

---

# Traductor morfológico del castellano y quechua

Indhira Castro Cavero  
Jaime Farfán Madariaga

---

## RESUMEN

El presente proyecto se orientó a desarrollar un sistema automático de traducción inteligente entre los idiomas castellano y quechua, el cual nace a raíz del Proyecto de Titulación presentado por Indhira Mayra Castro Cavero, egresada de la especialidad de Redes y Comunicación de Datos – Promoción 2004 – II (becaria de la Unión Europea).

Este sistema de traducción supone la automatización del proceso de traducción de palabras del castellano al quechua y viceversa, basado en la transferencia, que opera en tres fases (análisis, transferencia y generación) usando representaciones morfológicas para las palabras. Al traducir una palabra, el sistema no sólo devuelve la palabra convertida al otro idioma, sino también, muestra información lingüística de los componentes de la palabra.

Es importante destacar que para el desarrollo del sistema se usó herramientas tecnológicas de código abierto como es Java, MySQL y Apache. Así también este trabajo cumple un eminente rol informativo, que trae como consecuencia el enriquecimiento lingüístico del idioma quechua, que resulta más explicativo que un simple inventario de términos; y, sobre todo, la difusión de una lengua que se debe conocer y desarrollar. En la actualidad, el quechua es lengua materna de un 17% de la población en el Perú, lo que representa el 53% de los quechua hablantes en Sudamérica.

La red mundial de la Web representa, hoy en día, el modo de comunicación más utilizado, lo cual permitirá publicar la aplicación para que pueda ser fácilmente accedida. Esto indudablemente abre la posibilidad de incrementar el material lingüístico disponible para el aprendizaje del idioma y es, además, una buena herramienta para llegar a un grupo más amplio de individuos. El sistema puede ser utilizado desde cualquier parte y por cualquier persona.

## INTRODUCCIÓN

Durante los últimos años, la informática ha revolucionado la mayoría de las actividades profesionales y la traduc-

ción no es una excepción. La traducción automática (TA) es una disciplina que ha contribuido de manera determinante al desarrollo de la tecnología lingüística. Existe actualmente un gran abanico de herramientas de ayuda a la traducción: programas de interpretación asistida por ordenador, herramientas de gestión de terminología, sistemas de traducción automática, etc.

Si bien estas aplicaciones son empleadas habitualmente por un gran número de usuarios en todo el mundo, en la actualidad, no existe un traductor automático morfológico que utilice el idioma quechua, a pesar de ser este el mayor idioma indígena de América Latina, con cerca de 10 millones de usuarios.

Asimismo, en años recientes, y como consecuencia de la globalización, se percibe un creciente interés por el rescate, el mantenimiento y la revaloración de nuestras lenguas nativas, especialmente el quechua. A pesar de que esta lengua junto al aymara y el castellano, es lengua oficial en el Perú y Bolivia, ello sólo queda “en el papel”, pues aún no se da al quechua el estatuto oficial que por ley, merece; más aún, si existe la gran necesidad de conocimiento y aprendizaje del quechua para la comunicación con personas que tienen este idioma como lengua nativa.

Desde esta perspectiva este trabajo busca ser un aporte para nuestra sociedad a través de la difusión del idioma quechua. Es muy probable ahorrar tiempo y esfuerzo empleando este tipo de sistemas si tenemos que traducir del castellano al quechua, o viceversa, un informe técnico o administrativo. La traducción automática es hoy en día tema de trabajo de investigación de lingüistas, ingenieros y científicos de la información. Queda entonces, en nosotros profesionales, profundizar acerca del tema y más aún, si se trata del idioma quechua.

Este documento presenta los alcances más importantes del proyecto. En primer lugar se fundamentan los moti-

vos que llevaron al desarrollo del sistema traductor para el idioma quechua. A continuación se exponen los factores o dimensiones que condicionan la viabilidad del desarrollo del sistema. Luego, se presentan las principales características del idioma, como también su estructuración lingüística. De esta manera se podrá comprobar cómo se ha adecuado una metodología, que permite desarrollar un motor automático que traduce cada componente de una palabra de una lengua a la otra. Finalmente, se analiza con algo más de detalle las principales funcionalidades del sistema, poniendo un énfasis especial en cada una de sus facilidades como en sus limitaciones.

