

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
LICENCIATURA EN LENGUA Y LITERATURAS HISPÁNICAS

FENÓMENOS DE PÉRDIDA EN EL *CORPUS*
DIME PARA SU INCLUSIÓN EN UN
RECONOCEDOR DE HABLA

TESIS QUE, PARA OBTENER EL TÍTULO DE LICENCIADA
EN LENGUA Y LITERATURAS HISPÁNICAS, PRESENTA

ANA HELENA CEBALLOS MORA

Asesores: Dr. Luis Alberto Pineda Cortés
Mtro. Javier Octavio Cuétara Priede

Ciudad de México, 2007



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

صرت عبيدا لمن علمنى حرفا

Agradecimientos

Este fue un largo pero enriquecedor camino que no habría podido realizarse sin el apoyo incondicional de los que aquí menciono:

Mami, sabes que eres la luz que siempre ha guiado mis pasos y te agradezco enormemente que hayas decidido apoyarme en todas mis decisiones y que me hicieras el viaje tan ligerito. Eternamente estaré en deuda contigo.

Papi, desde allá, donde quiera que te encuentres, sabes que yo soy quién soy por ti, porque tú me enseñaste a ver y a oír el mundo; gracias por heredarme el placer de aprender y por construir esta amistad tan bonita que, ya comprobamos, va más allá de las fronteras de la vida. Te llevo siempre en mi corazón.

Grant, ¿De qué sirve construir y andar caminos si uno no tiene hacia donde dirigirlos? Todos mis pasos van hacia ti y hacia ese mundo que pronto comenzaremos a crear juntos. Te amo con todo mi corazón y te agradezco que hayas estado allí siempre para apoyarme y para darle sentido a todo esto.

A mis hermanos Gus y Fer por ser mis amigos de toda la vida, por escucharme, consentirme y quererme tanto.

Amiguita preciosa, Ale, fue un placer enorme emprender este esfuerzo contigo, todo juntas, todo al tiempo y el trabajo está terminado, y, sin embargo, lo más bonito que nos quedó fueron todas esas horas que pasamos juntas y que hicieron crecer algo muy bonito que está hecho para durar.

Luis, gracias por confiar en mí en el proyecto y por hacerme parte de él. También gracias por mover cielo, mar y tierra por procurarnos y por hacernos plácida la estancia. Y, además, por hacer que no fuera sólo una relación puramente profesional sino una estrecha amistad.

Javier, me salvaste la vida, gracias por toooodo lo que me enseñaste, por la paciencia, por los jalones de orejas, por los libros, las lecturas, las recomendaciones y por abrirme las puertas de tu casa y de tu corazón.

A mis amiguitos de la Fac que siempre me acompañaron, ya de salida, les agradezco que me hicieran pasar una de las etapas más divertidas y profundas de mi vida. Judith, Richi, Julio, Vero, Mariana, Ale, Blanca Vero, Abril, Nati.

Mis amigas de siempre: Alis y Mari.

شكر ل لاو و يمانويل و اريك و يوسف وساندره و مارسيلة و كلاودية :اصدقاء
وموسى واسكندرة واستاذي و ابريل و اسكارو ارايس و كيكيتو

Y a mis amigos y compañeros del IIMAS: César, Pati, Haydé, Janet, Varinia y Sergio.

Un agradecimiento especial al Proyecto NSF/CONACYT 39380-A DIME II:
Diálogos Inteligentes Multimodales en Español.

Y al Proyecto PAPIIT IN121206-3 “Diálogos Inteligentes Multimodales en
Español”

Ambos bajo la coordinación del Dr. Luis Alberto Pineda Cortés.

ÍNDICE

1. Introducción	9
1.1. Metodología	12
2. Tecnologías del habla	18
2.1. Las tecnologías del habla y la lingüística	18
2.1.1. Reconocimiento de habla	23
2.1.2. Síntesis de habla	25
2.1.3. Recursos lingüísticos para síntesis y reconocimiento de habla .	28
2.1.4. Aplicaciones de las Tecnologías del habla	30
2.2. El <i>Proyecto DIME</i>	34
2.2.1. El <i>Corpus DIME</i> y la importancia del habla espontánea	35
3. Fenómenos de elisión en el español	41
3.1. Tipología general de cambios de sonido	44
3.1.1. La elisión como fenómeno fonético productivo	45
3.2. Tipos de elisión	47
3.3. Relajamiento y pérdida en el español de la Ciudad de México	49
3.3.1. Vocales	51
3.3.2. Consonantes	53
4. Fenómenos fonéticos del español de México en el <i>Corpus DIME</i>	57
4.1. Resultados del análisis primario	58
4.2. Fenómenos de elisión en el <i>Corpus DIME</i>	60
4.2.1. Aféresis	62
4.2.2. Apócope	63

4.2.2.1. Apócope de vocales	64
4.2.2.2. Apócope de consonantes	65
4.2.3. Síncopa	71
4.2.3.1. Síncopa de consonantes intervocálicas	72
4.2.3.2. Síncopa de consonantes en posición implosiva	76
4.2.3.3. Síncopa en palabras específicas	82
4.2.3.4. Síncopa de vocales por reducción de diptongos	85
4.2.3.5. Silabificación	86
4.3. Reglas fonéticas	88
4.4. Aplicación de las reglas al diccionario de pronunciación	91
5. Conclusiones	94
6. Referencias bibliográficas	98
7. Apéndices	101
7.1. Apéndice 1. Imágenes espectrográficas de fonemas	101
7.2. Apéndice 2. Tabla de porcentajes	102

ÍNDICE DE FIGURAS

Figura 1. Funcionamiento de la herramienta <i>Speech Viewer</i>	14
Figura 2. Niveles de transcripción. Ejemplo de transcripción del segmento <i>es invisible</i>	15
Figura 3. Arquitectura de un reconocedor de habla (Llisterri, 2003a:15)	24
Figura 4. Ejemplo de selección de una palabra en el diccionario de pronunciación (Cuétara, 2004:27)	25
Figura 5. Esquema de un conversor de texto en habla (Llisterri, 2003:14) ...	27
Figura 6. Ejemplo de pronunciación de oración, con pausas, repeticiones e interjecciones (Oración 35 diálogo 22)	37
Figura 7. Interfaz gráfica del corpus. Panel de diseño de cocina	39
Figura 8. Tabla de porcentajes de fenómenos fonéticos en el <i>Corpus DIME</i>	58
Figura 9. Gráfica de porcentajes de elisiones en el <i>Corpus DIME</i>	61
Figura 10. Aféresis de /e/ ante /n/ en la frase <i>en la esquina</i> (Diálogo 10, oración 42, hablante 2)	62
Figura 11. Aféresis de /e/ ante /s/ en la palabra <i>espérame</i> (Diálogo 8, oración 151, Mago 1)	63
Figura 12. Apócope de /o/ en la palabra <i>acuerdo</i> (Diálogo 20, oración 3, Mago 2)	64
Figura 13. Apócope de /e/ en la palabra <i>parte</i> (Diálogo 20, oración 76, Mago 2)	65
Figura 14. Síncopa de /d/ final de palabra en la palabra <i>pared</i> (Diálogo 20, oración 88, Mago 2)	66
Figura 15. Apócope de /n/ en la palabra <i>bien</i> (Diálogo 22, oración 155, hablante 4)	68
Figura 16. Apócope de /r/ en la palabra <i>frigobar</i> (Diálogo 20, oración 17, Mago 2)	68
Figura 17. Apócope de /s/ en la palabra <i>tenemos</i> (Diálogo 26, oración 36, Mago 1)	69

