidad Nacional del Altiplano Puno. C.E. mcedwin@gmail.com

RESUMEN

Apertium es una plataforma de traducción automática libre y de código abierto que ha sido creado inicialmente para traducciones entre lenguas emparentadas, sin embargo por su evolución permite crear pares de lenguas divergentes. El objetivo del estudio fue la comparación traductora entre una lengua nativa aglutinante e idioma flexivo. La medición de la calidad del traductor automático fue entre español y quechua mediante el método Word Error Rate. La implementación del traductor automático correspondió al Quechua del Este de Apurímac. El estudio se realizó durante el año 2013 en la Región Apurímac-Perú. Las etapas en la traducción correspondieron a: incubación del sistema traductor automático, creación del diccionario monolingüe quechua y reutilización del diccionario monolingüe español, creación del diccionario bilingüe y creación de las reglas de transferencia estructural. El resultado del sistema de traducción automática presenta, más de 4000 palabras raíces, 5000 traducciones de palabras raíces entre español y quechua, reglas de transferencia estructural de quechua a español y reglas de transferencia estructural de español a quechua implementadas. Finalmente, la calidad del traductor automático aplicando el método WER, presentó un promedio de error de calidad de traducción de quechua a español de 19,48 y calidad de traducción de español a quechua con error de 24,19.

Palabras claves: Apertium, español, quechua, software libre, traductor automático, traducción automática.

ABSTRACT

Apertium is a translation automatic platform free and open code that has been initially created for translations between related languages; however its evolution creates pairs from divergent language. The aim of the study was the compared translation between a binder translator native language and inflectional language. The quality measurement translator was between Spanish and Quechua using the Word Error Rate method. The implementation of automatic translation corresponded to East Apurimac Quechua. The study was conducted during 2013 in the Apurimac Region, Peru. The stages in the translation corresponded to: incubation of the machine translator, creation of Quechua monolingual dictionary and to reuse a Spanish monolingual dictionary, creation of a bilingual dictionary and creation of structural transfer rules system. The result of automatic translation system has more than 4,000 root words, 5000 translations of root words between Spanish and Quechua, structural transfer rules from Quechua to Spanish and structural transfer rules from Spanish to Quechua were implemented. Finally, the quality of the automatic translator using the WER method had an error average of 19.48 translating from Spanish to Quechua and 24.19 of average error translating from Spanish to Quechau. .

Keywords: Apertium, Spanish, Quechua, free software, automatic translation, machine translation.

INTRODUCCION

El contexto actual es un mundo globalizado de múltiples culturas e idiomas, donde los traductores automáticos como aplicación del procesamiento de lenguaje natural han aportado significativamente en la interacción de culturas permitiendo la traducción de textos o habla de un lenguaje natural a otro. Sin embargo, dichos avances como los traductores automáticos poco trascienden todavía en las lenguas minoritarias como el quechua, por lo que es un vacío la creación de sistemas de traducción automática que incluyan a las lenguas minoritarias. Por lo que en esta investigación se implementa el traductor automático entre español y quechua midiendo la calidad de traducción automática con el método WER (Word Error Rate).

El traductor automático entre español y quechua aportará a la educación, beneficiará a los estudiantes, profesionales e investigadores de lenguas, quienes tendrán la posibilidad de traducir en línea textos del idioma español a la lengua quechua y de la lengua quechua al idioma español. Así mismo con el traductor automático se acortará la brecha de comprensión de textos entre el par de lenguas.

(Armentano *et al.* , 2007) sostienen en su investigación sobre "Apertium, una plataforma de código abierto para el

desarrollo de sistemas de traducción automática", presenta que la plataforma Apertium ha sido desarrollada por el grupo de investigación Transducens de la Universidad de Alicante en el marco de varios proyectos de colaboración con universidades y empresas de España en los que, además de los programas que conforman el motor de traducción, se han confeccionado datos lingüísticos abiertos para la traducción automática catalán-español, gallego-español, portugués-español, francés-catalán, inglés-catalán y occitano-catalán. Tanto la plataforma en la que se integra el motor de traducción como los datos para estos pares de lenguas están disponibles para su descarga en su sitio oficial de Apertium.

(Rios , 2011), afirma que los métodos de corrección de ortografía desarrollada para idiomas como el inglés por lo general dependen de una lista completa de formas de las palabras completas, el requisito de que no pueden ser satisfechas por idiomas morfológicamente complejas. Como resultado describe la implementación de un corrector ortográfico con métodos de estados finitos para la lengua aglutinante quechua.

Asimismo (Rios & Göhring, 2009), describen las características de la lengua quechua y su anotación morfológica y sintáctica, y demuestran cómo se alinean la lengua quechua con el idioma español mediante su análisis morfológico.

Enero - Junio 2009

Por otro lado (Rios, Göhring & Volk , 2009), experimentaron alineamientos paralelos entre los idiomas Alemán, Español y la lengua quechua, midiendo la calidad de alineamiento y comparando los resultados, utiliza las herramientas GIZA++ de traducción automática basados en métodos estadísticos, llegando a los resultados que español y quechua no se alinean para este método de traducción, mientras que español y alemán tienen un mejor alineamiento. (Rataj , 2005), habla acerca la influencia del quechua en el español andino, donde detecta fenómenos ajenos al español general y peculiares del español andino, dice se dan tanto en el plano fonético e incluso fonológico, como en la morfología y sintaxis, más tarde Rataj implementó un traductor automática en un sola dirección quechua español, tomando quechua de Cusco (quz), sistema en construcción. Además, (Tyers *et. al.* , 2010), refieren que los recursos disponibles dentro de la plataforma que toman estados finitos para el análisis morfológico y la generación de palabras.

