



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN LINGÜÍSTICA
MAESTRÍA EN LINGÜÍSTICA HISPÁNICA

FONÉTICA DE LA CIUDAD DE MÉXICO.
APORTACIONES DESDE LAS
TECNOLOGÍAS DEL HABLA

TESIS QUE, PARA OBTENER EL TÍTULO DE
MAESTRO EN LINGÜÍSTICA HISPÁNICA,
PRESENTA

JAVIER OCTAVIO CUÉTARA PRIEDE

ASESORA: DRA. CONCEPCIÓN COMPANY COMPANY

CIUDAD DE MÉXICO, 2004

Las páginas de agradecimientos y dedicatorias siempre me han puesto nervioso; me niego a parecer cursi o comprometido en la ventana subjetiva de un trabajo de tesis.

No es a través de una mención como se agradece el apoyo que brindan ciertas personas a lo largo de un proceso de investigación, sino a partir de una respuesta profesional y la promesa de una amistad que trascenderá. Tampoco es por obligación mencionar en primer lugar a mi tutora, asesora y amiga, la doctora Concepción Company Company, que me enseñó a crecer en la academia, y me apoyó y riñó cuando lo merecía.

Tampoco es compromiso agradecer a mi amigo el doctor Joaquim Llisterri Boix, y a su ayuda siempre inteligentísima y generosísima. Las páginas de este trabajo contienen en muy buena medida su mano.

El trabajo del doctor José G. Moreno de Alba es un modelo a seguir; el doctor Luis A. Pineda me ha abierto las puertas de su trabajo y amistad. Gracias a ambos por formar parte de esta etapa de mi trabajo profesional.

Al conjuntar esfuerzos con el doctor Luis Villaseñor-Pineda hemos creado una herramienta computacional lingüística. Jugar a la interdisciplina con alguien que te hace fácil la vida es muy motivante.

El doctor Ignacio Guzmán Betancourt (qepd) siempre fue un gran apoyo personal y profesional. Hay cosas que no se olvidan.

Georgina Barraza, Alejandra López, Margarita Palacios y Gerardo Sierra creen en mí, y yo en ellos.

A Alberto, Beatriz, Claudio, Eugenia, Graciela y a Silvestre, de manera muy especial.

Cerrando, la Chata y su familia; toda y grande.

Esta tesis se realizó, en parte, gracias a una beca para llevar a cabo estudios de maestría otorgada por el Consejo Nacional de Ciencia y Tecnología (CONACYT), con número de registro 163617.

ÍNDICE

1.	INTRODUCCIÓN	8
2.	UN CORPUS ORAL PARA EL ESPAÑOL DE LA CIUDAD DE MÉXICO: <i>DIME</i>	15
2.1.	La fonética acústica en los estudios hispánicos	15
2.2.	Lingüística y computación en nuestro país: los inicios	17
2.3.	Las tecnologías del habla y la interdisciplinariedad	20
2.3.1.	<i>El reconocimiento de habla</i>	23
2.4.	El Proyecto y el Corpus DIME	31
2.4.1.	<i>Los experimentos con el protocolo del Mago de Oz</i>	34
2.5.	Etiquetado fonético computacional	38
3.	UN ALFABETO FONÉTICO COMPUTACIONAL PARA EL ESPAÑOL DE MÉXICO: <i>MEXBET</i>	43
3.1.	Alfabetos fonéticos tradicionales	44
3.1.1.	<i>La tradición hispánica: el alfabeto de la Revista de Filología Española</i>	44
3.1.2.	<i>El Alfabeto Fonético Internacional</i>	46
3.2.	Alfabetos fonéticos computacionales	52
3.2.1.	<i>SAMPA</i>	53
3.2.2.	<i>OGIbet</i>	56
3.2.3.	<i>Worldbet</i>	59
3.3.	Mexbet	63
4.	ESTUDIO FONÉTICO DE LA CIUDAD DE MÉXICO Y SU APLICACIÓN A LAS TECNOLOGÍAS DEL HABLA: <i>TRANSCRÍBEMEX</i>	72
4.1	Estudios sobre la pronunciación del español de México	73
4.2.	Frecuencia de aparición de fonemas y alófonos	76
4.3.	Análisis fonético del Corpus DIME	93
4.4.	Consonantes	94
4.4.1.	<i>Consonantes oclusivas</i>	94
4.4.1.1.	Consonantes oclusivas sordas: /p, t, k/	95
4.4.1.1.1.	Alófono palatalizado: /k/ > [kʲ]	95
4.4.1.2.	Consonantes oclusivas sonoras: /b, d, g/	97
4.4.1.2.1.	Alófonos aproximantes: /b, d, g/ > [β, ð, ɣ]	98

4.4.2.	<i>Consonante africada sorda: /tʃ/</i>	100
4.4.3.	<i>Consonantes fricativas sordas: /f, s, x/</i>	101
4.4.3.1.	Alófono dentalizado: /s/ > [s̺]	103
4.4.3.2.	Alófono sonorizado: /s/ > [s̺]	106
4.4.4.	<i>Consonante fricativa sonora: /ʒ/</i>	108
4.4.4.1.	Alófono africado: /ʒ/ > [dʒ̺]	108
4.4.5.	<i>Consonantes nasales: /m, n, ñ/</i>	109
4.4.5.1.	Alófono dentalizado: /n/ > [n̺]	110
4.4.5.2.	Alófono velarizado: /n/ > [n̠]	111
4.4.6.	<i>Consonantes líquidas: /l, r, lʲ/</i>	113
4.4.6.1.	Asibilación de /r/ y /rʲ/	114
4.5.	Vocales	118
4.5.1.	<i>Vocal abierta: /a/</i>	118
4.5.1.1.	Alófonos velarizado (/a/ > [a̠]) y palatalizado (/a/ > [a+])	119
4.5.2.	<i>Vocales medias: /e, o/</i>	120
4.5.2.1.	Alófonos abiertos: /e/ > [e̞]; /o/ > [o̞]	120
4.5.3.	<i>Vocales cerradas: /i, u/</i>	122
4.5.3.1.	Paravocales: /i/ > [j]; /u/ > [w]	122
4.5.4.	<i>Otros fenómenos vocálicos</i>	123
4.6.	TranscribEMex	126
4.7.	Aplicaciones y perspectivas futuras	133
5.	CONCLUSIONES	136
6.	APÉNDICE. EQUIVALENCIAS ENTRE EL AFI, EL ALFABETO DE LA RFE Y MEXBET	141
7.	REFERENCIAS BIBLIOGRÁFICAS	146

ÍNDICE DE FIGURAS

Figura 1.	Fragmento del diccionario del Corpus DIME	27
Figura 2.	Ejemplo de los modelos de pronunciación del Corpus DIME	28
Figura 3.	Ejemplo del proceso de creación de un modelo acústico para la vocal /a/	29
Figura 4.	Funcionamiento de un reconocedor de habla	30
Figura 5.	Fragmento del diálogo s6-t1-g1 del Corpus DIME	33
Figura 6.	Aspecto del programa de diseño de cocinas utilizado por el Proyecto DIME para la grabación de su corpus	34
Figura 7.	Ejemplo de un diálogo entre una persona y un sistema multimodal adaptado de Allen <i>et al.</i> (2001)	36
Figura 8.	Sujeto 5: “Sf”. Segmentado y etiquetado fonético con Mexbet del archivo 2 del diálogo s5-t1-g1 del Corpus DIME, con el programa <i>SpeechView</i>	39
Figura 9.	El programa <i>SpeechView</i> muestra dos oscilogramas y dos espectrogramas, segmentados y alineados con el etiquetado fonético	40
Figura 10.	Archivo *.phn del etiquetado fonético, visualizado con un lector de texto	41
Figura 11.	Fragmento del “Alfabeto fonético” publicado por la <i>Revista de Filología Española</i> (1915)	45
Figura 12.	Consonantes (pulmonares) del Alfabeto Fonético Internacional	49
Figura 13.	El alfabeto fonético computacional SAMPA para el español castellano (Wells 1998b)	54
Figura 14.	Alófonos de procedencia indígena para el español de México, según Moreno y Mariño (1998)	55
Figura 15.	El alfabeto fonético computacional OGIbet para el español mexicano (Tlatoa 2000a)	57
Figura 16.	Anotación del alófono velarizado del fonema /a/ con Worldbet	60
Figura 17.	Algunos de los diacríticos de Worldbet, adaptables al español de México	60
Figura 18.	Worldbet (1994) para el español castellano	61
Figura 19.	Worldbet (1997) para el español mexicano	62
Figura 20.	Fonemas del español de México en el alfabeto Mexbet	68
Figura 21.	Alófonos del español de México en el alfabeto Mexbet	69
Figura 22.	Frecuencia de aparición de los fonemas del español de México, en 52 archivos del diálogo s6-t1-g1 del Corpus DIME	76
Figura 23.	Frecuencia de aparición de las vocales (en relación con el total de vocales) en 52 archivos del diálogo s6-t1-g1 del Corpus DIME	78
Figura 24.	Frecuencia de aparición de las consonantes (en relación con el total de consonantes) en 52 archivos del diálogo s6-t1-g1 del Corpus DIME	79
Figura 25.	Frecuencia de aparición de los alófonos del español de México en cuatro diálogos del Corpus DIME	82
Figura 26.	En orden descendente, frecuencia de aparición de los alófonos del español de México en cuatro diálogos del Corpus DIME	85

Figura 27.	Frecuencia de aparición de los fonemas y de los alófonos del español según varios autores	88
Figura 28.	Diferencias en el punto de articulación entre el nivel fonológico (superior) y el fonético (inferior) en las consonantes del Corpus DIME	90
Figura 29.	Diferencias en el modo de articulación entre el nivel fonológico (superior) y el fonético (inferior) en las consonantes del Corpus DIME	91
Figura 30.	Diferencias entre el nivel fonológico (superior) y el fonético (inferior) en las vocales del Corpus DIME	92
Figura 31.	Mago: “Okey”. Archivo número 35 del diálogo s5-t1-g1. Se observa la modificación en la transición del segundo formante de la vocal	96
Figura 32.	Sujeto 6: “Eh, ¿Me puedes mostrar los tipos de muebles que tengo?”. Archivo número 30 del diálogo s6-t1-g1. En la palabra “puedes”, el sujeto debilita tanto /d/, que prácticamente se elide	99
Figura 33.	Alófonos y reglas del fonema /s/, para el español de la ciudad de México	103
Figura 34.	Mago: “¿Así está bien?”. Archivo 27 del diálogo s6-t1-g1. En la contigüidad de /s/ y /t/ se aprecia la modificación del oscilograma en la /s/ y la pérdida de la oclusión del fonema /t/	106
Figura 35.	Sujeto 4: “Lo quiero junto al... refrigerador pequeño”. Archivo 28 del diálogo s4-t1-g1. Se aprecia el descenso en el segundo formante y la pérdida del cierre de la consonante oclusiva /t/ ante la nasal /n/	111
Figura 36.	Mago: “Ese es un gabinete”. Archivo 60 del diálogo s4-t1-g1. Se aprecia el ascenso en el segundo formante y la pérdida del cierre de la consonante oclusiva /g/ ante la nasal /n/	112
Figura 37.	Alófonos y reglas del fonema /n/ para el español de la ciudad de México	112
Figura 38.	Sujeto 6: “Arriba del frigobar”. Archivo 54 del diálogo s6-t1-g1. Se aprecia claramente, además de la asibilación, un alargamiento en la realización del fonema /r/ en posición final absoluta	116
Figura 39.	Mago: “¿Quieres un mueble como estos aquí? Archivo 65 del diálogo s6-t1-g1. Se observa claramente cómo lo que debería ser un hiato [-o.e-], se presenta como diptongo [-we-]	125
Figura 40.	Aspecto del programa TranscríbEMex	127
Figura 41.	Procedimiento para la alineación automática de la señal sonora con las transcripciones fonológica, fonética y silábica, con el programa TranscríbEMex .	131
Figura 42.	Mago: “Hemos terminado la tarea”. Archivo 116 del diálogo s6-t1-g1. Se aprecian (de arriba hacia abajo) dos oscilogramas, tres etiquetas en los niveles fonológico, fonético y silábico, y un espectrograma, alineados manualmente	132
Figura 43.	Representación de los fonemas del español de México con los símbolos del AFI, el alfabeto de la RFE y Mexbet	144
Figura 44.	Representación de los alófonos más comunes del español de México, con los símbolos del AFI, el alfabeto de la RFE y Mexbet	145

1. INTRODUCCIÓN

The belief that humans will be able to interact with computers in conversational speech has long been a favorite subject in science fiction. This reflects the persistent belief that spoken dialogue would be the most natural and powerful user interface to computer. With recent improvements in computer technology and in speech and language processing such systems are starting to appear feasible. There are significant technical problems that still need to be solved before speech-driven interfaces become truly conversational
Allen et al. 2001:1

La manera como el hombre se ha comunicado con las computadoras —a través de herramientas manuales— ha seguido un proceso exactamente contrario al de la comunicación humana, primordialmente oral. En computación, la electrónica no ha podido independizarse del elemento mecánico (ha necesitado de pantallas, teclados, *joy sticks*, o *mouses*, según la década). El presente trabajo se ubica en un ámbito interdisciplinario entre dos áreas de estudio que, aunque ante los ojos de algunos podría parecer todavía un hecho extraño, comparten muchos objetivos: la lingüística y la computación. La relación de ambas ciencias ha creado una disciplina relativamente joven: la lingüística computacional.

La práctica de la lingüística computacional ha sido realizada en nuestro país por ingenieros, matemáticos e informáticos principalmente, y en pocos y muy específicos centros de investigación.¹ Durante dos años y medio de colaboración en

¹ Hasta donde sé, en nuestro país los únicos grupos dedicados a la lingüística y la computación (en todos sus niveles: fonético y fonológico, morfológico, sintáctico, semántico, etc.) son cinco. De ellos, dos están en el seno de la Universidad Nacional Autónoma de México: en el Instituto de

un grupo de trabajo sobre tecnologías del habla, me he percatado de dos aspectos importantes: la oportunidad que ofrecen los avances tecnológicos para la investigación lingüística, y la necesidad de incorporar conocimientos fonéticos y fonológicos básicos que se desprendan de la teoría tradicional y del estudio instrumental del habla.

La presente tesis tiene como objetivo estudiar algunos aspectos fonéticos del español de México a partir del Proyecto DIME (Pineda *et al.* 2001) y del Corpus DIME (Villaseñor, Massé y Pineda 2001), un corpus oral que se inscribe dentro de las llamadas tecnologías del habla (síntesis de habla, reconocimiento de habla y sistemas multimodales, principalmente) para la creación de un sistema de reconocimiento de habla (un programa de cómputo a través del cual una persona se comunica con una computadora de manera oral). Esta investigación tiene varios intereses; por una parte, aprovecho los conocimientos de fonética articuladora ya existentes en México (Matluck 1951, Perissinoto 1975 y Moreno de Alba 1994, principalmente) y los complemento con los resultados del análisis acústico del Corpus DIME con el objeto de definir cuáles son los alófonos del español de México que tienen mayor repercusión en nuestro dialecto. Por otra parte, reviso y replanteo un alfabeto fonético computacional para el español de México, y propongo un listado de 37 alófonos que

Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) trabaja desde 1999 el Grupo DIME; en el Instituto de Ingeniería (IINGEN) está el Grupo de Ingeniería Lingüística (GIL). Ambos equipos colaboran de una u otra manera. Asimismo, hay una estrecha relación entre estos grupos y el Proyecto de Interacción Oral Hombre-Máquina, del Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), en Tonantzintla, Puebla. En el Departamento de Ciencias de la Computación, del Instituto Politécnico Nacional (IPN) existe también un importante equipo de investigadores dedicados a las tecnologías del habla. En la Universidad de la Américas, de Cholula, Puebla, se encuentra el Grupo Tlatoa.

sirven para la creación de modelos acústicos computacionales para perfeccionar el funcionamiento de un reconocedor de habla. Estos 37 alófonos fueron considerados por dos razones principales: por su frecuencia de aparición en la lengua y por la posibilidad de ser modelados por reglas. Así, los resultados de esta investigación aportan conocimientos tanto a la lingüística como a la lingüística computacional.

El Corpus DIME (Villaseñor, Massé y Pineda 2001) pertenece al Proyecto Diálogos Inteligentes Multimodales en Español (el Proyecto DIME) que se desarrolla en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM. Uno de los propósitos del Proyecto DIME es crear modelos de pronunciación para los sonidos del español de México, para la producción de un programa computacional de reconocimiento de habla, con posibles fines comerciales. De tal suerte y como objetivo último, pretendo mostrar que la aplicación de investigaciones lingüísticas sobre el español de México resulta primordial para las tecnologías del habla en nuestro país, y que esta combinación produce frutos para ambas disciplinas.

Esta tesis está organizada en 3 capítulos, además de esta introducción. En el capítulo 2 me ha parecido importante realizar una breve revisión del aprovechamiento que han obtenido diferentes autores (Gili Gaya 1950; Quilis 1981) a partir de la aplicación de la fonética acústica, y mostrar los trabajos de algunos investigadores que incursionaron en la lingüística computacional en nuestro país entre los años sesenta y ochenta (Grimes 1965; Manrique 1968; Ávila 1987). Menciono después los objetivos del Proyecto DIME y puntualizo las características del Corpus DIME, grabado en 1999 con un experimento *Mago de Oz*, un protocolo para el estudio de la

interacción entre personas y computadoras. En este mismo capítulo abordo las tecnologías del habla, y describo de manera específica el reconocimiento de habla, por ser ésta el área en la que se inscriben los resultados de este trabajo. Posteriormente, detallo de qué manera se elaboran las transcripciones fonéticas computacionales, que sirven para procesar y sistematizar los datos lingüísticos que alimentarán un programa de reconocimiento de habla a partir de la creación de modelos acústicos.

En el capítulo 3 presento Mexbet, el alfabeto fonético computacional para el español de México que se ha utilizado en el Proyecto DIME para transcribir fonéticamente el Corpus DIME. A partir de la revisión de la estructura original de Mexbet (Uraga y Pineda 2000 y 2002), replanteo su uso y presento una nueva versión. En ésta, procuré respetar al máximo sus primeros aciertos, pero puntalicé su uso al incorporar conocimientos fonéticos y fonológicos necesarios para delimitar el sistema fonológico del español de México, y conseguir un inventario fonético correcto para una transcripción clara y detallada de nuestro dialecto. Para ello, constituyen en la base de este apartado los conocimientos extraídos del análisis de dos alfabetos fonéticos tradicionales, el alfabeto fonético Internacional (AFI) y el de la *Revista de Filología Española (RFE)*; por otra parte, se revisan las características de SAMPA, OGIbet y Worldbet, otros alfabetos fonéticos computacionales que de una manera u otra tienen relación con Mexbet.

En el capítulo 4 analizo cómo los resultados del análisis fonético del Corpus DIME ayudan a matizar los conocimientos de la fonética del español de México. A partir de mi análisis, reviso qué alófonos se presentan en determinados contextos, y

comparo mis datos con los de diferentes autores (Navarro Tomás 1918; Alarcos 1950; Gili Gaya 1950; Quilis 1981, entre otros, para el español castellano, y Matluck 1951; Perissinotto 1975; Moreno de Alba 1994, principalmente, para el español de México).

Desde que inicié el análisis, observé que había algunos fonemas y alófonos que tenían una frecuencia de aparición muy alta; sin embargo, otros aparecían únicamente de manera esporádica, o casi nunca. Así, estudié las frecuencias de aparición de los fonemas y alófonos en el Corpus DIME, y comparé mis resultados con los de especialistas que han estudiado este tema (Alarcos 1950; Quilis 1981; Llisterri y Mariño 1993; Pérez 2003, entre otros). Esta labor me procuró datos sobre el impacto del uso en la lengua.

El alfabeto computacional Mexbet arrojó, tras el análisis, un inventario de 37 alófonos, los cuales por su frecuencia de aparición y por la posibilidad de ser modelados por reglas eran candidatos óptimos para ser incluidos en el programa de reconocimiento de habla del Proyecto DIME. Además, con este inventario de alófonos y reglas, ha sido posible crear un programa computacional que realiza transcripciones fonéticas de manera automática, el fonetizador TranscríbEMex (Cuétara y Villaseñor 2004). Este programa está siendo utilizado en el Proyecto DIME para concluir de manera rápida y eficiente no sólo el etiquetado final del Corpus DIME, sino también el de otros corpus orales. Asimismo, he utilizado TranscríbEMex como una herramienta didáctica con los alumnos del curso de Fonética y fonología y del seminario de Lingüística computacional, del Colegio de Letras Hispánicas de la Facultad de Filosofía y Letras de la UNAM.

Los resultados de esta tesis de maestría pretenden tener un utilidad práctica y variada. Las aportaciones principales de este trabajo son las siguientes:

- 1) Lograr un conocimiento matizado —de lo que ya se sabía— de la fonética del español de la ciudad de México, partiendo de las necesidades que se plantean desde las tecnologías del habla.
- 2) Dotar de un alfabeto fonético computacional claro, eficiente y práctico, Mexbet, al Proyecto DIME y a todos los grupos de tecnologías del habla interesados en él.²
- 3) Crear un transcriptor fonético automático para el desarrollo de las tecnologías del habla en nuestro país.
- 4) Incentivar a partir de este programa futuras investigaciones sobre el español hablado en México y proyectos educativos sobre fonética.

Merece la pena destacar también el hecho de que esta investigación toma como base un proyecto de tecnologías del habla, el Proyecto DIME, que se desarrolla en el seno de nuestra Universidad, y que cuenta con el apoyo del Consejo Nacional de Ciencia y Tecnología (CONACyT).

Es necesario señalar que en la presente investigación he optado por el uso del Alfabeto Fonético Internacional (el AFI; IPA por sus siglas en inglés: *International Phonetic Alphabet*). Como es sabido, la práctica lingüística, en especial los trabajos de fonética y fonología, requiere de alfabetos especiales para representar los sonidos de la lengua objeto de estudio. Por lo que respecta al español, la *Revista de Filología Española (RFE)* publicó en 1915 su alfabeto fonético (al que llamaré el alfabeto de la *RFE* o el alfabeto hispánico), el cual constituye una tradición académica de gran aceptación entre la comunidad lingüística iberoamericana. En el marco de las investigaciones en un nivel internacional, no obstante, el alfabeto más difundido y

² Mexbet ha sido adoptado ya por el Proyecto de Interacción Oral Hombre-Máquina, el grupo de trabajo sobre tecnologías del habla del INAOE.

conocido es el AFI. En una búsqueda por extender los resultados de sus investigaciones, este alfabeto ha empezado a ganar terreno entre los propios hispanistas, como bien lo señalan Gil y Llisterri (en prensa). De tal suerte, el uso del AFI en la presente investigación permitirá un diálogo más amplio con otros estudios de este ámbito. Así, en un apéndice presento las equivalencias entre el AFI, el alfabeto de la *RFE* y Mexbet que servirán como guía durante la lectura de esta tesis.

Por último, y como se podrá observar, las referencias bibliográficas representan una conjunción de tradición y novedad, al incluir a los autores más clásicos de la bibliografía lingüística hispánica y los artículos más actuales sobre tecnologías del habla.

2. UN CORPUS ORAL PARA EL ESPAÑOL DE LA CIUDAD DE MÉXICO: *DIME*

*If words were always pronounced in the same way,
automatic speech recognition (ASR) would be relatively easy.*
Strik y Cucchiarini 1999:226

En este capítulo presento el Corpus DIME, que es el objeto de estudio de esta tesis. Debido a que el trabajo se ha hecho con base en análisis acústicos, me ha parecido importante iniciar el capítulo presentando algunos aspectos sobre la fonética acústica en el marco de los estudios hispánicos. Posteriormente, y debido a que el Corpus DIME se inscribe en las llamadas tecnologías del habla, introduciré las aplicaciones de éstas, principalmente del reconocimiento de habla. Así, llegaremos al Proyecto DIME y a su Corpus DIME, el proyecto en el que se inscribe esta tesis, y el corpus que se utilizó en esta investigación.

2.1. La fonética acústica en los estudios hispánicos

Por dos razones importantes: por su reciente fallecimiento y porque siempre procuró mantener una estrecha relación entre los conocimientos fonéticos y fonológicos tradicionales y los avances tecnológicos, resulta apropiado iniciar este apartado con una cita de Antonio Quilis (1993:7), uno de los hispanistas más reconocidos del siglo XX:

Desde la publicación del *Manual de pronunciación española* de Tomás Navarro Tomás, hasta el momento actual, la evolución de la lingüística ha sido enorme y con ella, la de la vieja fonética: la consolidación de determinadas perspectivas teóricas, el nacimiento de otras, y un espectacular desarrollo tecnológico han originado una profunda transformación no sólo en la metodología que se debe aplicar hoy en la investigación del nivel de la expresión, sino también en su perspectiva didáctica; y, lógicamente, todos estos hallazgos han dado como resultado importantes modificaciones y precisiones en el contenido de esta materia.

Con estas palabras, Quilis nos presenta la visión de un lingüista abierto a la modernidad y dispuesto a aprovechar los avances tecnológicos en favor de su disciplina. Años antes, Gili Gaya (1950:13-24) había iniciado su famoso libro *Elementos de fonética general* con el capítulo “Nociones de acústica”, en el que relató las ventajas que los instrumentos ofrecían entonces: el fonógrafo para grabar y reproducir la voz, y el quimógrafo para obtener fotograbados de la onda sonora. Gili Gaya destaca, asimismo, la novedad del oscilógrafo de rayos catódicos y la del espectrógrafo (de tecnología costosa y uso complicado en su época)³. De hecho, a lo largo de todos sus capítulos, este autor muestra amplios y firmes conocimientos técnicos en su aplicación al estudio de la intensidad, timbre y cantidad en los sonidos del habla; de las características acústicas de vocales y consonantes, etc., mostrando

³ No olvidemos que a principios de los años sesenta en México, Lope Blanch (1963-1964:59) se lamentaba de la falta de un espectrógrafo: “Advierto que, en no pocas ocasiones, me vi precisado a escuchar 15 ó 20 veces la misma palabra para tratar de determinar el grado de debilitación o de ensordecimiento de una sola vocal. Y que, en algún caso, mi interpretación no era absolutamente segura. Espero poder contar en breve con el auxilio de un espectrógrafo, lo cual me permitirá resolver con absoluta certeza esos casos dudosos”.

cómo se explotaban en aquel entonces las ventajas de los instrumentos en favor de los estudios fonéticos.

Resulta curioso ver cómo el estudio del habla a partir de la onda sonora se ha hecho de diferentes formas a lo largo de las décadas. La tecnología avanza día a día y hoy cualquier estudiante puede disponer de un espectrógrafo en su propia computadora. Para Quilis (1981:9), el uso de los instrumentos acústicos en los estudios fonéticos era a principios de los años ochenta “prácticamente imprescindible”, y ya entonces señalaba los beneficios científicos que resultan de la unión entre la lingüística y la computación. Mentalidades como éstas han sido las que han promovido el avance en los conocimientos lingüísticos y tecnológicos en nuestra disciplina lingüística.⁴

2.2. Lingüística y computación en nuestro país: los inicios

La combinación de la lingüística y la computación en México no es nueva. En 1963, el Centro de Cálculo Electrónico de la UNAM organizó el “Simposio Latinoamericano sobre la utilización de las Calculadoras Electrónicas en la

⁴ En contraste, Alarcos Llorach (1950:55) llamó “artilugio” al espectrógrafo, calificativo que podría entenderse, de alguna manera, como peyorativo. El *Esbozo* de la Real Academia Española (RAE 1973:10) tampoco parece darle mucho crédito al espectrograma, que “parece representar con más fidelidad lo que entendemos por una sucesión de unidades de sonido discretas, o segmentos [...] La correlación con dichas unidades, sin embargo, tampoco es completa en un espectrograma. La representación no es constante; un mismo sonido puede estar representado por diversas variantes que no tienen un solo rasgo en común. Se dan también interferencias: formas espectrográficas que representan sonidos diferentes se parecen más entre sí que las variantes que representan un mismo sonido. Hay también encabalgamientos: la delimitación en el espectrograma no es tampoco precisa a veces y hasta se borra el carácter discreto de la representación”.

Investigación Lingüística” (Manrique 1968). A mediados de los años sesenta, Grimes (1965) destacó la manera como la lingüística se beneficiaba con el uso de las computadoras para la contabilización y el control de un gran volumen de datos, así como para el resumen y la recuperación de información. El entusiasmo que deja ver Grimes por esta nueva área muestra —a mi modo de ver— una mentalidad más vanguardista en la lingüística de hace casi cuatro décadas que la que actualmente se percibe en buena medida en nuestro país.⁵

En una ponencia presentada por Manrique (1968) en la *Reunión de Mayistas* de ese año, con entusiasmo y contento evidentes el autor expone los avances que ya se habían hecho en México para la automatización y modernización de diversos proyectos de lingüística, augurando grandes expectativas hacia el futuro. Por esas épocas, a instancias de Sergio Beltrán (del Centro de Cálculo Electrónico de la UNAM) y de Mauricio Swadesh, se creó la Sección de Investigaciones Lingüísticas que promovió dos proyectos relacionados con la computación: uno sobre traducción automática —en boga en ese entonces en muchas universidades de Europa y los Estados Unidos— y otro que buscaba crear un programa para el estudio de los códigos mayas (Manrique 1968). Los logros fueron muchos en esa época, según la relación de Manrique: el refinamiento de los métodos de transcripción, la captura de artículos y diccionarios relacionados con lenguas indígenas y el análisis matemático de la estructura de los tres primeros capítulos del *Chilam Balam de Chumayel*. Como

⁵ “Los equipos electrónicos para el tratamiento de datos no pueden ya considerarse como una incursión del materialismo en tierra sagrada, sino como un método de realizar los trabajos que todos tenemos que hacer, sin necesidad de perder tiempo en actividades rutinarias y en repeticiones mecánicas” (Grimes 1965:164).

podemos percibir, hubo muchos avances, pero parece ser que hasta cierto punto desaparecieron.⁶

En 1970, Bott mencionó la “influencia notable, y globalmente positiva” de la lingüística computacional en el desarrollo de la teoría lingüística, la cual había “obligado a los lingüistas que trabajan con computadoras a formular las reglas con más precisión” (227). Con este mismo espíritu, años mas tarde Ávila (1987) reportó en México la creación del sistema de cómputo ORTOFON, un programa que realizaba una retranscripción de la ortografía convencional a una nueva ortografía más sencilla y apegada a la pronunciación a la que Ávila llamó ALMA (Alfabeto Mexicano para Analfabetos).⁷ Este programa, o la idea de ese alfabeto, puede ser discutible para muchos, pero representa un ejemplo de las aplicaciones de la lingüística y la computación que se han hecho en nuestro país desde hace más de un cuarto de siglo.

Desde esos primeros pasos, el desarrollo de la tecnología computacional aplicada a la lingüística ha avanzado considerablemente, por lo cual la aplicación de los medios electrónicos para los estudios lingüísticos hoy en día es, pues, indiscutible.

⁶ “La Sección de Investigaciones Lingüísticas del Centro de Cálculo Electrónico ha prácticamente desaparecido; el interés que antes había en sus trabajos ya no existe [...] Debemos sin embargo referirnos al trabajo realizado por la Sección de Investigaciones Lingüísticas, considerándolo como una tarea suspendida temporalmente pero que debe continuarse [...] No podemos menos que manifestar nuestra pena por el hecho de que la computadora del Centro de Cálculo Electrónico para la cual fue elaborado el programa tenga uno de esos problemas «individuales» que ha impedido, junto con la actitud general de la institución, obtener todo el provecho que el esfuerzo del doctor Grimes nos hacía esperar” (Manrique 1968:3).

⁷ “La idea es que los adultos —dadas las dificultades que implican para muchos de nosotros las innovaciones— no tengan que aprender la nueva ortografía: basta con que escriban en una computadora y el programa ORTOFON —ortografía fonológica— se encarga del resto. Se pueden seleccionar, además, dos transcripciones por ahora: la mexicana (sin *z* ni *ll* pero con *sh*) y la hispánica (con inclusión de las tres letras citadas). Los niños, por su parte, podrían utilizar —tras aprender— a leer y a escribir con el ALMA” (Ávila 1987:17).

Al tomar como base el hecho de que éste es un trabajo que se inscribe en el nivel fonético y en las tecnologías del habla, el uso de los instrumentos más actuales es básico. Aunque todavía estamos lejos de ello, en un futuro los teclados, el *mouse* y otras herramientas para comunicarse con las computadoras serán cosa del pasado. De ello se desprende la necesidad de estimular estudios de lingüística que hagan uso de la computación; todavía más, habrá que favorecer la posibilidad de unir ambas ciencias.

2.3. Las tecnologías del habla y la interdisciplinariedad

Las tecnologías del habla tienen como objetivo la búsqueda de la comunicación oral entre las personas y las computadoras, e implican la interacción de los conocimientos y la práctica fonética y fonológica con los sistemas computacionales. Las tecnologías del habla son tres, principalmente: la *síntesis de habla*, el *reconocimiento de habla* y los *sistemas de diálogo*. En la síntesis, una computadora emite electrónicamente una voz que busca semejarse al habla humana (Llisterri 2002); por otra parte, en el reconocimiento de habla, la computadora, al recibir la señal acústica del habla, es capaz de reconocer los segmentos fónicos y decodificarlos. Los sistemas de diálogo enlazan a ambas y, por ende, a partir de ellos una persona puede interactuar con una máquina través del habla.⁸

⁸ Uno de los trabajos más amplios sobre la relación entre la lingüística y la computación aplicadas a las tecnologías del habla —así como a las tecnologías del texto: correctores ortográficos y gramaticales, analizadores morfológicos y sintácticos, etc.— se encuentra en el artículo de Llisterri “Lingüística y

Me parece interesante mencionar también otras tecnologías del habla, como el *reconocimiento de hablante* y la *fonética forense* —ésta última, aunque en un estricto sentido no está considerada como una tecnología del habla, pone de manifiesto el provecho que se obtiene a partir de la interacción de la práctica fonética y otras disciplinas. El reconocimiento de hablante tiene como propósito reconocer locutores, con aplicaciones en la banca telefónica o en sistemas de seguridad empresarial. La aplicación de la fonética forense es principalmente judicial; no ha logrado el mismo desarrollo en México que en otros países; actualmente sólo se realiza —hasta donde tengo conocimiento— en la Procuraduría del Distrito Federal y en la Procuraduría de la República.⁹

En España, entre los programas de aplicación comercial del reconocimiento de hablante se encuentra el sistema de verificación de locutor de la empresa Telefónica I+D (Rodríguez *et al.* 2001). Como parte de los proyectos de investigación más reconocidos están VILE, un estudio acústico de la variación inter e intralocutor para el español castellano (Battaner *et al.* 2003), y el corpus AHUMADA¹⁰ (Ortega *et*

tecnologías del lenguaje” (2002). Otro trabajo importante es el de Aguilar *et al.* (1997), en el que se presenta de manera muy completa el tema de las aplicaciones de las tecnologías del habla en diferentes desarrollos de lingüística computacional.

⁹ La fonética forense es todavía una disciplina a la que le falta mucho por resolver. Por una parte, los diferentes autores no se ponen de acuerdo en los parámetros que deben considerarse para la identificación de un hablante, y el margen de error es aún muy grande. El proyecto VILE (Battaner *et al.* 2003), por ejemplo, toma muy en cuenta los resultados del análisis acústico (las propiedades de la frecuencia fundamental, las frecuencias centrales y el ancho de la banda de los formantes vocálicos y la composición espectral de las consonantes). Por el origen tecnológico del proyecto, en contraste, los parámetros empleados por otros equipos dedicados al reconocimiento del locutor desde el ámbito de la ingeniería se relacionan de un modo menos directo con los habitualmente empleados en el análisis fonético. Además, por el propio carácter de las investigaciones, los adelantos en el área de la fonética forense están bastante vedados.

¹⁰ Destaca el espíritu interdisciplinario del grupo de trabajo implicado en la creación de este corpus, conformado por investigadores del Área de Tratamiento de Voz y Señales (<http://www.atvs.diac.upm>).

al. 1998), un corpus específicamente diseñado para el desarrollo de sistemas automáticos de reconocimiento de hablantes.