## FUNDAMENTACIÓN

- Existe gran necesidad de conocimiento y aprendizaje del quechua para la comunicación con personas que tienen el quechua como idioma nativo.
- Las fuerzas armadas y policiales, jueces, periodistas, médicos y otros profesionales tienen dificultades para dirigirse y comprender las necesidades y propuestas de la población quechua-hablante; que merece ser atendida en su propia lengua, por respeto a la diversidad cultural que existe en nuestro país.



- Un tercio de los niños indígenas peruanos abandonan la escuela, principalmente porque la educación no se adapta a su lengua materna. ([http://www.servindi.org/sp/noticias/Noticias\\_2005/actualidad\\_74.htm](http://www.servindi.org/sp/noticias/Noticias_2005/actualidad_74.htm))
- Carencia de traductores: Actualmente no se cuenta con un sistema de traducción para el quechua, solamente existen diccionarios y glosarios bilingües.
- La necesidad de una nueva herramienta que resulte más explicativa que un simple inventario de términos quechuas.
- Incremento del material a traducir.
- La comisión europea ha puesto un énfasis especial en el estímulo a las llamadas "industrias de la lengua" y a la tecnología necesaria para su desarrollo. ([http://www.foreignword.com/es/Technology/art/Abaitua/Abaitua\\_3.htm#sociedad](http://www.foreignword.com/es/Technology/art/Abaitua/Abaitua_3.htm#sociedad))

## ANTECEDENTES

- Actualmente los proyectos y programas educativos están trabajando en el rescate y difusión de la práctica de la lengua quechua.

- Programa de alfabetización (<http://alfa.minedu.gob.pe/alfa/>).
- Programa Huascarán (<http://www.huascarán.edu.pe/>).
- Escuela del Aire (<http://www.minedu.gob.pe/ocder/escueladelaire/>).
- DENEBI (Dirección Nacional de Educación Bilingüe Intercultural). ([http://www.minedu.gob.pe/gestion\\_pedagogica/dir\\_edubilingue/presentacion\\_bilingue.htm](http://www.minedu.gob.pe/gestion_pedagogica/dir_edubilingue/presentacion_bilingue.htm)).
- Proyecto Microsoft: Traducción del Windows XP y Office 2003 al quechua. (<http://www.microsoft.com/latam/prensa/2004/noviembre/quechua.asp>).
- Quechua Network: Diccionario quechua. (<http://www.quechuanetwork.org/dictionary.cfm?lang=s>).
- Intervida (Asoc. Solaris Perú): Proyecto "Escuela Andina Intercultural" en el que participan 775 niños de 13 escuelas andinas. (<http://www.intervida.org/>).

## LOS SISTEMAS DE TRADUCCIÓN



Existen dos tipos de sistemas de traducción:

- **Manual:** Uso de medios físicos (diccionarios y traductores humanos).
- **Automática:** Sistemas informáticos.
  - **La TA sin participación humana:** es la realizada por un computador.
  - **La TA con participación humana:** el programa es el que lleva a cabo la traducción, pero el traductor interviene cuando es necesario.

### Traducción automática (TA)

Son sistemas informáticos que llevan a cabo traducciones de una lengua a otra con o sin intervención humana. Podemos clasificar los sistemas de TA atendiendo a diferentes criterios:

- Número de lenguas: sistemas bilingües y sistemas multilingües.
- Dirección de la traducción: unidireccionales y bidireccionales.

- Enfoque: sistemas directos, sistemas de transferencia y sistemas de interlingua.

