ONOMAZEIN 3 (1998): 315-324

# DISEÑO E IMPLEMENTACIÓN DE UN TRANSCRIPTOR FONÉTICO AUTOMÁTICO DE TEXTOS GENERALES DEL ESPAÑOL

**Hernán Emilio Pérez**
**Thomas Armstrong**
Universidad de Concepción

**Resumen**

Este trabajo tiene como propósito describir el diseño, implementación y evaluación de un módulo de transcripción fonética automática que toma como entrada cualquier texto escrito en español y genera como salida una cadena de carateres ASCII equivalentes a los fonemas de la lengua.

***Abstract***

*The aim of this paper is to describe the design, implementation and evaluation of an automatic module for phonetic transcription. Such module can take as input any Spanish text and yields as output a string of ASCII characters equivalent to the phonemes of Spanish.*

Las *tecnologías del habla* se dividen en dos grandes campos de trabajo y estudio: 1) La *síntesis del habla, c*uyo objetivo es lograr la generación automática de una señal vocal. 2) El *reconocimiento del habla*, cuyo objetivo es lograr una representación simbólica automática de una señal vocal.

Dentro del ámbito de la síntesis del habla existen numerosas aplicaciones que circulan ya incluso comercialmente. Las más difundidas actualmente son las que realizan síntesis a partir de un texto, lo que se ha dado en llamar *conversión texto-habla* o *text to speech synthesis* (TtS). La conversión texto-habla como transformación de un texto escrito en su realización sonora se ha concebido como un proceso que se va dando en etapas, cada una de las cuales ha derivado en líneas de investigación independientes, pero paralelas. De este modo, se asume actualmente que un conversor texto-habla se compo-

ne de módulos, cada uno de los cuales se ocupa de un aspecto distinto del proceso de transformación de la cadena inicial hasta llegar a la señal sonora.

Existe consenso en que la primera etapa de la conversión texto-habla es la que se ha llamado preprocesamiento del texto. En esta etapa, se asumen ciertas convenciones para el tratamiento y conversión de elementos como las abreviaturas, las siglas, las fechas, las horas o cualquier símbolo especial que contenga el texto y se diseña e implementa un módulo que realice las transformaciones pertinentes.

Luego del preprocesamiento viene la etapa de transcripción fonética automática, paso imprescindible debido, obviamente, al carácter no-unívoco de la relación letra-sonido. Esta etapa consiste en la transformación del texto en una cadena de unidades fonéticas, ya sean fonemas o alófonos. Para ello, se asume, en primer lugar, una convención para homologar un alfabeto fonético con un inventario de caracteres ASCII para luego diseñar e implementar un módulo que contenga reglas de transformación de letras a alófonos o fonemas y, además, reglas de detección y asignación de variaciones suprasegmentales (entonación, ritmo, acento, etc.).

Cabe destacar que el tratamiento de los elementos suprasegmentales en la conversión texto-habla es uno de los problemas que, actualmente, más ocupa a quienes trabajan en las tecnologías del habla.

Luego de que se ha realizado la transcripción fonética, surge un problema teórico-práctico que debe resolverse antes de continuar el proceso conversión texto-habla. Esto es, la elección de una unidad de segmentación de la cadena de caracteres fonéticos que permita el pareamiento de dichas unidades con sus homólogas en versión de sonido codificado o digitalizado. Aguilar *et al*. afirman al respecto:

> La elección de la unidad que se va a utilizar en un sistema de síntesis es un compromiso entre la arquitectura del sistema por un lado, y la naturalidad, la calidad y la adecuación entre las unidades fonéticas y las unidades lingüísticas por otro: el tamaño, la flexibilidad y la posibilidad de automatización del procedimiento de obtención de la unidad son criterios básicos en la decisión final. Frases y palabras requieren poco esfuerzo de concatenación, pero el sistema carece de flexibilidad y el costo de almacenamiento es muy alto. Los fonemas, por el contrario, se configuran como una unidad natural que dota de gran flexibilidad al sistema, y que resulta económica desde el punto de vista del número de unidades; sin embargo, es una unidad abstracta sometida a variaciones contextuales, lo que origina problemas importantes de concatenación. Unidades más pequeñas como la semisílaba [...] o el difonema [...] reúnen las condiciones para constituir un inventario relativamente reducido de unidades, a la vez que se eluden los problemas de unión en las fronteras de los segmentos [...]. (Aguilar *et al*., 1994:1)