Así, podemos apreciar cómo las tecnologías del habla representan un área fértil en la que la interdisciplinariedad tiene cabida; la interacción entre los fonólogos y fonetistas y los expertos en computación resulta vital para el fortalecimiento de los proyectos de ciencias aplicadas al habla humana, en lo que Gil y Llisterri (en prensa) llaman “el futuro factor de muchas de las aproximaciones al aspecto fónico de las lenguas”.

En este repaso de la interacción de la lingüística y la tecnología en las últimas décadas (§2.1. y §2.2.), el propósito principal fue percatarnos de lo que parece haber sido una mentalidad más amplia en el pasado;¹¹ a finales de los años sesenta, Manrique (1968:2) insistía en la necesidad de la interdisciplinariedad entre la lingüística y la computación en México: “por supuesto, no se ha intentado convertir a los lingüistas en ingenieros en computación, sino solamente darles la información necesaria para que la cooperación entre los especialistas de ambos campos fuera fructífera”. Por su parte, y en cuanto a los problemas de lingüística computacional que preocupaban a principios de los años setenta, Bott (1970:237-238) mencionó que

es) de la Escuela Universitaria de Ingenieros de Telecomunicaciones de la Universidad Politécnica de Madrid, del Departamento de Lengua Española de la Universidad Nacional de Educación a Distancia y del Servicio de Policía Judicial de la Dirección General de la Guardia Civil Española.

¹¹ Sorprende enterarse de que “desde 1958, apenas un año después de haber modificado sus planes de estudio, la especialidad de Lingüística de la Escuela Nacional de Antropología e Historia, reconoció las posibilidades para la investigación que se abrían con el uso de las computadoras [...] se impulsó a los estudiantes a seguir los cursos que al efecto se impartían [...] en el Centro de Cálculo Electrónico de la UNAM; en varias ocasiones los cursillos y conferencias se tomaron sin crédito para los estudiantes, pero en otras se consiguió que tuvieran el mismo valor que uno de los cursos optativos o seminarios que la Escuela ofrece en su recinto” (Manrique 1968:14).

“la solución a nuestros problemas depende, al parecer, de los avances de otras disciplinas a las que concierne la manera en que el cerebro humano organiza su información”. Así, los diferentes modelos computacionales vinculados a la inteligencia artificial tratan de reproducir las funciones del cerebro en todos los niveles.

Dentro de las tecnologías del habla hasta aquí mencionadas, el área en la que se insertan los resultados de este trabajo fonético es el reconocimiento de habla, por lo cual conviene profundizar un poco más en la estructura y funcionamiento de este tipo de programas.

2.3.1. El reconocimiento de habla

Un reconocedor de habla es un programa de cómputo que, al recibir la señal sonora del habla, distingue los segmentos fónicos y decodifica el mensaje.¹² Por su espectro de acción y por sus alcances, los reconocedores de habla se clasifican en (Uraga 1999:10):

- 1) Reconocedores dependientes de locutor (identifican a un solo hablante) y reconocedores independientes de locutor (identifican a cualquier locutor).
- 2) Reconocedores de palabras aisladas (dígitos, nombres de personas, etcétera) y reconocedores de habla continua (capaces de reconocer y decodificar frases y oraciones).

¹² El artículo “Modeling pronunciation variation for ASR: A survey of the literature”, de Strik y Cucciarini (1999), ofrece un claro y completo panorama sobre el reconocimiento de habla (tipos de reconocedores, funcionamiento y propósitos, etcétera).

Algunas de las aplicaciones de los reconocedores de habla son: el dictado automático, la navegación por la red electrónica mundial a través del habla, mandos en cabinas de aviones, climatización de coches, dictado de informes médicos y apoyo a personas con discapacidades auditivas o motrices (Llisterri y Martí 2002). En el caso del Proyecto DIME, como veremos más adelante (§2.4.), se busca crear un programa de reconocimiento de habla a partir del cual un usuario pueda dar instrucciones vía oral a un programa para el diseño de cocinas.

Dentro de las muchas ventajas que ofrecen los reconocedores de habla está el hecho que de estos programas permiten liberar las manos y la vista, haciendo posible, por ende, la movilidad y la realización de tareas simultáneas. En el mundo moderno, los desarrollos del reconocimiento de habla se están aplicando cada vez en más áreas, sobre todo en la industria y el comercio. En el mercado hay dos programas comerciales para el dictado automático en español, que se pueden obtener en nuestro país a precios relativamente accesibles: *Via Voice*, de IBM¹³ y *Naturally Speaking*, de ScanSoft.¹⁴

Uno de los desarrollos más recientes del que tenemos noticia es el reconocedor de habla SCANMail, un programa que transforma los correos de voz en texto, para facilitar la búsqueda de información oral (Bacchiani *et al.* 2001). El

¹³ <http://www-306.ibm.com/software/voice/viavoice>.

¹⁴ <http://www.scansoft.com/naturallyspeaking>.

programa integra sistemas de reconocimiento de habla y de recuperación de información.¹⁵

En cuanto a los proyectos de investigación lingüística computacional y tecnológica para la lengua española, cabe mencionar el proyecto BASURDE,¹⁶ “Desarrollo de un Sistema de Diálogo Oral en Dominios Restringidos” (Bonafonte *et al.* 1998), que busca crear sistemas de comunicación entre personas y computadoras para la consulta de precios, horarios e itinerarios de trenes de la compañía de ferrocarriles españoles RENFE.¹⁷ En este mismo país, la empresa Telefónica I+D ha logrado producir un reconocedor bastante sofisticado, pues es incluso capaz de reducir ruidos ambientales y aislar y eliminar interferencias para centrarse exclusivamente en la voz del usuario (Rodríguez *et al.* 2001).

A pesar de los importantes avances, la investigación sobre reconocimiento de habla todavía tiene muchos problemas por resolver. Entre sus mayores limitaciones están las irregularidades que de manera natural contiene el habla continua. Las computadoras no tienen todavía la capacidad para distinguir y caracterizar en una sola forma, realizaciones fonéticas tan variadas como [ˈpwes, pus, pos, ps, pu, ˈpue, ˈpueh],

¹⁵ En los Estados Unidos y en Europa, el reconocimiento de habla ha tenido grandes avances y ha producido una gran variedad de programas con múltiples aplicaciones. Uno de los grupos de investigación más avanzados es el de los Laboratorios Bell (*Bell Labs*), en Nueva Jersey, que desde hace más de cinco décadas ha dedicado grandes esfuerzos al reconocimiento de habla. En 1952, por ejemplo, se creó en este laboratorio el primer dispositivo de reconocimiento capaz de discriminar con cierta precisión los diez dígitos ingleses pronunciados de forma aislada por un único locutor (Casacuberta y Vidal 1987:1).

¹⁶ <http://gps-tsc.upc.es/veu/basurde/Home.htm>.

¹⁷ Basurde tiene su continuación en dos proyectos más recientes, DIHANA I: Sistema de diálogo distribuido para el acceso a la información mediante habla natural, y DIHANA II: Sistema de diálogo para el acceso a la información mediante el habla espontánea en diferentes entornos (<http://www.dsic.upv.es/~dihana/index.html>); de cualquier manera, tanto BASURDE como DIHANA representan proyectos de diálogo multimodales más que de reconocimiento de habla.

por *pues*, por poner un ejemplo; aún los sistemas más robustos son frágiles ante la presencia de nuevas palabras o del ruido. Por ejemplo, Acero (1995) ha destacado el hecho de que durante los últimos 30 años los investigadores no han logrado resolver muchos de los problemas que conlleva la construcción de un reconocedor de habla, y predice que no será antes del año 2025 cuando un sistema de comunicación simulado podría pasar la prueba de Turing.¹⁸

El funcionamiento de un reconocedor de habla se lleva a cabo a partir de tres módulos principales, según lo señalan diversos autores (Bonafonte *et al.* 1998; Strik y Cucchiaroni 1999; Rodríguez *et al.* 2001):

- 1) Diccionario.
- 2) Modelos de pronunciación.
- 3) Modelos acústicos.

El diccionario de un reconocedor de habla contiene todas las palabras que necesita el sistema en un dominio específico. Este dominio se determina de acuerdo con las características que tendrá el futuro programa de reconocimiento de habla. Por ejemplo, para el dominio del diseño de cocinas, que es el que, como mencioné, se estableció en el Proyecto DIME para el estudio del español de México, se enlistan todas las palabras que se determina y que se espera que pronuncie el hablante, como *alacenas*, *anaquel*, *fregadero*, *refrigerador*, etc. En la figura 1 se muestra un fragmento del léxico que incluye el diccionario del Corpus DIME.

¹⁸ A mediados del siglo pasado en el famoso artículo “Computing machinery and intelligence”, Turing (1950) lanzó a la comunidad científica internacional la pregunta de si las computadoras podían o no pensar, y si tendrían la posibilidad de aprender a través de fórmulas de acierto-error y premio-castigo como los niños (*child machine*). Si en algún momento de la historia, un humano no puede distinguir entre la manera de expresarse de otro humano y la de una computadora, ésta habrá pasado la prueba de Turing.

A	ALGO	ARRIME	BUSQUE	CHICO
AQUÍ	ALGUNA	ASISTENCIA	BUSQUEMOS	CHICOS
ABAJO	ALGUNAS	ASISTENTE	BÓRRALO	CHIN
ABRAN	ALGUNO	ASOCIARLO	BÚSCAME	CHIQUITA
ABRE	ALGUNOS	ASÍ	CABE	CIENTO
ABRIR	ALGÚN	ATINÉ	CABER	CIERTO
ABRIRLA	ALGUNAS	ATRÁS	CABRÁ	CINCO
ABRIRÍA	ALINEADO	AY	CABÍA	CINCUENTA
ABSORBE	ALINEADOS	BAJAR	CAFÉ	CLARO
ACABAMOS	ALINEAR	BAJARLA	CAFÉS	COCINA
ACABAS	ALLÍ	BAJO	CAJONES	COCINAS
ACABO	ALMACENAR	BAJOS	CAJÓN	COINCIDIERA
ACCESORIOS	ALMACÉN	BAJÓ	CAMBIAMOS	COINCIDIR
ACCIONES	ALREDEDOR	BANCO	CAMBIAR	COLGADAS
ACCIÓN	ALTA	BANCOS	CAMBIARLE	COLOCA
ACOMODAR	ALTAS	BARRA	CAMBIARLO	COLOCADA
ACOMODARLO	ALTO	BARRAS	CAMBIARSE	COLOCADOS
ACOMODARLOS	ALTURA	BARRITA	CAMBIEMOS	COLOCAMOS
ACUERDO	AMABLE	BASE	CAMBIO	COLOCAR
ACÁ	ANAQUEL	BAÑO	CAMINAR	COLOCARLA
ADECUADA	ANCHA	BIEN	CAMPANA	COLOCARLO
ADELANTE	ANGOSTA	BLANCA	CAMPANAS	COLOCO
AFUERA	ANGOSTAS	BLANCAS	CARA	COLOQUE
AGACHADO	ANTERIOR	BLANCO	CASA	COLOQUES
AGREGARLE	ANTES	BLANCOS	CASI	COLOR
AH	APARTE	BORDE	CATÁLOGO	COLORES
AHORA	APRECIAR	BORRAR	CATÁLOGOS	COMEDOR
AHORITA	AQUEL	BORRE	CENTRADA	COMEDORES
AHÍ	AQUÍ	BORRES	CENTRARLO	COMO
AL	ARMARIO	BUEN	CENTRO	CON
ALACENA	ARMARIOS	BUENO	CENTÍMETROS	CONECTORES
ALACENAS	ARRIBA	BUSCANDO	CERCA	CONFUSIÓN
ALACENITA	ARRIBITA	BUSCAR	CERRAR	CONGELADOR
ALCANZA	ARRIMAR	BUSCAS	CHICA	CONJUNTO

Figura 1. Fragmento del diccionario del Corpus DIME

Una vez que se ha obtenido el diccionario, se crean los modelos de pronunciación mediante la realización de un listado con las transcripciones fonológicas y con las de todas las realizaciones fonéticas posibles para cada una de las palabras incluidas en éste. En la figura 2 se presenta un fragmento del archivo de los modelos de pronunciación del Corpus DIME. En la primera columna se encuentran las palabras del diccionario; en la segunda, la transcripción fonológica, y en la tercera algunas realizaciones fonéticas. Como se detallará en el capítulo 3, las transcripciones en formato digital se realizan con un alfabeto fonético computacional, en este caso Mexbet (§3.3.).

A	a	a a_j a_2
AQUÍ	a k i	a k i a k_j
ABAJO	a b a x o	a V a x o a V a x O
ABRAN	a b r(a -N	a V r(a n a V r(a n_[
ABRE	a b r(e	a V r(a N a V r(e
ABRIR	a b r(i -R	a V r(E a V r(i r(
ABRIRLA	a b r(i -R l a	a V r(i r a V r(i r(l a a V r(i r(l a_j a V r(i r(l a_2

Figura 2. Ejemplo de los modelos de pronunciación del Corpus DIME

En la fase de entrenamiento de un reconocedor de habla, se programa al sistema para que diseñe un modelo acústico para cada uno de los fonemas. Por ejemplo, con las *aes* de *alacenas*, *anaquel*, *fregadero* y *refrigerador* que hayan dicho diferentes locutores, el sistema modela una forma prototípica para determinar la cualidad acústica de /a/. Así, los modelos acústicos contienen la información sobre los fonemas combinados en su uso en la lengua (es decir, esas *aes* de diferentes voces y en diferentes contextos: inicial, medial, final, tónica, átona, los sonidos que la rodean, etcétera), y proporcionan información sobre las propiedades y características

de los sonidos del habla en sus contextos. Un ejemplo de ello se presenta en la figura 3.¹⁹

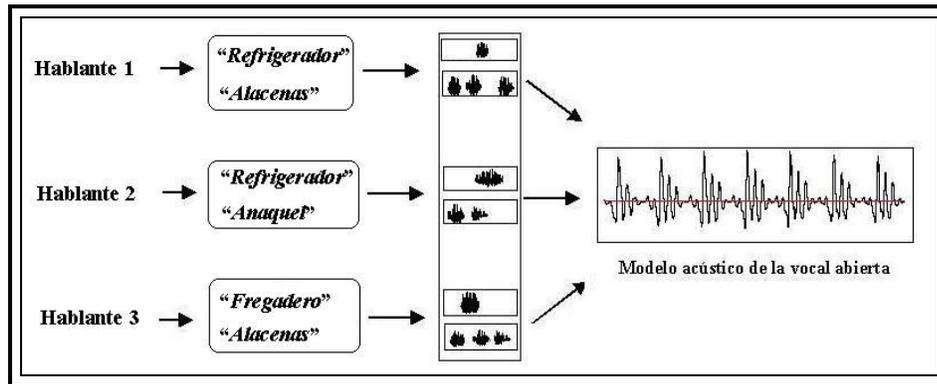


Figura 3. Ejemplo del proceso de creación de un modelo acústico para la vocal abierta

A partir de estos tres módulos principales, el reconocedor de habla funciona de la siguiente manera: como apreciamos en la figura 4, un hablante emite la palabra *refrigerador*; el programa recibe la señal sonora de la voz y busca uno a uno en los modelos acústicos aquellos que más se parecen a cada segmento fónico que emitió el usuario. Posteriormente, y apoyado en los modelos de pronunciación, el sistema encuentra en el diccionario la palabra emitida por el hablante.

¹⁹ Los métodos más comunes para crear los modelos acústicos de los reconocedores de habla, son los que utilizan modelos ocultos de Markov, redes neuronales, o la combinación de ambos. Para los modelos ocultos de Markov y las redes neuronales aplicados a experimentos de reconocimiento de habla, los siguientes textos resultan útiles: Casacuberta y Vidal (1987), Ali (manuscrito), Olivier (1999) y Uruga (1999).

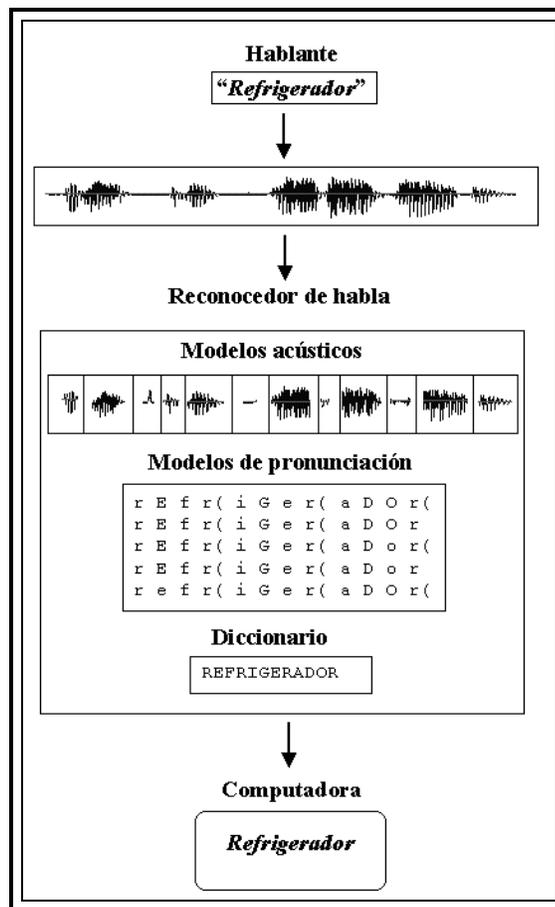


Figura 4. Funcionamiento de un reconocedor de habla

Para que se realicen los modelos acústicos, se cree el diccionario y se definan los modelos de pronunciación, es necesario contar con un corpus oral previamente etiquetado mediante un alfabeto fonético computacional. Se busca así proporcionar al programa la mayor cantidad posible de datos para que pueda diseñar los modelos acústicos. Los modelos de pronunciación se crean, asimismo, a través de reglas fonéticas motivadas por el contexto fónico. De tal suerte, parte del trabajo de esta tesis es determinar los alófonos del español de México que deberá incluir el

reconocedor de habla del Proyecto DIME, a partir del análisis fonético del Corpus DIME.

2.4. El Proyecto y el Corpus DIME

DIME es un proyecto de investigación en tecnologías del habla para el español de México que se desarrolla en el IIMAS desde 1999. En sus orígenes, el Proyecto comprendió tres objetivos para la construcción de un sistema conversacional (Pineda *et al.* 2001; Villaseñor, Massé y Pineda 2001):

- 1) La constitución de un corpus y su transcripción.
- 2) La construcción de un programa de reconocimiento de habla para el español de México.
- 3) La definición de una gramática para el español de México.

El producto del primer módulo es el Corpus DIME, un corpus oral en el dominio del diseño de cocinas. Al definir su aplicación futura, el Proyecto DIME decidió la creación de una interfaz oral asistida por computadora que sirviera para diseñar cocinas, porque este dominio es abundante en referencias espaciales, suficientemente complejo y a la vez concreto (Villaseñor, Massé y Pineda 2001).²⁰ El Corpus DIME se utiliza dentro del Proyecto para realizar los modelos acústicos del reconocedor de habla, y para definir la cobertura de la gramática del español y el diccionario que se desarrolla en el contexto del sistema (Pineda *et al.* 2001).

²⁰ A partir del Corpus DIME se determinó el léxico necesario, que como vimos incluye palabras como *alacena, anaquel, fregadero* y *refrigerador*, hace de lado referentes poco aplicables al dominio del diseño de cocinas, como *extinguidor* y *botella*, y elimina palabras como *calcetín, elevador* o *semáforo*, por resultar para el caso prácticamente inútiles.

Para la producción de un sistema de reconocimiento de habla, otro aspecto lingüístico que es necesario considerar es la determinación de quién hará uso del sistema, si un castellano o un mexicano, por ejemplo, no solamente para determinar si se incluye la palabra *refrigerador* o la palabra *nevera*, o ambas, sino también para la definición de la pronunciación de los usuarios. En el Proyecto DIME se ha definido que la cobertura lingüística de su programa de reconocimiento está centrada en el español de México.

El Corpus DIME consta de 31 diálogos entre una persona y un sistema de cómputo que le ayuda a diseñar una cocina. El corpus tiene un total de 7:10 horas de grabación digital²¹ (14 minutos en promedio por diálogo); contiene 27,459 palabras (886 en promedio por diálogo) y 5779 enunciados²² (185 en promedio por diálogo) (Pineda *et al.* 2001). Está integrado por 17 hablantes, con una media de edad de 30 años, la mayoría de los cuales estaba relacionada con las ciencias computacionales (Villaseñor, Massé y Pineda 2001). Respecto de la procedencia geográfica de los informantes, éstos fueron seleccionados al azar, lo cual coincide con uno de los modelos de muestreo enlistados por Llisterri (1991), tomando como base la idea de

²¹ El programa que se utilizó fue *Wave Studio*, de la compañía *Creative* (Villaseñor, Massé y Pineda 2001).

²² Los diálogos fueron grabados sin interrupciones. Para que fueran más manipulables, posteriormente se editaron, cortando ruidos, interjecciones o silencios muy largos, obteniéndose una segunda versión de cada uno, que fue ordenada en directorios y archivos. Posteriormente, se segmentaron en frases u oraciones que daban una idea completa, a las que se les llamó elocuciones (Villaseñor, Massé y Pineda. 2001) (es preciso señalar que en los desarrollos de lingüística computacional, la palabra *utterance* es el término inglés utilizado con más frecuencia para referirse a la unidad elocución). Finalmente, se realizó la transcripción ortográfica de todos los diálogos.

que el habla de la ciudad de México representa cuantitativamente al país.²³ Una muestra de estos diálogos se encuentra en la figura 5.

1	Sistema:	¿Quieres que desplace o traiga algún objeto a la cocina?
2	Usuario:	No.
3	Usuario:	¿Puedes mover la estufa hacia la izquierda?
4	Sistema:	¿Hacia dónde?
5	Usuario:	Hacia... ¡hacia la derecha!
6	Sistema:	Hacia la derecha.
7	Sistema:	Okey.

Figura 5. Fragmento del diálogo s6-t1-g1 del Corpus DIME

Así, durante la fase de grabación de los diálogos del Corpus DIME, los hablantes pedían de manera oral a la computadora que moviera y ordenara los diferentes objetos en la cocina. Los usuarios veían la disposición del espacio y las medidas de cada elemento en un plano de dos dimensiones y en otro de tres dimensiones. Asimismo, cada sujeto podía seleccionar los elementos que quería

²³ Aunque el punto de vista que expone López Chávez (1981) en su trabajo sobre la /s/ mexicana no está a favor de una mentalidad centralista que haga extensiva el habla de la ciudad de México al resto del país, varios autores han detectado la representatividad del habla de la capital. Canfield (1981:62) menciona que la población de las tierras altas de México representa a la gran mayoría de la nación. Más tarde, Perissinotto (1975:21) habla de la “fuerza de estandarización” de la ciudad de México. Moreno de Alba (1994:21) señala la idea de irradiación lingüística de las grandes ciudades. Ya para principios de los años sesenta, Lope Blanch (1963-1964:58) afirmaba que el habla de la ciudad de México era “la modalidad lingüística más importante, no sólo de la altiplanicie mexicana, sino de toda la República”; pocos años después, reafirmaba: “Como bien se sabe, las grandes ciudades han sido siempre focos lingüísticos con enorme poder de irradiación; y lo son todavía más en nuestro tiempo, ya que la concentración en las grandes urbes de los poderosos medios de difusión verbal modernos (radio, televisión, imprenta, cinematógrafo, etcétera) permiten que las modalidades lingüísticas urbanas lleguen rápidamente a los más alejados rincones del país. Por otro lado, las grandes ciudades de nuestro tiempo, por su inmenso poder de atracción, reúnen en su seno a hablantes de muy diversa procedencia regional, y actúan así como crisol en que se funden las distintas tendencias o peculiaridades idiomáticas de todo el país” (Lope Blanch 1969:46).

incorporar a la cocina desde un menú que se encontraba en la parte baja de su pantalla. Una muestra del diseño se presenta en la figura 6.

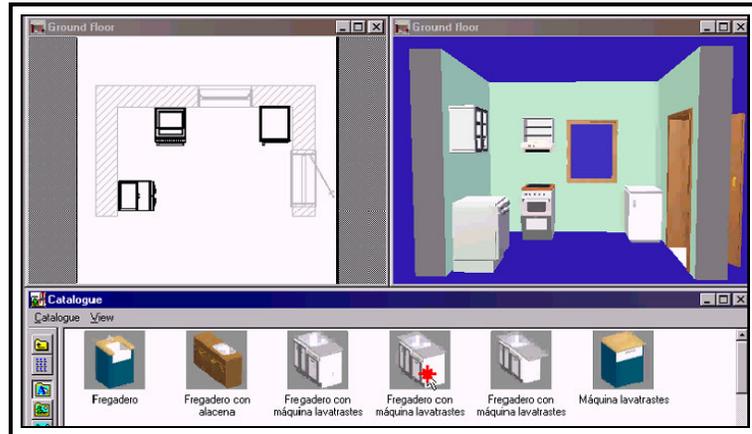


Figura 6. Aspecto del programa de diseño de cocinas utilizado por el Proyecto DIME para la grabación de su corpus

Durante la grabación del Corpus DIME, en la labor de diseño de la cocina, las funciones del sistema de reconocimiento de habla eran simuladas por una persona, que debía actuar como si fuera una computadora; este es el tipo de experimentos con el protocolo del *Mago de Oz*.

2.4.1. Los experimentos con el protocolo del *Mago de Oz*

Los experimentos realizados con el protocolo conocido como *Mago de Oz* se utilizan en los desarrollos sobre tecnologías del habla para estudiar el comportamiento humano durante la interacción entre una persona y una computadora (Dahlbäck *et al.* 1993). De hecho, otro de los aspectos importantes del Corpus DIME es que, hasta

donde he podido investigar, es el único corpus en nuestro país que ha sido grabado con este protocolo.

Para estudiar las necesidades de los futuros programas de cómputo, los experimentos *Mago de Oz* tienen como finalidad caracterizar el habla de los usuarios (qué tipo de oraciones se construyen, qué recursos estilísticos emplean, cuáles son las muletillas o frases recurrentes, etc.), delimitar el léxico apropiado para dicho programa (diccionario del dominio) y estudiar el comportamiento humano en interacción con un sistema computacional. En estos experimentos, se pide a un sujeto que desarrolle una tarea en la computadora a través de peticiones y órdenes. El sujeto debe comunicarse con el sistema oralmente, sin el uso del teclado o del *mouse*. Se le informa a este usuario que el sistema será capaz de escucharlo y entenderlo, y que le responderá con voz sintética.²⁴ Lo que el sujeto no sabe es que el funcionamiento del sistema está siendo simulado por otro humano, un *Mago de Oz*, que se encuentra oculto manipulando el sistema y viendo en un monitor lo que el sujeto realiza,²⁵ escuchando sus peticiones y órdenes, y cumpliéndolas. El comportamiento del *Mago* debe imitar perfectamente a una computadora: debe responder de la manera más rápida posible, y no cometer errores (Villaseñor, Massé y Pineda 2001). Durante el desarrollo de los experimentos con el protocolo del *Mago de Oz*, el *Mago* debe seguir también otros lineamientos: no puede interrumpir al usuario mientras éste habla; no

²⁴ Es preciso mencionar que en el caso del Corpus DIME, no se utilizó una voz sintética para el *Mago*.

²⁵ Recordemos al famoso personaje de la novela de Frank Baum: el *Mago de Oz*, que oculto, se presentaba ante Dorothy, ante el león, ante el hombre de paja y ante el de hojalata con diferente disfraz, pretendiendo ser una bola de fuego, una bella dama, etc., manipulando una máquina a la que podemos comparar con una computadora.

puede inferir más de lo que se ha dicho durante el habla; debe aparentar que no entiende oraciones largas, complejas o poco claras y, si hay lugar a ambigüedades, debe requerir confirmación de la información (Villaseñor, Massé y Pineda 2001). El resultado de los experimentos se transcribe y se analiza para el desarrollo futuro del proyecto.

El ejemplo de la figura 7 está adaptado a una situación en la ciudad de México a partir del artículo “Towards conversational human-computer interaction”, de Allen y sus colaboradores (2001) del Departamento de Ciencias de la Computación de la Universidad de Rochester. Este grupo ha realizado ya importantes avances en reconocimiento de habla, y su principal interés se dirige ahora hacia el problema de la interacción entre una persona y una máquina, así como el entendimiento entre ambas:

1	Usuario:	Debemos trasladar a una mujer de Tehuacán al “20 de noviembre”.
2	Sistema:	Entendido.
3	Usuario:	¿Qué vehículos están disponibles?
4	Sistema:	Hay ambulancias en Miguel Ángel de Quevedo y en Félix Cuevas .
5	Usuario:	Envíen una de Miguel Ángel.
6	Sistema:	¿Sabes que Universidad está bloqueada?
7	Usuario:	Tomemos entonces Insurgentes.
8	Sistema:	Correcto; enviaré al personal.

Figura 7. Ejemplo de un diálogo entre una persona y un sistema multimodal adaptado de Allen *et al.* (2001)

El sistema deberá ser capaz de reconocer uno a uno los segmentos fónicos; reconstruirlos, y entender que la mujer que está herida no es nativa del estado de Puebla, sino que *Tehuacán* es una calle, así como *Miguel Ángel de Quevedo* y *Félix Cuevas* no son personas. Debe discernir que *Universidad* es una avenida y que no

está bloqueada por alguna manifestación del CGH sino por obras de construcción; que *20 de noviembre* no es una fecha sino un hospital, y que *vehículos* se refiere específicamente a *ambulancias*. Toda esa información con la que un ser humano está familiarizado en su vida cotidiana es idealmente lo que la máquina necesita saber y procesar correctamente.

Las características de la conversación están siempre dadas por las situaciones comunicativas o por la forma del mensaje; es distinta la información lingüística que arroja una conversación a aquella que muestra un texto escrito. También es distinta la manera como una persona se dirige a sus superiores, a sus colegas de trabajo o a sus amigos: hay un tono profesional, un tono familiar, otro personal, etc.²⁶ En el acto lingüístico, es natural que los seres humanos adapten su modo de hablar a su interlocutor (Dahlbäck *et al.* 1993; Strik y Cucchiarini 1999). Ante una computadora, una persona se ubica en otra realidad y muestra una actitud totalmente diferente, más cuidada, como si le hablara a un niño pequeño o a un extranjero, buscando siempre la certeza de que un sistema artificial lo entienda plenamente.

Si he repasado brevemente el protocolo del *Mago de Oz* y las características de la interacción persona-máquina, ha sido para destacar que en las tecnologías del habla existen puertas abiertas para el estudio de todos los niveles de la lengua, dándose cabida para especialistas en sintaxis, semántica e, incluso, pragmática, además de fonetistas y fonólogos.

²⁶ Un acertado repaso de las diferentes conductas humanas que se presentan durante la interacción entre personas y computadoras es la tipología de situaciones comunicativas que hacen Aguilar y sus colaboradores (1994). Entre ellas, cabe destacar la imagen que el hablante desea proyectar, así como la que tiene del receptor y el tipo de información que se está comunicando, etcétera.

Después de la grabación de un corpus oral, de la segmentación en elocuciones y de la creación de directorios y archivos, se inicia la etapa de transcripción fonética y fonológica, para la creación de modelos acústicos, de modelos de pronunciación y del diccionario de pronunciación a los que he aludido (§2.3.1.).

2.5. Etiquetado fonético computacional

Como hemos visto (§2.3.1.), durante la fase de entrenamiento de los programas de reconocimiento de habla, la información lingüística debe extraerse de la señal sonora para proveer al sistema con datos precisos sobre los sonidos de las lenguas (Strik y Cucchiaroni 1999). Después de que se ha grabado un corpus oral, se lleva a cabo una segmentación en la que se señala dónde empieza y dónde acaba cada elemento fónico, y se le añade una etiqueta que especifica sus propiedades a partir de un símbolo fonético. El etiquetado fonético del Corpus DIME se empezó a realizar de manera manual en abril de 2002 por mí y por un equipo de trabajo conformado por estudiantes del Colegio de Letras Hispánicas de la Facultad de Filosofía y Letras de la UNAM.²⁷ En la figura 8 se muestra la manera de segmentar y transcribir una señal de habla con el alfabeto fonético computacional Mexbet y con la herramienta *SpeechView*.

²⁷ Participaron en el etiquetado Haydee Castellanos Vargas, Fernanda López Escobedo, Ivonne López Morales y Norma Ortega Bernal. Posteriormente, las transcripciones y alineaciones fueron revisadas y afinadas por mí para garantizar una uniformidad.

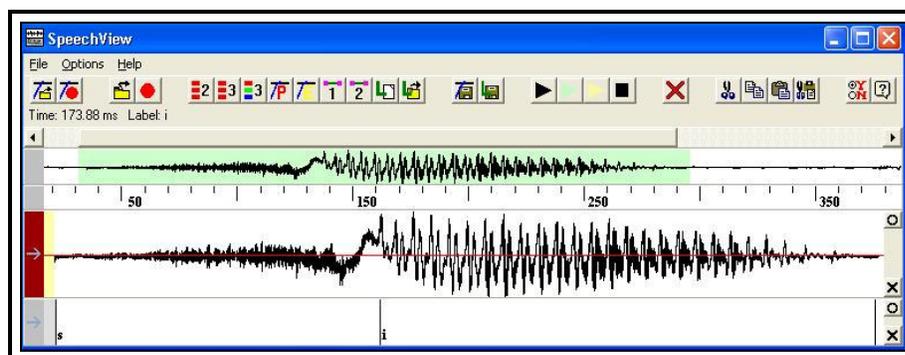


Figura 8. Sujeto 5: “Sí”. Segmentado y etiquetado fonético con Mexbet del archivo 2 del diálogo s5-t1-g1 del Corpus DIME, con el programa *SpeechView*

En esta fase de segmentado y etiquetado, el programa que ha elegido el Proyecto DIME es el *SpeechView*, una de las herramientas del paquete *CSLU Toolkit* elaborado por el *Center for Spoken Language Understanding*, del *Oregon Graduate Institute of Science and Technology* que se ofrece de manera gratuita a la comunidad científica internacional.²⁸ El *SpeechView* permite analizar la señal sonora del habla a partir de varias representaciones: dos oscilogramas (uno fijo y uno manipulable), un espectrograma de banda ancha (que distingue con mayor precisión los formantes vocálicos), y otro de banda estrecha (que muestra mejor los componentes armónicos, y además aparece en blanco y negro y a colores). Todos estos instrumentos se pueden desplegar fácilmente en una sola “ventana”, como se muestra en la figura 9.

²⁸ El programa se puede obtener actualmente en la red electrónica mundial en la página <http://www.cslu.ogi.edu/toolkit/download/index.html>.