Figura 18. Reducción vocálica en el diptongo [ej] en la palabra <i>okay</i> (Diálogo 26, oración 170, hablante 5)	70
Figura 19. Síncopa de /b/ intervocálica la frase <i>por favor</i> (Diálogo 8, oración 20, Mago 1)	73
Figura 20. Reducción del segmento /b/ a [l] entre vocales en la palabra <i>problema</i> (Diálogo 22, oración 162, hablante 4)	73
Figura 21. Síncopa de /r/ intervocálica en la palabra <i>sería</i> (Diálogo 22, oración 32, hablante 4)	74
Figura 22. Síncopa de <i>r</i> en el segmento /tr/ en la palabra <i>muestras</i> (Diálogo 26, oración 33, hablante 5)	75
Figura 23. Síncopa de /d/ intervocálica en la palabra <i>pegada</i> . (Diálogo 10, oración 92, hablante 2)	76
Figura 24. Reducción del segmento /ns/ en posición implosiva a [s] en <i>dimensiones</i> (Diálogo 8, oración 3, Mago 1)	78
Figura 25. Síncopa de /n/ ante /l/ en la oración <i>son las sillas</i> (Diálogo 8, oración 107, Mago 1)	79
Figura 26. Síncopa de /r/ final de palabra ante /f/ en la frase <i>por favor</i> (Diálogo 26, oración 179, hablante 5)	80
Figura 27. Síncopa de /s/ final de palabra ante /l/ en la frase <i>alacenas laterales</i> (Diálogo 8, oración 28, Mago 1)	81
Figura 28. Síncopa de /s/ final de palabra ante /r/ en la oración <i>podemos rotar</i> (Diálogo 20, oración 147, hablante 3)	81
Figura 29. Reducción del segmento /ks/ intervocálico a [s] en la palabra <i>exacto</i> (Diálogo 26, oración 141, hablante 5)	82
Figura 30. Reducción silábica de la palabra <i>puedes</i> a [pués] (Diálogo 10, oración 45, hablante 2)	83
Figura 31. Reducción silábica de la palabra <i>quieres</i> a [kjés] (Diálogo 22, oración 110, Mago 1)	84
Figura 32. Reducción silábica de la palabra <i>necesitaría</i> a [nesitaría] (Diálogo 22, oración 86, hablante 4)	84
Figura 33. Reducción del diptongo /je/ a [i] en la palabra <i>tienes</i> (Diálogo 8, oración 130, hablante 1)	86
Figura 34. Reducción del diptongo /je/ a [e] en la oración <i>está bien</i> .(Diálogo 10, oración 58, hablante 2)	86

Figura 35. Espectrograma de /e/ inicial ante /n/	103
Figura 36. Espectrograma de /d/ fricativa final	103
Figura 37. Espectrograma de /d/ oclusiva final	103
Figura 38. Espectrograma de /d/ fricativa intervocálica	104
Figura 39. Espectrograma de /o/ final	104
Figura 40. Espectrograma de /n/ final en la sílaba /en/	104
Figura 41. Espectrograma de diptongo /ie/	104
Figura 42. Espectrograma de /ks/	104
Figura 43. Espectrograma de /n/ intervocálica	105
Figura 45. Espectrograma de /n/ ante consonante /l/	105
Figura 46. Espectrograma de /n/ ante	105
Figura 47. Espectrograma de /r/ final asibilada	106
Figura 48. Espectrograma de /r/ final	106
Figura 49. Espectrograma de /r/ intervocálica	106
Figura 50. Espectrograma de /r/ ante /f/	106
Figura 51. Espectrograma de /s/ final	107
Figura 52. Espectrograma de /s/ intervocálica sorda	107
Figura 53. Espectrograma de /s/ sonora	107
Figura 54. Espectrograma de /s/ ante consonante dental	107
Figura 55. Espectrograma de /s/ ante /l/	108
Figura 56. Espectrograma de /s/ ante /r/	108
Figura 57. Espectrograma del segmento /tr/	108
Figura 58. Espectrograma de /b/ intervocálica	108
Figura 59. Espectrograma del segmento /b/	109

1. INTRODUCCIÓN

El lenguaje y las computadoras establecen cada vez una relación más estrecha en el área de la investigación académica y en el desarrollo de la tecnología. Desde que las máquinas existen se ha buscado siempre que faciliten y resuelvan las necesidades de la vida cotidiana. Por lo tanto, no podría quedar atrás una de las necesidades más básicas de los seres humanos, la comunicación. Ya que las computadoras han resultado de gran ayuda en la resolución de tareas, habilitarlas con mecanismos de comunicación oral y escrita parece ser el paso cumbre para el desarrollo de tecnologías avanzadas y de administración de tareas. Si una máquina, basada en sus propios recursos, poseyera la capacidad de comunicarse con seres humanos facilitaría muchas tareas mecánicas que se resuelven con diálogos prácticos y con fórmulas; actualmente son muchas las aplicaciones con las que se cuenta que ya resuelven este tipo de tareas; por ejemplo, en algunos países existen vendedores automáticos de boletos de tren o servicios de información vía telefónica (Poza, 1991 y Villarubia, 2002; ver §2.1.4).

Estas tecnologías que se desarrollan a través del tratamiento del lenguaje son las llamadas tecnologías lingüísticas, que se dividen en síntesis de habla, reconocimiento de habla y sistemas conversacionales. La primera consiste en crear una señal oral a base de textos o de parámetros acústicos; la segunda en convertir una señal oral en texto mediante métodos de interpretación, y la tercera pretende que un sistema sea capaz de producir y reconocer habla al mismo tiempo que puede interpretarla y predecir actos lingüísticos; es decir, participar en conversaciones naturales. Todas tienen métodos

complejos que se explicarán profundamente en los capítulos siguientes. Lo que resulta importante señalar es que, para el desarrollo de estas tecnologías, es necesario hacer estudios lingüísticos que permitan obtener datos del comportamiento de una lengua para poder aplicarlos a sistemas computacionales y así darles las características que requieren para tratar el habla óptimamente.

La presente investigación pretende hacer un estudio lingüístico de una muestra de habla espontánea para identificar fenómenos fonéticos característicos del español de la Ciudad de México, para dar razón de su comportamiento actual y para utilizar esta información como aplicación en las tecnologías del habla, con la intención de lograr mejoras en los sistemas actuales. Este estudio tendrá como fin alcanzar un mejor funcionamiento del reconocedor de habla del Proyecto *DIME* y aportar reglas acústicas que podrían emplearse en la síntesis de habla para otras investigaciones.

Para esto, el análisis exhibirá aspectos computacionales y lingüísticos como marco para interpretar el trabajo empírico realizado y las aplicaciones tecnológicas que tendrá. Así, el primer capítulo abordará el tema de las tecnologías del habla, explicando sus ramas de estudio, sus aplicaciones y haciendo un breve recorrido histórico desde su surgimiento hasta la actualidad. En cuanto a sus ramas de estudio, se explicarán los procedimientos de reconocimiento y síntesis de habla, ahondando en la arquitectura de los programas computacionales que se emplean en estos procedimientos, pues, como se mencionó, es precisamente en estos programas que se pretende aplicar los resultados del presente estudio lingüístico, para lograr mejoras en su funcionamiento. Se tratará también el estado actual de los avances en el campo de la lingüística computacional,

para dar un marco histórico del creciente interés por desarrollar investigaciones en esta área.

Además, se presenta el *Proyecto DIME*¹, en el cual se inserta la investigación aquí presentada. Se mostrará los recursos lingüísticos que sirven de base a diversas investigaciones multidisciplinarias y se explicará ampliamente el *Corpus DIME* de donde se toma la muestra de habla que se analizó en este estudio.