Como se puede apreciar, existen varias alternativas en la elección de la unidad a utilizar en el sistema de síntesis. En un primer momento se tiende a pensar que la unidad más adecuada es el segmento; sin embargo, esta elección va en detrimento de la naturalidad de la voz sintetizada al obviar todos los fenómenos de coarticulación que se producen en la lengua oral. Otra unidad que intuitivamente se tiende a considerar adecuada es la sílaba, si bien Aguilar *et al*. (1994) no la mencionan. El problema con la sílaba como unidad de pareamiento es más o menos el mismo que con el segmento, esto es, que, si bien, no se obvian todos los fenómenos de coarticulación, sí una considerable parte de ellos. Dos unidades que han dado mejores resultados son la semisílaba, considerada como "el fragmento entre el inicio de la sílaba y el centro del núcleo silábico, o entre dicho centro y el final de la sílaba", y el difonema, considerado como el "segmento que abarca desde la mitad de la zona estable de un fonema hasta la mitad de la zona estable del siguiente y que incluye la transición entre ambos fonemas consecutivos".

Una vez que se decide la unidad que se utilizará, se propone una serie de reglas que se implementan en un módulo que realiza la segmentación y luego el pareamiento. Finalmente, este módulo debe realizar la concatenación de las unidades sonoras que corresponden a cada unidad segmentada del texto en escritura fonética.

Cabe destacar que existe, además, una dilatada línea de investigación que se preocupa de la codificación y digitalización de la señal sonora. En este ámbito se destacan, por un lado, la *síntesis por formantes* (Klatt, 1980) y la *síntesis por codificación predictiva lineal* (LPC).

En este trabajo se describe el diseño e implementación de un módulo de transcripción fonética automática desarrollado en el *Laboratorio de Fonética de la Universidad de Concepción* en conjunto con el *Grupo de investigación de desarrollo de interfaces multimediales inteligentes* de la misma universidad.

Este módulo de transcripción fonética automática es la primera etapa de un proyecto de implementación de un conversor texto-habla para el español de Chile, aún en desarrollo.

El sistema de transcripción consiste en un conjunto de tres paquetes de reglas: 1) Reglas de asignación de acento. 2) Reglas de adecuación del texto.[1] 3) Reglas de transformación de letras en fonemas.

---

[1] Se habla de reglas de adecuación del texto y no de preprocesamiento debido a que en estricto rigor el preprocesamiento del texto involucra tener en cuenta muchos fenómenos que no se consideraron en esta primera etapa del módulo de transcripción fonética. El objetivo principal de este módulo es lograr una transcripción fonética adecuada de palabras completas del español, sin considerar aún abreviaturas, cifras, siglas, palabras extranjeras, etcétera.

Las reglas de asignación del acento permiten que el módulo detecte la vocal tónica de la palabra y le asigne una marca que la diferencie del resto. Debido a que la programación original se hizo en lenguaje C, bajo ambiente Unix y en un computador Sun Sparc Station, máquina cuyo teclado no considera los tildes del español, se adoptó la convención de que toda vocal tónica se expresara gráficamente con su versión en mayúsculas. Sin embargo, se debe consignar que el programa se puede compilar sin modificación en cualquier máquina, ya que el código C es estándar ANSI. En un PC, por ejemplo, sí se pueden utilizar los diacríticos del teclado.

Para la asignación del acento se consideraron dos principios básicos: 1) Toda palabra que lleve tilde se ignora y solamente se transforma su vocal tildada en mayúscula. 2) Las palabras que no lleven tilde se analizan aplicando las reglas clásicas de tildación del español para palabras graves y agudas, pero en su sentido inverso. Por ejemplo, en el caso de una palabra como *transcriptor* (figura 1) se considera que si no termina en *n* o *s* y no lleva tilde en la penúltima vocal se trata de una palabra cuya última vocal es tónica, o sea, es una palabra aguda; en consecuencia, el módulo procede a transformar la *o* en *O*. En resumen, se fuerza la acentuación gráfica de todas las palabras.