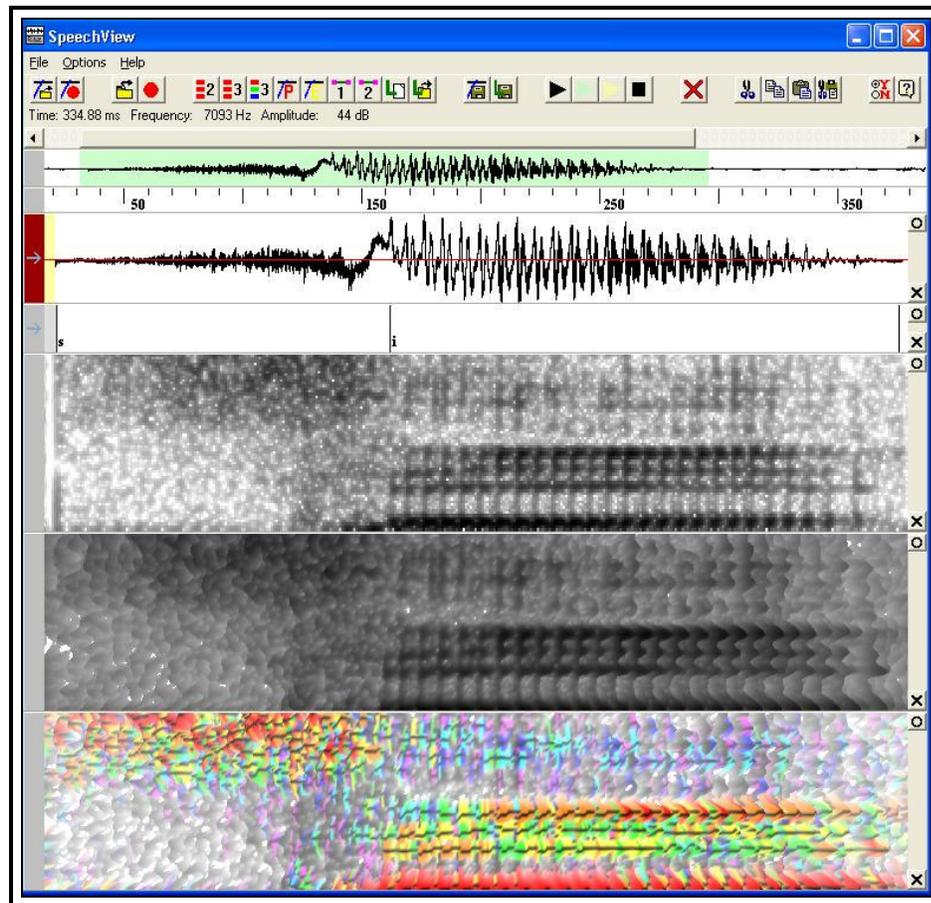
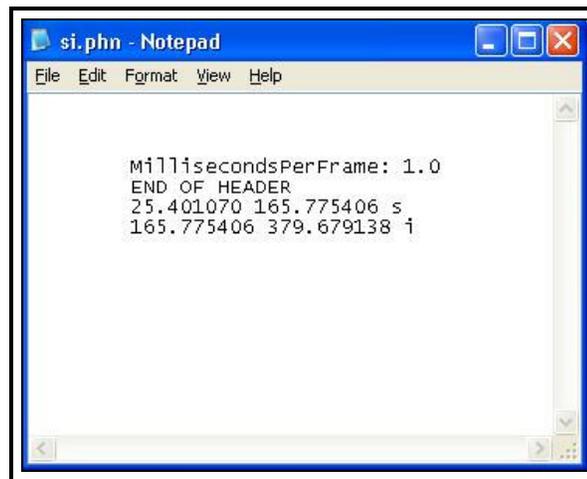


Figura 9. El programa *SpeechView* muestra dos oscilogramas y dos espectrogramas, segmentados y alineados con el etiquetado fonético

Además de este práctico formato visual, otra ventaja de etiquetar con el *SpeechView* —comparado con otros instrumentos— es que la elaboración de etiquetas resulta más sencilla, en cuanto a que es más fácil segmentar la señal e insertar los símbolos fonéticos computacionales. Quizá la utilidad más importante del *SpeechView* es que las etiquetas se pueden crear en un sistema operativo de uso más generalizado entre los lingüistas, como lo es *Windows*, y posteriormente generar compatibilidad para el intercambio de información con otros sistemas operativos, como *Linux*, por ejemplo.

El etiquetado fonético computacional tiene como propósito, pues, realizar una alineación fonética generando archivos que señalen las propiedades acústicas y la duración exacta en milisegundos de los segmentos fónicos en la señal de la voz (Llisterri 1997; Olivier 1999:vi). Con el *SpeechView* se generan etiquetas que se guardan en archivos con extensión *.phn, que posteriormente pueden también ser leídas con cualquier lector de texto, para cuantificar la duración exacta de cada segmento. En la figura 10 apreciamos que la duración de [s] del archivo 2 del diálogo s5-t1-g1 va del milisegundo 25.401070 al 165.775406, y la de [i] del milisegundo 165.775406 al 379.679138.



```

si.phn - Notepad
File Edit Format View Help

MillisecondsPerFrame: 1.0
END OF HEADER
25.401070 165.775406 s
165.775406 379.679138 i

```

Figura 10. Archivo *.phn del etiquetado fonético, visualizado con un lector de texto

Así, contar con corpus orales transcritos en formato electrónico permite la búsqueda de información de manera rápida, favoreciendo la cuantificación y el análisis de los datos (Llisterri 1997 y 1999). Una vez que un corpus oral ha sido segmentado y etiquetado, esta información se integra al sistema para entrenar al

reconocedor de habla en espera de que, al recibir la señal sonora de la voz humana, analice uno a uno cada segmento y reconstruya el mensaje.

En todos estos procesos, pues, se pone en evidencia a manera como se relacionan la lingüística y la computación en las tecnologías del habla: análisis acústico, transcripciones fonéticas y fonológicas computacionales, etc. De ello se desprende la necesidad de contar con un alfabeto fonético apropiado. Para nuestro caso es preciso detenernos en Mexbet, el alfabeto fonético computacional para el español de México que utiliza el Proyecto DIME.

3. UN ALFABETO FONÉTICO COMPUTACIONAL PARA EL ESPAÑOL DE MÉXICO: *MEXBET*

No deja de resultar hasta cierto punto paradójico que, siendo el habla el medio de comunicación por excelencia entre las personas, la interacción con los equipos informáticos se lleve a cabo casi siempre mediante la escritura.
Llisterri et al. 2003:1

Como he mencionado, en la práctica fonética y en la teoría fonológica existen alfabetos especializados para transcribir los sonidos del habla. En el entorno de nuestro país se trabaja principalmente con dos: el de la Asociación Fonética Internacional (AFI) y el de la *Revista de Filología Española*. En contraste con el AFI, que tiene una amplia difusión en todo el mundo, la aplicación del alfabeto hispánico se dirige principalmente a los estudios de lengua española.

Por otro lado, en la práctica lingüística computacional se han tenido que adaptar los alfabetos fonéticos tradicionales a los requisitos de los sistemas informáticos para evitar confusiones en el intercambio de información electrónica. El hecho es simple de explicar: símbolos como f , ϑ o η no son compatibles con los programas computacionales para transcripciones, mediciones y contabilizaciones. Estos símbolos tienen que ser reemplazados por caracteres universales como S, @ o N, según el código que se utilice.

En el presente capítulo veré primero las características generales de estos dos alfabetos fonéticos tradicionales. Posteriormente, revisaré algunos de los alfabetos

fonéticos computacionales que existen para el español y las particularidades de cada uno. Me detendré al final en Mexbet, un alfabeto fonético computacional para el español de México. Revisaré sus antecedentes (Uraga y Pineda 2000 y 2002) y presentaré una nueva propuesta, en la que puntualizo su uso e incluyo conocimientos lingüísticos básicos para perfeccionarlo.

3.1. Alfabetos fonéticos tradicionales

Antes de llegar a Mexbet, es oportuno detenerse con más detalle en el AFI y en el alfabeto hispánico para ver qué aporta cada uno.

3.1.1. La tradición hispánica: el alfabeto de la *Revista de Filología Española*

La *Revista de Filología Española*, que fue fundada en 1914 en el Centro de Estudios Históricos en Madrid, publicó en 1915 su propio alfabeto fonético (alfabeto de la *RFE* o alfabeto hispánico). Entre los creadores de éste se encontraba Navarro Tomás (1966-1967:5), quien señala que fue necesario elaborar las convenciones para la transcripción de los sonidos de la lengua española y para algunas de sus variantes dialectales (peninsulares: Andalucía y Asturias, por ejemplo; americanas: Chile, México, Puerto Rico, etcétera), debido a que se había advertido “la necesidad de disponer de un alfabeto fonético, para los artículos que hubieran de requerirlo y para los estudios de dialectología, fonética y geografía lingüística que el Centro habría emprendido”.

De acuerdo con Navarro Tomás (1918), el propósito original del alfabeto de la *RFE* no fue la invención de un sistema nuevo, sino la adaptación de las ventajas que ofrecían los sistemas de transcripción más conocidos en la época.²⁹ El alfabeto de la *RFE* establece los símbolos y los diacríticos para los alófonos de la lengua española y para algunas variantes dialectales. Una muestra de su formato puede verse en la figura 11.

ALFABETO FONÉTICO	
Para la transcripción fonética de los sonidos, la REVISTA DE FILOLOGÍA ESPAÑOLA se servirá del siguiente alfabeto:	
Bilabiales.	Dentales.
b esp. bondad. boṽdáṽ	d esp. ducho.. dúçho
p esp. padre.. pádre	t esp. tomar.. tomár
m esp. amar... amár	ṽ esp. monte.. móṽte
ṽ and. mismo.. míṽmo	ṽ esp. desde.. déṽde
ḃ esp. haba... ába	ḣ esp. hasta... áḣta
ḥ and. las botas la ḥótah	ḥ esp. falda... fáḥda

Figura 11. Fragmento del “Alfabeto fonético” publicado por la *Revista de Filología Española* (1915)

Como he mencionado, el alfabeto de la *RFE* ha gozado de gran aceptación en el mundo hispánico (Ríos Mestre 1999);³⁰ ha probado ser eficiente para representar

²⁹ Algunos autores que han tratado de determinar posteriormente los orígenes de las convenciones de este alfabeto no han llegado a un acuerdo preciso. Según Gili Gaya (1950:75), tiene sus antecedentes en el Internacional; para Martínez Celadrán (1984:147) se basa en el alfabeto de los filólogos romanistas. Para Gil (1988:110), este alfabeto se basó fundamentalmente “en el alfabeto empleado de forma general por los filólogos romanistas europeos”. De la lectura de Navarro Tomás (1966-1967:9) se desprende que algunas convenciones fueron tomadas también del Alfabeto Fonético Internacional.

³⁰ La gran aceptación del alfabeto de la *RFE* provocó, de hecho, que en algún momento fuera también utilizado para otras lenguas hispánicas, como el catalán y el gallego (Martínez Celadrán 1984:147), pero

los sonidos del español de México y todos los trabajos lingüísticos que he consultado —salvo los más recientes— basan sus convenciones en él. Además, es el alfabeto que se enseña en casi todas las universidades de nuestro país. Lo anterior ha provocado falta de entendimiento y aún confusiones en cuanto a la dificultad que tiene la comunidad lingüística internacional para comprenderlo.³¹ No obstante, en últimas fechas los hispanistas han comenzado a utilizar el Alfabeto Fonético Internacional, en una búsqueda por ampliar la difusión de sus trabajos.

3.1.2. El Alfabeto Fonético Internacional

La Asociación Fonética Internacional (la AFI)³² fue fundada en París en 1886 con el nombre *Dhi Fonètik Ticerz' Asóciécon* (IPA 1949). Estuvo conformada por un pequeño grupo de profesores de idiomas que, bajo la guía de Paul Passy, se reunieron para definir una notación fonética que ayudara en la adquisición de lenguas extranjeras (IPA 1999:195). En 1889 el nombre cambió por el de *Association Phonétique des Professeurs de Langues Vivantes*, y fue hasta el año de 1897 cuando se convirtió en la *Association Phonétique Internationale*. En sus orígenes, las metas de la AFI eran, entre otras, facilitar y mejorar la adquisición de lenguas extranjeras, dar las herramientas para la representación de los sonidos de lenguas que no tenían una ortografía y procurar el acercamiento amistoso entre las naciones (IPA 1949).

“resulta inadecuado para la descripción fonética de otras lenguas no románicas” (De la Mota y Ríos 1995:98).

³¹ He podido comprobar a través de alumnos extranjeros, que es difícil enseñarles las convenciones del alfabeto de la *RFE* una vez que ya están familiarizados con el AFI.

³² Para evitar confusiones, me referiré a la Asociación Fonética Internacional como “la AFI” y al Alfabeto Fonético Internacional como “el AFI”.

Actualmente, la AFI cuenta con una publicación periódica, el *Journal of the International Phonetic Association (JIPA)*. Además de esta revista, de múltiples artículos y de las actas de sus congresos, son dos sus publicaciones más importantes: a mediados del siglo pasado la Asociación presentó sus *Principios (The principles of the International Phonetic Association, 1949)* y, 50 años después, su *Manual (Handbook of the International Phonetic Association, 1999)*, que sustituyó a la primera versión.³³

Desde el principio, la AFI estableció un alfabeto para representar todos los sonidos de todas las lenguas del mundo. Cabe destacar el hecho de que muchos de los símbolos que ha creado para los sonidos específicos de ciertas lenguas, han sido incorporados al sistema ortográfico de ellas, principalmente para algunas lenguas africanas (IPA 1949). La idea original de establecer un alfabeto fonético que sirviera para todas las lenguas —con o sin sistema ortográfico— fue sugerida por el fonetista danés Otto Jespersen a Paul Passy en junio de 1886 en una carta que fue publicada posteriormente en *Dhi Fonètik títcer* (IPA 1999:194). En ella, Jespersen mencionaba la necesidad de un alfabeto internacional, en lugar de los alfabetos particulares que existían para las diferentes lenguas.³⁴

Periódicamente, el AFI es revisado y mejorado; con un espíritu de apertura y de permanente perfeccionamiento, la Asociación declara que su alfabeto no está necesariamente completo, y que sus propuestas son susceptibles de mejorarse (IPA

³³ La Asociación Fonética Internacional también tiene presencia en la red electrónica mundial, en la página <http://www2.arts.gla.ac.uk/IPA/ipa.html>.

³⁴ No obstante, para esos tiempos el mismo Jespersen había creado ya su propio alfabeto para el danés (IPA 1999:197).

1949:19). La última versión es la que se presenta en el *Manual* (1999), que fue elaborada en la Convención Internacional de 1989, en Kiel (IPA 1999:3) y publicada años después en el *JIPA* (1993). Una diferencia entre el AFI y el de la *RFE* es que el primero es más flexible en sus convenciones tipográficas (para marcar la distensión de las vocales, por ejemplo, el AFI ofrece el uso del diacrítico [̄] o del diacrítico [̆], mientras que el de la *RFE* únicamente da como opción [̄]); además, el AFI permite que los diacríticos se ubiquen por encima o por debajo del símbolo base).

Los cuadros que presenta el AFI no son solamente una lista de símbolos. Vemos en la figura 12 de las consonantes pulmonares cómo en las filas y columnas se representa un claro sistema de clasificación fonológica por modo de articulación y zona de articulación, además de la sonoridad (sordas a la izquierda; sonoras a la derecha).

CONSONANTS (PULMONIC)											
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ɾ					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Figura 12. Consonantes (pulmonares) del Alfabeto Fonético Internacional

Cabe mencionar que, entre muchas otras, los *Principios* de 1949 incluyen la lengua española, pero no señalan con claridad la cualidad fonética o fonológica de los símbolos enlistados. La publicación de 1999, el *Manual*, revisa otras lenguas ya presentes en la primera, pero no retoma el español.³⁵ Las últimas ilustraciones del AFI para el español fueron presentadas en el *JIPA* el año pasado (Martínez Celdrán *et al.* 2003).

La adaptación del Alfabeto Fonético Internacional a nuestra lengua ha sido objeto de muchas interpretaciones y, hasta cierto punto, de confusiones.³⁶ Contrástense por ejemplo las diferencias en las equivalencias de los alfabetos de la *RFE* y del AFI que establecen Martínez Celdrán (1984) y Quilis (1993); recuérdese,

³⁵ Describe, no obstante el uso del AFI en otras lenguas peninsulares, como el gallego y el catalán.

³⁶ Navarro Tomás (1974:181-187) habla ampliamente de los problemas que representan las transcripciones fonéticas y fonológicas para el español.

además, que los mismos autores modifican sus convenciones de texto a texto. “Todo ello no facilita, evidentemente, la labor de los [...] investigadores, que en ocasiones deben realizar un permanente ejercicio de traducción entre transcripciones” (Gil y Llisterri, en prensa). No obstante, cabe mencionar que los dilemas sobre las convenciones alfabéticas fonéticas no son privativos del español. Pullum y Ladusaw (1986:xvi) en su reconocido libro *Phonetic symbol guide* mencionan los problemas a los que ellos mismos se han enfrentado:

We know what it is like to confront traditions conflicting with the ones in which we were trained: one of us was trained in the tradition associated with the International Phonetic Association (IPA) and later learned American transcription practices on the streets, and the other had the converse experience.

Por otra parte, una de las ventajas que debemos reconocer en el AFI es su carácter vanguardista. En su introducción, el *Manual* (1999) aclara que esta nueva versión no sólo es para profesores de fonética o fonetistas, sino que también sirve para estudios de patologías del lenguaje, teoría fonológica y, con una visión muy actual y moderna, para los trabajos de tecnologías del habla.³⁷ Así, en el segundo apéndice la AFI incluye el apartado “Computer coding of IPA symbols”³⁸ (IPA 1999:161-185), que es, en esencia, una revisión de un artículo previo de Esling y Gaylord (1993), en el que se muestra el modo de asignar códigos computacionales a los símbolos del AFI para que dicho alfabeto pudiera ser utilizado como medio de intercambio electrónico en los desarrollos de las tecnologías del habla. Esta labor

³⁷ Este *Manual* reconoce medio siglo de avances tecnológicos para el análisis acústico del habla y, de hecho, da por sentado que “many readers will be familiar with, and quite possibly working with, speech as an acoustic signal” (IPA 1999:vii).

³⁸ Disponible en la página electrónica <http://www.arts.gla.ac.uk/IPA/IPANumberChart96.pdf>.

tuvo su origen en un grupo de trabajo especial durante la Convención de Kiel en 1989 (IPA 1999:161), que asignó a cada símbolo fonético un número de tres dígitos conocido como “número AFI” (*IPA Number*). Por ejemplo, los números para /a/ y /b/, respectivamente, son 304 y 102. El uso de estas convenciones no es suficientemente práctico, pues para el etiquetado fonético computacional de grandes volúmenes de datos resultan muy poco legibles.

Hoy en día, las fuentes fonéticas del AFI forman parte del equipo original de cualquier computadora o pueden instalarse con rapidez y economía.³⁹ Haciendo un poco de historia, no obstante, Pullum y Ladusaw (1986:xxiii) nos recuerdan la práctica de sacar y voltear la hoja para transcribir la *schwa* u otros símbolos fonéticos, cuando la máquina de escribir era la herramienta más común entre los investigadores. A mediados de los años ochenta, cuando estos autores publicaron su guía, las computadoras no eran aún capaces de transcribir los símbolos de los alfabetos lingüísticos. No obstante, Pullum y Ladusaw (1986:xxiv) intuyeron que en un futuro cercano eso sería posible:

Finally, the medium of the future: word processing on computing machines [...] the creative designing of new characters and fonts is being made available by modern software.

³⁹ Las fuentes del AFI se pueden obtener desde la página de la Asociación Fonética Internacional en la red electrónica mundial: <http://www2.arts.gla.ac.uk/IPA/ipafonts.html>. También se encuentran disponibles de manera gratuita en la página http://www.sil.org/computing/catalog/encore_ipa.html.

3.2. Alfabetos fonéticos computacionales

Para la investigación lingüística computacional se utilizan caracteres que representan símbolos fonéticos y que pueden ser intercambiables entre diferentes sistemas operativos (*Windows*, *Linux*, *Macintosh*, etcétera). Estos son los llamados alfabetos fonéticos computacionales. Una de las primeras tareas que requiere un proyecto o programa de tecnologías del habla es la determinación, creación y adopción de un alfabeto fonético computacional (Acero 1995; Aguilar *et al.* 1997), los cuales se diseñan en códigos ASCII⁴⁰ con los caracteres disponibles en el teclado de cualquier computadora —o máquina de escribir en su época—, sin importar el sistema o programa que se utilice. Estos símbolos alfanuméricos están conformados principalmente por las letras del alfabeto romano y los números arábigos, más algunos símbolos comunes (paréntesis, guiones, puntos, comas, etc.). El uso de códigos ASCII es indispensable cuando se emplea una aplicación (un espectrógrafo, un sintetizador, etcétera), y cuando se desea procesar datos fonéticos a través del intercambio de *bytes* en las computadoras. Esta necesidad de símbolos fácilmente representables se detectó desde hace décadas. En cuanto a los esfuerzos de la Sección de Investigaciones Lingüísticas a finales de los años sesenta, por ejemplo, Manrique (1968) menciona que para aquel entonces ya se ponía en evidencia la urgencia de diseñar una forma de codificación especial para las transcripciones fonéticas de lenguas indígenas. En este proyecto, además de símbolos para cada fonema, se

⁴⁰ ASCII: *American standard coding for information interchange*, un código estándar americano para el intercambio de información.

establecieron también claves para codificar las clases de sonidos: P (para el orden labial: /p, b, f, v/), T (dental), K (velar), S (dorsal), Y (palatal), H (glotal), etcétera.

Para el español existen varios alfabetos fonéticos computacionales; algunos específicos para el español de México, además de Mexbet. A continuación revisaremos las características de SAMPA, OGibet y Worldbet, por la importancia que puedan tener en su relación con éste.

3.2.1. SAMPA

El alfabeto computacional *Speech Assessment Methodology Phonetic Alphabet* (SAMPA) fue desarrollado como parte del proyecto ESPRIT 1541 en 1987-1989 por un grupo internacional de fonetistas. Se hizo primero para el danés, francés, inglés, holandés e italiano (1989); después se elaboraron las versiones para el noruego y el sueco (1992), y posteriormente para el español, griego y portugués (1993). Entre sus propósitos, el espíritu de SAMPA busca simplicidad en la transcripción y relativa facilidad de utilización por parte de personas con poca formación en fonética (Llisterri 1997). El principio que rige el alfabeto SAMPA es básicamente fonológico, al igual que el AFI, empleándose solamente símbolos distintos en el caso de segmentos con valor diferencial (Llisterri 1997). Así, además de los 24 fonemas castellanos, como se puede apreciar en la figura 13, esta versión de SAMPA⁴¹ incluye

⁴¹ Disponible en la página <http://www.phon.ucl.ac.uk/home/sampa/spanish.htm>.

las aproximantes de /b, d, g/ ([B, D, G], en SAMPA), el palatal africado sonoro ([j]) y las paravocales ([j, w]).⁴²



SAMPA for Spanish

Consonants

	Symbol	Word	Transcription	
plosives				
	p	padre	"paDre	
	b	vino	"bino	
	t	tomo	"tomo	
	d	donde	"donde	
	k	casa	"kaSa	
	g	gata	"gata	
affricates				
	tʃ	mucho	"mutʃo	
	ʝ	hielo	"ʝjelo	
fricatives				
	f	fácil	"faTil	
	B	cabra	"kaBra	(= /b/)
	T	cinco	"Tinko	
	D	nada	"naDa	(= /d/)
	s	sala	"sala	
	x	mujer	mu"xer	
	G	luego	"lweGo	(= /g/)
nasals				
	m	mismo	"mismo	
	n	nunca	"nunKa	
	J	año	"aJo	
liquids				
	l	lejos	"lexos	
	L	caballo	ka"baLo	(or as ʝj)
	r	puro	"puro	
	rr	torre	"torre	
semivowels				
	j	rei	rrej	
		pie	pje	
	w	deuda	"dewDa	
		muy	mwi	
Vowels:				
	i	pico	"piko	
	e	pero	"pero	
	a	valle	"baLe	
	o	toro	"toro	
	u	duro	"duro	

Figura 13. El alfabeto fonético computacional SAMPA para el español castellano (Wells 1998b)

⁴² El término paravocal engloba a los segmentos que tradicionalmente se conocen en la bibliografía hispánica como semivocales y semiconsonantes. Ha sido tomado de Whitley (2000), pues combate la confusión que crea el anglicismo "glides" o el término "deslizantes" con que las llaman algunos.

De acuerdo con Wells (1998a), SAMPA es el mejor sistema de codificación ASCII para la notación fonética de las lenguas, aunque goza más bien de difusión solamente entre los desarrollos de tecnologías del habla realizados en Europa. La versión española de SAMPA ha tratado de expandirse a América y, en el marco del proyecto SpeechDat Across Latin America (SALA),⁴³ Moreno y Mariño (1998) establecieron el inventario de los fonemas y alófonos de 6 dialectos americanos (argentino, colombiano, chileno, mexicano, peruano y venezolano), pero solamente se concentraron en formalizar los sonidos de procedencia indígena. En el caso de México, presentan la propuesta que aparece en la figura 14.

Alófono	Descripción
[dl] ⁴⁴	Lateral africada sonora
[ts]	Alveolar africada sorda
[S]	Palatal fricativa sorda

Figura 14. Alófonos de procedencia indígena para el español de México, según Moreno y Mariño (1998)

En un trabajo posterior, Mariño y Moreno (2000) elaboraron el inventario completo de los fonemas y de los alófonos del español americano. Este esfuerzo por adaptar SAMPA a todas las variantes del español hablado en América es, sin duda, loable; sin embargo, no incluye muchísimas variantes dialectales americanas, como podrían ser los distintos tipos de aspiración de las áreas de costa. La forma aspirada, por ejemplo, que Moreno y Mariño (1998) sólo incluyeron para las hablas caribeñas,

⁴³ <http://www.sala2.org>.

⁴⁴ Merece la pena destacar el hecho de que Moreno y Mariño (1998) otorgan el rasgo [+sonoro] al fonema de procedencia indígena /tɬ/, cuando en realidad se trata de un fonema sordo.

también debería considerarse para el español de México si se hacen estudios dialectales para Yucatán, Tabasco y Veracruz, por ejemplo.

3.2.2. OGIbet

El *Oregon Graduate Institute of Science and Technology* realizó su alfabeto fonético (*Oregon Graduate Institute Alphabet*, OGIbet) para etiquetar a nivel de fonema y alófono los corpus orales en inglés del *Center for Spoken Language Understanding* (CSLU) (Lander 1997). Por su parte Tlatoa, el Grupo de Investigación en Tecnologías del Habla (Tlatoa 2000b), de la Universidad de las Américas de Puebla, realizó una adaptación al español de México (Tlatoa 2000a), la cual se puede apreciar en la figura 15.

OGIbet Mexican Spanish

Worldbet Symbol	Example Word	Worldbet Pronunciation
/p/	peso	p e s o
/b/	beso	b e s o
/t/	tos	t o s
/d/	dos	d o s
/k/	casa	k a s a
/g/	gasa	g a s a
/f/	fifa	f i f a
/s/	sisá	s i s á
/r/	misma	m i s m a

Figura 15. El alfabeto fonético computacional OGIbet para el español mexicano (Tlatoa 2000a)

En un ejercicio por validar su propuesta de alfabeto fonético, el CSLU elaboró un cuadro de equivalencias entre el AFI, Worldbet (Hieronimus 1994; Hieronimus 1997) y OGIbet (CSLU 1999)⁴⁵ en el que encontramos datos muy interesantes. A todas luces resalta el hecho de que OGIbet adapta a sus necesidades el espíritu de

⁴⁵ Disponible en las direcciones electrónicas <http://colton.byuh.edu/courses/cs441/9504.refbet.pdf> y <http://dipaola.org/stanford/facade/lipsync/refbet.pdf>.

Worldbet (§.3.2.3), pero hay detalles de corte fonético con los que no se puede estar del todo de acuerdo. Por ejemplo, en el caso concreto de la adaptación de OGIBet para el español de México (Tlatoa 2000a) se descuidaron, a mi modo de ver, bases lingüísticas elementales, como veremos a continuación.

Las convenciones que propone Tlatoa (2000a), aclaran los propios autores, no están basadas en estudios lingüísticos del español, sino en la experiencia del etiquetado; es decir, con base en la práctica de transcribir corpus orales, han ido determinando los fonemas y alófonos necesarios. Además, aunque el mismo grupo Tlatoa reconoce el peligro de basar sus convenciones fonéticas y fonológicas en el inglés, comete errores serios, como los siguientes:

- 1) No distingue entre una transcripción fonológica y una fonética.
- 2) Considera semivocal a la consonante /l/.
- 3) Confunde el uso de la paravocal [j] con el de la consonante palatal fricativa sonora y su alófono africado.⁴⁶
- 4) Confunde las grafías “y” y “ll” y su correspondencia con los fonemas /ʒ/ y /ʎ/.
- 5) Incluye símbolos para los alófonos [S] y [rZ], pero no explica cómo ni cuándo ocurren (por las observaciones que los autores hacen sobre ellos, parece ser que se trata del alófono de procedencia indígena [ʃ], y de las variantes de los fonemas /r/ y /r/ que los estudios hispánicos tradicionales han llamado asibiladas).

En el apartado 3.3. veremos de qué manera la nueva versión de Mexbet resuelve este tipo de problemas.

⁴⁶ “The voiced palatal fricative /ʒ/ is one of the three options for pronouncing the letters “ll” or “y”, which are more commonly realized by the affricate /dʒ/ or the glide /j/” (Tlatoa 2000a). El problema se agrava cuando más adelante se menciona que “whether the differences between /dʒ/, /ʒ/, and /j/ can be labeled consistently in Spanish remains to be seen, as in most cases the phoneme is realized as some mixture of the three” (Tlatoa 2000a).

3.2.3. Worldbet

Worldbet fue creado por Hieronymus (1994 y 1997) en el marco de los desarrollos de reconocimiento y síntesis de habla de los laboratorios Bell, en los Estados Unidos de Norteamérica. El principio que llevó a la creación de Worldbet fue que muchas versiones del AFI en códigos ASCII habían sido hechas únicamente para lenguas europeas, por lo cual se excluían muchos sonidos de otras lenguas o se utilizaban incorrectamente algunos símbolos (Hieronymus 1994). Cuando en los proyectos llevados a cabo en Europa se trató posteriormente de etiquetar corpus en una gran cantidad de lenguas con estos símbolos, se observó que dichos alfabetos eran inadecuados. Así, Hieronymus pensó en la creación de un alfabeto que, al igual que el AFI, abarcara a todas las lenguas del mundo.

Worldbet goza de mucho prestigio en los Estados Unidos, donde se le considera un alfabeto robusto, capaz de captar los sonidos de todas las lenguas y de detallar distinciones fonéticas (Lander 1997:4). En Worldbet, al igual que en el AFI, los diacríticos se utilizan para modificar los fonemas y representar las variantes alofónicas. Así, los alófonos se marcan con el símbolo base del fonema de procedencia, unido por un guión bajo a un símbolo en un sintagma que resulte transparente. Pongamos por ejemplo una *a* velarizada. Para marcar la velarización, Worldbet propone el uso del número [2]. La representación sería la que se indica en la figura 16.

Símbolo base	+	Velarización		Alófono
a	-	2	=	[a_2]

Figura 16. Anotación del alófono velarizado del fonema /a/ con Worldbet

De la larga lista de diacríticos que propone Hieronymus para Worldbet, los que señalan la dentalización, la palatalización y la velarización están ya en uso en el Proyecto DIME. Si resulta necesario transcribir ciertos fenómenos con más detalle, el alargamiento, el ensordecimiento, la nasalización y la sonorización han probado ser útiles también en Mexbet; en cuanto a las variantes regionales, la aspiración y la glotalización resultarían necesarios. En la figura 17 aparece la lista de estos diacríticos.

Fenómeno	Símbolo
Alargamiento	_:
Aspiración	_h
Dentalización	_l
Ensordecimiento	_0
Glotalización	_?
Nasalización	_~
Palatalización	_j
Sonorización	_v
Velarización	_2

Figura 17. Algunos de los diacríticos de Worldbet adaptables al español de México

En la primera versión de Worldbet (1994), Hieronymus elaboró los alfabetos para el alemán, danés, francés, inglés, hindi, japonés y mandarín, entre otras lenguas, incluyendo el español castellano. En la revisión de 1997, detalló las versiones de algunas de las lenguas que ya había creado, y sólo incluyó una más: el español de

México. No obstante y a pesar de las expectativas que esto podría producir, Hieronymus no hizo ninguna aportación y se limitó a repetir su cuadro previo de fonemas del español de España, como se puede apreciar en las figuras 18 y 19. Este hecho produjo errores, como el hecho de que se incluyeran los fonemas interdental fricativo sordo y lateral palatal, entre otros.

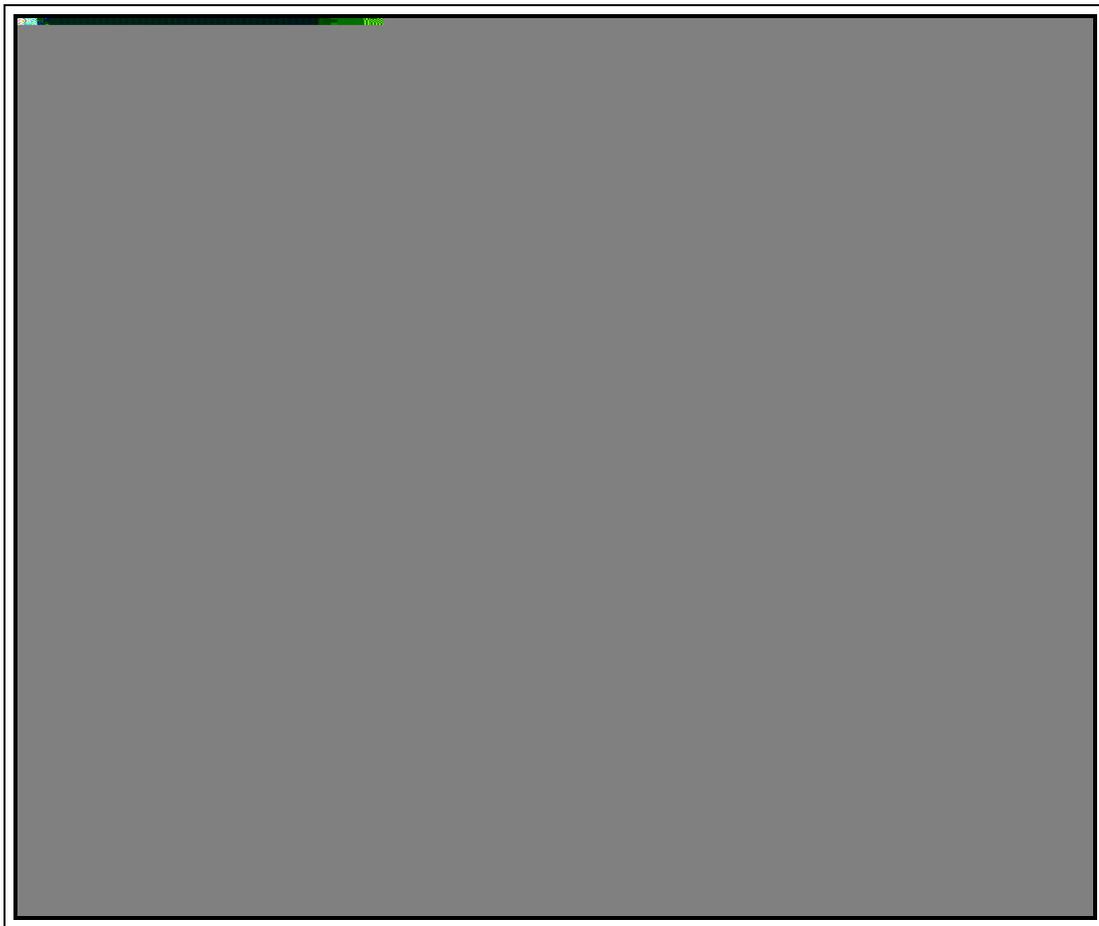


Figura 18. Worldbet (1994) para el español castellano

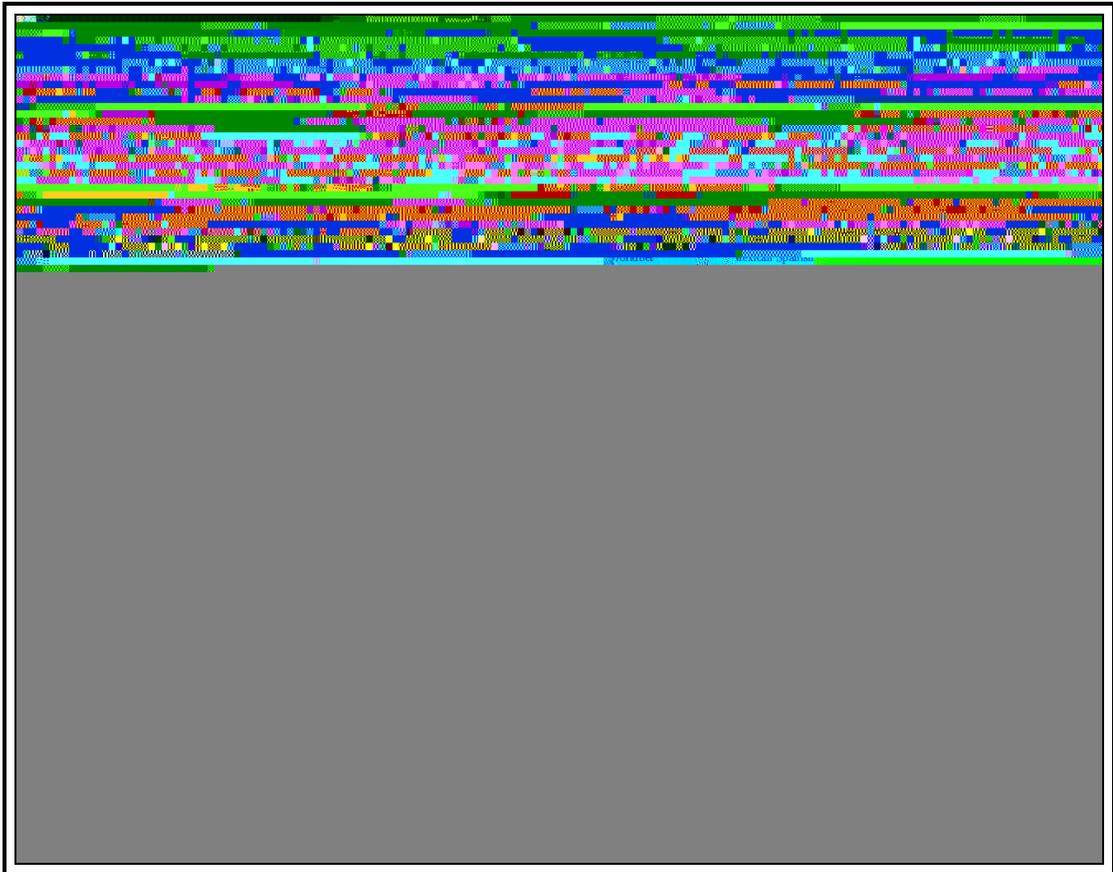


Figura 19. Worldbet (1997) para el español mexicano

Otro error de Worldbet fue el haber establecido un inventario fonológico para el español de México de 26 consonantes. Como vemos, en las figuras 18 y 19, además de los 19 fonemas del español de Castilla, se marcan los alófonos aproximantes de los fonemas /b, d, g/, la forma sonorizada del fonema /s/, la velarización de /n/ y las paravocales. Hieronymus omitió la consonante palatal fricativa sonora —seguramente por confusión con la paravocal [j]— y en su lugar incluyó al alófono africado, otorgándole carácter fonémico. También, de manera equivocada, y seguramente por influencia de la fonética inglesa, clasificó a las líquidas como semivocales.