## Traducción quechua

A través de la web podemos encontrar actualmente diferentes herramientas que ayudan a la traducción, pero no existe un traductor automático morfológico quechua:

- Diccionarios on-line.  
(<http://www.quechuanetwork.org/dictionary.cfm?lang=s>)
- Glosarios terminológicos.  
(<http://www.folkloredelnorte.com.ar/diccionario.htm#j>)  
([http://www.serindigena.cl/territorios/recursos/biblioteca/diccionarios/esp\\_a\\_quechu/](http://www.serindigena.cl/territorios/recursos/biblioteca/diccionarios/esp_a_quechu/))  
(<http://www.prodiversitas.bioetica.org/quecl-o.htm>)
- Diccionarios en la Web (diccionarios bilingües): sistemas que hacen uso de una base de datos, en donde almacenan las palabras asociadas a su respectiva traducción.
  - Se ingresa la palabra.
  - El sistema consulta a la base de datos por su equivalente.
  - El sistema muestra la traducción.

## Traducción automática de palabras

- Traductores morfológicos: estos tienen tres componentes básicos:
  - El diccionario, donde se almacena el contenido.
  - El motor morfológico, que permite aceptar palabras derivadas.
  - La interfaz gráfica del usuario, mediante la cual este interactúa con el diccionario.

## Necesidades básicas de un sistema de traducción

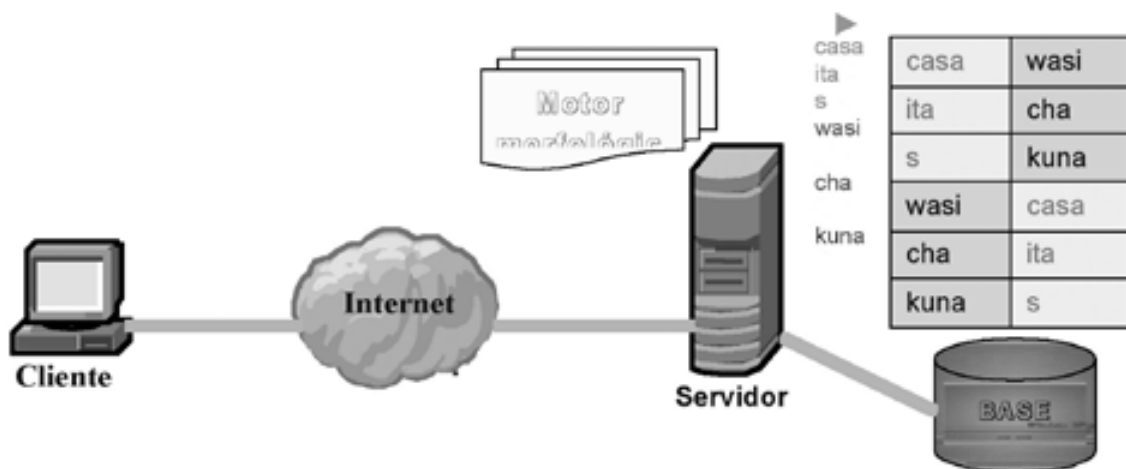
- Conocimiento de la lengua origen (LO).
- Conocimiento de la lengua meta (LM).
- Conocimiento de las distintas correspondencias entre LO y LM: en el más básico de los niveles, es el conocimiento de los equivalentes de traducción de las distintas palabras individuales.
- Conocimiento del dominio sobre el que se traduce: análisis y diseño del sistema, desarrollo del sistema y la implementación del sistema
- Conocimiento del entorno cultural: convenciones sociales, costumbres, etc., de los hablantes de LO y LM.

## Niveles de conocimiento lingüístico

- Conocimiento fonológico (conocimiento sobre el sistema de sonidos de un lenguaje).
- Conocimiento morfológico (conocimiento de la manera en que las formas de un determinado lema son construidas, así como de las posibles producciones de una determinada raíz).
- Conocimiento sintáctico (cómo las distintas palabras se combinan de forma lineal para construir frases y oraciones aceptables de una lengua. Por lo que respecta al tratamiento automático de una lengua es, básicamente, un problema de ordenación de cadenas de caracteres).
- Conocimiento semántico (el conocimiento del significado de palabras y frases y de las relaciones de significado entre una frase y sus unidades constituyentes).