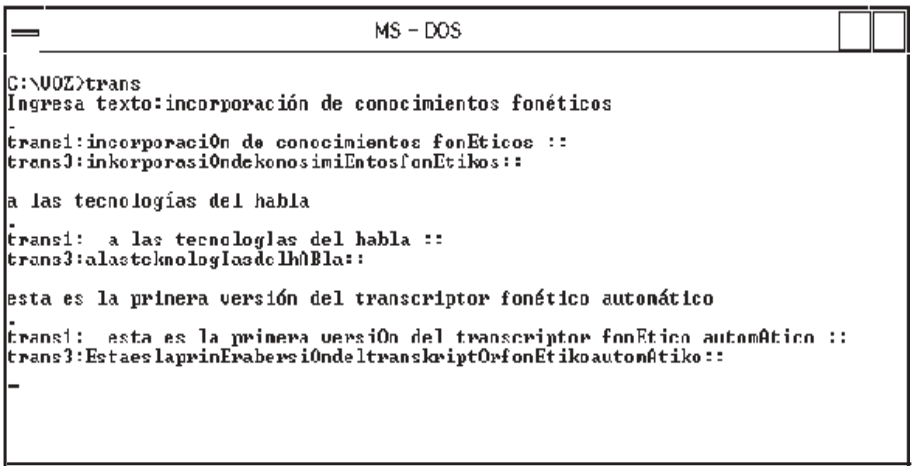
A partir de estos dos principios básicos se generó un paquete de reglas, las que se detallan en el anexo 1.

El paquete de reglas de adecuación del texto se compone de un pequeño inventario de reglas cuya finalidad es adecuar el texto para el análisis posterior. En primer lugar, más que una regla, contiene una simple instrucción para eliminar los espacios en blanco entre palabras del texto. Esto, por el consabido carácter continuo del habla. Esta instrucción para eliminar los espacios en blanco se complementa con un inventario de reglas que consideran fenómenos como el tratamiento de la *r* al inicio de palabra; la ocurrencia de *ll* y la concurrencia de *l* final de palabra con otra, inicial de palabra; la ocurrencia de *x* al inicio de palabra, y la presencia de la conjunción *y*. En el anexo 2 se detallan las reglas de adecuación del texto.

La transformación de letras en fonemas o alófonos trae consigo varios problemas teóricos y prácticos que deben ser resueltos antes de elaborar una propuesta de reglas al respecto.

Un primer problema es decidir qué alfabeto fonético se va a utilizar. Esto no resulta importante si se parte de la consideración de que el funcionamiento interno de un conversor texto-habla debiera ser, en teoría, transparente para el usuario y bastaría con que el diseñador estableciera un sistema de caracteres claro y coherente con respecto a las realizaciones fonéticas o fonemáticas del habla; sin embargo, esto resulta crucial si se piensa en la divulgación de los

Figura 1
Impresión de pantalla del módulo de transcripción fonética funcionando. El programa se inicia digitando *trans* en el prompt del sistema operativo. En seguida aparece el prompt *ingresa texto* en donde el usuario puede ingresar el texto hasta que inserte un punto. Inmediatamente el módulo realiza la transformación, la que aparece en el prompt *trans3*. Luego el módulo queda listo para recibir más texto




resultados y avances del sistema, así como también si se considera que las actuales aplicaciones generalmente traen un diccionario en donde el usuario puede ir agregando pronunciaciones excepcionales o que el sistema no logra reconocer o reproducir con exactitud.

De esto se deriva el segundo problema que, como bien afirman De la Mota & Ríos, consiste en que "Es preciso reconocer que los inventarios fonéticos que se conocen en la actualidad presentan numerosos problemas tanto de tipo conceptual como en el uso de símbolos y diacríticos" (1995: 97). Efectivamente, la utilización de cualquiera de los dos alfabetos fonéticos más difundidos en la tradición hispánica, como son el de la IPA y el de la RFE, o bien no logran dar cuenta plenamente de todas las realizaciones posibles de algunas variantes lingüísticas, o bien presentan una proliferación de diacríticos o una inconsistencia en el uso de éstos.[2]

En este trabajo se optó por utilizar el alfabeto fonético de la RFE, porque resulta más adecuado para representar las variaciones fonéticas más relevantes de la variante nacional.

Debido a que los computadores poseen un sistema de caracteres restringido llamado código ASCII (American Standard Code for

[2] Para más detalle sobre este tema, ver De la Mota & Ríos, 1995.