Quechua también denominada quichua es una familia de lenguas originaria de los Andes Centrales que se extiende por la parte occidental de Sudamérica. Es una macro lengua con una población hablante de más de 9'000,000 distribuidos en los países Perú, Argentina, Ecuador, Chile y Bolivia, es lengua co-oficial en Perú (SIL, 2013). Quechua como macro lengua se clasifica en 44 lenguas diferentes con

código propio de acuerdo al ISO 639-3 dada por SIL International, de los cuales 32 variantes de quechua se encuentran relacionadas con Perú, Quechua del Este de Apurímac (qve) es una de las variantes (SIL, 2013).

El idioma español es una lengua flexiva de tipo fusionante, hablado en diferentes países con más de 400'000,000 hablantes de acuerdo al SIL International, pertenece a las familias indoeuropeo, itálico e ibero-occidental, con código ES de acuerdo ISO 639-1 y SPA de acuerdo ISO 639-3 segunda lengua del mundo por el número de personas que tienen como lengua (Sil, 2013).

Las aplicaciones de procesamiento de lenguaje natural son: Síntesis del discurso, análisis del lenguaje, Comprensión del lenguaje, reconocimiento del habla, síntesis de voz, Generación de lenguajes naturales, traducción automática, recuperación de la información, dictado automático. Asimismo, existen múltiples aplicaciones el Procesamiento del Lenguaje Natural contempla elementos como: Análisis morfológico, análisis sintáctico, análisis semántico y análisis pragmático (Nils, 2004).

La lingüística computacional está considerada como una rama de la inteligencia artificial (IA), como todos los campos dentro de la IA, se ocupa de la investigación y sistematización de una capacidad cognitiva. En el caso de la

lingüística computacional, el objetivo central es la capacidad lingüística. Sin embargo, su preocupación no es necesariamente construir un modelo psicológicamente realista del comportamiento lingüístico humano; sino es identificar y caracterizar las clases de procesos y los tipos de conocimiento que están implicados en la habilidad de comunicar y asimilar información por medio del lenguaje natural, sin tomar en consideración su status psicológico. Una de las contribuciones de la lingüística computacional consiste en un conjunto de técnicas que capacitan al conocimiento lingüístico para guiar y constreñir el procesamiento lingüístico realizado por un sistema de procesamiento del lenguaje natural (Halvorsen, 1991).

La traducción automática, es una aplicación de procesamiento de lenguaje natural, también considerada como área de la lingüística computacional que investiga el uso de software para traducir texto o habla de un lenguaje natural a otro. El traductor automático debe analizar el texto original, interrelacionar con la situación referida y como resultado debe encontrar el texto correspondiente en el lenguaje destino (Rusell, 2004).

HFST (Helsinki Finite-State Transducer Toolkit) es software libre y de código abierto como herramienta para análisis morfológico, desarrollado por Xerox finite-state, incluye LEXC lenguaje

de alto nivel para especificar lexicones y TWOLC lenguaje de alto nivel para escribir reglas de fonología y morfologías (Kenneth, 2002).

Apertium es una plataforma de traducción automática de código abierto desarrollado por un grupo de investigadores de la Universidad de Alicante España, basado en reglas, cuya arquitectura usa transductores de estados finitos para el procesamiento léxico, modelos ocultos de Markov para la desambiguación léxica y procesamiento de patrones basado en estados finitos para la transferencia estructural. Actualmente esta plataforma de traducción automática por transferencia ha permitido implementar y poner en marcha a más de 35 pares de lenguas como sistemas de traducción automática (Armentano *et al.*, 2007).

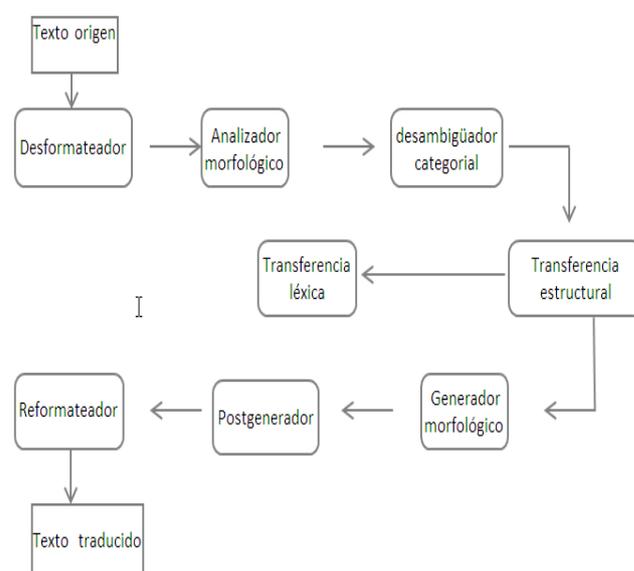


Figura 1. Módulos del sistema de traducción Automática de Apertium

Fuente: Armentano *et al* (2007).

Enero - Junio 2009

Descripción de los módulos del sistema de traducción automática de Apertium según Armentano *et al* (2007):

- *El desformateador.*- Encapsula las cadenas de texto en bloques de formato o superblancos poniendo delimitadores “[]” cada una de estas cadenas encapsuladas son tratadas como un blanco .
- *El analizador morfológico.*-Segmenta el texto en formas superficiales (FS) (las unidades léxicas tal como se presentan en los textos) y entrega para cada FS una o más formas léxicas (FL) consistentes en un lema (forma base usada en los diccionarios clásicos), la categoría léxica (nombre, verbo, preposición, entre otros) y la información de flexión morfológica (número, género, persona, tiempo, entre otros).Las unidades léxicas de más de una palabra (multipalabras) son tratadas como formas léxicas individuales y según su naturaleza.
- *El desambiguador léxico categorial.*- Está basado en modelos ocultos de Markov de primer orden, que representan categorías gramaticales y los observables son clases de ambigüedad, esto es, conjunto de categorías gramaticales, analiza una palabra ambigua de acuerdo con su contexto.
- *El módulo de transferencia léxica.*- Gestiona un diccionario bilingüe y es

invocado por el módulo de transferencia estructural, lee cada forma léxica (FL) en lengua origen (LO) y entrega la FL correspondiente en lengua meta (LM). El diccionario contiene un único equivalente para cada forma léxica de la LO.