## EL IDIOMA QUECHUA

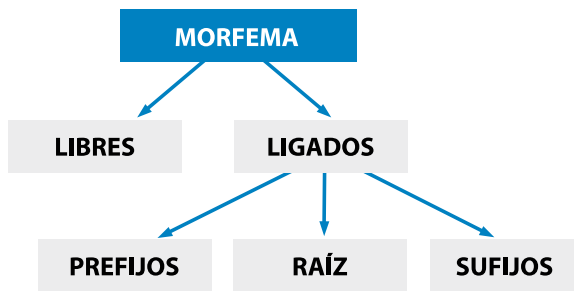
Es la lengua nativa americana más extendida en todo el mundo y la cuarta más extendida en el continente, le si-



güen el aymara y el guaraní, se habla en la zona occidental de Sudamérica y es idioma oficial en Ecuador, Perú y Bolivia, junto con el castellano y el aymara. A continuación se presenta las características más relevantes del idioma.

### Características del quechua

- Es aglutinante, es decir, reúne varias palabras en una: wasinmanta => de su casa
- Ausencia de prefijos y abundancia de sufijos.
- Ausencia de artículo y género gramatical.
- Conjugación muy perfeccionada (tiempos y modos completos).
  - Estoy cantando => takichkani.
  - Estoy comiendo => mikuchkani.
- Poseedor antes de cosa poseída:
  - wasipa punkun => de la casa su puerta.
- Topología SOV (sujeto + objeto + verbo).
- Ubicación del elemento modificador antes del núcleo.
  - Adjetivo antes del sustantivo: yuraq wasi
  - Adjetivo antes del verbo: kunanmi hamunqa



### Estudio morfológico:

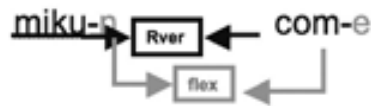
El morfema: unidad mínima portadora de significado.  
 Palabra: es el resultado de la unión de uno o más morfemas. La palabra está compuesta por raíz o por raíz más afijos.

- |                     |  |
|---------------------|--|
| • <u>Castellano</u> | * <u>Quechua</u>                               |
| • Adjetivo          | * Palabra Partícula                            |
| • Adverbio          | * Palabra Partícula                            |
| • Posesivo          | * Palabra Cualificante (adjetivos y adverbios) |
| • Preposición       |  |
| • Pronombre         | * Palabra pronombre                            |
| • Sustantivo        | * Palabra nominal                              |
| • Verbo             | * Palabra verbo                                |

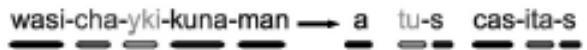
### La palabra quechua

Citamos a continuación dos ejemplos de clases de palabras y su sintaxis de composición en el idioma quechua.

- Palabra Verbo: PV > Rverb + (deriv) + flex + (disc)



- Palabra nominal: PN > Rnom + (deriv) + (flex) + (disc)



### Sufijos derivativos

- Sufijos de derivación:
  - Castellano**
  - **Diminutivos:** ito, ita, cico, cica, cillo, ecilla, uelo, cica => "cha"
    - piedrita rumicha
    - perrito allqucha
  - **Aumentativos:** oso, osa => "sapa" ote, ota => "su"
    - musculoso challwansapa
    - perrote allqusu

### Quechua

- |        |              |           |                            |
|--------|--------------|-----------|----------------------------|
| • cha  | diminutivo   | ito/ita   | (wasicha - casita)         |
| • yuq  | posesivo     | con       | (usayuyq - con piojo)      |
| • naq  | desposesivo  | sin       | (usanaq - sin piojo)       |
| • sapa | aumentativo  | oso/osa   | (usapa - piojoso(a))       |
| • su   | aumentativo  | ote/ota   | (wasisu - casota)          |
| • ntin | incorporante | todos/los | (laqtantin-todo el pueblo) |
| • lla  | limitativo   | solo      | (atuqlla - solo zorro)     |