El segundo capítulo será introducido por la teoría sobre cambios de sonido y, dentro de éstos, el concepto de elisión, su división y la importancia del fenómeno en la evolución de las lenguas. Se mostrará un panorama general sobre la teoría de los fenómenos de relajamiento y pérdida del español de México, para hacer una descripción de éste con base en las distintas investigaciones lingüísticas que se han realizado desde mediados del siglo pasado, hasta las más recientes. Algunos autores a abordarse son: Henríquez Ureña (1938), Matluck (1951), Boyd- Bowman (1960), Perissinotto (1975), Canfield (1981), Lope Blanch (1991) y Moreno de Alba (1994).

El tercer capítulo abarcará el análisis profundo de la muestra de habla que se utilizó para esta investigación. Se explicará el proceso empírico con el que se obtuvieron los datos y cómo éstos fueron clasificados de acuerdo con la sistematización que presentaron. Se hará, además, el estudio de cada fenómeno exponiendo los resultados y sus causas fonéticas. Finalmente, se explicará cuál fue el resultado del estudio y se mostrarán los fenómenos que resultaron relevantes para el reconocimiento de habla formalizados en reglas para su aplicación tecnológica. Será presentada también la teoría sobre la

¹ <http://leibniz.iimas.unam.mx/~luis/DIME>

aplicación de estas reglas a un programa computacional de reconocimiento de habla y se expondrá cuál sería su impacto, así como pruebas parciales de la aplicación de reglas fonéticas al reconocedor de habla del Proyecto *DIME*.

Al final de la investigación se presentan dos apéndices. El primer apéndice es una referencia del trabajo empírico realizado, en el que se presenta una tabla con el número total de oraciones para cada fenómeno, así como los porcentajes generales del estudio. El segundo apéndice es un compendio de figuras que muestran la realización óptima de los fonemas cuya pérdida se expone en el presente análisis. Este apéndice tiene la utilidad de mostrar al lector no familiarizado con los espectrogramas, cómo se ven las representaciones de estos fonemas, para contrastarlas con las figuras que muestran su pérdida y comprobar así, que éstas se han realizado efectivamente. Las imágenes de este segundo apéndice fueron tomadas del mismo corpus; para eso se seleccionaron oraciones en las que los fonemas fueran pronunciados claramente para que su representación visual fuera la óptima.

1.1 Metodología

Esta investigación nació del interés por implementar en un reconocedor de habla un sistema de reglas fonéticas basadas en análisis empíricos de muestras de habla espontánea. En principio de cuentas, con este análisis se pretendía obtener un panorama total de fenómenos fonéticos recurrentes en el habla con la intención de ver cuáles resultaban relevantes en cuanto a aplicación tecnológica. Con esta meta, se inició un trabajo empírico cuya duración fue aproximadamente de cinco meses.

Al inicio se seleccionó una muestra de habla perteneciente al *Corpus DIME*, un corpus oral que es parte de los recursos lingüísticos con los que cuenta el *Proyecto DIME* del Departamento de Ciencias de la Computación (DCC) del IIMAS de la UNAM (en §2.2 se abordará el tema más ampliamente). Como se pretendía que las reglas obtenidas se aplicaran en un reconocedor de habla que se utiliza en situaciones de habla natural, se consideró pertinente utilizar este corpus, ya que es de habla espontánea. Este corpus fue obtenido con un proceso similar al de *Mago de Oz*, el cual consiste en enfrentar a un hablante con una máquina en una situación comunicativa natural en la que el hablante no sabe que está hablando con una computadora. En este caso, el papel de la computadora o sistema era desarrollado por un hablante también y ambos estaban al tanto de la grabación. La espontaneidad se logró gracias a que la situación conversacional giraba en torno al diseño de cocinas y así la atención de ambos hablantes no se dirigía a la pronunciación.

De este corpus tomamos cinco diálogos, equivalentes a 828 oraciones de aproximadamente 20 minutos de duración cada una. El corpus es multimodal; es decir, cuenta con audio y video. En este caso, no consideré pertinente utilizar el video, por lo que trabajé con el audio solamente. Una vez que las oraciones fueron seleccionadas, comencé un proceso de transcripción de las mismas, haciendo uso de la herramienta *Speech Viewer*. Esta herramienta de transcripción es un programa computacional que permite estudiar emisiones de sonido en distintos niveles. En la Figura 1 vemos la herramienta y sus partes.

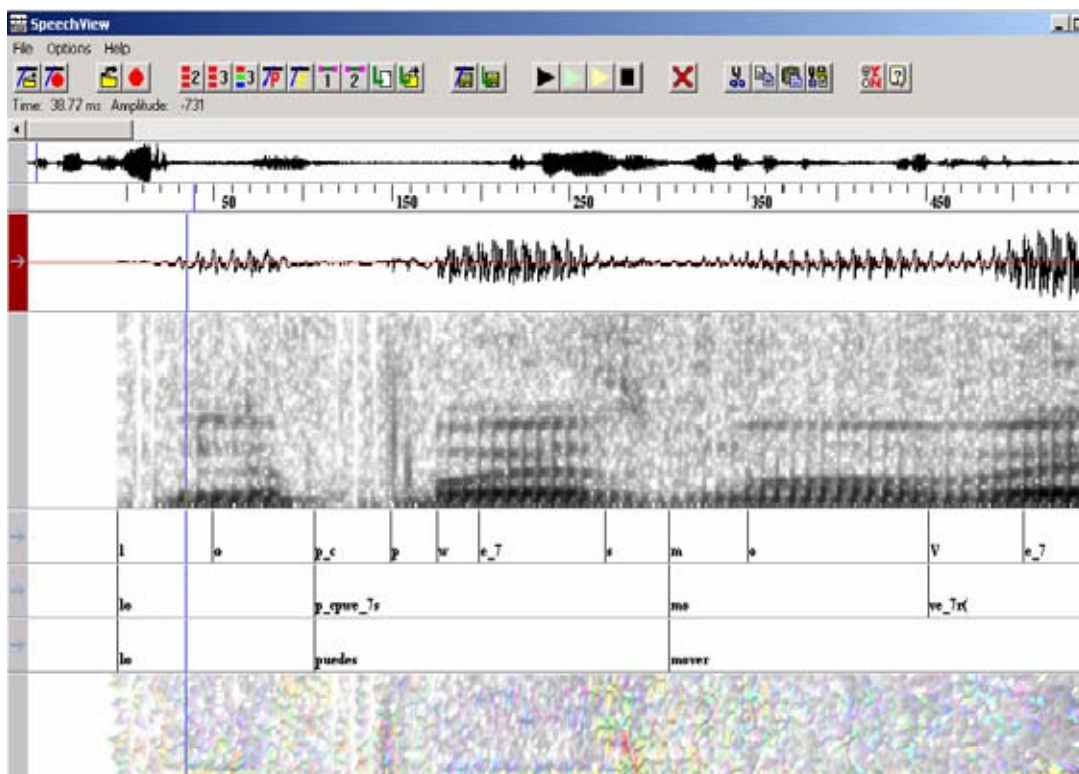


Figura 1. Funcionamiento de la herramienta *Speech Viewer*.

Como vemos, la herramienta cuenta con dos oscilogramas, uno fijo -el superior- y otro móvil -el inferior; dos espectrogramas, uno de banda estrecha y otro de banda ancha; etiquetas de texto cuya utilidad principal es permitir la alineación de las transcripciones con la señal de audio; con la herramienta se pueden obtener también la curva entonativa y el nivel de sonoridad de los fonemas, entre otras características; estas ventanas pueden utilizarse según el tipo de transcripción que se realice. En el ejemplo de la Figura 1, vemos en orden descendente la transcripción alofónica, la de sílabas fonéticas y la ortográfica.