- *El módulo de transferencia estructural.*-Detecta y trata patrones de palabras (sintagmas) que exigen un tratamiento especial por causa de las divergencias gramaticales entre las lenguas tales como: cambios de género, número, reordenamientos, cambios preposicionales etc. La transferencia estructural se aplica en sus tres niveles: chunker, interchunk y postchunk.
- *El generador morfológico.*- Genera a partir de la forma léxica en lengua meta una forma superficial flexionada adecuadamente. El resultado para la frase de ejemplo sería:

Los objetivos fueron: a) Diseño del sistema de traducción automática entre español - quechua y b) Evaluación de la calidad de traducción.

MATERIALES Y MÉTODOS

La investigación se ha realizado en la Universidad Nacional Micaela Bastidas de Apurímac en la Escuela Académico Profesional de Ingeniería Informática y Sistemas durante el año 2013. Se ha tomado para la investigación una de las variantes de

la macro lengua quechua la elegida es “Quechua del Este de Apurímac” cuyo código es “qve” de acuerdo al SIL Internacional.

La población para la lengua quechua son todas las palabras raíces. La muestra está conformada por 400 palabras raíces de la lengua Quechua e igual número de palabras raíces de la lengua Aymara las palabras son elegidas a criterio de los investigadores, entre nombres, adjetivos, verbos, interjecciones, pronombres, adverbios entre otras categorías gramaticales, las fueron insertadas a los diccionarios.

Procedimientos para el diseño del traductor automático

- Etapa de incubación del sistema de traducción automática.
- Etapa de Implementación de diccionarios monolingües(quechua y español)
- Etapa de implementación de diccionarios bilingüe
- Etapa de implementación de reglas de transferencia estructural
- Etapa de medición de la calidad del traductor automático
- Implantación del traductor automático en línea

El método de evaluación de la calidad de traducción automática

Para evaluación de la eficiencia del traductor automático de textos del idioma español a quechua y de quechua a español

se ha utilizado la métrica de evaluación de traductores automáticos WER (Word Error Rate):

$$WER = \frac{S + B + I}{N}$$

Dónde:

- S es el número de sustituciones,
- B es el número de borrados,
- I es el número de inserciones,
- N es el número de palabras que tiene la frase de referencia.

Infraestructura tecnológica utilizada

01 Oficina o ambiente de investigación y domicilio

01 Servidor de información implementado

06 computadoras personales de laboratorio

Servicios de Internet permanente

Software utilizado

- Sistema operativo Servidor Linux con servicios habilitados:ftp, ssh, domain, http, netbios, imap, micosoft-ds.
- Librerías instaladas: subversion, build-essential, g++, pkg-config, gawk, libxml2, libxml2-dev, libxml2-utils, xsltproc, flex, automake, autoconf, libtool y libpre3-dev.
- Apertium software de código abierto de Traducción Automática instalado: lttoolbox, apertium, apertium-lex-tools.
- Software especializado de análisis morfológico y software especializado de desambiguación instalado: OpenFST,

Enero - Junio 2009

Foma, Hilsinke Finite State Transducer,
y Constrain Grammar VISL3G3.

RESULTADOS Y DISCUSIÓN

Resultados del diseño del sistema de traducción automática entre español y quechua

- Compilando el diccionario morfológico español \$lt-comprlapertium-es-qve.es.dix es-qve.automorf.bin
- Dando formato con foma a apertium-es-qve.qve.lexc \$hfst-lexc -- formatfomaapertium-es-qve.qve.lexc qve.lexc.hfst
- Dando formato foma para apertium-es-qve.qve.twol \$hfst-twolc -- formatfomaapertium-es-qve.qve.twol -o qve.twol.hfst
- Composición lexc y twol \$hfst-compose-intersect -l qve.lexc.hfst -2 qve.twol.hfst -o qve.hfst
- Compilando diccionario morfológico quechua \$hfst-invertqve.hfst | hfstfst2fst -O -o qve-es.automorf.hfst
- Compilando diccionario bilingüe español-quechua \$lt-comprl apertium-es-qve.es-qve.dix es-qve.autobil.bin.
- Compilando diccionario bilingüe quechua-español \$lt-comprl apertium-es-qve.es-qve.dix qve-es.autobil.bin
- Compilando diccionario de auto-generación es-qve \$hfst-fst2fst -O qve.hfst -o es-qve.autogen.hfst
- Compilando diccionario de auto-generación qve-es \$lt-comprlapertium-es-qve.es.dix qve-es.autogen.bin
- Compilando reglas de desambiguación al diccionario monolingüe quechua \$csg-comp apertium-es-qve.es-qve.rlx es-qve.rlx.bin
- Compilando reglas de transferencia estructural de Español-Quechua nivel chunker \$apertium-preprocess-transfer apertium-es-qve.es-qve.t1x es-qve.t1x.bin
- Compilando reglas de transferencia estructural de Español-Quechua nivel interchunk \$apertium-preprocess-transfer apertium-es-qve.es-qve.t2x es-qve.t2x.bin
- Compilando reglas de transferencia estructural de Español-Quechua nivel postchunk \$apertium-preprocess-transfer apertium-es-qve.es-qve.t3x es-qve.t3x.bin
- Compilando reglas de transferencia estructural de Quechua-Español nivel chunker \$apertium-preprocess-transfer apertium-es-qve.qve-es.t1x qve-es.t1x.bin
- Compilando reglas de transferencia estructural de Quechua-Español nivel interchunk \$apertium-preprocess-transfer apertium-es-qve.qve-es.t2x qve-es.t2x.bin
- Compilando reglas de transferencia estructural de Quechua-Español nivel postchunker \$apertium-preprocess-

transfer apertium-es-qve.qve-es.t3x
qve-es.t3x.bin.