El propósito de Hieronymus es bueno pero, en mi opinión, ambicioso. Bueno, porque resuelve aspectos que otros alfabetos fonéticos no habían logrado; ambicioso, porque al querer abarcar todas las lenguas del mundo no pudo resolver ciertas particularidades que él mismo critica en otros. No obstante, y a pesar de todas las faltas que se le puedan imputar, cabe destacar algunos de los aciertos de su propuesta:

- 1) Los símbolos de Worldbet tienen en buena medida una similitud visual con los del AFI.
- 2) Las convenciones ofrecen la opción de elaborar transcripciones anchas o estrechas, ya que en sí mismo el espíritu de este código ASCII es bastante flexible.
- 3) Las propuestas se pueden adaptar a las necesidades particulares de cada proyecto de investigación en tecnologías del habla.

Todo esto propició que el Proyecto DIME eligiera Worldbet como base para la elaboración de las convenciones fonéticas y fonológicas de Mexbet.

3.3. Mexbet

Mexbet, el alfabeto fonético computacional para el español de México, tiene sus antecedentes en Uruga (1999), Uruga y Pineda (2000) y Uruga y Pineda (2002). Como he mencionado a la largo de este trabajo, el Proyecto DIME ha hecho uso de este alfabeto; no obstante, Mexbet requería de correcciones básicas y puntualizaciones ligüísticas, como se desarrolla a continuación. Cabe señalar que estas correcciones y ajustes constituyen algunas de las aportaciones de esta tesis.

En sus inicios, Mexbet se basó en las versiones para el español de México de OGIbet y, principalmente, de Worldbet. De tal suerte, copió sus aciertos y arrastró sus

errores. Después de un exhaustivo análisis de las diferentes versiones, podemos resumir estos hechos en los siguientes renglones:

- 1) Se confundía el carácter fonético y fonémico de los segmentos.
- 2) Se confundía la paravocal palatal [j] con el fonema consonántico palatal fricativo sonoro. Este problema se hizo extensivo a la paravocal velar [w], considerada a su vez como un fonema.
- 3) No se definió en qué contextos se presentan las variantes alofónicas que se marcaron.
- 4) Se filtraron los fonemas hispánicos /θ/ y /ʎ/, por contaminación con Worldbet y por una incorrecta interpretación de la teoría fonológica castellana.⁴⁷

Como es bien sabido, el sistema fonológico del español de México contiene 17 fonemas consonánticos y 5 vocálicos.⁴⁸ En una primera versión (Uraga 1999:23), Mexbet presentó un inventario de 23 fonemas para el español de México: cinco vocales más 18 consonantes (incluyó la lateral palatal /ʎ/). Esto indica que, al igual que OGIbet, en sus inicios Mexbet tuvo confusiones sobre el grupo conformado por los alófonos de la vocal cerrada palatal /i/ y los de la consonante palatal fricativa sonora, así como ha sucedido con otros autores, como veremos a continuación.

Vale la pena recordar en este momento que la confusión de estos fonemas y alófonos se ha presentado desde hace mucho tiempo en algunos estudios fonéticos y fonológicos del español. A lo largo de la revisión de diferentes textos, me parece que esto se debe a una visión de la fonología del español desde el punto de vista de teorías

⁴⁷ En sus primeras versiones, el texto básico para el establecimiento del inventario fonológico de Mexbet fue el *Esbozo de una nueva gramática de la lengua española* (RAE 1973); no se puntualizaron en estas versiones originales las características específicas para el español de México.

⁴⁸ Algunos autores (Matluck 1951 y Perissinotto 1975, entre otros) presentan al inicio de sus trabajos el sistema fonológico del español de México y sus convenciones para la transcripción de los mismos. En el apéndice de este trabajo, en la figura 43, se presenta el sistema fonológico del español de México con el AFI, el alfabeto hispánico y Mexbet.

extranjerías, principalmente de índole sajona o de lenguas en las que nuestras paravocales [j] y [w] son fonémicas.

Sobre esta problemática, es apropiado recordar que Alarcos (1950) mencionó que había un cierto parentesco entre los alófonos [i], [j] y [dʒ̃], provocando en algunos autores la falsa idea de que éstos son variantes del fonema consonántico palatal fricativo sonoro. Alarcos aclaró que la forma africada es consonántica y las demás vocálicas, excluyéndose necesariamente. El autor concluyó que es alófono de la vocal cerrada palatal toda realización que sea [i] o que en el contexto pueda alternar con esta vocal, y que es un alófono de la consonante palatal fricativa sonora toda realización que sea [dʒ̃] o que en tal contexto pueda alternar con ésta (1950:155). La confusión de formas fonéticas y fonémicas se había extendido también con la caracterización de la vocal velar cerrada. Alarcos combatió este hecho cuando resumió: “En suma, los sonidos [j], [i̯], [w], [u̯] de los diptongos son, en general, simples variantes de los fonemas /i/ y /u/, respectivamente” (1950:159).

Estas nociones, que hoy a nuestros ojos pueden parecer obvias, no lo fueron tanto en alguna época. Otro autor, Harris (1969:250), nos ayuda también a entender este problema y sus orígenes:

Estos segmentos se consideran “estrechamente relacionados” no sólo por su parecido articulatorio y acústico, sino también porque entran en alternancias unos con otros. [...]

“Ahora bien, en la década de los cincuenta, periodo de apogeo del estructuralismo posbloomfieldiano, el problema de la “fonemización” de los segmentos palatales aludidos suscitó gran interés entre los lingüistas que trabajaban en Estados Unidos. La mayoría de ellos estuvo de acuerdo en que se trataba de dos fonemas: la vocal [i] pertenece al fonema /i/; [y], [y̯] pertenecen a otro, generalmente escrito /y/. También se estuvo de acuerdo en que

la semivocal [j] y la semiconsonante [j] son alófonos de un mismo fonema, hubo fuerte divergencia de opinión sobre si éste era /i/ o /y/. Además, todos los estudios estructuralistas, a diferencia de Navarro Tomás, coincidieron en denominar “semivocal” o “glide” (que inglés incluye tanto las semiconsonantes como las semivocales) al fonema /y/.

Así, no es de sorprender que las primeras versiones de Mexbet tuvieran vacilaciones a este respecto. En la segunda versión de Mexbet (Uraga y Pineda 2002) se menciona que el sistema fonológico del español incluye, además de las 5 vocales, 20 consonantes (se elimina la confusión de las palatales, pero permanece erróneamente el carácter fonológico de la paravocal velar [w]). Creo que esto se debió a dos razones: la adaptación sin cuestionamientos del alfabeto de Hieronymus (1994 y 1997), y una influencia del *Esbozo de una nueva gramática de la lengua española*,⁴⁹ texto que, como mencioné, fue tomado en las primeras versiones de Mexbet como la base lingüística principal.

La siguiente versión de Mexbet (Uraga y Pineda 2002) introdujo, además de los símbolos originales, un número considerablemente mayor de alófonos —a mi parecer, debido a una búsqueda exhaustiva de formas—, pero no explica si se trata de variedades dialectales, libres o en distribución complementaria. Entre los problemas que presenta esta versión podemos detallar los siguientes:

- 1) Se enlistaron de nuevo los fonemas castellanos /θ/ y /ʎ/, y se incluyeron esta vez los alófonos interdentalizados de las consonantes /t/, /s/, /n/ y /l/.
- 2) Se registró, además de /x/, una forma uvular [χ] como variante del español de México.
- 3) Se incluyeron —sin puntualizar— los alófonos aspirados de los fonemas /s/ y /x/, el alófono palatal fricativo sordo [ʃ] (característico de la región norte de México

⁴⁹ En el *Esbozo*, la Real Academia (1973:34) incluye en el sistema fonológico del español al fonema “velar redondeado sonoro /tʃ/”.

o de procedencia indígena), y la variante labiodental fricativa sonora [v], inexistente en nuestra lengua desde hace muchos siglos.⁵⁰

Ante todo este panorama, una revisión detallada y una delimitación lingüística del alfabeto resultaba necesaria para:

- 1) Especificar el inventario fonológico para un alfabeto fonético computacional del español de México, para subsanar confusiones y marcar límites (como la eliminación de los fonemas hispánicos y la delimitación alofónica de las paravocales).
- 2) Establecer convenciones certeras y prácticas, para concretar las formas alofónicas a aquellas que tuvieran una presencia recurrente en el habla y que a la vez se pudieran modelar por reglas.
- 3) Eliminar formas inexistentes —como [v], por ejemplo—, e incluir alófonos con mayor pertinencia y ocurrencia en el habla de México —como la vocal /a/ velarizada y palatalizada, o la consonante /k/ palatalizada.
- 4) Procurar que este alfabeto fuera lo suficientemente sencillo y claro, para facilitar en la medida de lo posible el etiquetado fonético de corpus orales extensos y el análisis y procesamiento de grandes volúmenes de datos.

El inventario de fonemas del español de México que propongo para Mexbet en este trabajo de tesis, se presenta, pues, en la figura 20.

⁵⁰ Aunque [v] podría presentarse en el habla como una asimilación del fonema /f/ ante consonante sonora, o como un fenómeno de ultracorrección de los programas de noticias radiofónicos y televisivos mexicanos, esta versión de Mexbet (Uruga y Pineda 2002) pone como ejemplo para [v] la palabra *virtud*.

Consonantes	Labiales	Labiodental	Dentales	Alveolares	Palatales	Velares
Oclusivos sordos	p		t			k
Oclusivos sonoros	b		d			g
Africado sordo					tʃ	
Fricativos sordos		f		s		x
Fricativo sonoro					ʒ	
Nasales	m			n	n~	
Vibrantes				r (/ r		
Laterales				l		
Vocales				Anteriores	Media	Posteriores
Cerradas				i		u
Medias				e		o
Abierta					a	

Figura 20. Fonemas del español de México en el alfabeto Mexbet

Para marcar diferencias funcionales, busqué que los niveles de representación fonético y fonológico de Mexbet fueran claros. En un principio, para esta nueva versión de Mexbet se había pensado tener un símbolo único para cada fonema, y distinguir luego los alófonos a partir de un guión base seguido del diacrítico correspondiente. No obstante, Hieronymus ya había creado un inventario de formas básicas para algunos alófonos comunes del español (como la consonante nasal velarizada a la que, en lugar de marcarla como [n_2] la representó como [N], adoptando el símbolo del fonema nasal velar de lenguas como el inglés), y Mexbet ya había copiado estas convenciones. Así, para evitar una sobregeneración de formas y mantener una continuidad con las versiones originales de Mexbet, conservé para el nivel fonológico el símbolo /Z/ para el fonema /ʒ/, y mantuve algunos alófonos que ya estaban señalados ([V] para la forma aproximante de /b/, [z] para el alófono sonORIZADO de /s/ y [N] para la nasal velarizada —aunque me parecen más apropiadas

las formas [B], [s_v] y [n_2], respectivamente). Sin embargo, para los nuevos símbolos procuré seguir la convención de marcar el fonema base seguido del diacrítico que señala el proceso de asimilación fonética ([s_[] y [n_[] para los alófonos dentalizados de /s/ y /n/; [k_j] y [a_j] para las formas palatalizadas de /k/ y /a/, y [a_2] para la velarización de /a/). De esta manera se obtienen dos ventajas: en el caso de [a_j] y [a_2], por ejemplo, la persona que etiqueta puede detallar que una vocal abierta está tomando rasgos palatales o velares, y la persona que revisa el etiquetado puede darse cuenta a simple vista de los fenómenos que están sucediendo en la lengua. La lista de alófonos se presenta en la figura 21.

Consonantes	Labiales	Dentales	Alveolares	Palatales	Velares
Oclusivo sordo				k_j	
Africado sonoro				dZ	
Fricativos sordos		s_[]			
Fricativo sonoro	V	D	z		G
Nasales		n_[]			N
Vocales				Anteriores	Posteriores
Cerradas				j	w
Medias abiertas				E	O
Abiertas				a_j	a_2

Figura 21. Alófonos del español de México en el alfabeto Mexbet

Después de elaborar un alfabeto fonético computacional para el español de la ciudad de México, hemos podido percatarnos de la dificultad que implica y de los inconvenientes que hay que resolver cuando se pretende transcribir con un único conjunto de símbolos todas las lenguas del mundo, tal como pretenden el Alfabeto Fonético Internacional y Worldbet. En la propuesta del alfabeto Mexbet que presento,

no ha sido mi objetivo contribuir a la sobregeneración de alfabetos a la que alude Acero (1995), quien afirma que en un momento determinado será necesario unificar las convenciones en los alfabetos fonéticos computacionales. Más bien, la idea de diseñar Mexbet es simple y práctica: contar con un alfabeto específico para el español de la ciudad de México que pueda aplicarse a las tecnologías del habla, buscando conciliar en la medida de lo posible los principios de los alfabetos de la AFI y de la *RFE*. Finalmente, busco apoyo en la idea de Samuel Gili Gaya en cuanto a que “se comprende que no puede haber sistema alguno tan general que prevea todos los sonidos posibles. Por esto ninguno de los intentos realizados hasta ahora ha conseguido satisfacer a todos los lingüistas” (1950:75). Asimismo, una cita de De la Mota y Ríos (1995:109) nos lleva a reflexionar sobre muchas de las cuestiones que se han abordado en este capítulo:

Qué información lingüística ha de aportar y qué convenciones ha de contener un alfabeto fonético para que el sistema de representación sea “racional y operativo”, a lo que debemos aspirar si pretendemos que la transcripción fonética sea una herramienta eficaz en la docencia y en la investigación lingüística. La discusión permanece abierta (De la Mota y Ríos 1995:109).

En el siguiente capítulo veremos cómo se relacionan los puntos que hemos abordado (Mexbet y la transcripción fonética computacional del Corpus DIME, principalmente) para un estudio actual de la fonética del español de la ciudad de México: el análisis de los datos, la frecuencia de aparición de los fonemas y alófonos en la lengua y la posibilidad de determinar reglas para algunos fenómenos fonéticos. Con base en estos conocimientos lingüísticos, la aplicación de herramientas computacionales puede crear un transcriptor fonético automático, TranscríbEMex,

que puede aportar beneficios para la lingüística y para las tecnologías del habla en nuestro país.

4. ESTUDIO FONÉTICO DE LA CIUDAD DE MÉXICO Y SU APLICACIÓN A LAS TECNOLOGÍAS DEL HABLA: *TRANSCRÍBEMEX*

Formalizing theories mathematically is a crucial step in making them predictive.
Pierrehumbert, en prensa

En el capítulo anterior vimos de qué manera se delimitó el alfabeto fonético computacional Mexbet para el español de la ciudad de México, y el uso de los símbolos apropiados para su aplicación en las tecnologías del habla. Este capítulo se inicia con una revisión de los estudios dialectales del español de México. Me ha parecido particularmente importante abordar este punto debido a que las consideraciones fonéticas de este trabajo han tomado como referencia —además de los trabajos castellanos de distintos años de Alarcos, Navarro Tomás y Quilis, entre otros— a los investigadores que han hecho diversos estudios de manera particular sobre el español de México: Lope Blanch, Matluck, Moreno de Alba y Perissinotto, principalmente.

Después, con base en el análisis fonético del Corpus DIME, muestro de qué manera en este trabajo llegué a un inventario de 37 alófonos, que por su ocurrencia en el habla y por la posibilidad de ser modelados por reglas merecen considerarse como pertinentes en los trabajos de tecnologías del habla. Por último, mostraré la aplicación

que se desprende de estas nociones fonéticas para la creación de un fonetizador:⁵¹ TranscríbEMex, un programa que realiza transcripciones fonéticas automáticas a partir de una serie de reglas de conversión de grafema a fonema y de grafema a alófono; presentaré también las aplicaciones de este fonetizador en el Proyecto DIME, como una herramienta que facilita el proceso global de etiquetado fonético, y como un instrumento que puede resultar útil para los estudios de fonética del español de México.

4.1. Estudios sobre la pronunciación del español de México

Para Lope Blanch (1969:46) el estudio de la lengua culta en la ciudad de México, a todos los niveles, era ya a finales de los años sesenta una de las tareas más urgentes. En sí, fue en buena medida por ello que se estableció el grupo de trabajo para la creación del *Atlas Lingüístico de México (ALM)* (Lope Blanch 1970 y 1980). En cuanto a los aspectos de la pronunciación en nuestro país, y de manera particular en la ciudad de México,⁵² muchos de los trabajos del propio Lope Blanch dan cuenta de fenómenos fonéticos como el debilitamiento vocálico (1963-1964) o el cierre de /e/ y /o/ finales (1973). A principios de los años cincuenta, el trabajo de Matluck (1951), *La pronunciación en el español del Valle de México*, describe la lengua de los alrededores del Distrito Federal (Xochimilco, Texcoco y Tlanepantla, que hoy

⁵¹ He adoptado el término “fonetizador” a partir de Ríos Mestre (1999). En §4.6. desarrollaré con más detalle este tema.

⁵² Una insuperable revisión de los estudios que se han hecho sobre la fonética y la fonología del español de México y sus regiones está expuesta en el libro *La pronunciación del español de México*, de Moreno de Alba (1994). En la introducción se presenta una reseña, desde finales del siglo XIX hasta los últimos años, de cada uno de los trabajos de dialectología de autores nacionales y extranjeros.

pertenecen ya a la mancha urbana), en un texto que representa una “minuciosa descripción de las realizaciones fonéticas en el valle de México” (Moreno de Alba 1994:17).

Hacia mediados de los años setenta, resalta la *Fonología del español hablado en la Ciudad de México. Ensayo de un método sociolingüístico*, de Perissinotto (1975), “hasta ahora, a pesar de su brevedad, el más coherente y completo estudio sobre la fonética del español hablado en la ciudad de México” (Moreno de Alba 1994:17).

En 1994 Moreno de Alba publicó *La pronunciación del español de México*, libro en el que utiliza el *ALM* como herramienta de análisis para dar cuenta de los aspectos dialectales más relevantes de la pronunciación en nuestro país (relajación vocálica, cierre vocálico, diptongación de hiatos, relajación consonántica, relajación y pérdida de los fonemas /d/ y /s/ en diferentes contextos y asibilación de /r/ y /r/, entre muchos otros fenómenos). Después de este texto de Moreno de Alba, no tengo conocimiento de ningún otro trabajo general abarcador de fonética mexicana y centrado en la descripción sincrónica de los elementos segmentales, hasta principios de este siglo, cuando parece que ha surgido un nuevo interés por los estudios de fonética del español de México —si bien en forma de estudios parciales— a raíz de las oportunidades que ofrece el estudio instrumental computarizado.⁵³

⁵³ La publicación *Temas de fonética instrumental*, editada por Herrera (2001), incluye dos artículos sobre el español de México: uno de Madrid y Marín sobre la “Estructura formántica de las vocales del español de la ciudad de México” y otro de Rodríguez y Mora titulado “Las consonantes oclusivas en posición de coda silábica: estudio sobre el español de la ciudad de México”. De Herrera (2002) destaca su artículo “La asimilación de las nasales en español. Un estudio instrumental” aparecido en la *Nueva Revista de Filología Hispánica*. De muy reciente aparición es el texto *Variación lingüística y teoría*

Canfield (1981:60), conocedor del español de México, afirmó que todavía había mucho por estudiar, y Lope Blanch acusó en 1969 la carencia de estudios lingüísticos sobre el español de México.⁵⁴ Perissinotto (1975:21-22), a su vez, señaló la falta de información sobre un dialecto tan importante. En este trabajo pretendo dar una luz sobre la pronunciación del español en la ciudad de México en el marco más moderno de las tecnologías del habla, con el uso de los instrumentos más actuales, y con base en las descripciones fonéticas y fonológicas tradicionales, tomando como un hecho y un punto de partida sólido el valor y el peso de la tradición lingüística dialectológica fonética de nuestro país.

Después de la etapa de transcripción electrónica con el alfabeto Mexbet, realicé el análisis de los datos fonéticos del Corpus DIME. Posteriormente, compararé mis resultados con las teorías fonética y fonológica hispánicas en general (Alarcos, Gili Gaya, Navarro Tomás y Quilis), y con las de los autores que se han detenido de manera particular en el español de México (Lope Blanch, Matluck, Moreno de Alba y Perissinotto, principalmente). Los resultados que obtuve, como veremos más adelante, mostraron que, en los contextos que reportan estos autores, se producían en buena medida los fenómenos fonéticos que señalaban, por lo cual sus trabajos resultaron altamente útiles en la etapa de etiquetado y en la de análisis. Por otra parte, merece la pena señalar que durante estas etapas detecté en mi corpus muchos contrastes entre el gran volumen de datos que proporcionaban algunos fonemas y la

fonológica, de Martín Butragueño (2003), que incluye importantes datos sobre la geografía fónica de México.

⁵⁴ “Por la antigüedad e importancia de su Universidad, debería ser foco de interesantes investigaciones lingüísticas. Sin embargo, el castellano de México es una de las modalidades de la lengua española de la que sólo tenemos un conocimiento muy incompleto y defectuoso” (Lope Blanch 1969:7).

poca cantidad de muestras de otros. Por ejemplo, las vocales, principalmente /e/ y /a/, presentaban una frecuencia de aparición muy alta y arrojaban gran cantidad de alófonos, mientras que algunas consonantes, como /tʃ/ o /ɲ/, casi nunca aparecían.

4.2. Frecuencia de aparición de fonemas y alófonos

Para profundizar en el comportamiento de los fonemas y de los alófonos, realicé primero una rápida exploración de la frecuencia de aparición de los fonemas en el 10% de mi corpus (52 elocuciones del diálogo s6-t1-g1, con 1,684 segmentos). Los resultados de esta exploración se muestran en la figura 22.

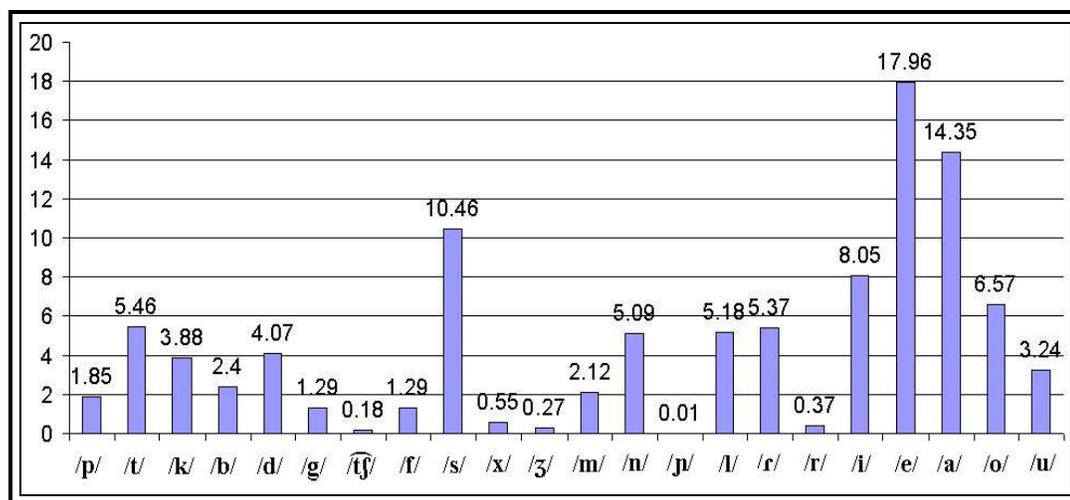


Figura 22. Frecuencia de aparición de los fonemas del español de México, en 52 archivos del diálogo s6-t1-g1 del Corpus DIME

Como podemos ver, hay fonemas con una alta incidencia y otros muy poco frecuentes; por ejemplo, los porcentajes de algunos fonemas estaban por arriba del

10%, como en el caso de /e/ (17.96%), /a/ (14.35%) y /s/ (10.46%). En contraste, el porcentaje de \widehat{tj} (0.18%), /x/ (0.55%), /ʒ/⁵⁵ (0.27%), /ɲ/ (0.01) y /r/ (0.37%), es inferior al 1%. Esto implicaba que durante el entrenamiento de modelos acústicos para el reconocedor de habla del Proyecto DIME, el Corpus DIME arrojaría una gran cantidad de datos sobre /e/, /a/ y /s/, pero, para otros fonemas, apenas si habría muestras.

Con el propósito de estudiar con más detalle el comportamiento estadístico de los fonemas por grupos, me detuve para comparar las 5 vocales y las 17 consonantes del corpus. La frecuencia de aparición de las vocales (50.17%) era prácticamente igual al de las consonantes (49.83%). Para conocer el comportamiento de las vocales como grupo, contabilicé y contrasté sus ocurrencias. La figura 23 muestra notables diferencias entre unas y otras: por una parte, destaca el porcentaje de /e/ (35.79%) y /a/ (28.61%) contra las demás, sobre todo frente al fonema /u/ (6.46%); por otra, si observamos los datos en cuanto a anterioridad y posterioridad, puede verse que las vocales anteriores, /i, e, a/, son considerablemente más numerosas (80.44%) que las posteriores, /o, u/, (19.56%).

⁵⁵ El uso de este símbolo para la fricativa palatal sonora ha sido adoptado de manera arbitraria, según se detalla en el apéndice de esta tesis (§6.).

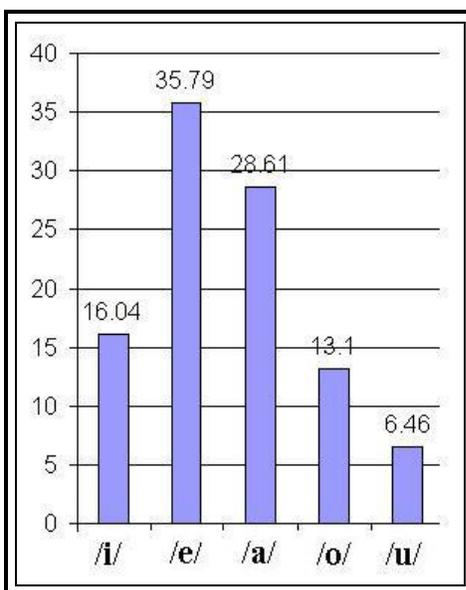


Figura 23. Frecuencia de aparición de las vocales (en relación con el total de vocales) en 52 archivos del diálogo s6-t1-g1 del Corpus DIME

El hecho de tener solo 5 segmentos —las vocales— con una alta ocurrencia en la lengua, me hizo pensar que, para fines de creación de modelos acústicos para el reconocedor de habla, sería muy importante definir con claridad los contextos en que se presenta cada uno de sus alófonos. Por otra parte, noté que era importante también encontrar si había cierta uniformidad en el comportamiento de las 17 consonantes en sus diferentes contextos, para procurar definir qué reglas fonéticas pueden utilizarse en el diseño y evaluación de corpus orales, y en la determinación de modelos de pronunciación y de modelos acústicos.

Al hacer un ejercicio similar con las consonantes (figura 24), sobresalió la frecuencia de aparición de /s/ (20.99%) por sobre todas las demás. Otros fonemas consonánticos que presentaban una alta considerable frecuencia de aparición dentro de su grupo —por arriba del 10%— fueron /t/ (10.94%), /n/ (10.21%), /l/ (10.39%) y

/r/ (10.81%). Por el contrario, las consonantes \widehat{tj} / (0.36%), /ʒ/ (0.54%), /ɲ/ (0.02%) y /r/ (0.74%), mantuvieron una frecuencia de aparición menor al 1%. Destaca a su vez el hecho de que los fonemas que se articulan en la zona palatal tuvieron como conjunto una frecuencia de aparición menor al 1% (los porcentajes de \widehat{tj} /, /ʒ/ y /ɲ/ suman sólo el 0.92%), y los fonemas de la zona alveolar representaron más de la mitad de todas las consonantes (/s/, /n/, /l/, /r/ y /r/ reúnen el 53.14%).

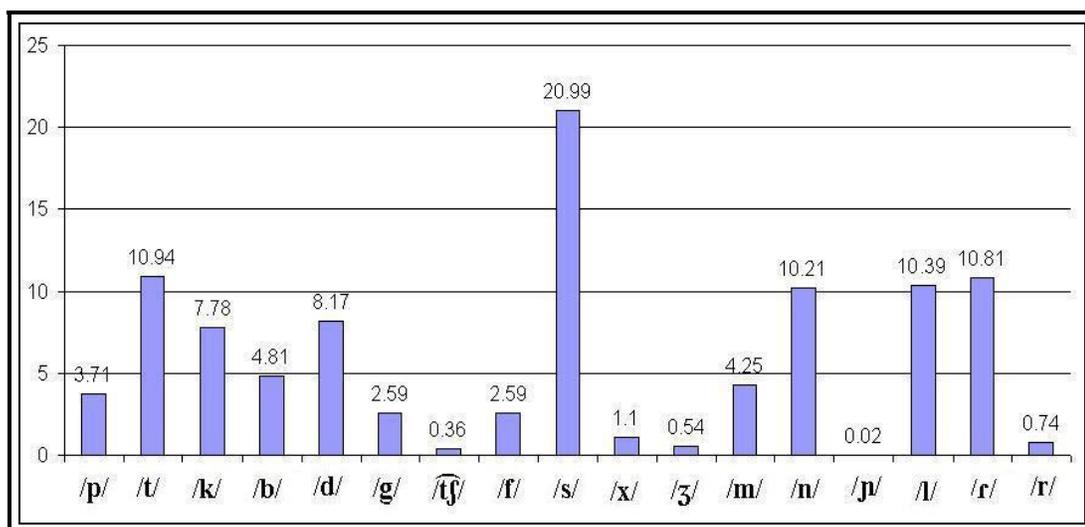


Figura 24. Frecuencia de aparición de las consonantes (en relación con el total de consonantes) en 52 archivos del diálogo s6-t1-g1 del Corpus DIME

Otro aspecto que resaltó durante la etapa de análisis fue la frecuencia de aparición de los fonemas en sus diferentes contextos: observé que los fonemas más frecuentes se encontraban, de manera lógica, en mayor cantidad de contextos que los menos frecuentes. Como detallaré más adelante (§4.5.), para /e/ y /a/, destacó el hecho de que estos fonemas mostraron una gran variedad de formas alofónicas, en

contraste con las vocales que presentaban una menor frecuencia de aparición: las variantes para /e/ fueron más de 12 (átona, tónica, nasalizada, relajada, alargada, etc.); /u/, en cambio, sólo presentó escasamente 4 formas (átona, tónica, semivocal y semiconsonante). Por su parte, y de manera obvia, lo mismo sucedió con las consonantes (§4.4.): el análisis de datos mostró la gran cantidad de ocasiones en que /s/ se encontraba ante /t/ > [s̺] frente a las mínimas ocasiones en los que /l/ estaba ante /d/ > [l̺]. Por el contrario, y de manera obvia, para /tʃ/, /ʒ/, /ɲ/ y /r/ no había muestras suficientes ni contextos variados. Todo esto puso de manifiesto un hecho fundamental: a mayor volumen de muestras, mayor cantidad de contextos y, por ende, un número más elevado de realizaciones alofónicas.

De tal suerte, resultaba necesario detallar más ampliamente estos datos, por lo cual realicé un conteo profundo de todo mi corpus (15,531 segmentos) para afinar dos hechos importantes que tendrían impacto sobre los modelos acústicos y el reconocedor de habla:

- 1) Hay alófonos que por su frecuencia de aparición y la consiguiente repercusión en la lengua son candidatos óptimos para ser considerados en los alfabetos fonéticos y para los modelos acústicos y, en general, para todas las aplicaciones de las tecnologías del habla.
- 2) Para fines prácticos, no es pertinente, al menos en esta etapa del Proyecto DIME, modelar los fenómenos alofónicos esporádicos o los alófonos de baja frecuencia.

En este punto, me propuse hacer algo similar al trabajo de Llisterri y Mariño (1993) a propósito de la versión para el castellano del alfabeto fonético SAMPA (§3.2.1.): para poder tener un número de unidades controlables y para garantizar una determinada cantidad de detalle fonético, procuré establecer una relación coherente

entre el número máximo de alófonos posibles en la lengua y la lista cerrada de fonemas.⁵⁶ Así, para un análisis más detallado, tomé los cuatro diálogos del Corpus DIME y saqué la frecuencia de aparición de los alófonos que surgieron del análisis del etiquetado fonético. Los resultados se muestran en la figura 25.

⁵⁶ Para llegar a ello, Llisterra y Mariño (1993) llevaron a cabo un estudio estadístico de la frecuencia de aparición de los alófonos españoles en un corpus de entrevistas orales con más de cien mil segmentos. Posteriormente, establecieron un inventario final, eliminando los alófonos con una frecuencia de aparición inferior al 0.10% en el corpus analizado y limitando el inventario de segmentos a los 24 fonemas del castellano, más siete segmentos considerados tradicionalmente como alófonos: las realizaciones aproximantes de /b, d, g/, el alófono sonorizado de /s/, el alófono nasal velar y las paravocales.