Cada una de las oraciones fue transcrita según las convenciones establecidas en el alfabeto Mexbet (Cuétara, 2004), un alfabeto fonético computacional capaz de

representar las variantes alofónicas más importantes para el español de México, además de marcas de oclusión y acentos; este alfabeto tiene caracteres compatibles con los sistemas computacionales y se utiliza para la transcripción fina de alófonos. Las oraciones fueron transcritas, además, por sílabas fonéticas; es decir, las sílabas reales de la pronunciación sin respetar límites de palabra. Por último, se transcribieron también ortográficamente para saber qué palabras se pronunciaron a pesar de que ocurrieran modificaciones; esta transcripción nos muestra también que los mensajes son captados íntegramente a pesar de sufrir variaciones en la pronunciación. En la Figura 2 tenemos el ejemplo de una oración transcrita en los tres niveles de segmentación mencionados.

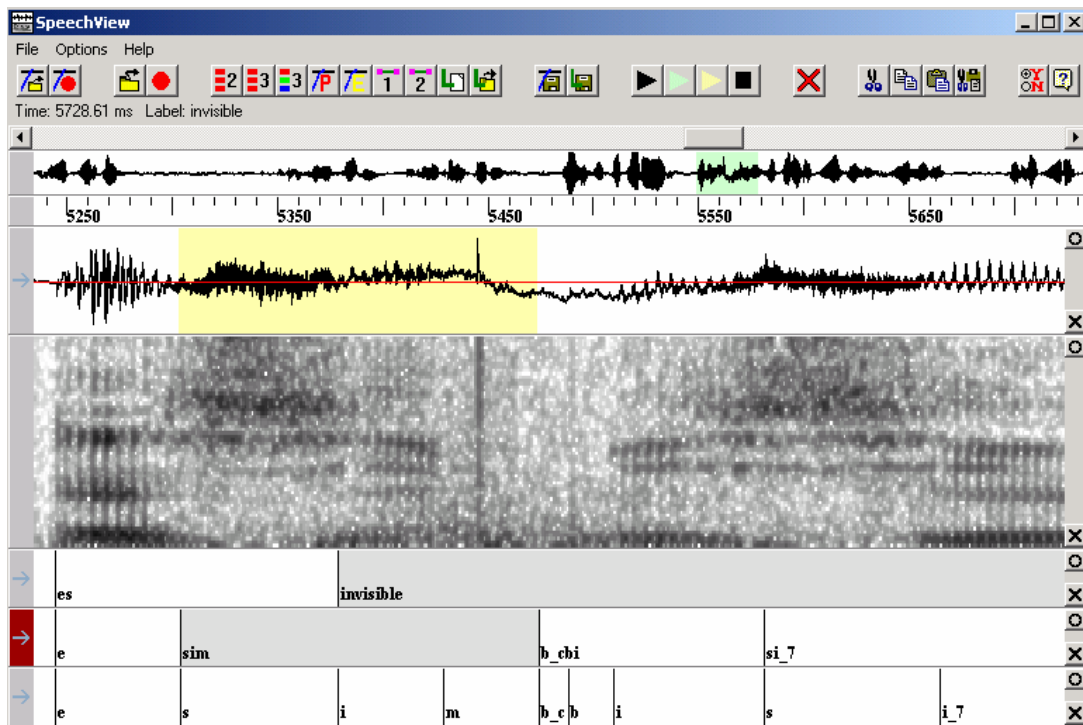


Figura 2. Niveles de transcripción. Ejemplo de transcripción del segmento *es invisible*.

En la Figura 2 vemos un ejemplo de transcripción -en orden descendente- de palabras, sílabas fonéticas y alófonos. Notamos que la transcripción de alófonos marca oclusiones ([b_c] oclusión de [b]) y acentos ([_7]), entre otros. Notamos también, en la segunda etiqueta, que la segmentación de sílabas fonéticas difiere de la canónica cuando se agrupa la /s/ con la sílaba siguiente.

El proceso de transcripción requiere de muchas horas de trabajo, de conocimiento fonético y de sensibilidad perceptiva, puesto que los fenómenos deben identificarse tanto visual como auditivamente. La transcripción tardó aproximadamente tres meses habiendo dedicado treinta horas de la semana. Una vez concluida esta etapa procedí al análisis de los datos. Para esto, se hizo necesario eliminar 157 oraciones por resultar irrelevantes para el objetivo de la investigación. Estas oraciones descartadas consistían en monosílabos como *sí* y *no*, interjecciones de un solo fonema y oraciones saturadas de ruido que impedía su interpretación adecuada. Al final, el corpus de análisis estaba compuesto por 671 oraciones. De éstas se obtuvieron 451 ejemplos de fenómenos fonéticos.

En una primera etapa todos los fenómenos fonéticos fueron separados por categorías y estudiados particularmente. Algunas categorías detectadas fueron: elisiones, inserciones, aperturas y cierres vocálicos, etc.; el análisis de estas categorías tenía como objeto encontrar los fenómenos que se daban con alta frecuencia en la muestra y los que revelaban patrones de causas y comportamiento. Posteriormente, se hizo evidente que, de los fenómenos analizados, las elisiones eran las más frecuentes y además presentaban un comportamiento más o menos uniforme y sistemático, por lo que decidí convertirlas en objeto de la investigación.

Los resultados del análisis y la explicación de cada uno de los fenómenos encontrados se presentan de manera amplia en el Capítulo 4. Por último, se exponen las conclusiones generales del análisis y una serie de apéndices que muestran parte del trabajo empírico realizado, así como ejemplos gráficos de la apariencia espectrográfica óptima de las consonantes y vocales, para que el lector identifique las pérdidas físicamente, contrastando las características visuales de un sonido en la imagen óptima y en la que ocurre un fenómeno de pérdida.

2. Tecnologías del habla

La presente investigación se inserta en el marco de las tecnologías del habla, por lo que considero necesario exponer aquí la teoría general sobre el procesamiento de lenguaje natural y textos, así como de las tecnologías del habla; de éstas últimas abordaré su desarrollo histórico, su impacto en la tecnología, y el estado actual de la materia.

Se hablará específicamente de las áreas de síntesis y reconocimiento de habla y, ya que este estudio buscará emplearse en esta última, se hablará ampliamente del proceso de reconocimiento de habla haciendo hincapié en los diccionarios de pronunciación, mostrando para esto la arquitectura del sistema computacional, puesto que es en estos programas computacionales en los que se aplicarán los resultados obtenidos de esta investigación.

2.1 Tecnologías del habla y lingüística

Desde mediados de la época de los cuarenta y hasta finales de los años cincuenta emerge una serie de teorías en el campo de la hoy denominada inteligencia artificial. La primera teoría surge con Turing (1950) que propuso la inteligencia artificial como respuesta a la cibernética. La segunda planteaba que los procesos del pensamiento pueden ser modelados como sistemas de control; la primera pretendía la creación e interpretación de representaciones de carácter simbólico, a la vez que conservaba los

sistemas reactivos en los que se enfoca la cibernética. Turing estableció además la interrogante sobre si las máquinas son capaces de pensar (Jurafsky, 2000), así como una prueba, conocida como “Prueba de Turing”, que se propone suficiente para determinar si una máquina es inteligente o no.

Surgieron otras propuestas como la Teoría matemática de la Información. En ésta, Shannon (1948) muestra que cualquier medio que sirva de transporte de información es medible y que la información sólo puede ser transmitida cuando el canal tiene una capacidad suficiente en cuanto a tamaño. Entre otras aportaciones, Shannon “applied probabilistic models of discrete Markov process to automata for language”¹ y estableció los cimientos sobre la corrección de errores y la exposición del modelo del canal ruidoso que permite la extracción de un mensaje a pesar de que se encuentre bajo condiciones de ruido (Jurafsky, 2000). Otra personalidad cuyas investigaciones resultaron fundamentales para el desarrollo de estas ciencias fue Noam Chomsky, quien contribuyó a la formulación de la teoría de los autómatas y de los lenguajes formales a la par que asentó las bases de la gramática generativa (Jurafsky, 2000).