Según Tyers *et al* (2010), en su investigación presentan “Recursos de código abierto en la plataforma Apertium para la investigación y desarrollo de traductores automáticos”, describe los recursos disponibles dentro de la plataforma que toman estados finitos para el análisis morfológico y la generación de palabras. En investigación confirma la utilización de recursos disponibles de la plataforma Apertium, ya que se ha tomado el diccionario monolingüe español del par de sistema de traducción automática entre español e inglés, así mismo las reglas de desambiguación son genéricas en las lenguas, de manera que los recursos de código abierto toman su mayor aporte para cualquier par de lenguas, siempre en cuando sean cuidadosamente utilizadas para incubar un nuevo sistema de traducción automática basada en la plataforma Apertium.

Por otro lado Rataj (2005), habla acerca la influencia del quechua en el español andino, donde detecta fenómenos ajenos al español general y peculiares del español andino, dice se dan tanto en el plano fonético e incluso fonológico, como en la morfología y sintaxis, más tarde Rataj implementó un traductor automática en una sola dirección quechua español, tomando quechua de Cusco (quz), sistema en construcción. Tomando la implementación de Rataj (2005) traductor automático en una sola dirección de quechua a español, su lógica de implementación también funciona en la dirección traducción automática de español a quechua de manera que en este experimento se ha incubado compilando en ambas direcciones, esto se logra implementando los diccionarios monolingües para cada par y utilizando cada diccionario para analizar morfológicamente cada palabra y además utilizando el mismo diccionario para generar cada palabra cuando viene en dirección contraria.

Cuadro 1. Estructura y contenido de los diccionarios morfológicos.

Diccionariomonolingüe QVEapertium-es-qve.lexc	Significado en español
%<n%>	! nombre
%<adj%>	! adjetivo
%<prnp%>	! pronombre
%<adv%>	! advervio
%<vblex%>	! verbo
%<m%>	! masculino
%<sg%>	! singular
%<pl%>	! plural
Otros	...
Nombres ;	!Nombres
Adjetivos ;	!Adjetivos

TRADUCTOR AUTOMÁTICO EN LINEA DEL ESPAÑOL A QUECHUA

Enero - Junio 2009

PrnPersonales ; Adverbios ; Verbos ; Otros	!PronombresP !Adverbios !Verbos ...
Lexicon Plural %<pl%>:%>kuna N-FLEX-Incl ; LEXICON Caso %<acc%>:%>ta %<abl%>:%>manta LEXICON Posv %<px 1sg%>	! plural s ! acusativo a ! ablativo (desde, de) ! posesivo 1ra persn singular
<u>Lexicon Nombres</u> wasi:wasi N ; t%'anta:t%'anta N ; <u>Lexicon Adjetivos</u> hatun:hatun ADJ; musuq:musuq ADJ; <u>LexiconPrnPerson</u> nuqa:nuqa PRNP; <u>Lexicon Adverbio</u> may:may ADV; <u>Lexicon verbos</u> mikhuy:mikhu V; munay:muna V; otros	Significado !casa !pan !grande !nuevo !yo !dónde !comer !querer ...

Corroborando con el estudio “Corrector ortográfico una lengua aglutinante: Quechua” presentada por Rios (2011), fundamenta que los métodos de corrección de ortografía desarrolladas para idiomas como el Inglés por lo general dependen de una lista completa de formas de las palabras completas, el requisito de que no pueden ser satisfechas por idiomas morfológicamente complejas. En cambio en esta investigación no siendo un corrector ortográfico pero si un traductor automático que también trabaja con un analizador morfológico se experimenta que funciona correctamente el analizador con pocas palabras insertadas en el diccionario, no necesariamente completa, se puede ir agregando las palabras en el

diccionario monolingüe para enriquecer el sistema, está claro también que solamente se analizarán las palabras insertadas.

De la misma forma que Tyers *et al* (2010), “Recursos de código abierto en la plataforma Apertium para la investigación y desarrollo de traductores automáticos”, que describe los recursos disponibles dentro de la plataforma que toman estados finitos para el análisis morfológico y la generación de palabras. En tanto en esta investigación confirma la utilización de recursos disponibles de la plataforma Apertium, ya que se ha tomado el diccionario monolingüe español del par de sistema de traducción automática entre español e inglés, así

mismo se ha tomado del par de sistema de traducción automática entre español e inglés las reglas de desambiguación de la parte de español, confirmando de esta manera que los recursos de código abierto toman su mayor aporte para cualquier par de lenguas, siempre en cuando sean cuidadosamente utilizadas para incubar un nuevo sistema de traducción automática basada en la plataforma Apertium.

Resultados de la medición de la calidad de traducción automática español y quechua

Siendo WER (Word Error Rate) el método utilizando para la medición de traductores automáticos, se ha aplicado en la traducción

automática entre español y quechua. La fórmula recibe como entrada una cadena de texto a traducir la cual está especificada en la columna (TEXTO A TRADUCIR), como resultado de la traducción automática se tiene la columna (TEXTO TRADUCIDO) marcada con * las palabras que no fueron reconocidas por el sistema, además la métrica WER requiere el texto referencia denominada traducción correcta esto es corroborado por el humano, esta información se muestra en la columna (traducido por el humano), aplicando la métrica WER se tiene los siguientes resultados:

Cuadro 2. Resultado de la medición del traductor automático Quechua a Español con la métrica WER.