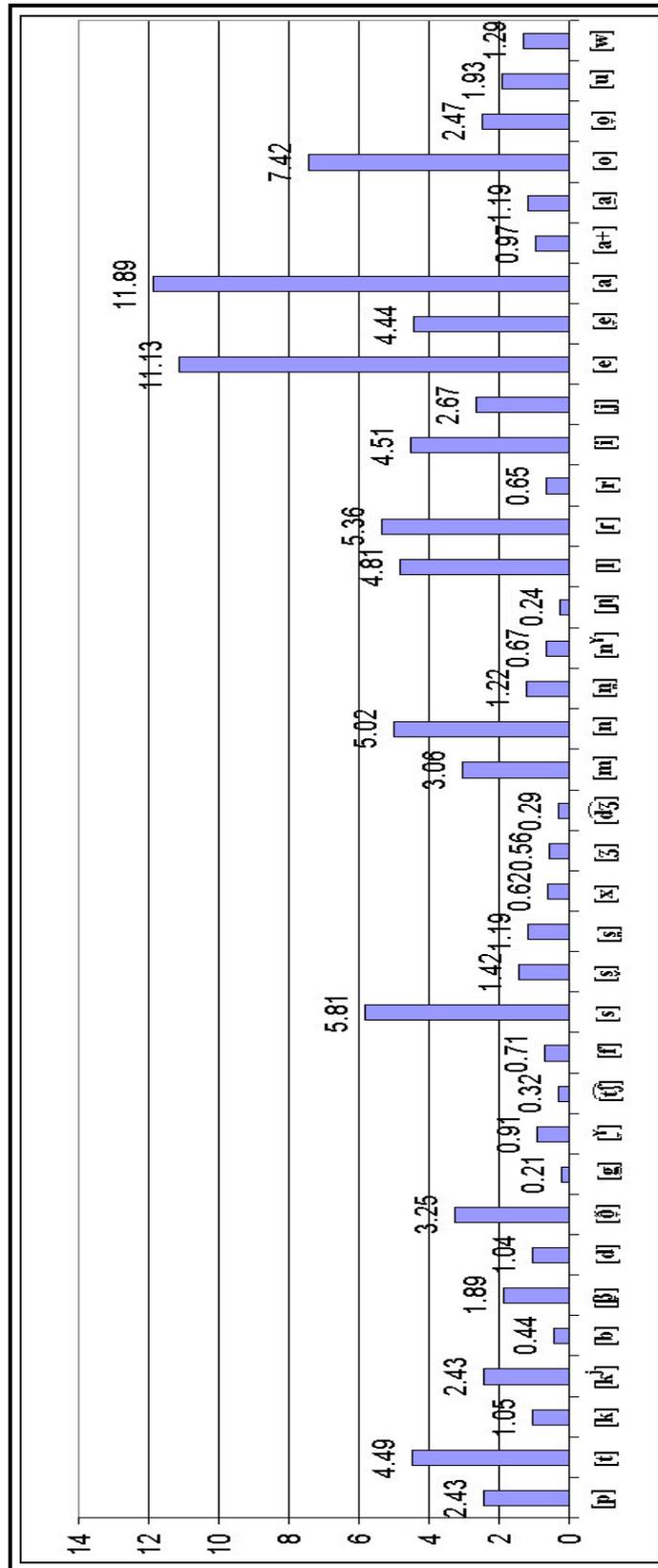


Figura 25. Frecuencia de aparición de los alófonos del español de México en cuatro diálogos del Corpus DIME

Para los fonemas de alta frecuencia de aparición se comprobó que hay una mayor cantidad de variantes alofónicas que ocurren de manera considerable. En el caso de las vocales, las medias presentaron un alófono abierto con una frecuencia de aparición bastante alta: para [e̞] fue del 4.44% y para [o̞] del 2.47%, cifras que, como hemos visto, son más altas que las de algunos fonemas. La vocal abierta tuvo dos realizaciones principales [a+] y [a̠], con porcentajes del 0.97% y 1.19%, respectivamente. Asimismo, las paravocales tuvieron una frecuencia de aparición considerable: 2.67% para [j] y 1.29% para [w].

En el caso de las consonantes, las de mayor frecuencia de aparición mostraron también mayor cantidad de alófonos: /s/ presentó los alófonos sonorizado [s̚] y dentalizado [s̺] con frecuencias de aparición del 1.42% y 1.19%, respectivamente; /n/, un alófono dental [n̪] y otro velar [n̠] con el 1.22% y 0.67%, respectivamente. El fonema /k/ presentó su alófono palatalizado [kʲ] (2.43%) con una frecuencia de aparición considerablemente mayor a la de su forma prototípica [k] (1.05%). Este mismo hecho se registró con las formas aproximantes de las consonantes fricativas sonoras: los porcentajes de [β] (1.89%), [ð] (3.25%) y [ɹ̠] (0.91%) fueron mucho más altos que los de [b] (0.44%), [d] (1.04%) y [g] (0.21%).

Al observar con detenimiento la figura 25, surgió la inquietud de visualizar los datos de otra manera para ver qué más información arrojaban. Así, ordené todos los

alófonos por el valor de su frecuencia de aparición, de más a menos, y dividí los 37 alófonos en 5 grupos (figura 26):

- 1) En negro, los alófonos que tienen una frecuencia de aparición mayor al 10% ([a] y [e]).
- 2) En blanco, aquellos con una frecuencia de aparición entre el 8 y el 4% ([o], [s], [r], [n], [l], [i], [t] y [ɛ]).
- 3) En negro, los alófonos que se encuentran entre el 4 y el 2% ([ð], [m], [j], [ɔ], [p] y [kʰ]).
- 4) En blanco, los que tienen una frecuencia entre el 2 y el 1% ([u], [β], [ʒ], [w], [ŋ], [ʃ], [a], [k] y [d]).
- 5) En negro, los alófonos con frecuencia de aparición menor al 1% ([a+], [ɲ], [f], [nʸ], [r], [x], [ʒ], [b], [tʰ], [dʒ], [ɲ] y [g]).

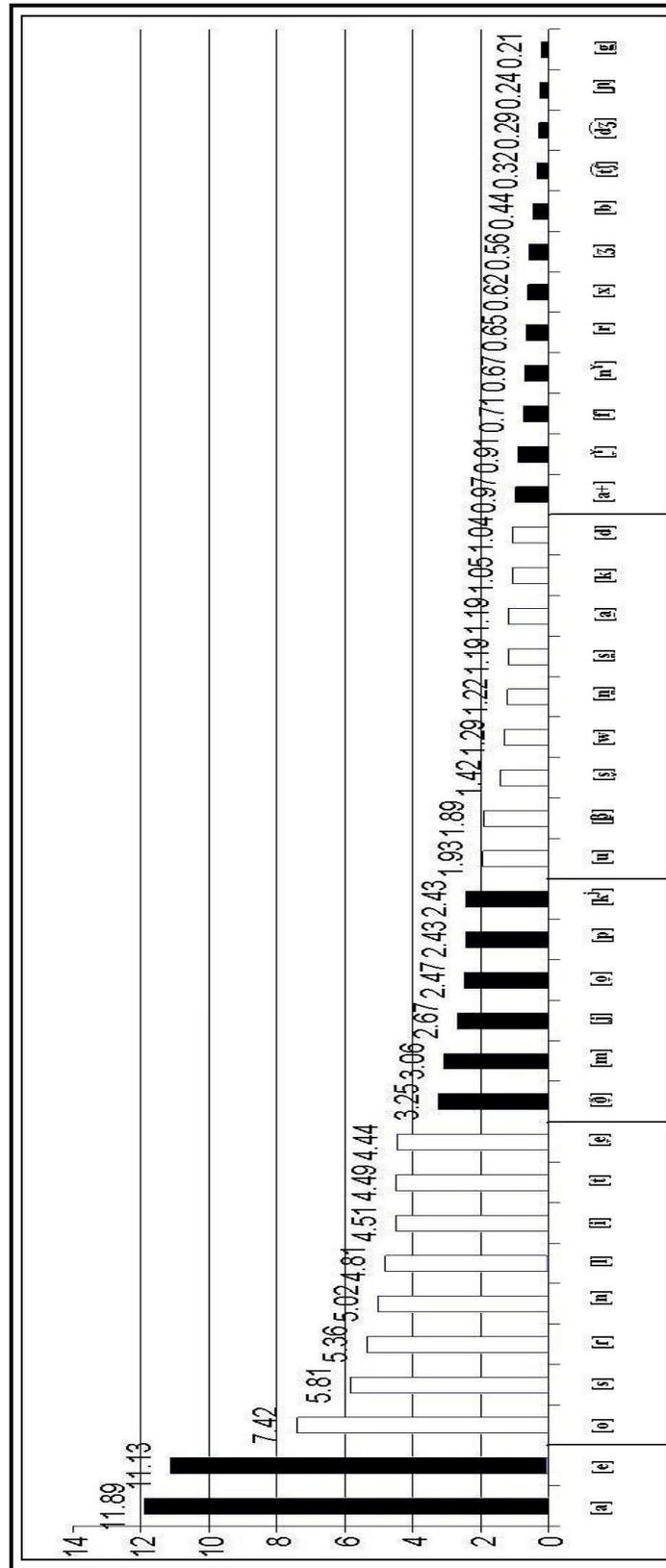


Figura 26. En orden descendente, frecuencia de aparición de los alófonos del español de México en cuatro diálogos del Corpus DIME

En cuanto al grupo 1, [a] y [e] suman casi la cuarta parte del valor de todos los alófonos (23.02%), poniendo las vocales de la zona media anterior a la zona baja en un lugar preferencial. Los alófonos del grupo 2 suman el 41.86%. En éste se encuentran 3 vocales, [o], [i] y [ɛ], que juntas alcanzan el 16.37%, cifra que representa poco menos de la mitad en su grupo. En este mismo grupo tienen una frecuencia considerable las consonantes alveolares [s], [r], [n] y [l], con un 21%, poco más de la mitad de este grupo. Aparece en él [t] con el 4.49%, la consonante oclusiva con la mayor frecuencia de aparición.

El grupo 3 representa el 16.31%; en él, hay dos segmentos vocálicos, [j] y [ɔ], dos alófonos oclusivos, [p] y [k^j], la nasal [m], y la aproximante con mayor frecuencia de aparición: [ɸ] (3.25%). Un hecho interesante es que hallamos dos segmentos labiales en este grupo, [p] y [m], aunque debido a que a partir de este tercer grupo las frecuencias empiezan a ser bajas, no es posible, no con las muestras de este trabajo, hacer generalizaciones más profundas. Así, el grupo 4 y 5, juntos, suman solamente el 18.81%. En el grupo 4 (12.22%), se encuentran tres alófonos vocálicos ([u], [a] y [w], que juntos suman el 4.41%), dos alveolares asimiladas a la zona dental ([ŋ] y [ʃ]), dos oclusivas ([k] y [d]), una alveolar sonorizada ([ʒ]) y una aproximante, [β]. El grupo 5, el de los alófonos con frecuencia de aparición menor al 1%, son los más numerosos (12 en total, 6.59%; así, notamos que los alófonos de mayor frecuencia de aparición son los menos, y los de menor frecuencia, los más). En

este último grupo se encuentran alófonos de muy variados puntos de articulación y modos de articulación: [a+], [a^y], [f], [n^y], [r], [x], [ʒ], [b], [tʃ], [dʒ], [ɲ] y [g]. Destaca una vez más un hecho importante: la gran mayoría de los segmentos consonánticos palatales, [ʒ], [tʃ], [dʒ] y [ɲ] (el 1.41% para los cuatro), se encuentra en este grupo.

Para confirmar la veracidad de mis datos, posteriormente comparé los resultados con los trabajos de algunos de los autores que han estudiado la frecuencia de aparición de los fonemas y de los alófonos del español: Zipf y Rogers (1939),⁵⁷ Navarro Tomás (1946),⁵⁸ Alarcos (1950), Quilis (1981), Llisterri y Mariño (1993) y Rojo (1991) para el español castellano; y Pérez (2003), para el español de Chile. Los resultados se presentan en la figura 27.

⁵⁷ *Apud* Alarcos (1950).

⁵⁸ *Apud* Alarcos (1950).

Fonema	Alófono	Zipf y Rogers 1939	Navarro 1946	Alarcos 1950	Quilis 1981	Rojo 1991	Llisterri y Mariño 1993	Pérez 2003	DIME
/p/	[p]	2.92	3.04	2.10	2.77	2.66	2.60	2.58	2.43
/t/	[t]	4.46	4.82	4.60	4.53	4.48	4.63	4.92	4.49
/k/	[k]	3.84	4.23	3.80	3.98	3.98	4.04	3.94	1.05
	[k']	-	-	-	-	-	-	-	2.43
/b/	[b]	3.26	2.54	2.50	2.37	2.66	0.45	1.92	0.44
	[β]	-	-	-	-	-	2.47	-	1.89
/B/		-	-	0.10	0.03	-	-	-	-
/d/	[d]	5.06	5.00	4.00	4.24	4.79	0.76	4.84	1.04
	[ð]	-	-	-	-	-	3.20	-	3.25
/D/		-	-	0.25	0.31	-	-	-	-
/g/	[g]	1.02	1.04	1.00	0.94	0.95	0.11	0.94	0.21
	[g']	-	-	-	-	-	0.79	-	0.91
/G/		-	-	0.25	0.28	-	-	-	-
/tʃ/	[tʃ]	0.30	0.30	0.40	0.37	0.28	0.40	0.32	0.32
/tʃ/	[tʃ]	0.72	0.72	1.00	0.55	0.68	0.51	0.75	0.71
/θ/	[θ]	1.74	2.23	1.70	1.45	1.68	1.53	-	-
/s/	[s]	8.12	8.50	8.00	8.32	7.58	6.95	9.61	5.81
	[s̺]	-	-	-	-	-	1.33	-	1.42
	[s̺]	-	-	-	-	-	-	-	1.19
/x/	[x]	0.58	0.51	0.70	0.57	0.73	0.63	0.74	0.62
/ɣ/	[ɣ]	2.40	0.40	0.40	0.41	0.22	0.19	0.69	0.56
	[d̺ɣ]	-	-	-	-	-	-	-	0.29
/m/	[m]	2.98	2.40	2.50	3.06	3.09	3.63	2.62	3.06
/n/	[n]	5.94	2.94	2.70	2.78	6.99	7.02	7.78	5.02
	[ɲ]	-	-	-	-	-	-	-	1.22
	[nʲ]	-	-	-	-	-	0.46	-	0.67
/ɲ/	[ɲ]	0.36	0.36	0.20	0.25	0.19	0.27	0.24	0.24
/N/		-	4.69	3.70	4.86	-	-	-	-
/l/	[l]	5.20	5.46	4.70	4.23	5.08	4.25	5.05	4.81
/ʎ/	[ʎ]	0.60	0.60	0.50	0.38	0.38	0.54	-	-
/r/	[r]	5.90	2.40	2.50	3.26	5.67	4.25	6.19	5.36
/r/	[r]	1.04	0.80	0.60	0.43	0.79	0.40	0.64	0.65
/R/		-	3.51	4.50	1.93	-	-	-	-
/i/	[i]	4.20	4.76	8.60	7.38	7.5	4.29	7.46	4.51
	[j]	-	-	-	-	-	2.60	-	2.67
/e/	[e]	12.20	11.75	12.60	14.67	13.51	13.73	14.13	11.13
	[e̺]	-	-	-	-	-	-	-	4.44
/a/	[a]	14.06	13.00	13.70	12.19	13.40	13.43	12.31	11.89
	[a+]	-	-	-	-	-	-	-	0.97
	[a]	-	-	-	-	-	-	-	1.19
/o/	[o]	9.32	8.90	10.30	9.98	9.57	10.37	9.28	7.42
	[o̺]	-	-	-	-	-	-	-	2.47
/u/	[u]	1.76	1.92	2.10	3.33	3.16	1.98	3.05	1.93
	[w]	-	-	-	-	-	1.35	-	1.29

Figura 27. Frecuencia de aparición de los fonemas y de los alófonos del español según varios autores

En cuanto a las consideraciones fonéticas y fonológicas, Alarcos, Navarro y Quilis reportan archifonemas. Navarro Tomás y Zipf y Rogers consideran a la paravocal palatal aparte, pero Zipf y Rogers la contabilizan como alófono de /ʒ/. El trabajo de Llisterra y Mariño, basado en un corpus oral, presenta mayor detalle fonético e incluye ciertos alófonos recurrentes en la lengua, como las realizaciones aproximantes de /b, d, g/, la /s / sonorizada, la nasal velarizada y las paravocales [j] y [w]. Pérez se basa en un corpus oral de lectura de noticias en la televisión chilena, pero señala únicamente los elementos fonémicos.

En términos generales, pude comprobar que mis cifras eran semejantes a las que presentan todos estos autores, lo cual dio un indicio bastante fiable de que los datos que arroja el Corpus DIME coinciden con los que se mencionan en la bibliografía. Antes de presentar el análisis de cada uno de los fonemas y alófonos del español en el Corpus DIME, me parece apropiado resumir y presentar visualmente algunos aspectos que he señalado con anterioridad. Para tal efecto, contrasto a nivel fonológico y fonético:⁵⁹ 1) la zona de articulación y el modo de articulación de las consonantes en las figuras 28 y 29, y 2) en la figura 30, los aspectos relacionados con la posición de las vocales.

La figura 28 muestra la distribución de los fonemas (imagen superior) y alófonos (imagen inferior) conforme a la zona de articulación. A nivel fonológico se aprecia, como hemos visto, que la zona alveolar es la que tiene una mayor cantidad de elementos. A nivel fonético, crece la zona dental (sombreada con gris), lo cual sucede

⁵⁹ El único propósito de estas figuras es mostrar visualmente los datos que ya he presentado a lo largo de este apartado. De tal suerte, no se muestran las cifras.

porque se agregan a este grupo los alófonos dentales de /s/ y /n/; la zona palatal (sombreada con gris), aumenta un poco su porcentaje debido a la inclusión del alófono palatalizado de /k/.

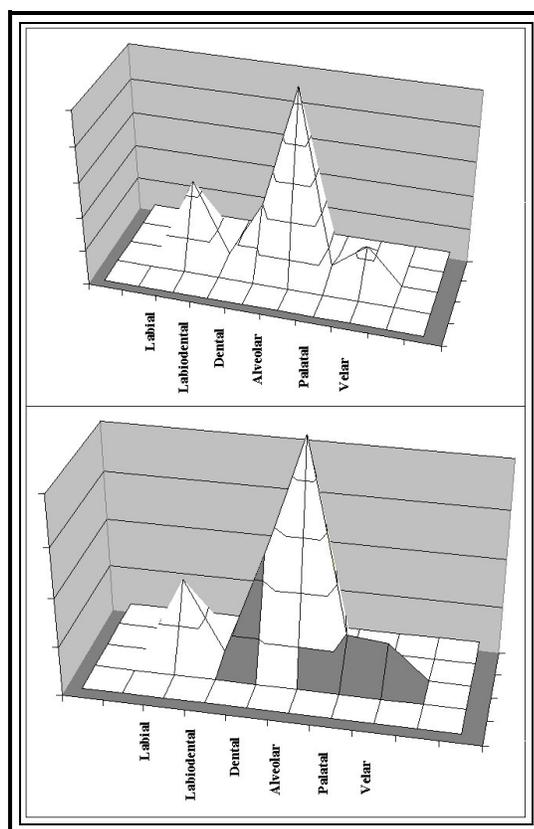


Figura 28. Diferencias en el punto de articulación entre el nivel fonológico (superior) y el fonético (inferior) en las consonantes del Corpus DIME

La figura 29 presenta, por modo de articulación, cómo a nivel fonológico (imagen superior) la parte de las consonantes fricativas sonoras es muy baja, y cómo a nivel fonético (imagen inferior), ésta aumenta considerablemente porque se le agregan los alófonos aproximantes de /b/, /d/ y /g/, y el alófono sonorizado de /s/. Por

otra parte, aparece a nivel fonético el modo de articulación de la consonante africada sonora.

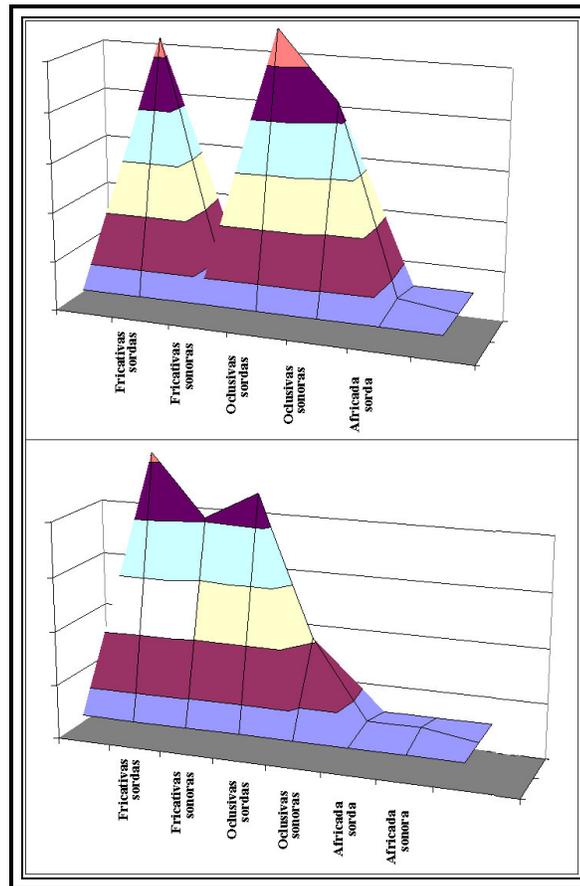


Figura 29. Diferencias en el modo de articulación entre el nivel fonológico (superior) y el fonético (inferior) en las consonantes del Corpus DIME

En cuanto a las vocales, en la figura 30 se simula la distribución de las vocales por punto de articulación y por altura de la lengua. Se aprecia cómo los contrastes del nivel fonológico (imagen superior), se diluyen en el nivel fonético (imagen inferior) al aumentar considerablemente el inventario de segmentos (de 5 fonemas a 11 alófonos).

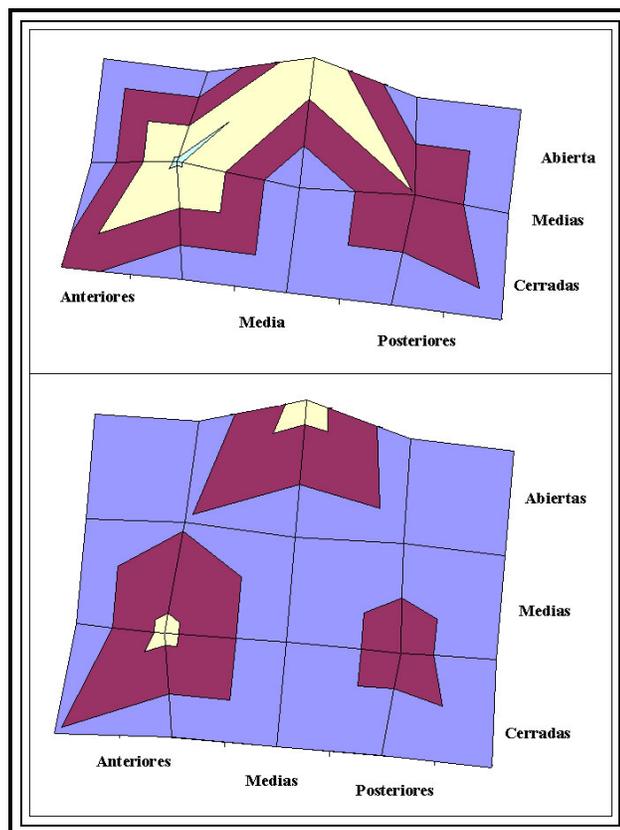


Figura 30. Diferencias entre el nivel fonológico (superior) y el fonético (inferior) en las vocales del Corpus DIME

Después de hacer estos ejercicios, hoy me parece que en los estudios fonéticos y fonológicos no se puede prescindir del estudio de la frecuencia de aparición de los fonemas y alófonos en su uso. De tal suerte, y a partir de los fenómenos fonéticos que mayor representatividad tienen en nuestra habla, presento a continuación los resultados del análisis fonético. Estos datos son precisamente los que esta tesis utiliza para proponer los 37 alófonos con una frecuencia de aparición considerable y las reglas que se desprenden de los datos, con las cuales se programó al transcriptor fonético automático TranscríbEMex.

4.3. Análisis fonético del Corpus DIME

Si tomamos en cuenta que los fenómenos de asimilación fonética en la lengua española pueden modelarse mediante reglas (Moreno y Mariño 1998), podemos ayudar a un programa de reconocimiento en el modelado de las formas del habla. Con base en la frecuencia de aparición de los fonemas del español y sus alófonos, detallaré a continuación los fenómenos considerablemente recurrentes, en los cuales se deberían concentrar los esfuerzos en la creación de modelos acústicos para una primera etapa en el Proyecto DIME.

Como señalé con anterioridad (§2.4.), las grandes ciudades representan focos lingüísticos. Para efectos de clasificación computacional y optimización de recursos, tomo en este trabajo el habla de la capital como la más representativa de México, especialmente en lo que se refiere al número de hablantes, lo que me permitirá definir los fenómenos fonéticos apropiados, aplicables en muchos casos al resto del país. Por otra parte, aunque esta habla no sea la común en algunas zonas, es comprensible y asimilable prácticamente por todos los mexicanos.

Antes de iniciar la revisión de las formas alofónicas de cada fonema, es pertinente destacar el hecho de que los diferentes autores que han realizado estudios de fonética en México se preocupan mucho por la descripción fina de los alófonos, como lo señalan Navarro Tomás (1966-1967) y López Chávez (1981). Debido a que en los modelos de pronunciación es necesario que un reconocedor detecte y codifique cada uno de los segmentos del habla humana, en el caso de este trabajo será necesario

eliminar descripciones muy finas y englobarlas en grupos. Por ejemplo, para el fonema /s/, por el momento no se incluirán en el programa de reconocimiento formas debilitadas o aspiraciones, que ocurren de manera muy esporádica en el habla de la ciudad de México, sino que se incorporarán los alófonos más frecuentes y modelables por reglas, para brindar la mayor ayuda al sistema de reconocimiento.

4.4. Consonantes

A continuación revisaré los fenómenos fonéticos consonánticos que se desprenden del análisis del Corpus DIME. Asimismo, propondré para cada fonema los alófonos que pueden incluirse en el fonetizador TranscríbEMex, y sus reglas. Para el desarrollo del análisis, me ha parecido apropiado agrupar los fonemas por clases: oclusivas, sordas y sonoras, africada sorda, fricativas, sordas y sonora, nasales y líquidas.

4.4.1. Consonantes oclusivas

En términos generales, la posición preferencial en que se presentan las consonantes oclusivas es la explosiva. Como veremos a continuación, de las tres consonantes oclusivas sordas, sólo /k/ presenta un alófono; en el caso de las sonoras, sus realizaciones aproximantes.

4.4.1.1. Consonantes oclusivas sordas: /p, t, k/

En el caso del fonema /p/, Perissinotto señala que “a diferencia de lo que se ha observado en el español peninsular estándar y en el de otras regiones de México es muy estable y resiste tanto la sonorización como el relajamiento en cualquier posición” (1975:44), aunque este investigador menciona después cambios esporádicos, como que el fonema /p/ en posición final de sílaba se puede presentar como [k]: /op 'sion/ > [ok 'sjon]; este tipo de fenómenos ocurrió en el Corpus DIME de manera tan esporádica (2 casos), que las muestras no bastaron para modelarlos por reglas. Navarro Tomás (1918:83) habla también de algunas variantes de /p/ según su posición en la sílaba, pero como su rendimiento es muy pobre, no es pertinente, en esta primera etapa, modelarlos matemáticamente para el Proyecto DIME. En cuanto al fonema /t/, también se mantiene bastante firme, hecho que coincide con lo que reporta Perissinotto.⁶⁰

4.4.1.1.1. Alófono palatalizado: /k/ > [kʲ]

De alguna manera, el fonema /k/ tampoco sufre mayores cambios según su posición en la sílaba o en la palabra, salvo por la presencia de un alófono que se mantuvo constante a todo lo largo del etiquetado. Se trata de una forma palatalizada, que se presenta regularmente ante [e, i, j] (*okey, anaquel, aquí, chiquita, izquierda, quieres*).

⁶⁰ “La articulación más común es, sin duda, la dental oclusiva sorda [t] [...] En posición final de sílaba el fonema conserva la misma articulación” (Perissinotto 1975:45).

Este alófono se refleja en el espectrograma con una elevación en la transición del segundo formante de la vocal, señalando la palatalización en la realización fonética de /k/ ante vocales palatales, como se aprecia en la figura 31.

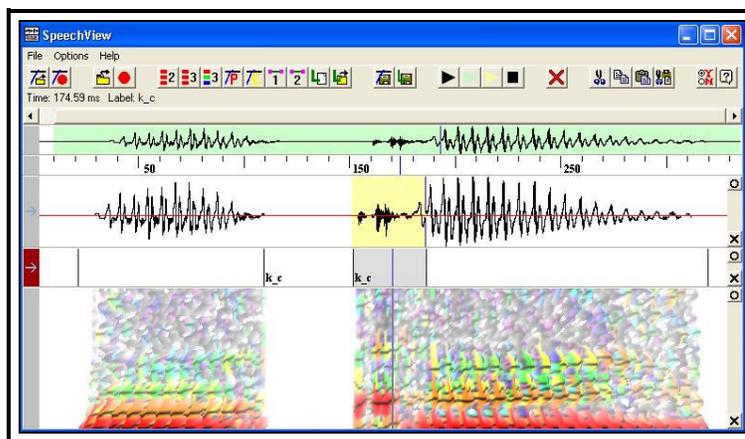


Figura 31. Mago: “Okey”. Archivo número 35 del diálogo s5-t1-g1. Se observa la modificación en la transición del segundo formante de la vocal

En su alfabeto de 1915, la *Revista de Filología Española* anotó las siguientes variantes postpalatales: [g´, k´, η´, ɣ´, x´]. Este fenómeno también fue mencionado por Navarro Tomás en su *Manual de pronunciación española* (1918:137) y en su *Cuestionario lingüístico hispanoamericano* (1945:37), donde aconseja a los encuestadores observar si ante /e, i/, el fonema /k/ presenta “avanzamiento”. Alarcos (1950) lo registró y llamó a estas consonantes “mojadas”;⁶¹ observación que también recoge Gili Gaya (1950).⁶²

⁶¹ “Éstas agregan al timbre particular de las correspondientes consonantes normales un matiz palatal semejante al de la vocal /i/, por tanto, una elevación de la elevación del timbre (reflejado en la elevación de la frecuencia de los formantes segundo y siguientes de la vocal adyacente)” (Alarcos 1950:80).

⁶² “El punto de articulación de la sorda *k* varía según la vocal siguiente. Si pronunciamos sucesivamente *ku, ko, ka, ke, ki*, observaremos que ante vocal velar el contacto se produce en el velo

En cuanto a los estudios de nuestro país, Matluck (1951:64) alude a esta forma prepalatal en el habla popular del valle de México. Perissinotto (1975) no la menciona.

En nuestro caso, la consonante velar oclusiva sorda mostró una palatalización constante ante vocales palatales en todos los diálogos; un ejercicio previo fue expuesto en Cuétara (2003) en el que, siguiendo las convenciones de Hieronymus (1994; 1997), se propuso el símbolo ASCII [k_j]. De tal suerte, esta forma ha sido incluida en el inventario de alófonos modelables por reglas para el fonetizador TranscríbEMex. Cabe mencionar que, hasta donde he podido investigar, este alófono no está considerado por ningún otro alfabeto computacional para la lengua española.

4.4.1.2. Consonantes oclusivas sonoras: /b, d, g/

En general, y conforme con lo que se ha expuesto tradicionalmente sobre las consonantes oclusivas sonoras, éstas tienen dos alófonos principales: su forma oclusiva y su realización aproximante.

del paladar; ante *a*, la *k* se articula aproximadamente en el límite posterior del paladar óseo; ante las vocales palatales *e*, *i*, la consonante es postpalatal. El latín antiguo empleaba a menudo tres letras distintas para expresar estas diferencias articulatorias: *q*, ante la semiconsonante *w* (*quatuor*, *quod*); *k* ante *a* (*kalendae*); *c* ante *e*, *i* (*centum*, *cinis*). La evolución románica refleja los diferentes puntos de articulación originarios latín *corpus* > fr. *corps*; lat. *caput* > fr. *chef*; lat. *centum* > fr. *cent*. En it., port. y esp. se mantiene la *k* latina ante *u*, *o*, *a* (*caballu* > *caballo*, *cavalho*, *cavallo*) y avanza en diversos grados el punto de articulación ante vocal de la serie anterior: latín *centum* > it. *cento* (š), port. *cento* (s), esp. *ciento* (θ)” (Gili Gaya 1950:132).

4.4.1.2.1. Alófonos aproximantes: /b, d, g/ > [β, ð, ɣ]

De manera constante en el español, y por consecuencia en todo el Corpus DIME, los fonemas /b/, /d/ y /g/ se presentan como oclusivos cuando se encuentran en posición inicial absoluta o ante pausa, y después de consonantes nasales (/d/, también después de /l/). No obstante, y como mencioné anteriormente (§4.2), se presentan más frecuentemente como aproximantes, tradicionalmente llamadas fricativas (Navarro Tomás 1918:84-85, 98-100; 137-139). Este fue un comportamiento constante a todo lo largo del corpus, por lo cual, en el inventario de Mexbet se suman a los alófonos oclusivos [b] (0.44%), [d] (1.04%) y [g] (0.21%) sus alófonos [β] (1.89%), [ð] (3.25%) y [ɣ] (0.91%), que se representan en Mexbet por los símbolos [V, D, G], respectivamente. Resulta obvio decir que, por su frecuencia y porque son modelables por reglas, estas seis formas serán incluidas en el fonetizador TranscríbEMex.

Es necesario señalar un hecho: los fonemas oclusivos sonoros, sobre todo sus formas aproximantes, presentan algunas otras variantes, pero éstas sólo aparecen de manera esporádica: se registraron algunos debilitamientos intervocálicos (2 casos para [β], 10 casos para [ð] y 1 caso para [ɣ]) y ensordecimientos (1 caso para [β] y 12 casos para [ð]), principalmente a final de sílaba. El debilitamiento de /d/ en posición intervocálica, como vemos, es considerablemente más alto que los demás fenómenos,

lo cual coincide con Moreno de Alba (1994:67-68).⁶³ Vemos en la figura 32 uno de los ejemplos que encontramos en el Corpus DIME.

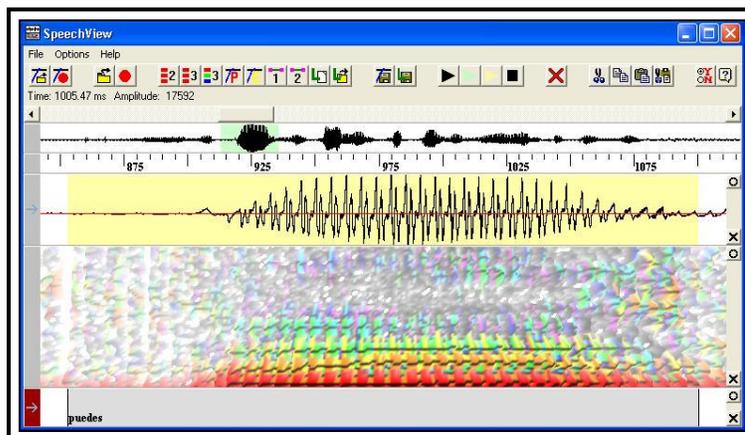


Figura 32. Sujeto 6: “Eh, ¿Me puedes mostrar los tipos de muebles que tengo?”. Archivo número 30 del diálogo s6-t1-g1. En la palabra *puedes*, el sujeto debilita tanto /d/, que prácticamente se elide

De igual manera, podríamos aludir a la pérdida de la consonante en la terminación [-aðo], señalada por muchos autores (Navarro Tomás 1918:100-103, por mencionar sólo uno). No obstante, es preciso decir que estos fenómenos no se presentan de manera constante y que su frecuencia de aparición es tan baja con relación a los demás cambios fonéticos, que solamente a partir de un estudio específico futuro se podría determinar la consideración de incluirlos o no en el inventario de alófonos de Mexbet, para el reconocedor de habla.⁶⁴

⁶³ En contraste, Matluck (1951:55) señala que el fonema /d/ en posición intervocálica casi nunca se pierde

⁶⁴ Moreno de Alba (1994: 67 ss.) trata el asunto de la relajación y pérdida, o del mantenimiento de la -d- intervocálica, y concluye que “puede decirse que la -d- intervocálica, en todo el país, de conformidad con los resultados del ALM y con la mayor parte de las referencias bibliográficas, tiende a conservarse. Las relajaciones no son muy notables y casi siempre se reducen a la terminación -ado” (73).