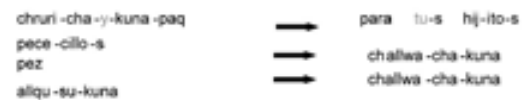
- Sufijos de Flexión:

- Persona
  - 1ra\_s -y
  - 2da\_s -yki
  - 3ra\_s -n
  - 1ra\_p -nchik/nchis
  - 1ra\_p -yku
  - 2da\_p -ykichik/ykichis
  - 3ra\_p -nku

- Número : kuna

- Ejemplos:

Sufijo	Función	Significado
-ta	ACU acusativo	a (deter.al objeto)
-man	ALA dativo (alativo)	a/hacia (prep)
-paq	BEN benefactivo	para
-pa	GEN genitivo	agente poseedor
-manta	ABL ablativo	de/desde (prep)
-pi	LOC locativo	en (prep.)
-wan	ASO asociativo/conminativo	con (prep.)
-kama	TER ilativo	hasta (prep.)
-pura	INT interactivo (palabra)	entre (prep.)



- rayku CAU causal por / a causa de

## EL SISTEMA TRADUCTOR MORFOLÓGICO DE CASTELLANO Y QUECHUA

### Tecnologías utilizadas en el sistema:

- **Java**

- Lenguaje de programación orientado a objetos, desarrollado por la empresa Sun Microsystems en 1995, y que se ha extendido ampliamente en World Wide Web.



- Robustez al verificar errores en tiempo de ejecución como en tiempo de compilación.
- El principal aporte de Java a la Comunidad de Usuarios y Programadores es la portabilidad, que permite que los usuarios se puedan mover fácilmente entre sistemas operativos y plataformas diferentes.

- **MySql**

- Base de datos que consume pocos recursos, tanto de CPU como de memoria.



- Utilidades de administración (backups) y Control de acceso (qué usuarios tienen acceso a qué tablas y con qué permisos).

### Requerimientos del sistema:

- Servidor

- Hardware

- Procesador: Pentium IV 2GHz
- RAM: 512 Mb
- Disco Duro: 10MB

- Software

- Sistema Operativo: Windows 200X / XP / Linux
- Servidor Web: Tomcat (jakarta-tomcat-4.1.30)
- Base de datos: MySql 4.1.7
- Java: j2sdk1.4.2

- Cliente

- Hardware

- Procesador: Pentium
- RAM: 128 Mb o superior

- Software

- Sistema Operativo: Windows95/98 o superior
- Acceso a Internet.

- Sistema público de fácil accesibilidad vía Web.
- Internet



- Bajo costo de implementación.
- Uso de software libre (Freeware).



- Sistema Multiplataforma:

- Gracias a la portabilidad que brinda Java.



### Codificación

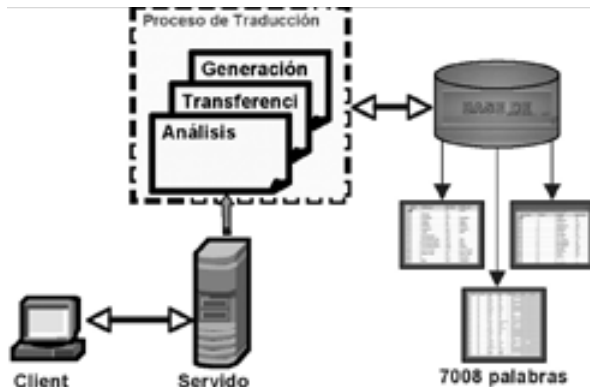
Se realizó una clasificación de las palabras según el criterio semántico-sintáctico, para su uso codificado dentro del sistema. Cabe mencionar la importancia que tiene la siguiente tabla debido a su original y particular elaboración, de acuerdo con las reglas de correspondencia entre ambos idiomas. A continuación se muestra una fracción de la tabla de codificación.