No	Texto a traducir	Traductor automático	Traducido por el humano	WER
1	pukawasinchismanrisun	a nuestra casa roja iremos	iremos a nuestra casa roja	0,00
2	sumaqalquchachinkamusqawasiykimanta	Perro lindo perdió de tu casa	un perro lindo se perdió de tu casa	33,50
3	mihusun chiri lawata	comeremos *a sopa fría	comeremos una sopa fría	25,00
4	chirimantawañusunkunanp'unchaw	de frío moriremos hoy día	de frío moriremos hoy día	0,00
5	yuraqmankaykitaapamusaq	a tu olla blanca traeré	traeré tu olla blanca	25,00
6	MaríawanJaimewanmihushankulawata	María y Jaime comen a sopa	María y Jaime comen una sopa	16,67
7	apasunrumitawasiykipirqanapaq	llevaremos piedra tu casa construir muro	llevaremos piedra a tu casa para construir muro	25,00
8	mamaywatukuyrirqaniAndahuaylasta	mi madre visitar fui a Andahuaylas	fui a visitar a mi madre a Andahuaylas	25,00

TRADUCTOR AUTOMÁTICO EN LINEA DEL ESPAÑOL A QUECHUA

Enero - Junio 2009

9	Pedrowan Juliawanripusqa Limamanku nanp'unchaw	Pedro y Julia ido a Lima hoy día	Pedro y Julia han ido a Lima hoy día	11,11
10	wawqiykuna Estados Unidos watukuyhamushan	mis hermanos Estados Unidos visitar viene	mis hermanos vienen de Estados Unidos a visitar	33,50
Promedio				19,48

Cuadro 3. Resultado de la medición del traductor automático español a quechua con la métrica WER.

No	Texto a traducir	Traductor automático	Traducido por el humano	WER %
1	yo quiero comprar una casa para mi hijo	ñuqamunanirantiy hukwasi *para *mío churi	ñuqawasimunan irantiyhukchuri	33,33
2	la casa blanca está bonita	*La wasiyuraqkashans umaq	yuraqwasisuma qkashan	25,00
3	esas cinco piedras son para ti	Chay pisqarumikunakan kuqanpaq	chay pisqarumikunaq anpaqkanku	20,00
4	quiero mucho azúcar para mí café	munaniaskhamisk'i ñuqapaq *café	askhamisk'imun aniñuqapaq *café	20,00
5	nosotras comemos pan todo el día	ÑuqanchisMikhuy kut'antap'unchawn intin	Ñuqanchist'anta tap'unchawninti nMikhuyku	25,00
6	ellos dan pasto al cuy	paykunaqunkuway lla *el quwi	paykunaquwiw ayllaqunku	25,00
7	quiero tomar agua con azúcar hoy	munaniupiyunu *con misk'ikunan	kunanunumisk'i upiymunani	20,00
8	hagan pan para comer hoy día	ruwankichist'anta *para mikhuykunap'unchay	t'antaruwankich iskunanp'uncha ymikhuy	20,00
9	la silla amarilla está rota	*La tiyanaq'illukashan p'akiy	q'illutiyanap'aki ykashan	25,00
10	Él está comiendo mucho, así engordará demasiado	PayKashanMikhus paaskha, aknawiranqanishi wta	Pay askhaMikhuspa Kashan, aknanishiwtawiranga	28,57
Promedio				24,19

En el cuadro 2 se tiene el promedio de error (WER) 19.48%, esto indica que de los

100% de traducciones que se haga de quechua a español, el 19.48% más o menos

traducciones se deben corregir puesto que es el índice de error en la traducción automática. De la misma manera en el cuadro 3 se tiene el promedio de error (WER) 24.19%, es decir que de los 100% traducciones que se haga de español a quechua, el 24.19% más o menos traducciones se deben corregir teniendo en cuenta que es el índice de error de este método.

Contrastando con (APERTIUM, 2013) se puede observar en el cuadro 4 su aplicación con el mismo método WER, el porcentaje de error es similar, debiendo que los traductores automáticos basados en el modelo de transferencia evolucionan insertando precisamente las correcciones conociendo el índice de error.

Cuadro 4. Resultados de la medición de la traducción automática de diferentes pares implementadas a base de Apertium.

Translator	Date	Version	Direction	Unknown words	WER
apertium-eo-fr	11th February 2011		fr → eo	Yes	22.4 %
			eo → fr		-
			en → mk		-
apertium-mk-bg	31st August 2010	0.1.0	mk → bg	Yes	26.67 %
			bg → mk		-
apertium-nn-nb	12th October 2009	0.6.1	nn → nb	Yes	-
			nb → nn		32.5%, 17.7%
			fr → br		-
apertium-sv-da	12th October 2009	0.5.0	sv → da	Yes	30.3 %
			da → sv		-
apertium-eu-es	2nd September 2009		eu → es	Unknown	72.4 %
			es → eu		-
apertium-cy-en	2nd January 2009		cy → en	Unknown	55.7 %
			en → cy		-
apertium-eo-en	8th May 2009	0.9.0	en → eo	Unknown	21.0 %
			eo → en		-
apertium-es-pt	15th May 2006		es → pt	Unknown	4.7 %
			pt → es		11.3 %
apertium-oc-ca	10th May 2006		oc → ca	Unknown	9.6 %
			ca → oc		-
apertium-pt-ca	28th July 2008		pt → ca	Unknown	16.6%
			ca → pt		14.1%
apertium-en-es	May 2009		en → es	Unknown	-
			es → en		-

Enero - Junio 2009

En cuanto a (Rios *et. Al.*, 2009), experimentaron alineamientos paralelos entre los idiomas Alemán, Español y la lengua quechua, midiendo la calidad de alineamiento y comparando los resultados, utiliza las herramientas GIZA++ de traducción automática basados en métodos estadísticos, llegando a los resultados que español y quechua no se alinean para este método de traducción, mientras que español y alemán se alinean mejor. Por tanto esta investigación apoya a la alternativa de traducción automática por transferencia además que la investigación está basada en la plataforma de código abierto y libre Apertium siendo su arquitectura utilizando el método de traducción automática por transferencia, finalmente decir que la traducción automática por transferencia sería la alternativa más cercana e eficiente para ser utilizado para las lenguas minoritarias.