De éstas y más investigaciones nació el interés por crear modelos de computarización del lenguaje y por desarrollar la comunicación entre humanos y máquinas. A partir de la década de los sesenta, libros y películas de ciencia ficción jugaban con la idea de máquinas inteligentes capaces de hablar y pensar como humanos. Surgieron así disciplinas como las tecnologías lingüísticas en las cuales se insertan las tecnologías del habla.

¹ Aplicó modelos probabilísticos del proceso discreto de Markov a autómatas para el lenguaje.

Las tecnologías del habla son una rama de la lingüística computacional que centra su estudio en el procesamiento del habla. Tienen como objetivo hacer posible y eficiente el diálogo entre los individuos y los sistemas computacionales, logrando con esto que se sustituyan los métodos tradicionales de comunicación entre ambos, como el teclado o el ratón, por conversaciones más o menos naturales.

Dichos objetivos tienen como principal interés la aplicación de estos métodos al entorno comercial, en el que cada vez son más necesarios y utilizados este tipo de recursos. Es común actualmente enfrentarse con sistemas comerciales que utilizan la síntesis de habla y que son capaces de reconocerla también. Por ejemplo las compañías telefónicas han sustituido los operadores humanos en la resolución de tareas simples como búsqueda o redirección de clientes a los departamentos correspondientes; los teléfonos celulares y sistemas de seguridad son capaces de reconocer el habla de modo limitado y marcar un número o permitir el acceso a los sistemas respectivamente. Así, vemos que las aplicaciones provienen del interés de lograr que las máquinas faciliten las tareas cotidianas.

Si las máquinas son capaces de interactuar oralmente o de recibir órdenes por medio de la lengua oral en lugar de forzar al usuario a instalarse frente a una máquina y utilizar el ratón o el teclado, el esfuerzo de realización de tareas sería mínimo y además, “el uso de la lengua oral como modo de comunicación con los sistemas informáticos libera las manos y la vista, facilitando la recepción de información desde cualquier punto y haciendo posible la movilidad; al mismo tiempo, permite efectuar otras actividades simultáneas” (Martí, 2002:20).

Los estudios en tecnologías del habla necesitan de una serie de disciplinas tales como la lingüística, la ingeniería, la computación, entre otras. Puesto que el habla es su objeto de estudio, estas ramas de pensamiento son necesarias para comprender cada aspecto de la materia y así lograr estudios profundos, útiles y aplicaciones óptimas de éstos a la tecnología (Martí Roca, 2002).

Aunque hace un par de décadas era difícil concebir que en sistemas comerciales informáticos la lingüística resultara una parte importante, Llisterri (2003) señala que el papel del lingüista se ha vuelto indispensable en cuanto que los sistemas pretenden reproducir el habla y, por tanto, es necesario conocerla profundamente. En su trabajo de 2003 (p.43), Llisterri cita también a Cabré y Payrató en su afirmación de que el campo de la lingüística computacional ha abierto nuevas aplicaciones para el profesional lingüística que participa en la base de los estudios realizados en el área.

Siempre que se busca fabricar un sistema de cualquier tipo basado en el habla es necesario crear recursos lingüísticos que sirvan de base a las investigaciones. Estos son conocidos como *corpus* y pueden ser lo mismo orales que escritos, según el tipo de investigación que se realice. Estos corpus y su análisis son trabajo del lingüista, quien se encargará de crear recursos confiables y adecuados para cada proyecto, además de extraer de ellos información relevante para aplicaciones tecnológicas. El tipo de estudio lingüístico dependerá de la rama de las tecnologías del habla en la que desempeñe su labor, así como de los objetivos particulares de cada estudio; del habla podrían estudiarse no sólo sus propiedades acústicas, de producción y de percepción sino también aquellas pertenecientes a los niveles léxico, sintáctico, semántico y

pragmático.

Uno de los niveles que aporta muchos resultados a las tecnologías del habla es la fonética; los estudios que se centren en este campo deben ser realizados por lingüistas y su aplicación a los sistemas informáticos será tarea de ingenieros. Así, los estudios fonéticos han permitido muchos avances en el marco de las tecnologías del habla, pero también es cierto que la fonética es una de las ramas lingüísticas que más se ha beneficiado de los avances tecnológicos (Llisterri *et al*, 1999), con los que ha podido desarrollarse y optimizar sus análisis.

Hablemos primero de la importancia de los estudios fonéticos en el desarrollo de las tecnologías del habla; ya que los sistemas informáticos esencialmente buscan crear y reconocer habla, deben apoyarse en las características de ésta. Todos los niveles lingüísticos pueden ser objeto de estudio, pero, dependiendo de la aplicación que se pretenda ejecutar, sólo algunos se vuelven relevantes. Sin embargo, la fonética participa en todos los estudios puesto que es base de la producción y la percepción del habla.

Como veremos en su momento (§1.1.1 y 1.1.2) los procesos en que se dividen las tecnologías del habla tienen como base modelos acústicos que se obtienen de estudios fonéticos. En cuanto a los métodos de los que la fonética se vale para sus investigaciones, podemos mencionar los oscilogramas y los espectrógrafos, que son las herramientas que sirven para identificar las propiedades acústicas de las señales sonoras y para trabajar con valores como la duración, la intensidad, la entonación, etc.

Según su estudio y sus aplicaciones, las tecnologías del habla se dividen en síntesis y reconocimiento de habla. Ambos procesos se explican a continuación.

2.1.1 Reconocimiento de habla

Un sistema reconocedor de habla se construye en dos etapas. La primera es la etapa de entrenamiento, en la que al sistema se le asignan las características acústicas que es necesario que almacene en la memoria. La segunda etapa es aquella de reconocimiento, en la que el sistema identifica un mensaje interpretando la pronunciación más adecuada según las propiedades (Pérez Pavón, 2006:12). Esta segunda etapa se explica a continuación.

El reconocimiento de habla es un proceso mediante el cual se convierte una señal de audio -es decir, habla- en texto, mediante un mecanismo complejo de descodificación. Los reconocedores de habla son programas de computadora cuyo funcionamiento consiste en captar una señal acústica y convertirla en analógica y posteriormente digitalizarla; cuando la señal es identificada como voz es analizada para la extracción de sus características relevantes, como son los formantes, la imagen espectral, el tono y la entonación; estas características son comparadas con los modelos acústicos y diccionarios de pronunciación para que posteriormente se convierta la palabra o la frase reconocidas en texto (Tapias Merino, 2003). Vemos pues, que el proceso de reconocimiento de habla se basa en un aprendizaje de las características de un corpus amplio con el cual se compararán las señales recibidas

(Llisterri, 2003).