4.4.2. Consonante africada sorda: $\widehat{tʃ}$

Es preciso recordar que la consonante palatal africada sorda, como mencioné (§4.2.), presenta una frecuencia de aparición bajísima en nuestra lengua (menor al 1% para todos los autores). Consecuentemente, su ocurrencia en nuestro corpus fue muy baja (0.32%). En nuestro caso, las muestras para modelar reglas aplicadas a la creación de modelos acústicos son insuficientes, y la única forma que se propone es la prototípica. No obstante, y siendo conscientes de la variedad de formas que puede presentar este fonema,⁶⁵ en un futuro se debería evaluar la probabilidad de realizar un estudio específico para ella, para considerar, por ejemplo, la inclusión del alófono fricativo, muy representativo del habla del norte de México y presente también en muchas palabras del léxico indígena (aunque hay que mencionar que en el lexicón del Corpus DIME no apareció ninguna palabra de procedencia indígena).

En cuanto a las diferencias que registran los autores sobre las características del fonema $\widehat{tʃ}$, Perissinotto discute que su articulación “difiere significativamente de la que describe Navarro para el español peninsular” y hace ver su preferencia por la descripción de Matluck para el español de México. Lo que llama la atención es que, a

⁶⁵ Baste recordar a Navarro Tomás (1918:126) cuando señala que en la pronunciación dialectal la articulación del fonema $\widehat{tʃ}$ presenta muchas variantes. Para Moreno de Alba (1994:118), “la combinación de variantes articulatorias de estos dos tipos de rasgos [palatal y africado] produce una amplia gama de alófonos [...] y hacen de la articulación de /ch/ uno de los fenómenos fonéticos más productivos en el español mexicano”. Sobre el español de América, Vaquero (1996:54) señala también las variaciones que se pueden presentar en este fonema que “incluye realizaciones que van, desde la articulación africada propiamente dicha, con dos momentos de igual duración [oclusión + fricación], hasta el debilitamiento total de la articulación, con la pérdida del momento oclusivo y la articulación fricativa”.

diferencia de los muchos autores que se detienen a observar las variantes de $\widehat{t\text{ʃ}}$, Perissinotto (1975:50-51) sostiene que la única articulación del fonema es “dorso-prepalatal africada sorda”.

4.4.3. Consonantes fricativas sordas: /f, s, x/

En términos generales, el comportamiento que mostraron las consonantes fricativas sordas en el Corpus DIME coincide en lo esencial con los fenómenos de asimilación que los autores reportan. La frecuencia de aparición de /f/ y /x/ es baja (0.70% y 0.62%, respectivamente). En cuanto al fonema labiodental, éste no manifestó muchas modificaciones en nuestros datos. De tal suerte, la única forma que se propone para el modelado computacional es [f].

Por su carácter anecdótico y sorprendente, quiero citar lo que Perissinotto reporta sobre este fonema: encontró que la variante más común es la bilabial,⁶⁶ lo cual no concuerda en lo absoluto con los datos del Corpus DIME.

El modelado matemático de /f/ no representa mayor problema para el español de México. No obstante, hay que mencionar que en el caso de un dialecto como el castellano, en cuyo sistema fonológico se encuentra /θ/, se presentan confusiones

⁶⁶ “En la mayor parte de las ocasiones el fonema /f/ se pronuncia como bilabial [ϕ]: los labios se juntan para producir fricción, y sólo esporádicamente se retrocede el labio inferior para juntarse con los incisivos superiores. La variante [ϕ] puede encontrarse en cualquier posición, pero siempre ante /ue, ui/” (Perissinotto 1975:55).

acústicas entre ambos, lo que obliga a modelar con muchísima mayor precisión cada una de estas formas.⁶⁷

En cuanto a /x/, su baja ocurrencia en la lengua española (por debajo del 0.74% para todos los autores), provoca que no se presente ni frecuentemente ni en muchos contextos. De tal suerte, la única forma que se propone para el modelado computacional es la velar fricativa sorda [x]. Ocurrieron en el corpus, no obstante, dos fenómenos: algunas palatalizaciones (4 casos en total: ['x^ji ra], *gira*, [su^j x^jje res] *sugieres*) ante las vocales /e, i/, con el mismo comportamiento que vimos para el fonema /k/ (§4.4.1.1.1.), y alguna aspiración (2 casos en total: ['i xo le], *hijole*, [rɛ fri x^je ra 'dɔr], *refrigerador*). Por insuficiencia de datos, no podemos considerar, al menos no por el momento, ninguna de estas variantes.⁶⁸

El fonema /s/ es la consonante que más frecuencia de aparición tiene en el español.⁶⁹ Por esta razón, es de esperar que, a su vez, sea uno de los fonemas que muestre mayor variedad de alófonos. En efecto, el comportamiento de /s/ en el Corpus DIME presentó alargamientos, dentalizaciones, sonorizaciones, aspiraciones, debilitamientos e, inclusive, elisiones. No obstante, las formas que mostraron un comportamiento constante y una ocurrencia considerable son tres: el alófono

⁶⁷ Para conocer de qué manera el modelado computacional debe considerar diferentes aspectos en diferentes dialectos, puede consultarse el interesante trabajo de Santiago Fernández y sus colaboradores (1998) en el que buscaron indicios acústicos contextuales para diferenciar /f/ de /θ/ en un sistema de reconocimiento de habla. Se encontraron con que, para ello, se debe considerar la región baja del espectro donde el contexto fónico, principalmente el vocálico, marca la diferencia.

⁶⁸ El alófono palatalizado de /x/ fue también observado por Perissinotto (1975:60); las aspiraciones propias de Yucatán y la costa del Golfo de México son ampliamente conocidas (Canfield 1981:62).

⁶⁹ De hecho, para todos los autores es la consonante de mayor frecuencia de aparición; mayor al 8% en todos los casos.

dentalizado [s̺], el sonorizado [s̻] y su forma prototípica [s]. Las reglas y símbolos en Mexbet se pueden ver en la figura 33.

Alófono	Contexto	Símbolo Mexbet
Dentalizado	- Ante consonantes dentales	s_[]
Sonorizado	- Ante consonantes sonoras - Entre vocales	z
Alveolar	- En todos los demás contextos	s

Figura 33. Alófonos y reglas del fonema /s/, para el español de la ciudad de México

4.4.3.1. Alófono dentalizado: /s/ > [s̺]

En la descripción fonética del español, tradicionalmente se ha considerado la dentalización que sufre el fonema /s/ cuando se encuentra a final de sílaba ante consonante dental (*estante, estufa, lavatrastes*). Desde Navarro Tomás (1918:104) lo registran prácticamente todos los autores. No obstante, este fenómeno de asimilación ha sido objeto de discusión desde que a mediados de la década de los sesenta Antonio Quilis publicó en la *Revista de Filología Española* el importante artículo “Sobre los alófonos dentales de /s/” (1966/1968), en el que pone en tela de juicio la dentalización del fonema /s/.⁷⁰

⁷⁰ “Desde hace tiempo, desde que empezamos a trabajar en el dominio de la fonética con el sonógrafo, veníamos dudando de la existencia real de estos dos alófonos dentales ante /t, d/, pues observábamos en los sonogramas que la frecuencia a la que aparecían los resonancias del ruido de la fricación eran

En 1995, en la misma revista, aparecieron dos artículos que retoman este tema. El primero, “Nuevos datos sobre la dentalización de -s- en español”, de Martínez Celadrán, critica a Quilis y afirma que sí se deben considerar los alófonos dentalizados de /s/.⁷¹ Para la comprobación de su hipótesis, Martínez Celadrán (1995) se valió de la electropalatografía. Para la defensa de sus argumentos originales, Quilis (1995) criticó fuertemente el experimento de Martínez Celadrán y resaltó el argumento de que la electropalatografía no aporta datos sobre la zona de los dientes.⁷²

Nos hemos detenido en estas cuestiones porque resulta interesante, y anecdótico, ver de qué manera los dos autores utilizan defensas a favor de sus trabajos y ataques en contra del otro. En realidad, el número de informantes y la cantidad de datos que reportan los trabajos de Martínez Celadrán (1995) y Quilis (1966/1968, 1995) son bajos.⁷³

Ante todas estas cuestiones, y específicamente para la síntesis de habla, Llisterri y sus colaboradores (1999) determinan que no es pertinente transcribir los alófonos dentales de /s/ si se toman como base los resultados acústicos. No obstante, para este trabajo considero necesario tomar en cuenta varios factores que obligan a

iguales aproximadamente para [s] y [ʃ]” (Quilis 1966/1968:335-336). En nota, incluso, afirma que en la segunda edición (1966) de su *Curso de Fonética* ya no incluyó este alófono.

⁷¹ “El punto de vista acústico es más complejo de lo que Quilis plantea; [...] existen diferencias significativas en la articulación de la s ante /t, d/, respecto de la intervocálica y ante otras consonantes” (Martínez Celadrán 1995:302).

⁷² “De estos datos se deduce fácilmente la imposibilidad de comprobar con este sistema [la electropalatografía] la dentalización o no de [s] ante [t] o [d]” (Quilis 1995:314).

⁷³ Cabe señalar que el estudio articulatorio de Quilis se basa en un informante —comprobado posteriormente con los datos de un segundo locutor— y el análisis acústico en la producción de pares mínimos por parte de tres hablantes; el de Martínez Celadrán analiza acústica y articulatoriamente 20 secuencias vocal-s/st/sd-vocal, producidas por cuatro informantes. Independientemente de las diferencias de fondo en los conceptos de anticipación articulatoria y asimilación considerados por ambos investigadores, puede verse que en esta, y en otras discusiones similares, los datos de los que se dispone no son precisamente abundantes” (Gil y Llisterri, en prensa).

mantener el alófono dentalizado de /s/ en nuestro inventario fonético. El más importante es que la /s/ mexicana tiene unas características muy particulares, diferentes a la /s/ castellana.

En efecto, las diferencias entre la /s/ castellana y la mexicana son notables. La /s/ alveolar mexicana tiene un carácter más estridente y largo, según lo registran varios autores (Matluck 1951:72-73; Lope Blanch 1963/1964; Perissinotto 1975:56); para Moreno de Alba (1994:13-14), el fonema /s/ es “de timbre agudo, singular entre todas las del mundo hispánico”. Esto provoca que, entre las formas alveolar y dental de /s/, sí haya diferencia. Creo que los argumentos de Quilis y Martínez Celdrán no son aplicables a la realidad mexicana; podemos observar en el oscilograma y en el espectrograma de la figura 34 que el punto de unión entre /s/ y /t/ provoca cambios acústicos apreciables. Durante el etiquetado computacional fue evidente siempre que la onda sonora de /s/ tiene un comportamiento muy diferente en este contexto: en la unión de ambos alófonos y en la pérdida del cierre en el fonema /t/; me parece que este hecho puede resultar una ayuda útil para el reconocedor de habla.

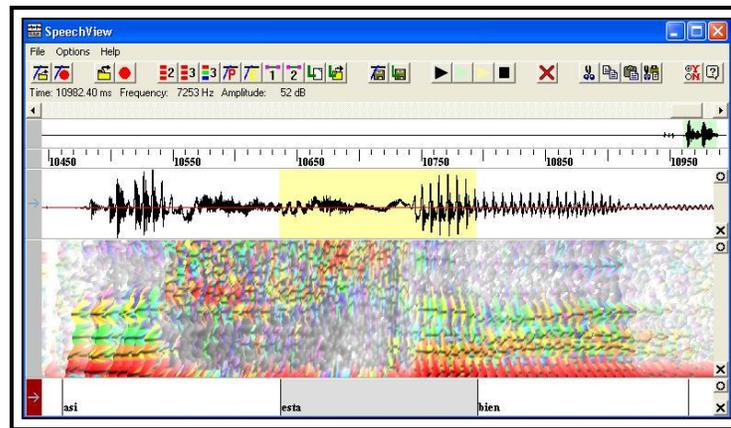


Figura 34. Mago: “¿Así está bien?”. Archivo 27 del diálogo s6-t1-g1. En la contigüidad de /s/ y /t/ se aprecia la modificación del oscilograma en la /s/ y la pérdida de la oclusión del fonema /t/

4.4.3.2. Alófono sonorizado: /s/ > [s̥]

Algunos estudiosos del tema han registrado la sonorización del fonema /s/ con diferentes matices,⁷⁴ pero prácticamente todos los autores coinciden en señalar su aparición ante consonantes sonoras y entre vocales. En los alfabetos fonéticos computacionales se ha representado tradicionalmente este alófono con el símbolo [z]. Estamos de acuerdo con Quilis (1966/1968) en considerarlo un alófono sonorizado y no sonoro, lo que estrictamente lo representaría como [s_v] en Mexbet. No obstante, para no sobregenerar formas y para evitar confusiones, he optado por conservar el símbolo [z] que ya había heredado Mexbet de Worldbet, como vimos en el capítulo anterior (§3.3.).

⁷⁴ Por tomar sólo a tres autores, destaquemos el hecho de que para Navarro Tomás (1918:107) “su presencia es anormal y esporádica”; Matluck (1951:75-76) la registra solamente ante consonantes nasales y oclusivas sonoras; Perissinotto (1975:58) menciona la sonorización en posición intervocálica, pero apunta que es un fenómeno que solamente “sucede con alguna frecuencia”.

Debemos destacar el hecho de que para los diferentes autores el fonema /s/ presenta muchas variantes articulatorias en el español de México.⁷⁵ Otros fenómenos esporádicos son las aspiraciones y elisiones. Sobre éstas últimas, en el Corpus DIME se presentaron únicamente un par de casos excepcionales, hecho que coincide con las observaciones de Moreno de Alba (1994:91) en cuanto a que no hay relajamiento de /s/ implosiva en la ciudad de México; sin embargo, estos fenómenos serían parte de un trabajo para variantes como las que registra Moreno de Alba (1988:161) para Guerrero, Tabasco y Veracruz. Por el momento, no obstante, adopto la idea de López Chávez (1981) —a propósito de la gran variedad de realizaciones del fonema /s/ como las que se desprenden del *Atlas Lingüístico de México*— en cuanto a que resulta necesario reducir el inventario de alófonos y tratar de englobar en la menor cantidad de símbolos posibles a todas ellas; por ello, la frecuencia de aparición nos ayuda a saber cuáles son las variantes alofónicas pertinentes: “acumular variaciones en forma exhaustiva y describir sin definir nos puede llevar a una situación caótica en la que se invierta mucho esfuerzo y se obtenga poco fruto. El estudio de la s exige un diseño que agote todas las posibilidades y que tenga un alto rendimiento funcional” (López Chávez 1981:782).

⁷⁵ Raúl Ávila (1973), por ejemplo, distingue realizaciones tensas, flojas y aún africadas. Las conclusiones a las que llegó muestran que, en términos generales, las realizaciones tensas predominaron sobre las flojas y las africadas en todos los contextos, pero que éstas no dejaron de presentarse ocasionalmente (las africadas, tras pausa, /n/ o /l/). En mi corpus, estas últimas no se presentaron.

4.4.4. Consonante fricativa sonora: /ʒ/⁷⁶

El fonema /ʒ/ es otra de las consonantes que tienen una bajísima frecuencia de aparición en la lengua (por debajo del 0.69% para todos los autores). En el Corpus DIME su frecuencia de aparición fue un poco más alta que para los demás autores (0.85%). Los alófonos del fonema /ʒ/ español son dos principalmente: el fricativo y el africado.

4.4.4.1. Alófono africado: /ʒ/ > [dʒ]

La forma africada del fonema /ʒ/ ha sido registrada en los estudios fonéticos hispánicos tradicionales, precedida de /n, l/ y en posición inicial absoluta (Alarcos 1950:163; RAE 1973:26; Gili Gaya 1950:155; Perissinotto 1975:52, entre muchos otros).⁷⁷ También resulta apropiado comentar que para Matluck, la realización de la consonante palatal fricativa sonora es “bastante abierta” y su pronunciación en México se acerca más al alófono paravocálico [j] de la vocal /i/ (1951:99), asunto con

⁷⁶ Este símbolo que he adoptado para la transcripción del fonema palatal fricativo sonoro, se detalla en el apéndice (§6.)

⁷⁷ Contrariamente a lo que todos los estudios fonéticos hispánicos han señalado hasta el momento, en la última ilustración del AFI para el español (Martínez Celadrán *et al.* 2003:258) se marca al alófono palatal africado sonoro [ʒ] como el fonema base: “In Spanish there are two voiced palatal variants of the phoneme /ʒ/ [...]: an affricate variant [ʒ], after a nasal, [l] or a pause, and an approximant [j] in all other contexts: [elⁱ ¹ʒa te] *el yate* «the yacht», [mi ¹ja te] *mi yate* «my yacht». Quizás esto es producto de la confusión que siempre ha existido para este fonema, como lo mencioné en la sección dedicada a Mexbet (3.3.) y como lo detallo en el apéndice (§6.), pero desde mediados del siglo pasado este asunto quedó aclarado por Alarcos (1950:163) desde el punto de vista funcional: “Los sonidos [ʒ] y [y], sonoros mediopalatales, africado y fricativo, respectivamente, son realizaciones fonéticas de un mismo y único fonema /y/”.

el que no coincido, pues nos habla de una debilitación que me parece más bien representativa de las hablas del norte de México —variantes registradas por Canfield (1981:62)— y no precisamente del Altiplano Central. Así, siguiendo la tradición de Mexbet y Worldbet, para [ʒ] y [dʒ] los símbolos ASCII serán [Z] y [dZ], respectivamente. Es preciso indicar, no obstante, que la ocurrencia de ambos alófonos fue bajísima (0.56% y 0.29%, respectivamente).

4.4.5. Consonantes nasales: /m, n, ɲ/

Hemos visto el comportamiento que presentaron las consonantes obstruyentes⁷⁸ en el Corpus DIME. A continuación revisaremos qué fenómenos presentaron las sonantes (nasales y líquidas).

De las consonantes nasales, /m/, con una frecuencia de aparición del 3.06%, no presentó modificaciones en nuestro corpus (algún relajamiento o ensordecimiento esporádicos).⁷⁹ En el caso de /ɲ/, su bajísima frecuencia de aparición (0.24%) provocó que sólo surgieran un par de formas, y éstas sin cambios.

En cuanto al fonema /n/, al ser, junto con /s/, una de las consonantes de mayor ocurrencia en el español (con una frecuencia de aparición del 6.91% en los cuatro diálogos), no es de sorprender que haya presentado un número considerablemente

⁷⁸ De acuerdo con la clasificación de la Real Academia, el término obstruyente “agrupa a todas las consonantes no sonánticas, clase que comprende las consonantes oclusivas (entre ellas suelen incluirse las consonantes africadas) y las fricativas, llamadas también constrictivas” (RAE 1973:18).

⁷⁹ Cabe mencionar que en el léxico de los cuatro diálogos del Corpus DIME que se estudiaron en este trabajo, nunca se presentó la combinación de una consonante labiodental precedida por una consonante nasal, hecho que seguramente hubiera mostrado el alófono nasal labiodentalizado.

alto de formas. Herrera (2002) se refiere a la inestabilidad del punto de articulación de las nasales, que asimilan su punto de articulación al de la consonante que le antecede. En mi análisis de los 4 diálogos del Corpus DIME, fueron constantes y suficientemente frecuentes la dentalización y la velarización. Así, los datos nos permiten considerar, además de su forma prototípica, estos dos alófonos.

4.4.5.1. Alófono dentalizado: /n/ > [n̪]

La variante dental [n̪] (1.22% en el Corpus DIME) se presentó ante consonantes dentales, lo cual coincide con los reportes de Navarro Tomás (1918:104) y de Perissinotto (1975:61). Acústicamente, /n/ ante consonantes oclusivas pierde su cierre, y el segundo formante presenta un ligero descenso, como se puede apreciar en la figura 35.

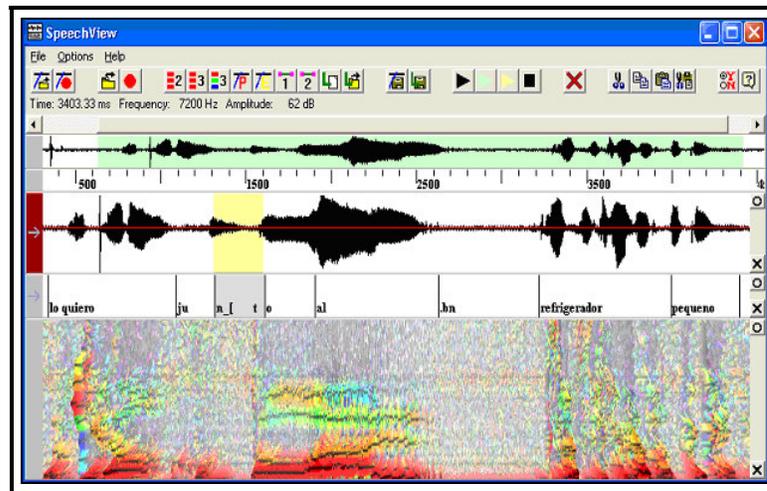


Figura 35. Sujeto 4: “Lo quiero junto al... refrigerador pequeño”. Archivo 28 del diálogo s4-t1-g1. Se aprecia el descenso en el segundo formante y la pérdida del cierre de la consonante oclusiva /t/ ante la nasal /n/

4.4.5.2. Alófono velarizado: /n/ > [n^v]

La velar [n^v] (0.67% en el Corpus DIME) se presentó ante velares, como lo señalan Navarro Tomás (1918:141) y Perissinotto (1975:61). En el análisis espectrográfico se mostró un descenso en las frecuencias del segundo formante e inalteración del primero, coincidiendo con los señalamientos de Quilis (1981); asimismo, la consonante oclusiva perdió su cierre, como se puede apreciar en la figura 36.

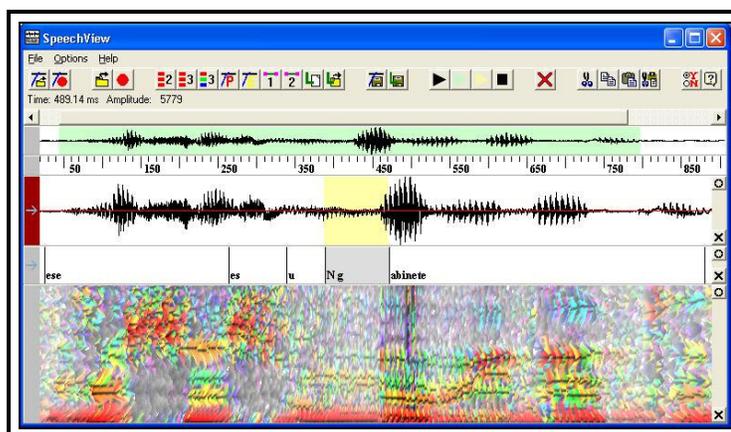


Figura 36. Mago: “Ese es un gabinete”. Archivo 60 del diálogo s4-t1-g1. Se aprecia el ascenso en el segundo formante y la pérdida del cierre de la consonante oclusiva /g/ ante la nasal /n/

Con base en estos datos, en la figura 37 se presentan los contextos de aparición y los símbolos para los alófonos del fonema /n/:

Alófono	Contexto	Símbolo Mexbet
Dentalizado	- Ante consonantes dentales	n_[]
Velarizado	- Ante consonantes velares	N
Alveolar	- En todos los demás contextos	n

Figura 37. Alófonos y reglas del fonema /n/ para el español de la ciudad de México

La palatalización del fonema /n/ en posición final de sílaba o de palabra ante consonante palatal también es bien conocida (Navarro Tomás 1918:133), pero la bajísima frecuencia con que aparece en la lengua este contexto fónico no hizo

candidato al alófono [n^j] para el modelado computacional en este trabajo.⁸⁰ Un investigación futura deberá determinar su inclusión o no, aunque se debe recordar que para Perissinotto (1975:61) esta asimilación no ocurre “tan regularmente como se podría esperar”.

4.4.6. Consonantes líquidas: /r, r, l/

En el caso de las líquidas, no consideré por el momento la inclusión de ningún alófono más allá de sus formas prototípicas. En cuanto a la lateral, el fonema /l/ tomó en algunos casos el punto de articulación dental ante /t/, como mencionan Navarro Tomás (1918:113-114) y Perissinotto (1975:63). No obstante, el alófono dentalizado de /l/ se presentó solamente en 5 ocasiones en los cuatro diálogos (*alto, falta, faltaría, volteada, vuelta*), por lo cual no resulta oportuno invertir espacio para este fenómeno en el reconocedor de habla. Otros contextos, como /l/ ante /d/ o /ʒ/, no ocurrieron.

Por lo que a las consonantes /r/ y /r/ respecta, las variantes que se presentaron fueron ensordecimientos (2 casos en mi corpus), debilitaciones (33 casos) y asibilaciones (23 casos). No obstante, debe resaltarse el hecho de que estos fenómenos mostraron un comportamiento sumamente inestable y tampoco fue posible establecer reglas precisas para considerarlos. Por el momento, pues, los dos

⁸⁰ En todo el Corpus DIME, las dos únicas palabras en que ocurre la combinación de /n/ ante consonante palatal son las palabras *ancha* y *enchufes*, mismas que no aparecieron en los diálogos s2-t1-g1, s4-t1-g1, s5-t1-g1 y s6-t1-g1.

alófonos que propongo para ser incluidos en los modelos computacionales para el reconocedor, son [r] y [r] (en Mexbet, [r()] y [r], respectivamente).

Lo que se deberá tomar en cuenta para desarrollos y modelos posteriores es la posición de /r/ y /r/ en la sílaba: al final de ésta, prácticamente en todos los casos sufrirán una u otra modificación (relajamiento, reforzamiento, ensordecimiento o asibilación). Una de estas modificaciones, la que creemos que merece un trabajo de investigación completo, actual y profundo, es el fenómeno de la asibilación en la ciudad de México, y aún en el país, como veremos a continuación.

4.4.6.1. Asibilación de /r/ y /r/

La asibilación de /r/⁸¹ se ha descrito, desde Navarro Tomás, como un fenómeno propio de alguna región de España (Álava, Navarra, Rioja y Aragón), y “muy extendido en la América de la lengua española” (Navarro Tomás 1918:120). Matluck la registró únicamente como de efecto esporádico en ciertos contextos muy particulares (1951:88 ss.) y concluyó que “las variedades fricativas y asibiladas pertenecen, en el Valle, más bien al habla de las personas incultas y, en menor grado, a las semicultas” (Matluck 1951:98). Para Canfield (1962:87-88), este fenómeno se

⁸¹ Navarro Tomás (1918:46) definió la asibilación de /r/ de la siguiente manera: “En algunos lugares la *r* se pronuncia con sonido relativamente tenso, fricativo y rehilante. Esta variedad de *r* puede resultar más o menos sorda detrás de las consonantes *p*, *t*, *k*. La *r* asibilada detrás de la *t* ofrece un timbre semejante al de la *r* inglesa en *tree*. Nótese también la posible influencia de estar *r* en la articulación de la *t* (haciéndola alveolar en vez de dental)”.

encuentra muy extendido en América y en camino de convertirse en variante complementaria y constante.

Uno de los trabajos que ha estudiado la asibilación de /r/ y /r/ en la ciudad de México con más detalle es el de Perissinotto (1975). Para /r/, Perissinotto afirma que la asibilación es un rasgo constante en posición final absoluta y esporádicamente en posición final de sílaba y en el grupo /tr/, en cualquier grupo social de la población (109); para /r/, concluye que presenta asibilaciones en cualquier contexto fonético y en todos los estratos de la población (63 ss.). Asimismo, indica que las mujeres asibilan con más frecuencia que los hombres (77).

En un estudio previo publicado en 1972, Perissinotto adelantó sus notas sobre la asibilación de /r/ y /r/ en la ciudad de México. En éste, el autor revela que la asibilación de /r/ “se da con frecuencias muy altas en el habla en todos los sectores de la población”. El mismo año y en la misma revista, Moreno de Alba (1972) presentó otro análisis de la asibilación de /r/ y /r/, pero con dos particularidades: 1) revisó además el contexto en que se presentan las asibiladas a final de sílaba, y 2) lo registró en todo el país. Este autor afirma también que la asibilación es más común en /r/ que en /r/, y que las mujeres asibilan más que los hombres. Destaca en su trabajo la observación de que este fenómeno “es menos frecuente a nivel «nacional» que a nivel «capitalino», de manera que, desde un punto de vista geográfico, podría verse en la ciudad de México un foco irradiador del fenómeno” (Moreno de Alba 1972:370). Mas recientemente, y con base en el *Atlas Lingüístico de México*, Moreno de Alba

(1994:132) dibuja una vez más el centro del país como principal zona de asibilación. Por otra parte, registra un 75% de porcentaje promedio de asibilación para las mujeres, y un 52% para los hombres (131).

En el caso del Corpus DIME, los cinco hablantes —hombres y mujeres— presentan en algún momento asibilación de estos fonemas. A manera de ejemplo, tomemos el caso del Sujeto 6, un hombre que cuando asibila /r/ en posición final, realiza además alargamiento,⁸² tal como se puede apreciar en la figura 38.

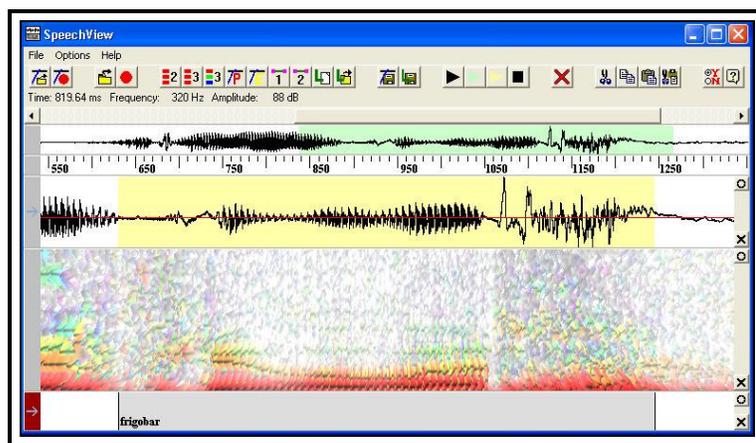


Figura 38. Sujeto 6: “Arriba del frigorbar”. Archivo 54 del diálogo s6-t1-g1. Se aprecia claramente, además de la asibilación, un alargamiento en la realización del fonema /r/ en posición final absoluta

No obstante, si contabilizamos los casos de asibilación y no asibilación en los mismos contextos, nos encontramos con que no hay una consistencia para la formalización de este fenómeno. Por ejemplo, en toda su participación en el diálogo s6-t1-g1, este sujeto pronunció 21 veces palabras que tienen el fonema /r/ en posición

⁸² Quizás otro factor que habría que replantearse en el estudio de la asibilación de /r/ y /r/ en México, es el hecho de que ya no haya una correspondencia clara entre los términos vibrante simple/asibilada breve y vibrante múltiple/asibilada larga que estableció Lope Blanch (1978).

final absoluta: de esos casos, en 9 ocasiones (42.86%) se presentaron asibilaciones, mientras que 12 casos (57.14%) no las presentaron.⁸³ Se observó, además, alternancia de formas en palabras recurrentes, como *extractor* y *favor*.

Sobra decir que el fenómeno de asibilación en nuestro corpus es aún más inconsistente en el caso de /r/, lo cual es el resultado de su bajísima ocurrencia en la lengua española; como vimos en la figura 27, la frecuencia de /r/ está entre el 0.40% de Llisterra y Mariño (1993) y el 1.04% de Zipf y Rogers (1939). En los cuatro diálogos del Corpus DIME analizados, su frecuencia fue solamente del 0.65%. Debido a la importancia que la asibilación representa, lo que he decidido por lo pronto —al menos para el nivel fonológico— es la inclusión del archifonema /-R/, como detallaré (§4.6.) que ayudaría al reconecedor de habla a encontrar fenómenos diversos en posición final de sílaba, que es cuando se debilitan, se ensordecen, se pierden o se asibilan /r/ y /r/. No podemos dar por concluido el tema de la asibilación sin tomar en consideración una cita de Perissinotto (1975:115): “creo que se puede proyectar la asibilada hacia el futuro y afirmar que irá adquiriendo más popularidad debido al impulso recibido por las mujeres de todas las edades y grupos socioeconómicos, por la clase media de ambos sexos, y con el consentimiento tácito de los hombres de la clase alta”. En cuanto a asibilación de /r/ y /r/, me parece que siguen vigentes muchas de las observaciones de Perissinotto y de Moreno de Alba; quizá el fenómeno merece la realización de una nueva investigación, pues parece ser

⁸³ Con asibilación: *extractor* (elocuciones 107 y 110), *favor* (20 y 43), *frigobar* (54 y 70), *mejor* (104), *refrigerador* (48), y *ver* (37). Sin asibilación: *extractor* (elocución 25), *favor* (43), *mostrar* (33, 59, 61 y 86), *mover* (3 y 48), *poner* (41 y 64), *quitar* (78), y *superior* (102).

que se han ampliado considerablemente la frecuencia de aparición y los contextos, y que ahora se escuchan asibilaciones en cualquier hablante de cualquier sexo, nivel socioeconómico y educación.

4.5. Vocales

En este trabajo señalé con anterioridad (§4.1.) cómo las vocales representan prácticamente la mitad de las ocurrencias en el español, y cómo hay un correlato entre el aumento en la frecuencia de aparición de los fonemas y la variedad de formas fonéticas. A continuación reviso el comportamiento de las vocales en el Corpus DIME, y hago las observaciones apropiadas para incluir en el fonetizador TranscríbEMex ciertas formas modelables por reglas, que son básicamente los desplazamientos hacia la zonas palatal y velar de /a/, la abertura de /e/ y /o/ y las formas paravocálicas de /i/ y /u/.⁸⁴

4.5.1. Vocal abierta: /a/

Debido a su alto índice de ocurrencia en la lengua —en todos los autores por arriba del 12%: el menor del 12.19% (Quilis 1981) y el mayor del 14.0% (Zipf y Rogers 1939), y el 14.05% en el Corpus DIME—, la vocal abierta presenta todas las

⁸⁴ Debido a que en uno de los desarrollos sobre tecnologías de habla más avanzados —el de los laboratorios Bell— se determina que la tonicidad en las vocales no es un factor necesario en el proceso de reconocimiento de habla (Bacchiani *et al.* 2001), cuando inicié el análisis fonético del Corpus DIME no consideré el factor acentual; sin embargo, hoy me parece que en un futuro cercano el acento deberá formar parte de los modelos acústico del Corpus DIME.

variedades de realizaciones propias de las vocales (relajamientos, alargamientos, nasalizaciones, etcétera). No obstante, para optimizar el volumen de alófonos considerados en el inventario fonético básico para un reconocedor de habla, se han diseñado las reglas para tres formas principales: una vocal central prototípica, una velarizada y una palatalizada.

4.5.1.1. Alófonos velarizado (/a/ > [a̠]) y palatalizado (/a/ > [a+])

La palatalización y la velarización de la vocal abierta central han sido documentadas por varios autores. En cuanto a la velarización, los investigadores coinciden en términos generales en la ocurrencia de esta variante ante la consonante velar fricativa sorda, trabada por lateral y ante vocales posteriores (Navarro Tomás 1918:55-56 y 1945:23; Alarcos 1950:148; RAE 1973:31, y Perissinotto 1975:24),⁸⁵ lo cual se manifestó en el Corpus DIME. Así, al considerar que los dos primeros formantes de la vocal abierta se encuentran muy próximos (Quilis 1981:138), el desplazamiento hacia la parte inferior que se refleja en el segundo formante podría darle indicios al reconocedor de habla de que el segmento siguiente es una consonante velar o una vocal posterior. En el caso del fenómeno de palatalización, éste ha sido registrado ante consonantes palatales y en el diptongo /ai/ por Navarro Tomás (1918:55;1945:23), Alarcos (1950:148), Matluck (1951:6) y Perissinotto (1975:24).

⁸⁵ Por el carácter anecdótico que representa el contraste con prácticamente todo lo que se ha dicho a ese respecto, menciono aquí que para Matluck la vocal abierta tónica ante /x/ no es completamente velar, pero trabada por /l/ “es menos velar aún” (1951:6). Todavía más sorprendente resulta la afirmación de que, ante vocales posteriores, la vocal abierta “tiende a un timbre medio y a veces hasta palatal” (1951:7).

4.5.2. Vocales medias: /e, o/

Como he señalado anteriormente, las vocales presentan múltiples realizaciones, hecho que concuerda con su alta frecuencia de aparición. En el caso de las vocales medias, éstas se presentan nasalizadas, relajadas, etcétera). No obstante, por su frecuencia de aparición y sistematicidad, en este trabajo considero solamente dos alófonos para las vocales medias: la realización prototípica y la abierta.