Contrastando con (Armentano *et al.*, 2007) que sostienen en su investigación sobre "Apertium, una plataforma de código abierto para el desarrollo de sistemas de traducción automática", presenta que la plataforma Apertium ha sido desarrollada por el grupo de investigación Transducens de la Universidad de Alicante en el marco de varios proyectos de colaboración con universidades y empresas de España en los que, además de los programas que conforman el motor de traducción, se han confeccionado datos lingüísticos abiertos para la traducción automática catalán-

español, gallego-español, portugués-español, francés-catalán, inglés-catalán y occitano-catalán. Tanto la plataforma en la que se integra el motor de traducción como los datos para estos pares de lenguas están disponibles para su descarga en su sitio oficial de Apertium (<http://www.apertium.org>).

Esta investigación demuestra la traducción entre español y quechua, siendo la lengua quechua de la familia de aglutinantes que expresan conceptos y relaciones gramaticales mediante la adición de sufijos, sin embargo el idioma español una lengua flexiva de tipo fusional es considerado diferente a las lenguas andinos, por lo que en contraste esta plataforma si se puede usar entre lenguas bien diferenciadas no solamente entre pares emparentadas, pues el avance y la madurez de esta plataforma permite crear sistemas de traducción automática entre pares de lenguas divergentes como dice el mismo autor.

(Rios & Göhring, 2009), en su investigación describen las características de la lengua quechua y su anotación morfológica y sintáctica, y demuestran cómo se alinean la lengua quechua con el idioma español mediante su análisis morfológico. La investigación anterior por ser demostrada su alineamiento entre español y quechua, se confirma mediante esta investigación ya que los resultados de la traducción medida por el método WER son aceptables.

Resultados verificables de la investigación

- El sistema traductor automático del idioma español a la lengua quechua y de la lengua quechua al idioma español en su versión beta se ha puesto en funcionamiento en el sitio <http://www.lenguasandinas.org>, realizado por los autores de esta investigación.
- Diccionario monolingüe de quechua implementado con más de 4,000 palabras raíces
- Diccionario monolingüe español adaptada de otro sistema de traducción automática
- Diccionario bilingüe español y quechua implementada con más de 5,000 traducciones de palabras raíces
- Reglas de transferencia estructural de quechua a español y reglas de transferencia estructural de español a quechua implementadas.
- Sistema traductor automático evaluado con el método WEB (Word Error Rate), teniendo la tasa de error aceptable en la comprensión de textos.

Pruebas de entrada y salida del sistema de traducción automática entre español y quechua

Análisis morfológico de la palabra “wasiykuna”

Entrada: \$echo "wasiykuna" | hfst-proc -x qve-es.automorf.hfst

Traducción: Casas

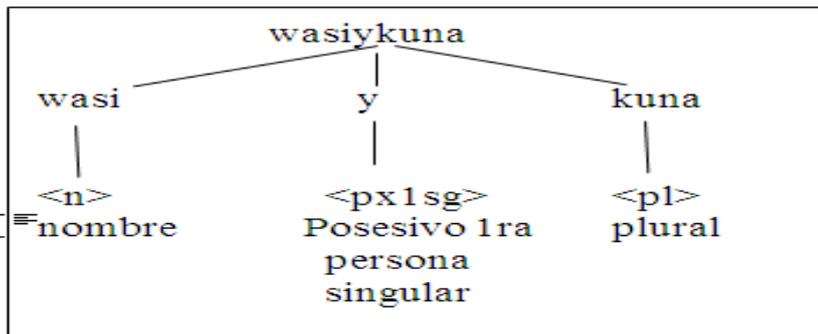


Figura 2 Resultado de análisis morfológico de la palabra wasiykuna

Análisis morfológico de la palabra “maypiraq”