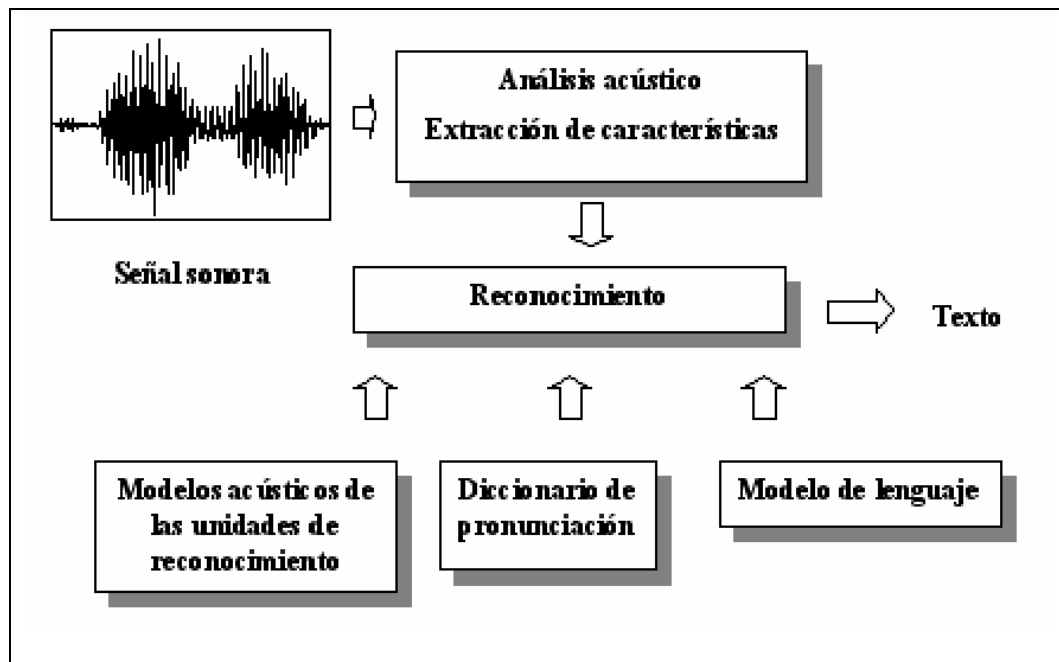


Figura 3. Arquitectura de un reconocedor de voz (Llisterri, 2003a:15).

Como se aprecia en la Figura 3, el reconocedor de habla cuenta con dos elementos principales en los que se basa todo el proceso de descodificación, los modelos acústicos y los diccionarios de pronunciación. Los modelos acústicos son creados a partir de muestras de unidades fonéticas y contienen información sobre las propiedades que caracterizan cada sonido. En la misma Figura 3 podemos notar que cuando un sonido es captado se analizan sus características acústicas principales y se contrastan con aquellas contenidas en los modelos acústicos, los diccionarios de pronunciación y los modelos del lenguaje; así se interpretan los sonidos pronunciados y se convierten en texto.

Los diccionarios de pronunciación son listas del léxico que el reconocedor de habla

es capaz de interpretar; indican la secuencia exacta de los sonidos que debe interpretar el reconocedor. Estos sonidos son representados por símbolos que pueden estar basados en alfabetos fonéticos y/o computacionales, como el Alfabeto Fonético Internacional (AFI) o el alfabeto de la Internacional Phonetic Association (IPA), entre otros. Las listas pueden ser creadas mediante hipótesis o pueden obtenerse, también, a partir del estudio de corpus orales. Estos diccionarios contienen la pronunciación canónica de cada palabra, así como pronunciaciones alternativas que los hablantes son capaces de realizar. En la Figura 4 podemos observar un ejemplo de las distintas pronunciaciones que se tienen para la palabra *refrigerador*, tomando en cuenta las distintas aperturas de las vocales. En esta muestra [E] y [O] son [e] y [o], respectivamente, de acuerdo con el alfabeto fonético computacional en código ASCII Mexbet, según Cuétara (2004).



Figura 4. Ejemplo de selección de una palabra en el diccionario de pronunciación (Cuétara, 2004: 27)

2.1.2 Síntesis de habla

La síntesis del habla es el proceso inverso al reconocimiento; y busca crear una señal oral automática a partir de ciertos parámetros que le han sido dados al sistema como reglas, o también transformar un texto en habla. Esta última es la aplicación más utilizada (Llisterri, 2001). Se emplea ampliamente en el mundo comercial para economizar la cantidad de información que se transmite a través de las vías de telefonía; por ejemplo, para crear la voz sintética es necesario generarla “por concatenación (y modificación) de las unidades necesarias para componer la secuencia de sonidos que se va a producir” (Villarrubia, 2002:13).

En cuanto a su función, el conversor de texto en habla trabaja de forma similar a la del reconocimiento porque es una estructura por módulos; es decir, en el proceso se tienen diversas etapas que poseen, cada una, una función distinta desde la percepción hasta la emisión de la señal oral.

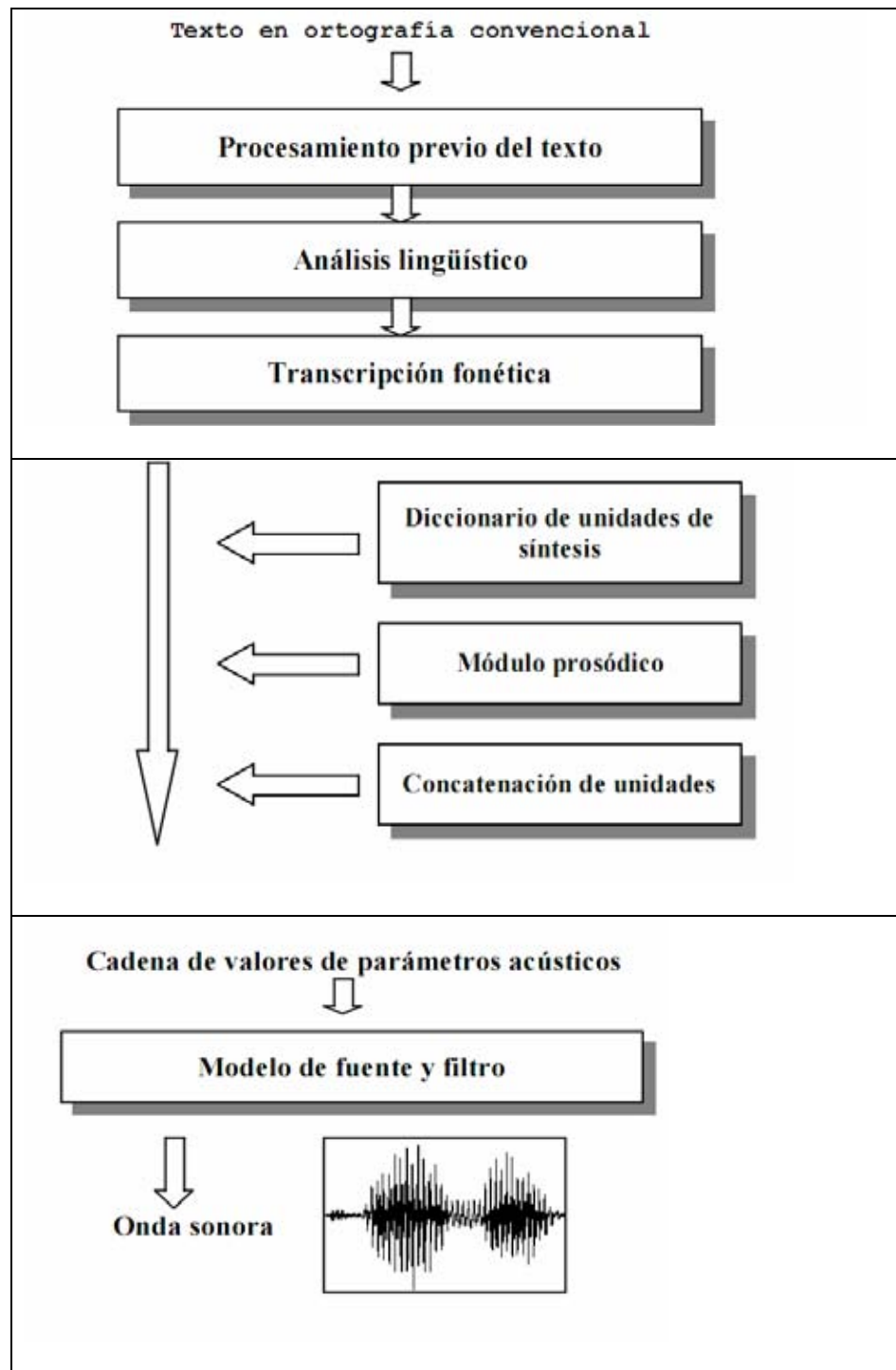


Figura 5. Esquema de un conversor de texto a voz (Llisterri, 2003:14).