4.5.2.1. Alófonos abiertos: /e/ > [e̞]; /o/ > [o̞]

La abertura de las vocales medias (y aún de las cerradas), en términos generales, ha sido resaltada por diferentes investigadores (Navarro Tomás 1918:52-53 y 59; 1945:24; Matluck 1951:8-9; RAE 1973:29 y 31; Perissinotto 1975:25), que la identifican en contacto con /r/ (*corredor, cerrar, rojo, borrar*) y en posición trabada por una consonante (*delgada, izquierda, control, cuántos*).⁸⁶

Es preciso mencionar que la sistematicidad en la abertura de las vocales medias ha sido cuestionada por algunos autores,⁸⁷ pero considero que un primer

⁸⁶ La excepción es que las nasales, la alveolar fricativa sorda y la dental oclusiva sonora no abren a la vocal media palatal.

⁸⁷ «Una de las cuestiones sujetas a discusión en la caracterización del vocalismo es la de la sistematicidad de los alófonos abiertos o cerrados en función del contexto que, en su momento, describió Navarro Tomás (1918). Monroy (1980), empleando datos propios y de otros autores, concluye que «en español académico, y por lo que respecta al cuadro de monoptongos, no parece que se registren los alófonos que se les han venido atribuyendo» (p. 55); Álvarez González (1981) y Martínez Celdrán (1984a) llegan a similares conclusiones. La posición de estos autores coincide con la de Quilis (1981a:145) —«en condiciones normales, las realizaciones de los fonemas vocálicos

ejercicio debe incluir estos alófonos, y posteriormente validarse mediante estudios más profundos la aplicabilidad de las reglas que se proponen aquí. La elevada frecuencia de aparición de la vocal /e/ provocó que en las muestras del Corpus DIME esta vocal manifestara muy diversas formas llegando, incluso, a combinarse en una misma realización más de un fenómeno fonético (por ejemplo, se inicia como nasalizada y al final se debilita o termina como abierta). Precisamente por su alta frecuencia de aparición y por la gran diversidad de fenómenos que pueden influir para sus múltiples emergencias, parece necesario revisar profundamente algunos aspectos de las vocales medias en el español de México señalados por Lope Blanch (1973) y otros autores, temas que no es posible abordar en este trabajo.⁸⁸

La abertura de las vocales medias se manifiesta en una elevación frecuencial en el primer formante (Quilis 1981:148). A este respecto, se han hecho algunos acercamientos recientes, como el trabajo instrumental de Madrid y Marín (2001) en el que tratan de determinar los valores típicos de las frecuencias de los tres primeros formantes de las vocales medias en el español de México, y si el tipo de sílaba influye en la vocal. Los autores concluyen que las vocales medias del español tienden a articularse como más abiertas en sílaba trabada.⁸⁹

españoles no se producen en distribución complementaria»—, y es también la que aceptan, entre otros, D’Introno *et al.* (1995) y Veiga (2002b) en sus análisis fonológicos del sistema vocálico español” (Gil y Llisterri, en prensa).

⁸⁸ Por ejemplo, “el frecuente cierre de *e* y *o* átonas en diversas posiciones” (Lope Blanch 1973:34). El autor indica cómo el carácter distensivo de los sonidos finales tiende a cerrar las vocales, cierre que es además favorecido por un contexto palatal. Pone como ejemplo un corpus con hablantes de Michoacán, pero señala que este fenómeno se puede hacer extensivo a muchas otras regiones de nuestro país.

⁸⁹ El contraste entre vocales medias y sus alófonos abiertos en el español, está claro, no es tan grande como en las lenguas en las que /e/ y /ɛ/, y /o/ y /ɔ/ son fonémicas (catalán e inglés, por ejemplo). En

4.5.3. Vocales cerradas: /i, u/

Las vocales /i/ y /u/, al igual que las demás vocales, presentan cierta variedad de realizaciones (nasalizaciones y debilitamientos, entre otros fenómenos). Para ambas vocales, hubo formas abiertas; formas que se presentaron en los mismo contextos que reportan los trabajos lingüísticos tradicionales, como Navarro Tomás (1918) y Quilis (1981), pero en la propuesta de reglas no fue oportuno incorporarlas debido a su baja frecuencia de aparición. Sin embargo, son constantes y modelables sin cuestionamientos sus realizaciones prototípicas y las paravocálicas.

4.5.3.1. Paravocales: /i/ > [j]; /u/ > [w]

Las formas que se proponen para el fonetizador TranscríbEMex, además de [i] y [u], son las paravocales en diptongos crecientes y decrecientes. Por el momento, se ha continuado con la tradición de Mexbet de transcribir las paravocales, ya sea al inicio o al final de diptongo, con una misma forma: [j] y [w],⁹⁰ lo cual es eficiente para señalar su carácter silábico no nuclear. No obstante, en un estudio futuro resultará pertinente para el Proyecto DIME determinar si se conservan estos dos únicos símbolos, o si resulta necesario determinar el carácter prenuclear o postnuclear de las

éstas, la diferencia de frecuencia en el primer formante de unas y otras suele superar los 150 Hz. En el ejercicio de Madrid y Marín no hubo diferencias por encima de los 64 Hz.

⁹⁰ He notado que en todos los alfabetos fonéticos computacionales se consideran únicamente las formas [j] y [w] sin distinguir el carácter semivocálico o semiconsonántico que detallan el alfabeto y los estudios hispánicos.

paravocales, en cuyo caso los símbolos en Mexbet serían [j] y [w] al inicio de diptongo, y [i(] y [u(] al final de diptongo. Actualmente se elabora otra investigación en el proyecto DIME relacionada con estos aspectos.⁹¹

4.5.4. Otros fenómenos vocálicos

En el desarrollo de este capítulo he dado cuenta de los fenómenos fonéticos que arroja el análisis del Corpus DIME y he distinguido aquellos que son recurrentes y definibles por reglas. En algunos casos señalé también algunos fenómenos que merecen una mención aparte. Por su importancia, a continuación detallaré tres más, que pueden ser de relevancia para un trabajo posterior:

- 1) Debilitamiento de vocales átonas.
- 2) Diptongación de hiatos.
- 3) Nasalización de vocales.

El debilitamiento de las vocales átonas del español ha sido documentado prácticamente por todos los autores (Navarro Tomás 1918:44; Lope Blanch 1963-1964 y 1966; Perissinotto 1975:26-29; Canfield 1981:62; Moreno de Alba 1988 y 1994:36). En el español de México, principalmente en el Altiplano, el debilitamiento de vocales es un fenómeno muy común. Lope Blanch (1963-1964:61) lo describe minuciosamente y llega a distinguir cuatro grados “que van desde la pérdida aparentemente completa hasta una simple relajación intensiva de las vocales”. Quizás

⁹¹ López Escobedo (2004) realizó una tesis de licenciatura en la que estudió las paravocales del español de México (ocurrencias de diptongos, frecuencias de diptongación, etcétera). Los resultados tendrán una aplicación directa en el Proyecto DIME.

la aportación más importante de Lope Blanch es que delimitó los contextos consonánticos en los que se producen el debilitamiento y la elisión vocálica —siendo los fonemas sordos, principalmente /s/, los que más propician este fenómeno, así como la pausa. Contrariamente a lo que se desprende de Navarro Tomás (1918:48, 50, 54, 57, 60 y 62) y de Matluck (1963:5), la posición silábica con respecto al acento no parece ser relevante.

La unión de dos vocales fuertes en diptongo (por ejemplo, *teatro* > [ˈtja tro]) ha sido reportada como un fenómeno de altísima ocurrencia en la lengua española en general: Lope Blanch (1969:61-66), Quilis (1981:179), Alarcos (1994:50) y Moreno de Alba (1994:50-62), entre otros. A pesar de no contar en las transcripciones del Corpus DIME con muestras suficientes, resulta clara la necesidad del estudio de la diptongación de hiatos, para definir con detalle su comportamiento. Pudimos ver en muchos casos del Corpus DIME cómo este fenómeno repercute claramente en el habla, donde no sólo se unen dos vocales fuertes en una sola sílaba, sino que la vocal media /o/ sufre un cierre y se vuelve una paravocal [w] (figura 39), mostrando la transición de los formantes propia de las labiovelares que señala Quilis (1981:177).

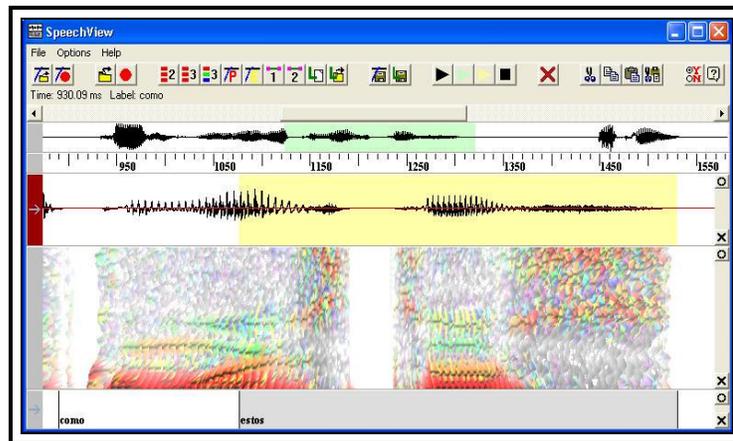


Figura 39. Mago: “¿Quieres un mueble como estos aquí? Archivo 65 del diálogo s6-t1-g1. Se observa claramente cómo lo que debería ser un hiato [-o.e-], se presenta diptongo [-we-]

En cuanto a la nasalización de las vocales en el español, prácticamente todos los autores han señalado esta resonancia suplementaria y los contextos en los que sucede (Alarcos 1950:65; Quilis 1981:165; Vaquero 1996:11). Su comportamiento en el corpus es, en términos generales, constante, pero habrá que definir con mayores datos la probabilidad de que estas reglas apoyen un programa de reconocimiento de habla.

He mencionado sólo tres fenómenos que tienen que ver con las vocales, pero son desde luego muchos otros los que se deberán estudiar posteriormente, como el debilitamiento o conservación del fonema /d/ en posición intervocálica, las sinalefas o la reducción, la fusión de vocales idénticas contiguas y si presentan alargamientos compensatorios o no (Alarcos 1994:), entre otros posibles fenómenos fonéticos relativos a la caracterización de las vocales. Asimismo, un trabajo dialectal para las consonantes debería considerar la realización fricativa [ʃ] del fonema palatal africado

/tʃ/, típica del norte del país,⁹² y los contextos y distribución geográfica de las aspiraciones de /s/ y /x/.⁹³

Sin duda, los estudios fonéticos del español de México que se pueden hacer son muchos, pero los elementos con los que se cuenta son favorables para la búsqueda de una mayor interdisciplinariedad entre la lingüística y la computación.

4.6. TranscríbEMex

Después de redefinir el uso y los alcances del alfabeto fonético computacional Mexbet, y de sistematizar los resultados del análisis fonético del Corpus DIME a través del estudio del comportamiento de los alófonos y de los fonemas en el español de México en un corpus oral de habla espontánea, con base en su frecuencia de aparición y en la posibilidad de ser modelados por reglas, a partir de esta tesis y de la oportunidad de continuar trabajando en un grupo de investigación interdisciplinario, se propiciaron los elementos para que Luis Villaseñor —del INAOE— y yo creáramos una herramienta computacional para el español de México que realiza transcripciones automáticas en el nivel fonológico y fonético, y permite, además, obtener una división automática en sílabas. Al programa lo llamamos

⁹² Resultaría interesante tomar en cuenta también la observación de López Chávez (1986:329) en cuanto a que “en una región donde la /ç/ tiende a la continuidad, el fonema /s/ se realizará de cualquier modo, excepto coronal o con tendencia a la palatalización y viceversa, en una región donde abundan las realizaciones de /s/ con resonancia palatal la /ç/ tenderá a las articulaciones centrales o a las variantes adelantadas, pero muy oclusivas”. El investigador llega a establecer una división dialectal: “predominio de la [s] en el Sur, y notoria alta frecuencia de [ç] en el Norte”.

⁹³ En cuanto a las aspiraciones, que podrían establecer neutralizaciones entre /s/ y /x/, López Chávez da claves para su delimitación. Entre ellas, dice que “La aspiración de /s/ intervocálica se da en el noroeste de México; mientras que la aspiración de /x/ abunda en las costas del Sur” (López Chávez 1986).

TranscríbEMex, señalando su carácter de transcriptor fonético automático (Transcrí-) para el español (-E-) de México (-Mex). En la figura 40 podemos apreciar su diseño y funcionamiento, con una muestra de la transcripción fonológica y de la transcripción fonética en Mexbet, así como la división silábica, de la palabra *fonetizador*.

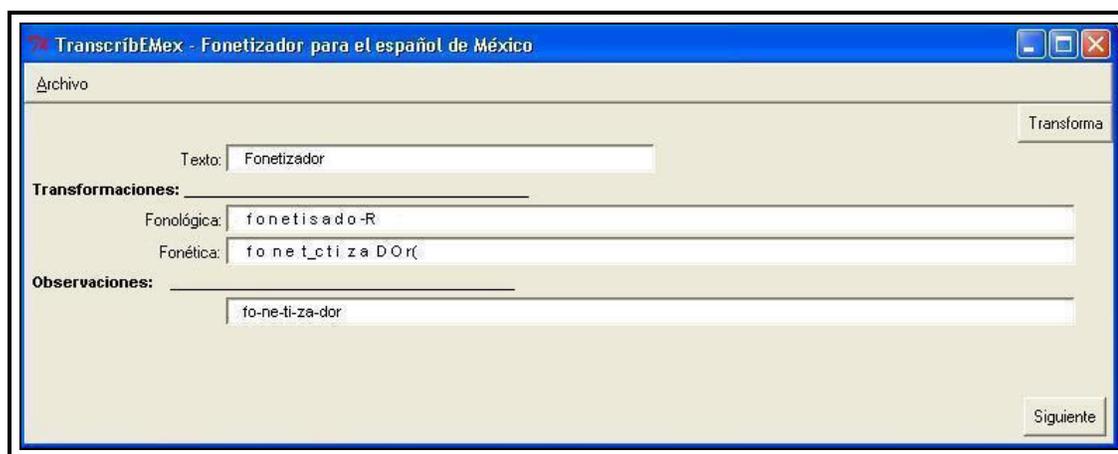


Figura 40. Aspecto del programa TranscríbEMex

En la ventana “Fonética:” aparece la transcripción fonética de la palabra *Fonetizador* en Mexbet: [f o n e t_c t i z a D O r(], con los alófonos que se han definido en esta tesis, donde destacan el alófono sonorizado de /s/ ([z]), el alófono aproximante de /d/ ([D]) y el alófono abierto de /o/ ([O]). Las reglas de conversión de grafema a alófono fueron extraídas de mi investigación, a partir del análisis fonético y del estudio de las frecuencias de aparición de los fonemas y alófonos en el Corpus DIME. Con estas reglas de conversión de grafema a alófono, Villaseñor, investigador del Proyecto de Interacción Oral Hombre-Máquina, del Laboratorio de Tecnologías del Lenguaje del INAOE, generó con el lenguaje de programación *Perl* una serie de herramientas que hacen la transformación .

TranscríbEMex realiza también una transcripción fonológica. Este módulo fue incorporado al programa a partir de algunos trabajos sobre reglas de conversión de grafema a fonema (Moreno y Mariño 1998; Ríos 1999; Uruga y Pineda 2000) y de otras investigaciones previas realizadas en el marco del Proyecto DIME. Para este nivel fonológico, a partir de este trabajo definí las reglas que se aplicaban estrictamente al español de México, como la eliminación del carácter fonémico que se le había dado a la paravocal [w], y la exclusión de nuestro sistema fonológico de los fonemas castellanos, entre otros aspectos que revisamos anteriormente (§3.3.). Además, y dado el comportamiento de ciertos fonemas en posición final de sílaba ante determinados segmentos, propuse para las transcripciones fonológicas la inclusión de los archifonemas /-B/, /-D/, /-G/, /-N/ y /-R/ para las parejas neutralizables b/p, d/t, g/k, n/m y r/r. Las reglas para la división silábica se incluyeron a partir de un ejercicio previo de Villaseñor.

En España se han elaborado diversos transcriptores fonéticos automáticos.⁹⁴ Dos de los que tengo conocimiento son *Segre* (Pachès *et al.* 2000; De la Mota y Riera 2000), para el catalán, y *ort2fon* (Castro *et al.* 2001), para el castellano. El transcriptor Segre realiza transcripciones para cuatro variantes dialectales del catalán (el dialecto central, el occidental, el balear y el valenciano).⁹⁵ El porcentaje de precisión de las transcripciones que realiza Segre, en comparación con

⁹⁴ Un panorama completo sobre los sistemas de conversión de grafema a fonema o alófono para el español y otras lenguas —aplicados principalmente a la síntesis del habla— se encuentra en Ríos (1999).

⁹⁵ Segre representa un desarrollo interdisciplinario entre lingüistas y computólogos de la Universidad Politécnica de Cataluña, la Universidad Autónoma de Barcelona y la Universidad de Barcelona en el marco de Centro de Referencia en Ingeniería Lingüística (CREL), de la *Generalitat de Catalunya*.

transcripciones manuales, es del 99.1% para palabras aisladas y del 99.39% para oraciones; Segre ha mostrado ser útil para la síntesis y para el reconocimiento de habla.⁹⁶ Al igual que TranscríbEMex y Segre, la utilidad de ort2fon (Castro *et al.* 2001) es el entrenamiento de sistemas de reconocimiento de habla, y encuentra también aplicaciones en la síntesis de habla. El fonetizador ort2fon realiza también silabificación. Asimismo, además de las transcripciones prototípicas, ort2fon presenta fenómenos de debilitamiento (*abogado, abogao*) y de habla enfática (*sicólogo, psicólogo*), y da como opción transcripciones con seseo y yeísmo, o sin ellas. No obstante, el trabajo de Castro *et al.* (2001) tiene errores básicos de terminología lingüística (llama fonemas a los alófonos y confunde el término aproximante con sonoro);⁹⁷ además, llama mucho la atención el hecho de que ort2fon creó sus propios símbolos ASCII para el español castellano, cuando el alfabeto SAMPA, como vimos (§3.2.1.), está ampliamente reconocido y difundido en toda Europa.

Al igual que Segre (Pachès *et al.* 2000) TranscríbEMex ordena las reglas para asegurar que su interacción se lleve a cabo de manera correcta. Asimismo, durante el proceso de transcripción automática, TranscríbEMex considera —como ort2fon (Castro *et al.* 2001)— los efectos de articulación al interior y en frontera de palabra. No obstante, TranscríbEMex presenta algunos aspectos bien identificados que no se

⁹⁶ Los conversores de texto a alófonos o de texto a fonemas son de uso imprescindible en el proceso de la síntesis del habla (Llisterri 2002).

⁹⁷ El transcriptor fonético automático ort2fon fue elaborado por el Departamento de Sistemas Informáticos y Computación de la Universidad Politécnica de Valencia y por el Departamento de Lenguajes y Sistemas Informáticos de la Universidad Jaime I. Como hemos visto en diversas ocasiones a lo largo de este trabajo, en los estudios desarrollados hechos solamente por computólogos se ponen en evidencia errores de terminología lingüística básicos.

han resuelto todavía (como la silabificación y los correlatos fonéticos de la grafía *x*, entre otros aspectos).⁹⁸

El fonetizador TranscríbEMex ofrece un beneficio doble: tiene una aplicación inmediata en los desarrollos de tecnologías del habla en nuestro país y puede resultar beneficioso para los estudios de la pronunciación del español de México. La utilidad de este fonetizador ya ha sido demostrada: actualmente es la herramienta base en el Proyecto DIME para etiquetar del Corpus DIME; a diferencia de las primeras etapas de etiquetado manual del Corpus DIME, TranscríbEMex puede elaborar un rápido etiquetado automático provisional del oscilograma y el espectrograma de una señal sonora, a partir de una previa transcripción ortográfica, con las etiquetas fonológica, fonética y silábica. Podemos apreciar un ejemplo de este proceso en la figura 41.

⁹⁸ Para *x*, por el momento sólo se ha incluido la equivalencia fonética [ks] (*exacto, existe*). A partir de un trabajo que se desarrolla hoy en día en el Proyecto DIME, estamos diseñando un diccionario de excepciones para que TranscríbEMex presente de manera correcta las equivalencias de *x* como [x] (*México*) y [s] (*extractor*), y aún como [ʃ] (*Xola*). Asimismo, se incluirá cierto léxico extranjerizante que está relacionado con la computación o el dominio de cocinas, como *mouse* y *sink*.

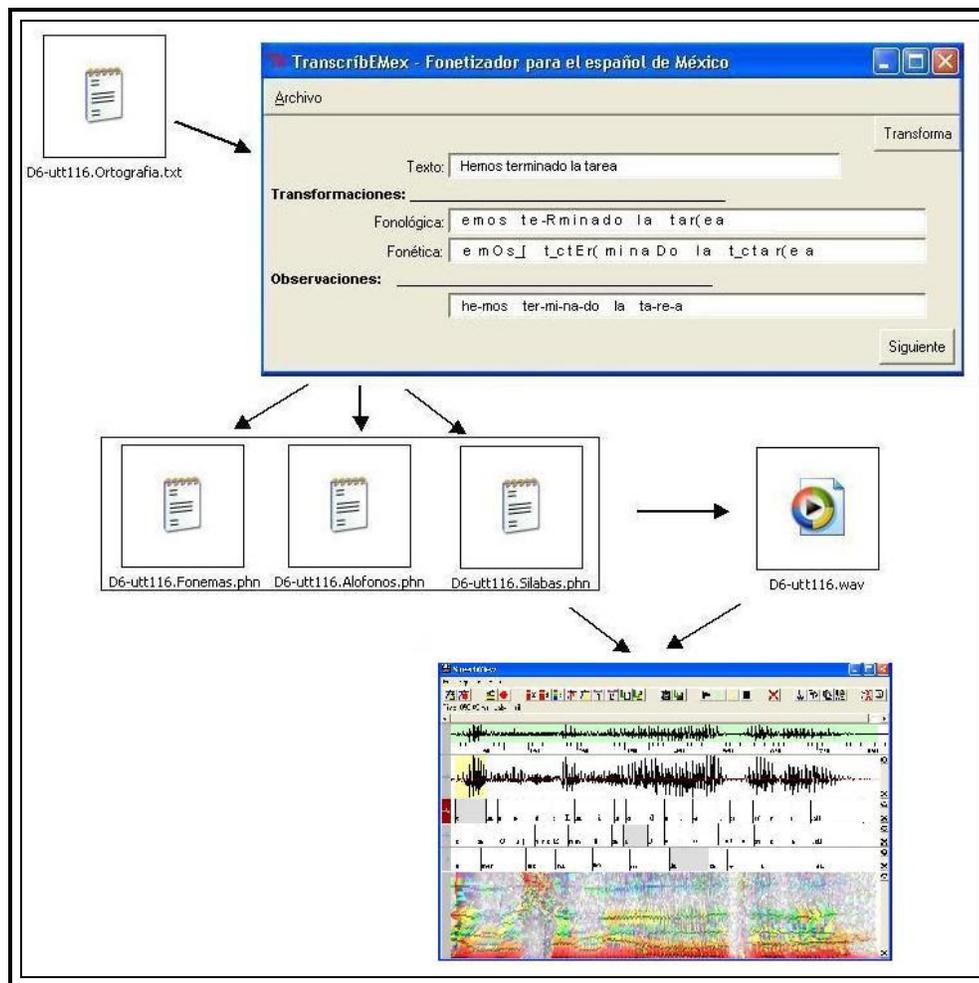


Figura 41. Procedimiento para la alineación automática de la señal sonora con las transcripciones fonológica, fonética y silábica, con el programa TranscribEMex

El funcionamiento es el siguiente: para cada uno de los archivos de audio de las elocuciones del Corpus DIME se ha elaborado una transcripción ortográfica en un archivo de texto (en este caso, el archivo “D6-utt116.Ortografia.txt”). Se programa TranscribEMex para que genere tres archivos con extensión *.phn; uno con la transcripción fonológica (archivo “D6-utt116.Fonemas.phn”), otro con la fonética (“D6-utt116.Alofonos.phn”) y otra con la silábica (“D6-utt116.Silabas.phn”). Posteriormente, el programa realiza una alineación temporal aproximada, alineando

estos tres archivos con el audio (archivo “D6-utt116.wav”), que pueden ser desplegados en el *SpeechView*, para hacer después un ajuste temporal de manera manual. Una muestra de las etiquetas finales se presenta en la figura 42.

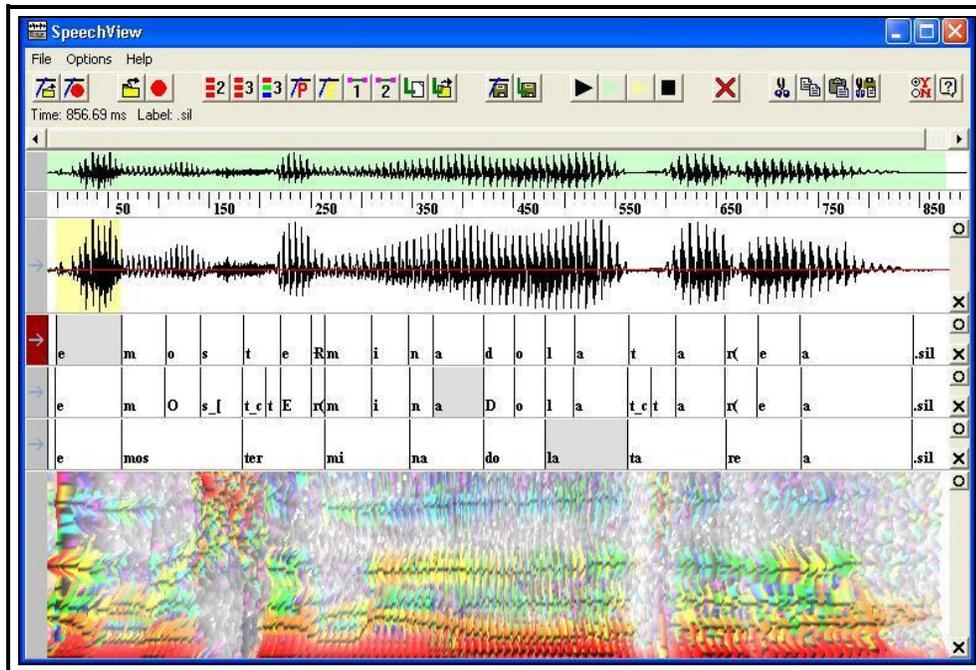


Figura 42. Mago: “Hemos terminado la tarea”. Archivo 116 del diálogo s6-t1-g1. Se aprecian (de arriba hacia abajo) dos oscilogramas, tres etiquetas a nivel fonológico, fonético y silábico, y un espectrograma, alineados manualmente

La necesidad de programas computacionales que realicen transcripciones fonéticas y fonológicas automáticas ha sido señalada por diversos autores (Llisterri 1997; Olivier 1999; Uraga 1999). Desde el punto de vista de las tecnologías del habla, entre las ventajas que ofrece TranscríbEMex se encuentra el hecho de que ofrece rapidez para etiquetar grandes volúmenes de corpus orales. Otra virtud de TranscríbEMex es que favorece la uniformidad y elimina ambigüedades en las transcripciones fonéticas, pues es bien sabido que aún a través del análisis

instrumental éstas pueden verse afectadas por decisiones subjetivas de los etiquetadores. Otro aspecto favorable se encuentra en la posibilidad de uso que pueden tener transcriutores no especializados: en el Proyecto DIME, TranscríbEMex ha sido utilizado para realizar transcripciones fonéticas y fonológicas incluso por personas no expertas.

Desde el punto de vista lingüístico, la aplicación de este tipo de sistemas a los estudios fonéticos del español de la ciudad de México permitirá sistematizar el conocimiento de los procesos fonéticos y fonológicos, como lo señala acertadamente Llisterri (2003). Además, este programa ofrece en un futuro inmediato la posibilidad de abordar una mayor cantidad de fenómenos fonéticos y dialectales, como he insistido.

4.7. Aplicaciones y perspectivas futuras

Además del fonetizador TranscríbEMex, otra aplicación que ya ha dado frutos a partir de mi análisis fonético del Corpus DIME es el diseño y la grabación de un corpus oral rico y completo desde el punto de vista fonético: el corpus DIMEx100 (Pineda *et al.* 2004). Rico porque se diseñó procurando suficientes muestras de cada contexto fónico; completo, porque incluyó todas las formas alofónicas detalladas aquí, buscando obtener un corpus valioso para la creación de modelos acústicos.⁹⁹ Este

⁹⁹ “La función principal de un corpus, tanto textual como oral, es establecer la relación entre la teoría y los datos; el corpus tiene que mostrar a pequeña escala cómo funciona una lengua natural; pero para ello es necesario que esté diseñado correctamente sobre una bases estadísticas apropiadas que aseguren que el desarrollo sea efectivamente un modelo de la realidad” (Torruella y Llisterri 1999:45-46).

nuevo corpus se ha realizado de manera coordinada entre el Proyecto DIME y el grupo Interacción Oral Hombre-Máquina, del INAOE.¹⁰⁰ Con base en las reglas y los alófonos propuestos en esta tesis, se diseñó un corpus inicial de 6,000 oraciones extraídas de textos procedentes de la red electrónica mundial. En el Proyecto DIME este corpus ha sido grabado procurando la mayor calidad acústica con el programa *Wave Labe*, con un micrófono de condensación con diafragma *Behringer B-1*, y en cámara aislada. Estas grabaciones digitales de altísima calidad proporcionarán un doble beneficio: servirán para hacer más estudios fonéticos sobre el habla de la ciudad de México y ayudarán a crear modelos acústicos más precisos para el módulo del reconocedor de habla del Proyecto DIME.

En una primera fase de este nuevo corpus, se consideró la grabación del habla culta de la ciudad de México en sujetos entre 18 y 38 años de edad. Se ha concluido la etapa de grabación y se cuenta ya con el registro de cien hablantes (51% mujeres y 49% hombres). Cada uno grabó 60 oraciones, 10 de las cuales fueron las mismas. De tal suerte, este corpus podrá ser utilizado para crear modelos acústicos más certeros y completos, y también para realizar estudios sobre reconocimiento de hablante.

Existen muchas otras opciones de trabajo futuro en el Proyecto DIME: actualmente se están diseñando las bases para el etiquetado prosódico, de manera que a partir de la entonación se obtenga información para el reconocedor de habla; esta labor ha ampliado las expectativas en el Proyecto DIME.

De manera personal, deseo resaltar que resulta particularmente agradable unir esfuerzos en dos áreas tan distintas y tan semejantes a la vez, como la lingüística y la

¹⁰⁰ Algunos antecedentes sobre el diseño de este corpus se encuentran en Villaseñor *et al.* (2003).

computación; las ventajas de la interdisciplinariedad, creo, han quedado puestas de manifiesto.

5. CONCLUSIONES

While there are still serious technical issues remaining to be overcome, dialogue-based user interfaces are showing promise. Once they reach a certain level of basic competence, they will rapidly start to revolutionize the way the people interact with computers, much like the direct manipulation interfaces (using menus and icons) revolutionized computer use in the last decade.
Allen et al. 2001:8

A partir de un trabajo interdisciplinario, en esta tesis aproveché la relación entre la lingüística y la computación para favorecer ambas disciplinas. Después de introducir los objetivos y la metodología de esta trabajo, inicié el capítulo 2 haciendo algunas consideraciones sobre la necesidad del análisis acústico en los estudios fonéticos, resaltando el hecho de que Quilis (1981) mencionaba hace más de 20 años que el uso de los instrumentos en la fonética experimental era imprescindible. Repasé después algunos aspectos de la relación entre la lingüística y la computación en nuestro país en las últimas cuatro décadas —Grimes (1965), Manrique (1968) y Ávila (1987), entre otros autores—; a partir de la labor de éstos, parece ser que la disposición de los lingüistas hacia la computación era más abierta en México hace 40 años que lo que parece ser hoy. Actualmente, no obstante, las tecnologías del habla ofrecen un campo fértil para el trabajo interdisciplinario.

Posteriormente, mencioné qué son las tecnologías del habla y, por ser el reconocimiento de habla la tecnología en la que se inscribe el Proyecto DIME, me detuve de manera específica en el funcionamiento, las aplicaciones y aún las

limitaciones de un sistema de cómputo que al recibir la señal sonora de la voz humana reconoce la información lingüística. Vimos los tres módulos de un reconocedor de habla: el diccionario de pronunciación, los modelos de pronunciación y los modelos acústicos, y resalté el hecho de que para la arquitectura de un reconocedor de habla los conocimientos fonéticos resultan indispensables.

Inmediatamente después abordé el Proyecto Diálogos Inteligentes Multimodales en Español (Proyecto DIME), del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM, que tiene como propósito el modelado lingüístico computacional del español de México. Presenté el Corpus DIME, que fue mi objeto de estudio; señalé el propósito que llevó a su creación y puntalicé sus propiedades. Además, señalé las características de los corpus creados con el protocolo del *Mago de Oz*, como el Corpus DIME, el único corpus creado bajo estos conceptos en nuestro país, hasta donde tengo noticia. Posteriormente, detallé lo que es el etiquetado fonético en las tecnologías del habla y señalé la necesidad de transcribir de manera eficaz corpus orales para poder sistematizar la información que se utiliza en la producción de programas de síntesis de habla, de sistemas conversacionales y, en este caso, de reconocimiento de habla.

En el capítulo 3 hice un repaso de la tradición y la novedad en los alfabetos fonéticos; mencioné primero las características y aportaciones de dos alfabetos fonéticos de tradición lingüística en México, el Alfabeto Fonético Internacional y el de la *Revista de Filología Española*, y procuré mostrar las ventajas que ofrece el uso del AFI, que representa hoy en día una herramienta más amplia y efectiva para la práctica fonética. Posteriormente, mencioné qué son los alfabetos fonéticos

computacionales y la necesidad de su uso en la práctica lingüística de las tecnologías del habla; presenté tres que de una u otra manera se relacionan con el alfabeto que propone este trabajo: SAMPA (Llisterra y Mariño 1993), Worldbet (Hieronymus 1994 y 1997) y OGibet (Tlatoa 2000a). Después introduje Mexbet, un alfabeto fonético computacional para el español de México; señalé primero sus características originales (Uruga y Pineda 2000 y 2002) y presenté las bases lingüísticas para puntualizar su carácter:

- 1) La distinción entre fonemas y alófonos.
- 2) La delimitación de los fonemas que pertenecen al sistema fonológico de nuestro dialecto mexicano.
- 3) La consecuente eliminación de los fonemas castellanos /θ/ y /ʎ/.
- 4) Los contextos en los que se presenta cada alófono.
- 5) El uso correcto de los diacríticos para señalar las asimilaciones.

Una de las aportaciones de este trabajo es esta nueva versión de Mexbet, que representa ahora un alfabeto fonético computacional correcto, concreto y claro para el español de México.

En el capítulo 4 presenté los resultados del análisis fonético del Corpus DIME y comparé su veracidad con los estudios de Navarro Tomás (1918), Alarcos (1950), Matluck (1951), Perissinotto (1975), Quilis (1981) y Moreno de Alba (1994), principalmente. Estudié las frecuencias de aparición de los fonemas y alófonos del español de México y contrasté mis resultados con diferentes autores, como Zipf y Rogers (1939), Navarro Tomás (1946), Alarcos (1950), Quilis (1981), Rojo (1991), Llisterra y Mariño (1993) y Pérez (2003). Así, se puso en evidencia que en los

estudios fonéticos es imprescindible considerar la frecuencia de aparición de los fonemas y alófonos en la lengua.

Con base en los fenómenos encontrados en el análisis fonético del Corpus DIME, establecí posteriormente un inventario de 37 alófonos para el español de la ciudad de México, que por su frecuencia de aparición y por la sistematicidad con que se presentan, pueden ser modelados por reglas. A continuación, presenté el programa TranscribEMex, un fonetizador o transcriptor fonético automático para el español de México que funciona a partir de las reglas de conversión de grafema a fonema y de grafema a alófono que se desprenden de esta investigación. Mencioné también la ventajas que la aplicación de este fonetizador ha probado ya para el Proyecto DIME, para la transcripción automática del Corpus DIME y del Corpus DIMEx100 (Pineda *et al.* 2004).