Entrada: \$echo "maypiraq" | hfst-proc -x qve-es.automorf.hfst

Traducción: Donde estará

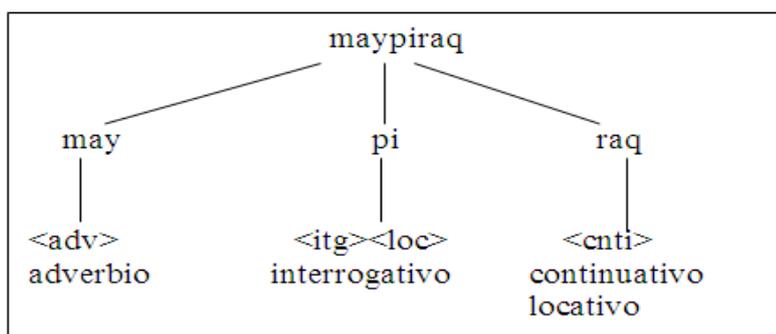


Figura 3. Resultado de análisis morfológico de la palabra maypiraq

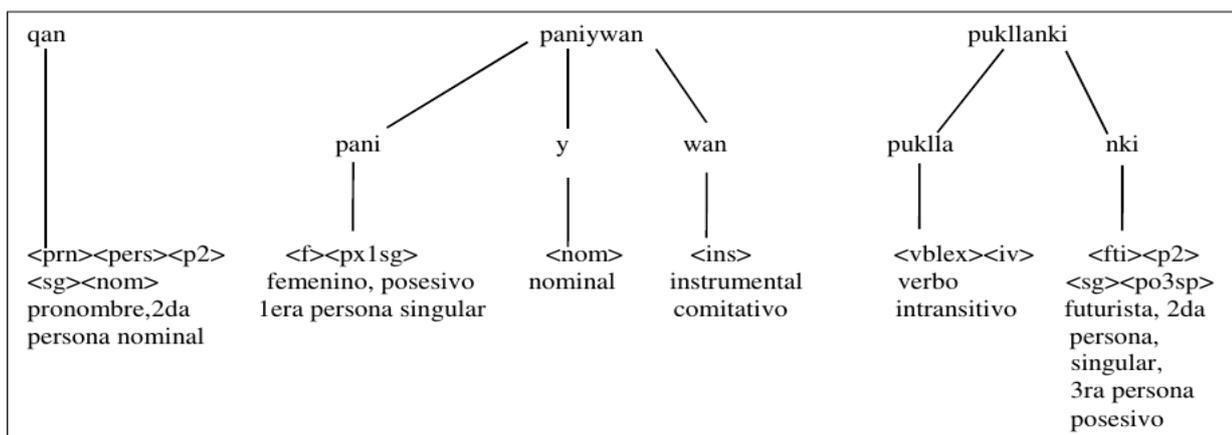


Figura 4. Análisis morfológico y traducción de la frase “qanpaniywanpukllanki”.

Entrada: # echo "qanpaniywanpukllanki" | hfst-proc -x qve-es.automorf.hfst

Traducción: Tú vas a jugar con mi hermana

Análisis morfológico y traducción de la frase "nuqaqilqaytayachani"

Entrada: \$echo "nuqaqilqaytayachani" | hfst-proc -x qve-es.automorf.hfst

Traducción: Yo se escribir

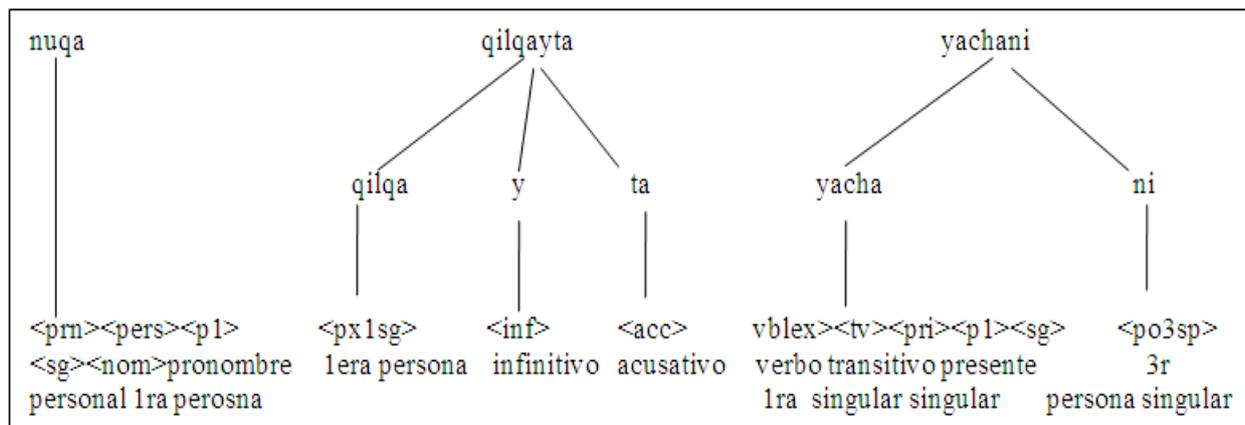


Figura 5. Resultado de análisis de la frase “nuqayachaniqilqayta”.

Resultados de la Traducción de Quechua a Español

Cuadro 5. Resultados de la traducción de frases de quechua a español.

<ul style="list-style-type: none">• Ingresando la palabra "wasi" al sistema para su traducción en la dirección de que-es quechua a español <u>Entrada:</u> \$echo "wasi" apertium -d .qve-es <u>Salida:</u> → casa.• Teniendo la palabra raíz "wasi" se agrega el sufijo "yki" <u>Entrada:</u> \$echo "wasiyki" apertium -d .qve-es <u>Salida:</u> → tu casa.• Aglutinando mas sufijos en quechua "wasi+yki+man" <u>Entrada:</u> \$echo "wasiykiman" apertium -d .qve-es <u>Salida:</u> → a tu casa.• Ingresando sufijos sobre sufijo "wasi+yki+manta" <u>Entrada:</u> \$echo "wasiykimanta" apertium -d .qve-es <u>Salida:</u> → de tu casa.• Sucesivamente quechua puede tener mas grande las palabras aglutinadas "wasi+yki+kuna+manta" <u>Entrada:</u> \$echo "wasiyikunamanta" apertium -d .qve-es <u>Salida:</u> → de tus casas.• Ingresando una frase adjetivo+nombre+verbo para la traducción <u>Entrada:</u> \$echo "sumaqalquchachinkamusqawasiykimanta" apertium -d .qve-es <u>Salida:</u> → un perro lindo se perdió de tu casa• Ingresando una frase verbal <u>Entrada:</u> \$echo "mamaywatukuyrirqaniAndahuaylasta" apertium -d .qve-es <u>Salida:</u> fui a visitar a mi madre a Andahuaylas
--

Fuente: Elaboración propia.

Traducción de Español Quechua

Cuadro 6. Resultados de la traducción de frases de español a quechua.

<ul style="list-style-type: none">• De forma analógica en la dirección de español-quechua, se ingresa la palabra "casa" <u>Entrada:</u> \$echo "casa" apertium -d .es-qve <u>Salida:</u> →wasi.

Enero - Junio 2009

- En este caso se ingresa la frase corta "esta casa"
Entrada: \$echo "esta casa" | apertium -d .es-qve
Salida: →kaywasi.
- Otra frase desde español a quechua
Entrada: \$echo "esta casa nueva" | apertium -d .es-qve
Salida: → kaywasimusuq.
- Traducción de una frase verbal
Entrada: \$echo "nosotras comemos pan todo el día" | apertium -d . es-qve
Salida: →Ñuqanchist'antatap'unchawnintinMikhuyku
- Traducción de frases de dos frases cortas
Entrada: \$echo "Él está comiendo mucho, así engordará demasiado" | apertium -d . es-qve
Salida: → PayaskhaMikhuspaKashan, aknanishiwawiranqa

Fuente: Elaboración propia.