TIPO	NOMBRE_QUE	NOMBRE_CAST	ABREV	SUBTIPO	NOMBRE_SUBTIPO	ABREV
3	Pronombre	Pronombre	pron			
				3	personal	pers
				44	interrogativo	int
				55	posesivo	pos
4	Adjetivo	Adjetivo	adj			
				7	calificativo	cal
				8	numeral o cuantitativo	num
				9	demonstrativo	dem
				10	indefinido	ind
5	Adverbio	Adverbio	adv			
				11	lugar	lug
				12	tiempo	tpo
				13	modo	mod
				14	cantidad	can

### Funcionamiento del sistema

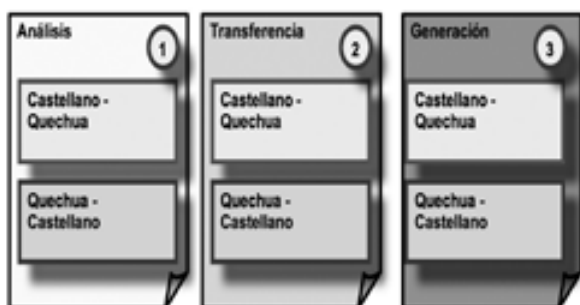
En principio, el sistema recibe del cliente la palabra que se quiere traducir y el método de traducción (castellano

a quechua o viceversa). Luego, el servidor realiza el proceso de traducción en tres fases e interactuando con la base de datos.

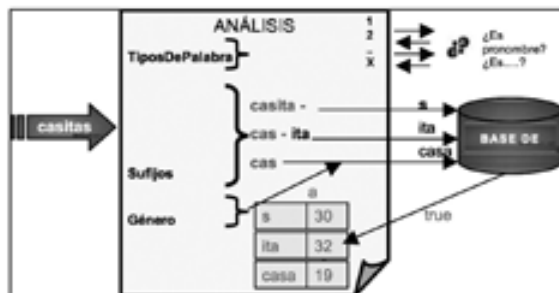


### Proceso de traducción

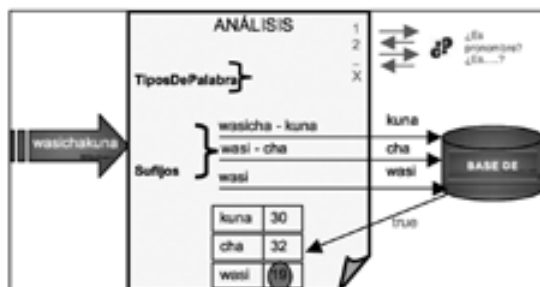
El sistema de traducción realiza tres procesos principales: **Proceso de análisis.-** En primer lugar, la palabra origen se compara con las diferentes clases de palabras existentes en la base de datos. Para este proceso el sistema realiza un detallado análisis morfológico (descomposición de la palabra: "raíz y sufijo(s)"), que se hace en función a las reglas gramaticales de la lengua en cuestión. Luego se les asigna el correspondiente tipo de palabra a cada componente de la palabra de origen (según codificación de raíces y sufijos). Finalmente esta estructura pasa al siguiente proceso. Se resalta la importancia de que la fase de análisis se diseñó de tal manera que sólo precisa de información de la lengua origen. A continuación se muestra un ejemplo del proceso de análisis para la traducción de una de las lenguas a la otra y viceversa.



- Castellano – Quechua: la palabra se analiza en función a las reglas del castellano.

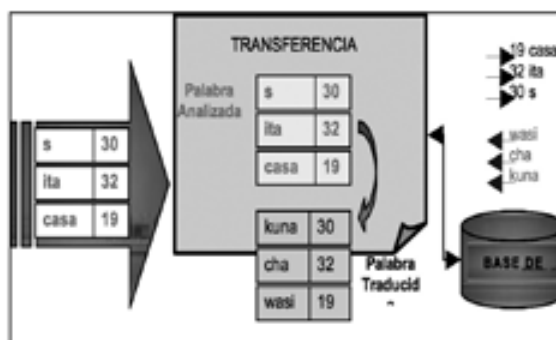


- Quechua - Castellano: la palabra se analiza en función a las reglas del quechua.