La Figura 5 explica cómo en una primera etapa se descodifican los símbolos

complejos del texto (como los monetarios, las abreviaturas de organizaciones, etc.) deletreándolos. Por ejemplo, si el texto dice “se pagaron \$425 000 al Dr. Ibáñez, militante del PRD”, la descodificación se hará de la siguiente manera: “se pagaron cuatrocientos veinticinco mil pesos al doctor Ibáñez, militante del perredé”. Una vez que el texto se encuentra ortográficamente uniforme se busca convertirlo fielmente en su representación sonora. Para esto, se tiene que consultar los diccionarios de unidades acústicas que pueden estar basados en una representación de gráfica-sonido o en reglas fonológicas y fonéticas (Llisterri, 2003).

La voz sintética se producirá basándose en los modelos de concatenación de sonidos, por lo tanto es necesario que las bases sean confiables. En este proceso es útil el estudio de corpus orales que permitan la identificación de las características acústicas de una lengua, así como la identificación del proceso de concatenación de unidades en el habla espontánea (I Fontanals *et al*, 2003:55).

Algunos problemas en la creación de los diccionarios de características para la conversión de texto a habla surgen cuando las lenguas tienen una correspondencia abstracta y disforme entre gráficas y sonidos, o cuando los sintetizadores tienen que pronunciar nombres en otras lenguas cuya estructura acústica o articulatoria es distinta. Por ejemplo, los anglicismos de la tecnología en países hispanohablantes, como *e-mail*, *scanner*, etc.

2.1.3 Recursos lingüísticos en síntesis y reconocimiento

Cuando los sistemas de síntesis y reconocimiento se emplean para ambientes más

espontáneos, surgen algunos problemas que tienen su origen en la naturaleza del habla misma. Desde su inicio, ambos procesos han tratado de sistematizarse para que los resultados sean óptimos y naturales. Lo que se busca con un sistema de reconocimiento es que sea capaz de captar una gran diversidad de señales e interpretarlas con el menor número de errores posible o sin error alguno; lo que se pretende con la síntesis es crear señales orales que sean lo más parecidas posible a las enunciadas por el hombre; es decir, se busca que los sistemas sean igual de sofisticados que los que el hombre posee de manera natural para la comunicación, o que aparenten serlo.

Los modelos acústicos se obtienen a partir de la extracción de las características de cada sonido en una lengua. Con estas características se crea un inventario; aunque los sonidos generalmente suelen ser estables en una lengua y se esperaría que fueran uniformes, en realidad sufren variaciones de pronunciación. Por ejemplo, la palabra *pues* puede pronunciarse con /s/ sonora, si antecede a una palabra que inicie con vocal, por lo que es importante que esos modelos se obtengan de grandes muestras de habla que manifiesten muchas posibles variaciones.

De igual forma, los diccionarios de pronunciación, aunque puedan en primera instancia construirse por hipótesis con las pronunciaciones canónicas, dependerán enteramente de la pronunciación de los hablantes. Cada persona es capaz de pronunciar una palabra de varias formas en distintas ocasiones, por ejemplo, la misma palabra *pues* puede pronunciarse: *pos*, *pus*, *ps* o *puesn*, por lo que el diccionario debe enriquecerse de tantas pronunciaciones como sea posible emitir. Además, las palabras no se pronuncian de forma aislada en una conversación, por lo

que es necesario ver qué tanto la estructura de las palabras varía en contacto con otras palabras o sonidos. Si se intenta que un reconocedor de habla funcione en ambientes naturales -aunque con un dominio específico- es necesario que su diccionario de pronunciación sea capaz de reconocer, dentro de su inventario léxico, la mayor cantidad de pronunciaciones posibles para que el reconocimiento sea más exacto. De aquí la necesidad de realizar estudios empíricos de habla que nos muestren el comportamiento fonético de un sector de la población que sirva de base para el robustecimiento del diccionario de pronunciación y para los modelos que utiliza la conversión de texto a habla. Más adelante especificaré el tipo de muestra que fue analizada en esta investigación (§3.1).

Una diferencia sustancial entre los procesos de síntesis y reconocimiento de habla radica en la complejidad de las señales orales de cada uno. En la conversión de texto en habla, las características acústicas que determinen la producción del sonido pueden ser extraídas de un solo hablante; es decir, que la voz puede modelarse para que sea infantil, masculina, grave, etc. En el proceso inverso, la conversión de habla a texto, los sistemas tienen que estar capacitados para interpretar una gran variedad de hablantes (Llisterri, 2003), lo cual hace de éste un proceso más complejo, cuyas bases lingüísticas deberán ser más amplias para poder satisfacer las necesidades de los entornos en los que pretenda utilizarse. Esto dependerá de las aplicaciones que tenga cada sistema.

2.1.4 Aplicaciones

Sobre las aplicaciones de las tecnologías del habla, hay que tener en cuenta que estos sistemas pretenden ser utilizados por usuarios específicos en tareas concretas, por lo

que su funcionalidad varía de acuerdo con los objetivos que se marcan para cada sistema (Rodríguez, 2003:131). La dificultad de la interacción entre el ser humano y las computadoras radica en el hecho de que a éstas se les exige un amplio rendimiento sin tener en cuenta sus realizaciones sobre todo en cuanto a la comunicación; Rodríguez (2003:133) explica que “el usuario tampoco acostumbra a tener claros los límites de la cobertura conceptual de la interfaz, y tiende, favorecido por la apariencia humana del proceso comunicativo, a suponer que la máquina es más inteligente de lo que realmente es”.

Los procesos de síntesis y reconocimiento de habla son aplicables principalmente a entornos informativos o a cualquiera que pretenda establecer la comunicación oral entre individuos y máquinas. Actualmente ambos procesos se aplican en entornos privados, en investigación y en sistemas comerciales. Si los sistemas son utilizados por investigadores o empresas a nivel interno serán menos complejos puesto que su entorno será restringido. Un sintetizador o un reconocedor que sean utilizados por pocos usuarios o para tareas específicas son más sencillos que aquéllos que pretenden emplearse en ambientes comerciales, donde la cantidad de usuarios es infinitamente más amplia y por tanto el funcionamiento debe ser más preciso y más complejo.

La síntesis se utiliza desde hace varios años en los sistemas telefónicos de asistencia a clientes, que son capaces de brindar menús e información de servicios y costos. Empresas como Telefónica I+D, en España², han dedicado muchos años de investigación para lograr no sólo que su sistema sea óptimo en reconocimiento y síntesis de habla sino, además, que la voz que generada tenga naturalidad y se adapte

² <http://www.tid.es/presencia/publicaciones/comsid/esp/articulos/vol23/habla/habla.html> (julio, 2007)

dialectalmente a los distintos países de habla hispana y de otras lenguas de España. Ésta y otras empresas utilizan actualmente los portales de voz, desde los cuales se puede acceder a servicios de información de todo tipo -cine, deportes, noticias- y a agendas y correos electrónicos.

Otra de las aplicaciones de la síntesis es la lectura en voz alta de textos en la computadora, como los correos electrónicos y documentos; en Internet existe una gran variedad de páginas que emplean la voz tanto para la simple comodidad del usuario como para tutoriales interactivos, como por ejemplo, la enseñanza de lenguas a distancia.

Un problema grande para la síntesis, que muchas empresas han buscado solucionar, es la naturalidad en las emisiones del habla. Si bien requiere de estudios más profundos, una voz natural traerá mayores beneficios a las compañías. Es bien sabido que las personas rara vez gustan de las voces electrónicas, porque la comunicación oral es una de las formas más estrechas de relacionarse entre los individuos; por tanto, los sistemas que utilicen el proceso de síntesis deben tener especial cuidado en las unidades del habla que aportan más naturalidad; es decir, deben intentar ser lo más parecidas al habla de un lugar particular para que los usuarios se sientan cómodos utilizando los servicios.