Information Interchange), en el cual muchos de los símbolos fonéticos no tienen representación, fue necesario establecer una convención de equivalencias, que se detallan en la siguiente tabla:

| Símbolo ortográfico | Carácter ASCII | Símbolo RFE |
|---|---|---|
| a | a | a |
| e | e | e |
| i | i | i |
| o | o | o |
| u | u | u |
| p | p | p |
| t | t | t |
| c, k | k | k |
| b, v | b | b |
| d | d | d |
| g | g | g |
| j, g | x | x |
| s, z, c | s | s |
| f | f | f |
| ch | C | ĉ |
| ll, y | y | y̌ |
| r | r | r |
| rr | R | r̄ |
| l | l | l |
| m | m | m |
| n | n | n |
| ñ | ñ | n̂ |

Otro problema teórico que se suscita al momento de proponer reglas de transformación de texto a escritura fonética es decidir sobre si deben reflejarse en la transcripción las variaciones de registro y las variantes diatópicas. Por el momento no es el propósito de este escrito discutir sobre este tema, ya que operativamente y por ser una primera aproximación al fenómeno de la transcripción fonética automática, se optó por que el módulo fuera capaz de realizar una transcripción fonémica solamente y se espera ir paulatinamente incorporando fenómenos dialectales, si es que a partir de la reflexión y discusión profunda se estima que es pertinente.[3] En ese momento

[3] De la Mota & Ríos (1995) realizan una primera aproximación a este problema teórico que sin duda debe ser discutido por los expertos, ya que la decisión que se tome al respecto puede trascender en la aceptación o rechazo que los usuarios manifiesten ante el conversor texto-habla.

habrá que clarificar algunos aspectos como la extensión de la incorporación de fenómenos dialectales, o sea, si se van a considerar todos o sólo algunos, esto a partir de si se considera que dicho fenómeno es definitorio o no de identidad de la comunidad, de si todos los hablantes lo realizan del mismo modo y de si es considerada por los hablantes como una manifestación prestigiosa o estigmatizada.

Una vez resueltos todos estos problemas, se propuso un inventario de reglas de conversión de la escritura ortográfica corriente en escritura fonética, las que se detallan en el anexo 3.

Finalmente se realizó una primera evaluación del módulo con estudiantes de fonética de la Universidad de Concepción. Los sujetos ingresaron textos elegidos por ellos mismos y luego examinaron las transcripciones. Se puede afirmar que, en principio, el transcriptor fonético automático cumple con su objetivo, que es generar una cadena de caracteres ASCII equivalentes a un alfabeto fonético (o más bien fonémico) a partir de un texto escrito en español.

Aún falta trabajo por hacer para, en primer lugar, perfeccionar el módulo de transcripción fonética y, en segundo lugar, implementar un conversor texto-habla para el español de Chile.

Durante la evaluación se detectaron algunos pequeños problemas como el no haber considerado grupos consonánticos como *sc* o *xc;* también, al parecer, por un problema en la programación, el módulo no realiza la eliminación de la *h* (como se puede apreciar en la figura 1 en la palabra *habla*); otro problema es que el paquete de reglas de asignación de acento no considera o, más bien, resulta imposible asignar acentuación a las palabras monosílabas, a menos que éstas traigan expresamente marcado el diacrítico.

## BIBLIOGRAFÍA

AGUILAR, L. *et al*. (1994). "Incorporación de conocimientos fonéticos a las tecnologías del habla". http://liceu.uab.es/~joaquim/publicacions/valencia_94.html.

DE LA MOTA, C. & Ríos, A. (1995). "Problemas en torno a la transcripción fonética del español: los alfabetos fonéticos propuestos por IPA y RFE y su aplicación a un sistema automático". *Estudios Hipánicos* IV. Pp. 97-109.

GURLEKIAN, J. *et al*. (1984). "El hombre dialoga con la máquina". *Quid* 14(2). Pp. 119-134.

KLATT, D. (1980). "Software for cascade/parallel formant synthesizer". *Journal of Acoustical Society of America* 67(3). Pp. 971-995.

_____ 1987. "Review of text-to-speech conversion for English". *Journal of Acoustical Society of America* 82(3). Pp. 737-793.

RODRIGUEZ, M. *et al*. (1984). "Visión panorámica de la respuesta oral de máquinas". *Mundo Electrónico* 144. Pp. 57-66.