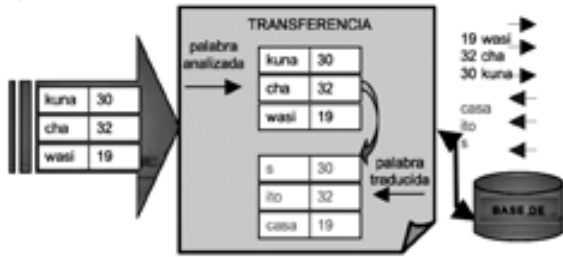


**Proceso de transferencia.-** Esta es la fase verdaderamente bilingüe del sistema. Primero se recibe la estructura de la palabra que envía el proceso de análisis. Luego se realiza la transferencia de significados de raíces y sufijos, es decir, se traduce cada componente de la palabra de acuerdo con su codificación de tipos, utilizando el diccionario bilingüe. Este subproceso devuelve el equivalente del idioma origen en el idioma destino de cada componente. Finalmente se forma una nueva estructura con su correspondiente codificación, pero ya en el idioma destino. A continuación se citan dos ejemplos de este proceso.

- Castellano – Quechua: se realiza la traducción de cada componente de la palabra en función al código de palabra asociado (en base al sistema de codificación).

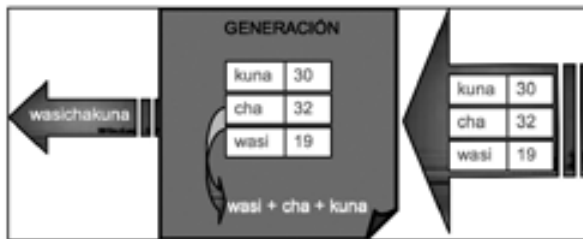


- Quechua - Castellano: se realiza la traducción de cada componente de la palabra en función al código de palabra asociado (en base al sistema de codificación).

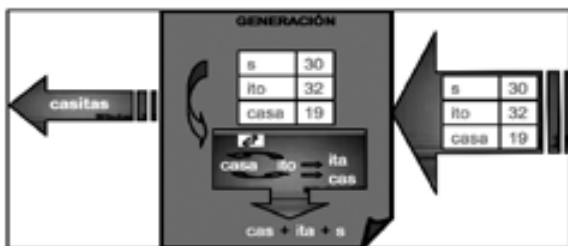


**Proceso de generación.-** En esta fase básicamente se realiza la composición de cada componente de la palabra: "raíz + sufijo(s)", formando la nueva palabra en lengua destino. En esta fase se realiza la comparación de acuerdo con las reglas morfológicas de la lengua destino.

- Castellano – Quechua: se realiza la generación (concatenación) de la palabra de acuerdo con las reglas morfológicas del quechua.



- Quechua - Castellano: se realiza la concatenación de la palabra de acuerdo con las reglas morfológicas del castellano.



### Componentes del sistema

- Traducción de palabras, principal funcionalidad del sistema.

Castellano - Quechua	
Palabra	Traducción
gato	chaca

Quechua - Castellano	
Palabra	Traducción
chaca	gato

- Descripción de la composición morfológica de la palabra, además de mostrar la palabra traducida el sistema es capaz de describir cómo está compuesta la palabra origen y las equivalencias de estos componentes en la lengua destino.

Palabra	Traducción
wasichakuna	animaleschakuna
Composición Morfológica:	
TAMAL: Sustantivo común => IRUMINTA: Sustantivo común	
ITO: Suf. Derivación diminutivo => CHA: Suf. Derivación diminutivo	
S. Plural => KUNA: Suf. Flexión Número	

- Sinónimos, el sistema también muestra los sinónimos (en lengua destino) de la palabra traducida.