CONCLUSIONES

Se ha incubado el sistema de traducción automática entre español y quechua haciendo pruebas es funcional el sistema compilado con sus diccionarios respectivamente.

Diccionarios morfológicos del idioma español fue reutilizado de otro par de lenguas y el diccionario morfológico de la lengua quechua fue implementada, así mismo se ha implementado el diccionario bilingüe del par español y quechua, de la misma manera se han definido reglas de transferencia para la traducción de idioma español a la lengua quechua.

Finalmente se ha evaluado la calidad del Traductor Automático con la métrica WER, el promedio de error en la traducción de español a quechua es 19.48% siendo esta aceptable la comprensión de textos

traducidos por el traductor automático; el promedio de error en la traducción de quechua a español es 19.48% siendo aceptable la comprensión de textos traducido por el traductor automático.

El sistema Traductor Automático Español-Quechua basado en Apertium es funcional y se pone al servicio de la comunidad, además es software libre.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia Tecnología e Innovación Tecnológica del Perú, por haber co-financiado la investigación en beneficio de la Región Apurímac.

Especial agradecimiento a Vlastimil Rataj, por su apoyo constante en la implementación del Traductor Automático, quién ha escrito el LEXC de quechua Cusco.

REFERENCIAS BIBLIOGRÁFICAS

- APERTIUM, Plataforma libre de Traducción. 2013, recuperado el 30 de diciembre de 2013, de SIL, Summer Institute of Linguistics 2013, ISO 639 Code Tables, recuperado el 20 de mayo de 2013, de <http://www.sil.org>.
- Armentano C., Corbí A., Forcada M., Ginestí M., Montava. & Ortiz M. 2007. Una Plataforma de Código Abierto Para el Desarrollo de Sistemas de Traducción Automática, Departamento de Lingüística y Sistemas Informáticos de la Universidad de Alicante.
- Breña R. 2003. Autómatas y Lenguajes. Tecnológico de Monterrey, Campus Monterrey. Editorial McGraw-Hill México. ISBN: 9781456210779
Url:
[<http://homepages.mty.itesm.mx/rbr/ena/AyL.html>].
- Cerrón P. 1987. Lingüística Quechua – Centro de Estudios Rurales Andinos Bartolomé de las Casas Noviembre.
- Forcada M., Boyan V. & Ortiz S. 2012. Documentación del sistema de código abierto OpentradApertium de Traducción Automática de transferencia sintáctica superficial. Departament de Llenguatges i Sistemes Informàtics Universitat d'Alacant.
- Gonzales J. 2009. Aprendizaje de Transductores Estocásticos de Estados Finitos y su Aplicación en Traducción Automática. Tesis Doctoral. Universidad Politécnica Valencia.
- Halvorsen K. 1991. Las Aplicaciones Informáticas de la Teoría Lingüística Traducción J. Gómez Guinovart y A. Tusón Valls. Edición Eguren.
- Kenneth B. & Karttunen L. 2002. Finite-State Morphology Xerox Tools and Techniques.
- Louden C. 2004. Construcción de Compiladores Principios y Practica. Thomson.
- Mooney A. & Raymond J. 2003. Fundamentals, Parte I Caps. II, III, IV, V. Oxford Handbook of Computacional Linguistics. Oxford University Pres. (Ruslan Mitkow Ed.).
- Manteca C. 1987. Lingüística General. Madrid, Cátedra.
- Moreno A. 2009. Estudios de Lingüística Española: Diseño e Implementación de un Lexicón Computacional para Lexicografía y Traducción

Enero - Junio 2009

- Automática. Facultad de Filosofía y Letras Universidad de Málaga. Volumen 9, Url: [http://elies.rediris.es/elies9/index.htm]. ISSN: 1139-8736.
- Nida A. 1949. Morphology. The Descriptive Analysis of Words, Michigan Ann Arbor, University of Michigan Press.
- Nils N. 2004. Inteligencia Artificial. Madrid. McGraw Hill/Interamericana S.A. Masarykova Univerzita Filozofická Faculta Latedra Romanistiky.
- Rataj V. 2005. La Influencia del Quechua en el Español Andino.
- Rich E. & Knight K. 1994. Inteligencia Artificial. Segunda Edición. España. McGraw-Hill/Interamericana S.A.
- Ríos A. & Volk M. 2012. Parallel Treebanking Spanish-Quechua, Linguistic Issues in Language Technology – LiLT.
- Ríos A. & Volk M. 2009. A Quechua-Spanish Parallel Treebank. University of Zurich – Zurich Open Repository and Archive.
- Ríos A. 2011. Spell Checking an Agglutinative Language: Quechua. University of Zurich. Zurich Open Repository and Archive.
- Rusell S. & Norvig M. 2004. Inteligencia Artificial un enfoque moderno. Segunda Edición. Madrid. Pearson Educación S.A.
- SIL Summer Institute of Linguistics 2013, ISO 639 Code Tables, recuperado el 20 de mayo de 2013, de http://www.sil.org.
- Tyers F., Sánchez F., Ortiz S. & Forcada M. 2010. Recursos de código abierto en la plataforma Apertium para la investigación y desarrollo de traductores automáticos.
- Torero A. 1983. La familia lingüística quechua. América Latina en sus lenguas indígenas. Caracas: Monte Ávila. ISBN 9233019268.
- Torero A. 1974. El quechua y la historia social andina. Lima: Universidad Ricardo Palma, Dirección Universitaria de Investigación. ISBN 9786034502109.
- UANCV, Universidad Andina Néstor Cáceres Velásquez. 2009. Morfología Contrastiva Quechua/Aymara/Castellano. Escuela de Postgrado de la, Segunda Especialización en Educación Bilingüe Intercultural.,