## Anexo 1
## Reglas de asignación de acento

1. Palabra con tilde, se ignora
2. Palabra sin tilde, se analiza
   - 2.1. Si termina en consonante distinta de *n* o *s* se analiza la última vocal
     - a. Si esta última vocal es *a*, *e* u *o*, se marca tilde sobre ella
     - b. Si es *i* o *u* se analiza la letra que la antecede
       - – Si esta letra antecedente es *a*, *e* u *o*, se marca tilde sobre ella
       - – Si esta letra antecedente es distinta de *a*. *e* u *o*, se vuelve a la vocal *i* o *u* y se marca tilde sobre ella
   - 2.2. Si termina en *n* o *s*, se analiza la penúltima vocal
     - a. Si esta penúltima vocal es *a*, *e* u *o*, se marca tilde sobre ella
     - b. Si es *i* o *u* se analiza la letra que la antecede
       - – Si esta letra antecedente es *a*, *e* u *o*, se marca tilde sobre ella
       - – Si esta letra antecedente es distinta de *a*, *e* u *o*, se vuelve a la vocal *i* o *u* y se marca tilde sobre ella
   - 2.3. Si termina en vocal se analiza la letra que la antecede
     - a. Si esta letra antecedente es consonante, *a*, *e* u *o* se analiza la penúltima vocal
       - – Si esta penúltima vocal es *a*, *e* u *o*, se marca tilde sobre ella
       - – Si es *i* o *u*, se analiza la letra que la antecede
       - – Si esta letra antecedente es *a*, *e* u *o*, se marca tilde sobre ella
       - – Si esta letra antecedente es distinta de *a*, *e* u *o*, se vuelve a la vocal *i* o *u* y se marca tilde sobre ella.
     - b. Si esta letra antecedente es *i* o *u*, se analiza la antepenúltima vocal
       - – Si esta antepenúltima vocal es *a*, *e* u *o*, se marca tilde sobre ella
       - – Si es *i* o *u*, se analiza la letra que la antecede
       - – Si esta letra antecedente es *a*, *e* o *u*, se marca tilde sobre ella
       - – Si esta letra antecedente es distinta de *a*, *e* u *o*, se vuelve a la vocal *i* o *u* y se marca tilde sobre ella

## Anexo 2
## Reglas de adecuación del texto

1. Si una palabra comienza con *r*, se debe agregar otra *r* al texto, es decir, la *r* se transforma en *rr*
2. Si una palabra termina con *l*, se revisa la primera letra de la siguiente palabra
   - Si esta letra es una *l* se revisa la siguiente letra
   - Si esta letra es una *l* se reemplazan ambas por "y" y se elimina el espacio en blanco; si no, se elimina una *l* y el espacio en blanco
   - Si no, se elimina sólo el espacio en blanco
3. Si una palabra comienza con *x*, se transforma en "s"
4. Si una palabra es *y* o termina en *y*, esta palabra o letra se transforma en "i"
5. Si una palabra termina en *s*, se revisa la primera letra de la siguiente palabra
   - Si esta letra es una *s*, se elimina una de ellas y el espacio en blanco
   - Si no, se elimina sólo el espacio en blanco

## Anexo 3
## Reglas de transformación de letras a escritura fonética

1. Las siguientes letras no necesitan análisis, se dejan igual
   *a, b, d, e, f, i, k, m, ñ, o, p, s, t, u*
2. Si la letra es *c*, se revisa la letra siguiente
   Si la letra es *i* o *e*, la *c* se transforma en "s"
   Si la letra es *h*, se elimina la *h* y la *c* se transforma en "C"
   Si la letra es distinta de *i*, *e*, o *h*, la *c* se transforma en "k"
5. Si la letra es *h*, se revisa la letra siguiente
   Si la letra es *i* se revisa la siguiente letra
   Si esta letra es vocal, la *h* se transforma en "y"
   Si esta letra es *u* se revisa la siguiente letra
   Si esta letra es vocal, la *h* se transforma en "g"
   Si esta letra es distinta de *i* o *u*, se elimina la *h*
6. Si la letra es *j* se transforma en "x"
7. Si la letra es *n*, se revisa la letra siguiente
   Si la letra siguiente es *p-b-v*, la *n* se transforma en "m"
   Si no, la *n* se deja igual
8. Si la letra es *q*, se transforma en "k"
9. Si la letra es *r*, se revisa la letra siguiente
   Si la letra siguiente es *r*, se elimina una *r* y se transforma la otra en "R"
   Si no, la *r* se deja igual
10. Si la letra es *v*, se transforma en "b"
11. Si la letra es *w*, se transforma en "gu"
12. Si la letra es *x*, se transforma en "ks"
13. Si la letra es *z*, se transforma en "s"
14. Si la letra es *g*, se revisa la letra siguiente
    Si esta letra es *i* o *e*, la *g* se transforma en "x"
    Si esta letra es *u*, se revisa la letra siguiente
    Si esta letra siguiente es *i* o *e*, se elimina la *u* y la *g* se transforma en "g"
    Si la letra es distinta de *i* o *e* , la *g* se transforma en "g"
15. Si la letra es *ü*, se transforma en "u"
16. Si la letra es *l*, se revisa la letra siguiente
    Si la letra siguiente es *l*, se elimina una *l* y se transforma la otra en "y"
    Si no, la *l* se deja igual