Palabra	Traducción	Sinónimo
pequeña	uchuy	hachuy
Composición Morfológica:		
PEQUEÑA: Adjetivo calificativo => UCHUY: Adjetivo calificativo		

### Características funcionales del sistema

- Reconoce y especifica la clase de palabra.
  - Sustantivo común, adjetivo calificativo, etc.
- Realiza la traducción de palabras en plural.
  - ríos => mayukuna
  - urqkuna => cerros
- Traduce palabras con sufijos derivativos.
  - perrito => allqucha
  - florcita => tikacha
- Traduce palabras con sufijos derivativos y flexivos.
  - Mallki-cha-kuna => arbol-ito-s
- Reconoce los artículos del castellano que no existen en el quechua.
- Realiza la traducción de verbos en infinitivo.
  - Barrer => pichay
  - mirar => jhaway
  - abrir => qhaway
- Realiza la traducción de verbos en participio.
  - amado => munasqa
  - temido => manchasqa
- Realiza la traducción de verbos en presente progresivo y gerundio.
  - rompiendo => pakichka
- Realiza la conjugación de los verbos regulares (terminados en "ar", "er" o "ir") y en Modo Indicativo.
  - Tiempo Simple: Presente.

- tem-en => mancha-niku
- cant-an => taki-niku
- abr-en => picha-niku
- Tiempo Simple: Pretérito Perfecto Simple.
  - Tem-ieron => mancha-rqaku
  - cant-aron => taki-rqaku
  - abr-ieron => kicha-rqaku
- Tiempo Simple: Pretérito Imperfecto.
  - tem-ían => mancha-rqaku
  - cant-aban => taki-rqaku
  - abr-ían => kicha-rqaku
- Tiempo Simple: Futuro Imperfecto.
  - tem-erán => mancha-nqaku
  - cant-arán => taki-nqaku
  - abr-irán => kicha-nqaku

## CONCLUSIONES

- Es importante considerar la complejidad de las lenguas y, especialmente, los principios que rigen la lengua quechua, que no tiene por qué ser compartidos por el castellano y viceversa.
- Este proyecto estudió las posibilidades de explotación de un sistema de traducción automática para el quechua, el cual podría ser de gran repercusión, por la utilidad como herramienta de ayuda al conocimiento del quechua, gracias a la información morfológica que nos brinda.
- Al tratar de la traducción automática estamos hablando de una tecnología muy desarrollada, pero que trabaja con algo tan complejo como la lengua; por lo cual es importante tener en cuenta que las computadoras no poseen el bagaje cultural que suele tener un traductor y, por este motivo, sus traducciones no son

perfectas.

- La traducción automática se genera de forma instantánea y, por lo tanto, es ideal para situaciones en las que no es posible que una persona haga la traducción.
- Este sistema servirá como prototipo para la optimización del sistema y el desarrollo de un sistema de traducción de frases, oraciones y posteriormente textos.

## REFERENCIAS BIBLIOGRÁFICAS

- 1 Gustavo Solis Fonseca, Jorge Chacón Sihuay. Lingüística y Gramática Runasimi-Chanka.
- 2 Nilda Velarde de Grgicevic, Universidad Tecnológica de los Andes. Quechua.
- 3 Mabel Tarazona Espíritu. Lecciones Básicas para aprender Quechua.
- 4 Demetrio Tupaq Yupanqui. Curso de Quechua Red Científica Peruana. Se consigue en URL: ([www.quechuanetwork.com](http://www.quechuanetwork.com))
- 5 El lexicón en la Traducción Automática. Se consigue en URL: (<http://elies.rediris.es/>)
- 6 La Linterna del Traductor. Se consigue en URL: (<http://traduccion.rediris.es/>)
- 7 Joseba Abaitua. Universidad de Deusto. Traducción automática: Presente y Futuro. Se encuentra en URL: ([http://foreignword.com/es/technology/art/Abaitua/Abaitua\\_4.htm](http://foreignword.com/es/technology/art/Abaitua/Abaitua_4.htm))
- 8 Jesús Valdés. La traducción automática. Se encuentra en URL: (<http://www.dgbiblio.unam.mx/servicios/dgb/publicdgb/bole/fulltext/volIV3/traduccion.htm>)
- 9 Arturo Trujillo. Estrategias de traducción automática/ Machine translation strategies. Se encuentra en URL: (<http://www.prbb.org/Quark/19/019053.htm>)