En un apéndice presento las equivalencias entre el Alfabeto Fonético Internacional, el alfabeto de la *Revista de Filología Española* y Mexbet. Me ha parecido particularmente importante incluir estos datos por dos motivos principales: porque sirve como guía de consulta para la lectura de esta tesis, y porque representa un esfuerzo por poner en evidencia y aclarar algunos aspectos de transcripción fonética que han causado confusiones o problemas para los fonetistas y para los lingüistas en general en nuestro país.

A lo largo de todas estas páginas he insistido en la retroalimentación que se establece entre la lingüística y la computación; uno de los aspectos más importantes de este trabajo es que se vincula a un importante proyecto que se lleva a cabo actualmente en la UNAM estableciendo lazos firmes entre la Facultad de Filosofía y

Letras y el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas. Asimismo, esta tesis proporciona al Proyecto DIME un alfabeto fonético computacional, un fonetizador y una serie de reglas de conversión de grafema a alófono. Todo ello favorece el análisis y la evaluación de los estudios fonéticos del español hablado en México. El uso de los datos del análisis lingüístico de este trabajo

- 1) Aporta información y arroja luz sobre el habla de la ciudad de México.
- 2) Corrobora y matiza información sobre la fonética del español de México.
- 3) Ofrece algunas bases y marca directrices para futuros trabajos de investigación.

Estoy convencido de que el habla será el instrumento para comunicarse con las computadoras en un futuro no muy lejano. De tal suerte, los estudios fonéticos del español de México para su aplicación en las tecnologías del habla son cada vez más necesarios. En este trabajo quise aprovechar esa coyuntura y obtener datos lingüísticos sobre nuestra pronunciación, para brindar a la vez beneficios a las tecnologías del habla en nuestro país.

6. APÉNDICE. EQUIVALENCIAS ENTRE EL AFI, EL ALFABETO DE LA *RFE* Y MEXBET

Para la realización de las transcripciones fonéticas y fonológicas, en esta tesis he optado por el uso del Alfabeto Fonético Internacional. No obstante, y en el entendido de que el alfabeto de la *Revista de Filología Española* y el AFI “dejan sin resolver algunos de los problemas de transcripción” para el español y de que el de la *RFE* “es ya poco utilizado” (De la Mota y Ríos 1995:97), me ha parecido necesario mostrar en las figuras 43 y 44 de este apéndice las convenciones que he utilizado para representar los sonidos del español de México. Asimismo, he establecido en cada tabla las equivalencias para cada uno de los fonemas y alófonos del español de México con el alfabeto de la *Revista de Filología Española* y Mexbet, hecho que podrá facilitar la lectura de este trabajo.

Cabe mencionar que básicamente y en su mayoría, estos símbolos y convenciones del AFI han sido tomados de las tablas para la transcripción de los sonidos del español de Garrido *et al.* (1998:102-106). Esta decisión se debe, en muy buena medida, a que los autores resuelven acertadamente confusiones que aún perduraban en los estudios fonéticos y fonológicos; además, proponen una clara diferenciación entre el carácter fonémico o fonético de cada segmento.

A este respecto, es preciso señalar que, en el uso de cada símbolo, he procurado distinguir entre transcripciones fonéticas y fonológicas, para responder a los *Principios* de la Asociación Fonética Internacional (1949:14-15) que recomiendan

el uso de diacríticos para señalar las variantes contextuales y mantener un número limitado de notaciones para los fonemas de cada lengua. Por ejemplo, para la forma sonorizada del fonema /s/, el uso del diacrítico [̣] bajo la forma base: [ṣ], muestra con transparencia su carácter alófonico, distinguiéndolo de lenguas donde /z/ es fonológico.

Para los investigadores que están más familiarizados con el uso del alfabeto hispánico, las diferencias de representación a nivel fonológico del AFI se encontrarán en los fonemas /t̃/, /z/, /ɲ/, /r/ y /r̃/ (/t̃/, /y/, /ɲ/, /r/ y /r̃/, respectivamente, en el alfabeto de la *RFE*).

En el nivel fonético, los alófonos fricativos tradicionales [b], [d] y [g], se representarán con las formas aproximantes [$\begin{smallmatrix} \beta \\ \tau \end{smallmatrix}$], [$\begin{smallmatrix} \delta \\ \tau \end{smallmatrix}$] y [$\begin{smallmatrix} \gamma \\ \tau \end{smallmatrix}$], respectivamente. Se utilizarán los diacríticos [̣] para las dentalizaciones, y [̠] para las palatalizaciones. El símbolo para la nasal velarizada, tradicionalmente representada en el alfabeto hispánico como [ŋ], será representado con el símbolo de velarización [ʷ]: [nʷ]. El diacrítico para los alófonos semiabiertos de las vocales medias será [̜], y para los desplazamientos anteriores y posteriores de la vocal abierta, los diacríticos [+] y [-], respectivamente.

Es preciso señalar que la representación del fonema palatal fricativo sonoro en el español ha sido siempre un problema. A lo largo de las décadas, los diferentes autores han señalado distintos detalles en el punto de articulación y han utilizado

diferentes símbolos para su forma fricativa y africada.¹⁰¹ Ante la dificultad que este hecho presenta y para ser coherente con las convenciones que Mexbet había establecido en sus versiones originales (Uraga y Pineda 2000 y 2002), extraídas a partir de Worldbet (1994 y 1997) —[Z] para la fricativa y [dZ] para la africada—, en este trabajo he optado de manera arbitraria por el uso del símbolo [ʒ] para el fonema palatal fricativo sonoro, sin asumir el punto de articulación postalveolar que el AFI le asigna.

¹⁰¹ Por poner sólo algunos ejemplos, el alfabeto de la *RFE* (1915:376) y Navarro Tomás (1918:82) utilizan para los alófonos de este fonema los símbolos [y] y [ÿ], describiéndolos como palatal fricativo sonoro y africado, respectivamente. Alarcos (1950:163) utiliza los mismos símbolos, y puntualiza su punto de articulación como mediopalatal. Para Matluck (1951:3), [y] y [ÿ] son prepalatales; Perissinotto (1975), las llama unas veces palatales (51) y otras prepalatales (19). Para Canellada y Madsen (1987) se trata del fonema mediopalatal aproximante sonoro /y/ y de su alófono africado [j] (36), o fonema palatal fricativo /y/ y su alófono palatal africado [j] (21). Para Quilis (1981:256 y 1993:291) son el fonema fricativo linguopalatal central /j/ y su alófono palatal africado sonoro [dʒ]. Garrido y sus colaboradores (1998:105) consideran una forma palatal oclusiva sonora [j], una palatal afrificada [dʒ] y una palatal fricativa [j]. Martínez Celdrán *et al.* (2003:258) dicen que el fonema es el palatal africado sonoro [j̥], y su alófono es el aproximante [j̥]. Para Martínez Celdrán y Planas (2001:175) la afrificada palatal sonora del español “constituye un segmento controvertido por cuanto los autores discrepan acerca de su naturaleza”, y afirman que se trata “de una articulación doble con punto de articulación plenamente palatal cuyo segundo elemento es aproximante; por tanto, su transcripción más adecuada es [j̥j̥]”.

Fonemas del español de México			
Descripción	AFI	RFE	Mexbet
Bilabial oclusiva sorda	p	p	p
Dental oclusiva sorda	t	t	t
Velar oclusiva sorda	k	k	k
Bilabial oclusiva sonora	b	b	b
Dental oclusiva sonora	d	d	d
Velar oclusiva sonora	g	g	g
Palatal africada sorda	tʃ	ç	tS
Labiodental fricativa sorda	f	f	f
Alveolar fricativa sorda	s	s	s
Velar fricativa sorda	x	x	x
Palatal fricativa sonora	ʒ	y	Z
Labial nasal	m	m	m
Alveolar nasal	n	n	n
Palatal nasal	ɲ	ɲ	n~
Alveolar lateral	l	l	l
Alveolar vibrante simple	r	r	r(
Alveolar vibrante múltiple	r	ř	r
Vocal cerrada palatal	i	i	i
Vocal media palatal	e	e	e
Vocal abierta	a	a	a
Vocal media velar	o	o	o
Vocal cerrada velar	u	u	u

Figura 43. Representación de los fonemas del español de México con los símbolos del AFI, el alfabeto de la RFE y Mexbet

Alófonos más comunes del español de México			
Descripción	AFI	RFE	Mexbet
Velar oclusiva sorda palatalizada	k ^h	k´	k_j
Bilabial aproximante sonora	β ·	b	V
Dental aproximante sonora	ð ·	d	D
Velar aproximante sonora	ɣ ·	g	G
Alveolar fricativa sorda dentalizada	ʃ̺	ʃ	s_[
Alveolar fricativa sorda sonorizada	ʒ	z	z
Palatal africada sonora	ɟ̺	ɟ	dZ
Alveolar nasal dentalizada	ɲ̺	ɲ	n_[
Alveolar nasal velarizada	n ^v	ŋ	N
Paravocal palatal	j	j / j̲	j
Vocal media palatal abierta	e̞	e̞	E
Vocal abierta palatalizada	a ⁺	a̟	a_j
Vocal abierta velarizada	a̠	a̠	a_2
Vocal media velar abierta	o̞	o̞	O
Paravocal velar	w	w / w̲	w

Figura 44. Representación de los alófonos más comunes del español de México con los símbolos del AFI, el alfabeto de la RFE y Mexbet

7. REFERENCIAS BIBLIOGRÁFICAS*

- ACERO, ALEJANDRO. 1995. "The role of phoneticians in speech technology", en *European studies in phonetics and speech communication*, G. Bloothoof, V. Hazan, D. Huber y J. Llisterri (Eds.), Utrecht: OTS Publications, pp. 170-175.
- AGUILAR, LOURDES, JOAQUIM LLISTERRI y MARÍA JESÚS MACHUCA. 1994 (Ms.). "Some phonetic data on speech produced in different experimental situations", ponencia presentada en *ESPRIT BRA VOX Workshop*, Orsay: LIMSI. [Cito por: <http://liceu.uab.es/~lourdes/paris.ps>]
- AGUILAR, LOURDES, JUAN M. GARRIDO y JOAQUIM LLISTERRI. 1997. "Incorporación de conocimientos fonéticos a las tecnologías del habla", en *Panorama de la investigació lingüística a l'Estat Espanyol*, E. Serra, B. Gallardo, M. Veyrat, D. Jorques y A. Alcina (Eds.), Valencia: Universidad de Valencia, pp. 5-13. [Cito por: http://liceu.uab.es/~joaquim/publicacions/valencia_94.html]
- ALARCOS LLORACH, EMILIO. 1950/1981. *Fonología española*, Madrid: Gredos.
- . 1994. *Gramática de la lengua española*, Madrid: Espasa-Calpe y Real Academia Española.
- ALI, AHMED M. ABDELATTY. Ms. "Acoustic-phonetic feature-based signal processing for automatic speech recognition: Brief results". [Cito por: <http://www.seas.upenn.edu/~ahm/results.pdf>]
- ALLEN, JAMES F., DONNA K. BYRON, MYROSLAVA DZIKOVSKA, GEORGE FERGUSON, LUCIAN GALESCU y AMANDA STENT. 2001 (Ms.). "Towards conversational human-computer interaction", Universidad de Rochester.
- ÁVILA, RAÚL. 1973. "Realizaciones tensas de /s/ en la ciudad de México", *Anuario de Letras*, 11, pp. 235-239.
- . 1987/1992. "Un alfabeto fonológico práctico para el español: pros y contras y pros y contras...", en *Scripta philologica. In honorem Juan M. Lope Blanch, 2: Lingüística española e iberoamericana*, E. Luna (Coord.), México: Universidad Nacional Autónoma de México, pp. 7-19.
- BACCHIANI, MICHIEL, JULIA HIRSCHBERG, AARON ROSENBERG, STEVE WHITTAKER, DONALD HINDLE, PHIL ISENHOUR, MARK JONES, LITZA STARK y GARY ZAMCHICK. 2001 (Ms.). "SCANMail: Audio navigation in the voicemail domain", en *Proceedings of the Workshop on Human Language Technology*. [Cito por: <http://acl.ldc.upenn.edu/H/H01/H01-1064.pdf>]
- BATTANER, ELENA, JUANA GIL, VICTORIA MARRERO, JOAQUIM LLISTERRI, CARMEN CARBÓ, MARÍA JESÚS MACHUCA, CARMEN DE LA MOTA y ANTONIO RÍOS. 2003. "VILE: Estudio acústico de la variación inter e intralocutor en español", en *Actas del Congreso de la Sociedad Española de Acústica Forense*, Barcelona: Sociedad Española de Acústica Forense, pp. 59-70. [Cito por: http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_SEAF03.pdf]

* La validez de las direcciones electrónicas fue verificada el 4 de julio de 2004.

- BLAHETA, DON. Ms. "Representation of IPA with ASCII", Providence: Brown University, Department of Computer Science.
- BONAFONTE, ANTONIO, PABLO AIBAR, NÚRIA CASTELL, EDUARDO LLEIDA, JOSÉ B. MARIÑO, EMILIO SANCHIS y M. INÉS TORRES. 1998 (Ms.). "Desarrollo de un sistema de diálogo oral en dominios restringidos", ponencia presentada en *First Meeting on Language Engineering*, Sevilla: Universidad de Sevilla. [Cito por: http://gps-tsc.upc.es/veu/basurde/download/Bon00a_sevilla.pdf]
- BOTT, M.F. 1970/1975. "Lingüística computacional", en *Nuevos horizontes de la lingüística*, J. Lyons (Comp.), Madrid: Alianza, pp. 227-240.
- CANELLADA, MARÍA JOSEFA y JOHN KUHLMANN MADSEN. 1987. *Pronunciación del español. Lengua hablada y literaria*, Madrid: Castalia.
- CANFIELD, DELOS LINCOLN. 1962. *La pronunciación del español de América. Ensayo histórico-descriptivo*, Bogotá: Instituto Caro y Cuervo.
- . 1981/1992. *Spanish pronunciation in the Americas*, Chicago: The University of Chicago Press.
- CASACUBERTA, FRANCISCO y VIDAL ENRIQUE. 1987. *Reconocimiento automático del habla*, Barcelona: Boixareu Editoriales.
- CASTRO, MARÍA JOSÉ, SALVADOR ESPAÑA, ANDRÉS MARZAL e ISMAEL SALVADOR. 2001. "Transcriptor ortográfico-fonético para el castellano", en *Procesamiento del lenguaje natural*, 27:241-246 [Cito por: <http://www.sepln.org/revistaSEPLN/revista/27/27-articulo28.pdf>]
- CENTER FOR SPOKEN LANGUAGE UNDERSTANDING. 1999 (Ms.). "IPA; Worldbet, and OGIbet English broad phonetic labels", Oregon: Oregon Graduate Institute of Science and Technology. [Cito por: <http://colton.byuh.edu/courses/cs441/9504.refbet.pdf> y <http://dipaola.org/stanford/facade/lipsync/refbet.pdf>]
- CUÉTARA PRIEDE, JAVIER. 2003. "Aplicación de la fonética instrumental para el mejoramiento de un reconocedor de voz del español de México", en *Estudios de lingüística y filología hispánicas en honor de José G. Moreno de Alba. Memoria del 4º Encuentro de Lingüística en Acatlán*, I. Guzmán Betancourt y P. Máynez (Coords.), México: Universidad Nacional Autónoma de México, pp. 229-240.
- DAHLBÄCK, NILS, ARNE JÖNSSON y LARS AHRENBERG. 1993. "Wizard of Oz studies – Why and how", en *Intelligent users interfaces*, Linköping: Laboratorio de Procesamiento de Lenguaje Natural, pp. 193-200.
- ESLING, JOHN H. y HARRY GAYLORD. 1993. "Computer codes for phonetic symbols", en *Journal of the International Phonetic Association*, 23:2, pp. 83-97.
- FERNÁNDEZ, SANTIAGO, SERGIO FEIJÓO, RAMÓN Balsa y NIEVES BARROS. 1998 (Ms.). "Acoustic cues for the auditory identification of the Spanish fricative /f/", en *Proceedings of ICSLP'98. The Fifth International Conference on Spoken Language Processing*, Rundle Mall: Causal Productions.
- GARRIDO, JUAN MARÍA, MARÍA JESÚS MACHUCA y CARMEN DE LA MOTA . 1998. "La transcripción de los sonidos del español", en *Prácticas de fonética. Lengua española I*, Bellaterra: Universidad Autónoma de Barcelona, pp. 102-106.
- GIL, JUANA. 1988/1993. *Los sonidos del lenguaje*, Madrid: Síntesis.

- GIL, JUANA y JOAQUIM LLISTERRI. En prensa. “Fonética y fonología del español en España (1978-2003)”, *Lingüística Española Actual*, 25. [Cito por: http://liceu.uab.es/~joaquim/publicacions/Gil_Llisterri_LEA_2003.pdf]
- GILI GAYA, SAMUEL. 1950/1975. *Elementos de fonética general*, Madrid: Gredos.
- GRIMES, JOSEPH E. 1965. “La computadora en las investigaciones humanísticas”, en *Anuario de Letras*, 5, pp. 5-34.
- HARRIS, JAMES W. 1969/1973. *Fonología generativa del español*, Barcelona: Planeta.
- HERRERA, ESTHER. 2002. “La asimilación de las nasales en español. Un estudio instrumental”, *Nueva Revista de Filología Hispánica*, 1:1, pp. 1-14.
- HIERONYMUS, JAMES L. 1994 (Ms.). “ASCII phonetic symbols for the world's languages: Worldbet”, Nueva Jersey: AT&T y Laboratorios Bell. [Cito por: <http://www.ling.gu.se/~jimh/courses/ipa.ps> y <http://dipaola.org/stanford/facade/lipsync/worldbet.pdf>]
- . 1997 (Ms.). “Worldbet phonetic symbols for multilanguage speech recognition and synthesis”, Nueva Jersey: AT&T y Laboratorios Bell. [Cito por: <http://www.ling.gu.se/~jimh/courses/ipa.recog.unicode.ps>]
- INTERNATIONAL PHONETIC ASSOCIATION. 1949/1971. *The principles of the International Phonetic Association*, Londres: University College.
- . 1993. “IPA Chart, revised to 1993”, en *Journal of the International Phonetic Association* 23,1. [Cito por: <http://www.arts.gla.ac.uk/IPA/ipachart.html>]
- . 1999. *Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet*, Cambridge: Cambridge University Press.
- LANDER, TERRI. 1997. *The CSLU labeling guide*, Oregon: Oregon Graduate Institute of Science and Technology.
- LOPE BLANCH, JUAN M. 1963-1964/1983. “En torno a las vocales caedizas del español mexicano”, en *Estudios sobre el español de México*, México: Universidad Nacional Autónoma de México, pp. 57-77.
- . 1969/1989. *La filología hispánica en México. Tareas más urgentes*, México: Universidad Nacional Autónoma de México.
- . 1970. “Las zonas dialectales de México. Proyecto de delimitación”, *Nueva Revista de Filología Hispánica*, 19:1, pp. 1-11.
- . 1973/1990. “Sobre el tratamiento de *-e*, *-o* finales en el español de México”, en *Investigaciones sobre dialectología mexicana*, México: Universidad Nacional Autónoma de México, pp. 34-40.
- . 1978. “Una nota sobre los sonidos vibrantes”, *Anuario de Letras*, 11, pp. 247-250.
- . 1980/1986. “La información fonética en los atlas lingüísticos”, en *Estudios de lingüística española*, México: Universidad Nacional Autónoma de México, pp. 33-39.
- LÓPEZ CHÁVEZ, JUAN. 1981/1988. “La *s* mexicana. Algunas consideraciones para su estudio”, en *Actas del Sexto Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina*, México: Universidad Nacional Autónoma de México, pp. 777-784.

- . 1986. “/s/, /ç/ y /x/ intervocálicas en la República Mexicana”, en *Actas del Segundo Congreso Internacional sobre el español de América*, México: Universidad Nacional Autónoma de México, pp. 327-329.
- LÓPEZ ESCOBEDO, FERNANDA. 2004. *El estudio de los diptongos del español de México para su aplicación en un reconocedor de habla*, tesis de licenciatura inédita, México: Universidad Nacional Autónoma de México.
- LLISTERRI, JOAQUIM. 1991. *Introducción a la fonética: el método experimental*, Barcelona: Anthropos.
- . 1997 (Ms.). “Transcripción, etiquetado y codificación de corpus orales”, en *Etiquetación y extracción de información de grandes corpus textuales*, curso impartido en *Seminario de Industrias de la Lengua*, Soria. [Cito por: <http://liceu.uab.es/~joaquim/publicacions/FDS97.html>]
- . 1999. “Transcripción, etiquetado y codificación de corpus orales”, en *Panorama de la investigación en lingüística informática*, J. Gómez, A. Lorenzo, J. Pérez y A. Álvarez (Eds.), *Revista Española de Lingüística Aplicada*, volumen monográfico, pp. 53-82. [Cito por: http://liceu.uab.es/~joaquim/publicacions/RESLA_99.pdf]
- . 2002 (Ms.). “La conversión de texto en habla: aspectos lingüísticos”, en *Avances en el tratamiento computacional del lenguaje y del habla*, curso impartido en *Seminario de Industrias de la Lengua*, Soria. [Cito por: http://liceu.uab.es/publicacions/Linguistica_CTH_FDS02.pdf]
- . 2003. “Lingüística y tecnologías del lenguaje”, *Lynx. Panorámica de Estudios Lingüísticos*, 2, pp. 9-71. [Cito por: http://liceu.uab.es/~joaquim/publicacions/TecnoLing_Lynx02.pdf]
- LLISTERRI, JOAQUIM, CARMEN CARBÓ, MARÍA JESÚS MACHUCA, CARMEN DE LA MOTA, MONTSERRAT RIERA y ANTONIO RÍOS. 2003 (Ms.). “El papel de la lingüística en el desarrollo de las tecnologías del habla”, en *Séptimas Jornadas de Lingüística*, Cádiz: Servicio de Publicaciones de la Universidad de Cádiz, pp. 137-191. [Cito por: http://liceu.uab.es/publicacions/Lingüística_TH_Cadiz_02.pdf]
- LLISTERRI, JOAQUIM y JOSÉ B. MARIÑO. 1993 (Ms.). “Spanish adaptation of SAMPA and automatic phonetic transcription”, Barcelona. [Cito por: http://liceu.uab.es/~joaquim/publicacions/SAMPA_Spanish_93.pdf]
- LLISTERRI, JOAQUIM, LOURDES AGUILAR, JUAN MARÍA GARRIDO, MARÍA JESÚS MACHUCA, RAFAEL MARÍN, CARMEN DE LA MOTA y ANTONIO RÍOS. 1999. “Fonética y tecnologías del habla”, en *Filología e informática. Nuevas tecnologías en los estudios filológicos*, J.M. Blecua, G. Clavería, C. Sánchez y J. Torruella (Eds.), Barcelona: Milenio, pp. 449-479. [Cito por: http://liceu.uab.es/~joaquim/publicacions/Fonetica_TecnolHabla.pdf]
- LLISTERRI, JOAQUIM y MARÍA ANTONIA MARTÍ. 2002. “Las tecnologías lingüísticas en la sociedad de la información”, en *Tratamiento del lenguaje natural*, M.A. Martí y J. Llisterri (Eds.), Barcelona: Fundación Duques de Soria y Edicions Universidad de Barcelona, pp. 13-28.

- MADRID, EDGAR A. y MARIO A. MARÍN. 2001. “Estructura formántica de las vocales del español de la ciudad de México”, en *Temas de fonética instrumental*, E. Herrera (Ed.), México: El Colegio de México, pp. 39-58.
- MANRIQUE CASTAÑEDA, LEONARDO. 1968 (Ms.). “Lingüística computacional en México”, ponencia presentada en la *Reunión de Mayistas*, México.
- MARIÑO, JOSÉ B. y ASUNCIÓN MORENO. 2000 (Ms.). “Spanish SAMPA set”, SALA (SpeechDat across Latin America). [Cito por: <http://www.sala2.org/SAMPALLEX.ps> y <http://www.sala2.org/SALASAMPA.rtf>]
- MARTÍN BUTRAGUEÑO, PEDRO. 2003. “Más sobre la evaluación global de los procesos fonológicos: la geografía fónica de México”, en *Variación lingüística y teoría fonológica*, México: El Colegio de México, pp. 63-104.
- MARTÍNEZ CELDRÁN, EUGENIO. 1984/1994. *Fonética (con especial referencia a la lengua castellana)*, Barcelona: Teide.
- . 1995. “Nuevos datos sobre la dentalización de -s- en español”, *Revista de Filología Española*, 75:3-4, pp. 301-312.
- MARTÍNEZ CELDRÁN, EUGENIO y ANA MARÍA FERNÁNDEZ PLANAS. 2001. “Propuesta de transcripción para la palatal africada sonora del español”, *Estudios de fonética experimental*, 11, pp. 175-190. [Cito Por: <http://www.ub.es/spub/ebooks/fonetica.zip>]
- MARTÍNEZ CELDRÁN, EUGENIO, ANA MARÍA FERNÁNDEZ PLANAS y JOSEFINA CARRERA-SABATÉ. 2003. “Illustrations of the IPA. Castillian Spanish”, *Journal of the International Phonetic Association*, 33:2, pp. 255-260.
- MATLUCK, JOSEPH H. 1951. *La pronunciación en el español del Valle de México*, México: Edición de autor.
- . 1963. “La é trabada en la ciudad de México: Estudio experimental”, *Anuario de Letras*, 3, pp. 5-34.
- MORENO, ASUNCIÓN y JOSÉ B. MARIÑO. 1998 (Ms.). “Spanish dialects: Phonetic transcription”, en *Proceedings of ICSLP'98. The Fifth International Conference on Spoken Language Processing*, Rundle Mall: Causal Productions.
- MORENO DE ALBA, JOSÉ G. 1972. “Frecuencias de la asibilación de /r/ y /rr/ en México”, *Nueva Revista de Filología Hispánica*, 21:2, pp. 363-370.
- . 1988. *El español en América*, México: Fondo de Cultura Económica.
- . 1994. *La pronunciación del español de México*, México: El Colegio de México.
- MOTA, CARME DE LA y ANTONIO RÍOS. 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”, en *Acta Universitatis Wratislaviensis*, Wrocław: Universidad de Wrocław, pp. 97-109.
- MOTA, CARME DE LA y MONTSERRAT RIERA. 2000 (Ms.). “El transcriptor automàtic Segre”, ponencia presentada en las *Jornadas del Centro de Referencia en Ingeniería Lingüística*, Barcelona.
- NAVARRO TOMÁS, TOMÁS. 1918/1970. *Manual de pronunciación española*, Madrid: Consejo Superior de Investigaciones Científicas.

- . 1945. *Cuestionario lingüístico hispanoamericano*, 1: *Fonética, morfología y sintaxis*, Buenos Aires: Universidad de Buenos Aires.
- . 1946. “Escala de frecuencia de los fonemas españoles”, en *Estudios de fonología española*, Syracuse, pp. 15-30.
- . 1966-1967. “El alfabeto fonético de la *Revista de Filología Española*”, *Anuario de Letras*, 6, pp. 5-10.
- . 1974. “Transcripción estrecha”, *Anuario de Letras*, 12, pp. 181-187.
- OLIVIER, MARÍA ALEJANDRA. 1999. *Evaluación de métodos de determinación automática de una transcripción fonética*, tesis de maestría inédita, México: Universidad de la Américas.
- ORTEGA, JAVIER, JOAQUÍN GONZÁLEZ, VICTORIA MARRERO, JESÚS DÍAZ, RAMÓN GARCÍA, JUAN JOSÉ LUCENA y JOSÉ ANTONIO SÁNCHEZ. 1998 (Ms.). “AHUMADA: A large speech corpus in Spanish for speaker identification and verification”, en *Proceedings of ICAPSSP-98*, Madrid. [Cito por: <ftp://www.atvs.diac.upm.es/pub/publicaciones/ICSSP98/AhAIISSP98.pdf>]
- PACHÈS, PAU, CARME DE LA MOTA, MONTSERRAT RIERA, MARÍA PILAR PEREA, ALBERT FEBRER, MÒNICA ESTRUCH, JUAN MARÍA GARRIDO, MARÍA JESÚS MACHUCA, ANTONIO RÍOS, JOAQUIM LLISTERRI, IGNASI ESQUERRA, JAVIER HERNANDO, JAUME PADRELL y CLIMENT NADEU. 2000. “SEGRE: an automatic tool for grapheme-to-allophone transcription in Catalan”, en *Second International Conference on Language Resources and Evaluation*, Atenas, pp. 52-61. [Cito por: http://liceu.uab.es/~joaquim/publicacions/Paches_et_al_2000.pdf]
- PÉREZ, HERNÁN EMILIO. 2003 (Ms.). “Frecuencia de fonemas”, Concepción: Universidad de Concepción.
- PERISSINOTTO, GIORGIO. 1972. “Distribución demográfica de la asibilación de vibrantes en el habla de la ciudad de México”, *Nueva Revista de Filología Hispánica*, 21:1, pp. 71-79.
- . 1975. *Fonología del español hablado en la Ciudad de México. Ensayo de un método sociolingüístico*, México: El Colegio de México.
- PIERREHUMBERT, JANET, MARY E. BECKMAN y D. ROBERT LADD. En prensa. “Conceptual foundations of phonology as a laboratory science”, en *Phonological knowledge: Its nature and status*, N. Burton-Roberts, P. Carr y G. Docherty (Eds.), Cambridge: Cambridge University Press.
- PINEDA, LUIS A., ANTONIO MASSÉ, IVÁN MEZA, MIGUEL SALAS, ERIK SCHWARZ, ESMERALDA URAGA y LUIS VILLASEÑOR. 2001 (Ms.). “El Proyecto DIME”, México: Universidad Nacional Autónoma de México.
- PINEDA, LUIS A., LUIS VILLASEÑOR, JAVIER CUÉTARA PRIEDE, HAYDE CASTELLANOS e IVONNE LÓPEZ. 2004 (Ms.). “DIMEx100: a new phonetic and speech corpus for Mexican Spanish”, México: Universidad Nacional Autónoma de México.
- PULLUM, GEOFFREY K. y WILLIAM A. LADUSAW. 1986. *Phonetic symbol guide*, Chicago: The University of Chicago Press.
- QUILIS, ANTONIO. 1966/1968. “Sobre los alófonos dentales de /s/”, *Revista de Filología Española*, 49:1-4, pp. 335-343.

- . 1981/1988. *Fonética acústica de la lengua española*, Madrid: Gredos.
- . 1993/1999. *Tratado de fonología y fonética españolas*, Madrid: Gredos.
- . 1995. “De nuevo sobre el alófono dental de /s/ en español”, *Revista de Filología Española*, 75:3-4, pp. 313-317.
- REAL ACADEMIA ESPAÑOLA. 1973. *Esbozo de una nueva gramática de la lengua española*, Madrid: Espasa-Calpe.
- REVISTA DE FILOLOGÍA ESPAÑOLA. 1915. “Alfabeto fonético”, *Revista de Filología Española*, Madrid, 2, pp. 374-376.
- RÍOS MESTRE, ANTONIO. 1999. *La transcripción fonética automática del diccionario electrónico de formas simples flexivas del español: estudio fonológico del léxico*, en *Estudios de Lingüística Española*, 4. [Cito por: <http://elies.rediris.es/elies4/>]
- RODRÍGUEZ, YOLANDA y ARMANDO MORA. 2001. “Las consonantes oclusivas en posición de coda silábica: estudio sobre el español de la ciudad de México”, en *Temas de fonética instrumental*, E. Herrera (Ed.), México: El Colegio de México, pp. 81-108.
- RODRÍGUEZ, MIGUEL ÁNGEL, ISMAEL CORTÁZAR, DANIEL TAPIAS y JOSÉ RELAÑO. 2001. “Estado del arte en tecnologías de voz”, *Comunicaciones de Telefónica I+D*, 20, pp. 117-136. [Cito por: <http://www.tid.es/presencia/publicaciones/comsid/esp/20/8XX.PDF>]
- ROJO, GUILLERMO. 1991. “Frecuencia de fonemas”, en *Homenaje ó profesor Constantino García*, M. Brea y F. Fernández Rei (Coords.), Santiago de Compostela: Universidade de Santiago de Compostela, pp. 451-467.
- STRIK, HELMER y CATIA CUCCHIARINI. 1999. “Modeling pronunciation variation for ASR: A survey of the literature”, *Speech Communication*, 29:2-4, pp. 225-246.
- TLATOYA. 2000a. “ASCII phonetic symbols for Mexican Spanish”, México: Universidad de las Américas. [Cito por: <http://mailweb.udlap.mx/~sistemas/tlatoa/documentation/labels.html>]
- . 2000b. “How to hand label a speech corpus”, México: Universidad de las Américas. [Cito por: http://mailweb.udlap.mx/~sistemas/tlatoa/howto/hand_label.html]
- TORRUELLA, JOAN y JOAQUIM LLISTERRI. 1999. “Diseño de corpus textuales y orales”, en *Filología e informática. Nuevas tecnologías en los estudios filológicos*, J.M. Bleca, G. Clavería, C. Sánchez y J. Torruella (Eds.), Barcelona: Milenio, pp. 45-77. [Cito por: http://liceu.uab.es/~joaquim/publicacions/Torruella_Llisterri_99.pdf]
- TURING, ALAN M. 1950/1995. “Computing machinery and intelligence”, en *Computers and thought*, E.A. Feigenbaum y J. Feldman (Eds), Menlo Park: American Association of Artificial Intelligence Press y Massachusetts Institute of Technology Press, pp. 11-35.

- URAGA, ESMERALDA. 1999. *Modelado fonético para un sistema de reconocimiento de voz continua en español*, tesis de maestría inédita, México: Instituto Tecnológico y de Estudios Superiores de Monterrey.
- URAGA, ESMERALDA y LUIS A. PINEDA. 2000 (Ms.). "A set of phonological rules for Mexican Spanish", México: Universidad Nacional Autónoma de México.
- . 2002. "Automatic generation of pronunciation lexicons for Spanish", en *Computational Linguistics and Intelligent Text Processing. Third International Conference, CICLing 2002*, A. Gelbukh (Ed.), Berlin, Heidelberg, Nueva York, Barcelona, Hong Kong, Londres, Milán, París, Tokio: Springer, pp. 300-308.
- VAQUERO DE RAMÍREZ, MARÍA. 1996. *El español de América I. Pronunciación*, Madrid: Arco Libros.
- VILLASEÑOR, LUIS, ANTONIO MASSÉ y LUIS A. PINEDA. 2001 (Ms.). "The DIME Corpus", en *Tercer Encuentro Internacional de Ciencias de la Computación ENC-01*, Aguascalientes: SMCC-INEGI, pp. 591-600.
- VILLASEÑOR, LUIS, MANUEL MONTES-Y-GÓMEZ, DOMINIQUE VAUFREYDAZ y JEAN-FRANÇOIS SERIGNAT. En prensa. "Elaboración de un Corpus Balanceado para el Cálculo de Modelos Acústicos usando la Web", Tonantzintla, Instituto Nacional de Astrofísica, Óptica y Electrónica.
- WELLS, JOHN. 1998a. "SAMPA. Computer readable phonetic alphabet", Inglaterra: University College London. [Cito por: <http://www.phon.ucl.ac.uk/home/sampa/home.htm>]
- . 1998b. "SAMPA. Spanish", Inglaterra: University College London. [Cito por: <http://www.phon.ucl.ac.uk/home/sampa/spanish.htm>]
- WHITLEY, M. STANLEY. 2000. "Las paravocales españolas, el hiato y la abertura de la conjunción", en *Panorama de la fonología española actual*, J. Gil (Ed.), Madrid, Arco Libros, pp. 129-157.
- ZIPF, GEORGE K. y F.M. ROGERS. 1939. "Phonemes and variphones in four present-day Romance languages and Classical Latin from the viewpoint of dynamic philology", *Archives Néerlandaises de Phonétique Expérimentale*, 15, pp. 111-147.