Como mencioné arriba, la empresa Telefónica I+D ha buscado llegar a su público no sólo creando un sintetizador de voz de la lengua española, sino adaptando el sintetizador a cada dialecto de los países americanos que disfrutan del servicio, sosteniendo que “el idioma materno suscita fuertes sentimientos en la mayoría de los

usuarios, que ven con desagrado los servicios en los que claramente se nota que son una «traducción directa» de otros similares en otros idiomas o incluso el «acento» de la máquina se percibe como de otro país, aun siendo el mismo idioma. Es, por tanto, una parte importante en la aceptación de muchos servicios automáticos”³.

Para el reconocimiento de habla, las aplicaciones son más y, como vimos, de bases más complejas puesto que los reconocedores deben ser capaces de distinguir entre un sinnúmero de usuarios con pronunciaci3nes distintas, sobre todo en los servicios comerciales de gran extensi3n.

Muchas aplicaciones se tienen tambi3n en la telefonía celular, en la que se puede hacer marcaciones autom3ticas de n3meros telef3nicos con tan s3lo mencionar el nombre de la persona a la que se quiere contactar. Los sistemas de ventas de boletos o de informaci3n de men3s que son capaces de redirigir a las personas hacia submen3s o de aceptar n3meros de tarjetas de cr3dito, de tel3fonos, nombres, etc. En cuanto a las aplicaciones en la computadora, la m3s utilizada probablemente sea la de dictado que genera textos o que realiza acciones.

Existen adem3s sistemas de di3logo que emplean ambos procesos para su funcionamiento y estos son m3s complejos puesto que deben adaptarse a la voz de cada usuario y responder a solicitudes y3rdenes. Estos sistemas permiten la realizaci3n de transacciones vía telef3nica (Martí *et al*, 2002) o funcionan para la enseñaanza asistida de lenguas.

³ www.tid.es/documentos/productos_servicios/2100_01.pdf. (julio, 2007)

Los investigadores que desarrollan estas tecnologías se han esforzado además por aplicarlas a entornos en los que no sólo son útiles sino indispensables, como es el caso de los usuarios que tienen alguna discapacidad que les impide usar los sistemas de forma tradicional. La síntesis puede ser utilizada por personas impedidas visualmente y también para aquéllas que tienen un trastorno físico, que les impida utilizar los métodos tradicionales de comunicación con una máquina, como el teclado, por ejemplo.

Actualmente y desde hace varios años, profesores de la Universidad de Los Andes (ULA) de Venezuela trabajan arduamente en investigaciones que permitan desarrollar estas tecnologías para el beneficio de personas con minusvalías visuales o auditivas. Al mismo tiempo, estos estudios permiten desarrollar investigaciones sobre entonación y comportamiento del dialecto venezolano, para el cual han ya construido un sintetizador que sustituirá a los que provienen de España, cuyo acento ha impedido su aceptación total por parte de la población local⁴.

En México, universidades de distintos estados trabajan en diversos proyectos de sistemas de reconocimiento, síntesis y sistemas de diálogo. Uno de ellos es el *Proyecto DIME*.

2.2 El Proyecto DIME

⁴ www.saber.ula.ve/.../alexandr/db/ssaber/Edocs/pubelectronicas/investigacion/inum12/avance-001.pdf (julio, 2007)

El *Proyecto DIME*⁵ -Diálogos Inteligentes Multimodales del Español- pertenece al Departamento de Ciencias de la Computación (DCC) del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la Universidad Nacional Autónoma de México. Está a cargo del Dr. Luis A. Pineda Cortés quien, reunido con diversos investigadores y alumnos participantes, fundó el proyecto en 1998 con el objetivo principal de desarrollar sistemas conversacionales multimodales para el español hablado en México en contextos lingüísticos especializados. Estos sistemas conversacionales están basados en modelos de diálogo que necesitan del estudio de situaciones conversacionales, actos de habla y retóricos, intenciones comunicativas, etc. Como objetivo particular del Proyecto se busca construir un reconocedor de habla para el español de México, tarea que ha sido parcialmente satisfecha, pero que precisa aún de algunos ajustes.

Desde su surgimiento se han propuesto y concretado varias investigaciones relacionadas entre sí por su utilidad para el logro de los objetivos del proyecto, pero enfocadas a distintos niveles de análisis según su aplicación; para el campo fonético, por ejemplo, se realizó una investigación que propone la segmentación de los diptongos como una sola unidad fonética (López, 2004). Estas investigaciones son multidisciplinarias, pues el proyecto mismo conjunta profesionales de distintas ramas del conocimiento que contribuyen de forma diferente a las investigaciones.

El grupo además trabaja en conjunto con el Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE) al interior de la República Mexicana, con la Universidad de Rochester, con el Institute for Human and Machine Cognition (IHMC) en los Estados

⁵ <http://leibniz.iimas.unam.mx/~luis/DIME/dimex-index-frame.html> (agosto, 2007)

Unidos y con la Universidad Autònoma de Barcelona, en España. Dichos vínculos se han establecido con el propósito de mostrar los avances del *Proyecto DIME* a investigadores que trabajen estudios similares, además de mantenerse al tanto de los avances que éstos logran y que podrían servir para el proyecto. Algunos de estos investigadores han fungido incluso como miembros activos del proyecto.

El proyecto tiene recursos computacionales y lingüísticos que sirven de base a las investigaciones que se realizan. En la parte computacional, cuenta con diversos reconocedores de habla robustos y capaces de reconocer a varios locutores, así como con el *Proyecto Golem*⁶, la navegación de un robot que responde al español mexicano. Para la parte lingüística se cuenta con dos corpus orales, gramáticas para el español de México y se trabaja actualmente en el desarrollo de una teoría para la interpretación de actos de habla y su inclusión en los sistemas conversacionales.

Los dos corpus orales arriba mencionados son el *Corpus DIMEx100* y el *Corpus DIME*⁷, los cuales sirven como base para investigaciones de todos los niveles lingüísticos que tienen aplicaciones tecnológicas. El *Corpus DIMEx100* es un corpus de habla controlada en la que cien hablantes grabaron sesenta oraciones de extensión variable cada uno. De estas sesenta oraciones, diez eran comunes a todos los hablantes, esto para probables estudios futuros sobre reconocimiento de locutores. Para la construcción de este corpus se seleccionaron oraciones buscando incluir todos los contextos fonéticos posibles en el español de México, para así estudiar las variaciones alofónicas de los mismos. El corpus está siendo transcrito con la herramienta Speech Viewer de la forma en que se mencionó en la introducción. Con

⁶ <http://leibniz.iimas.unam.mx/~luis/golem/> (agosto, 2007)

⁷ <http://leibniz.iimas.unam.mx/~luis> (agosto, 2007)

base en estas transcripciones, se han construido varios reconocedores de habla con los que actualmente cuenta el Proyecto.

2.2.1 El *Corpus DIME* y la importancia del habla espontánea

Abordaré aquí el *Corpus DIME* por separado porque es de éste que se tomó la muestra para esta investigación. El *Corpus DIME* es un corpus de habla espontánea que, como mencioné anteriormente (§1.1), se obtuvo de grabaciones realizadas en un modelo similar al del Mago de Oz. Los diálogos que el corpus contiene están insertos en el dominio del diseño de cocinas, por lo que las situaciones están enfocadas a la resolución de tareas y el vocabulario está restringido a este entorno (Villaseñor *et al*, 2001).

Sin embargo, un corpus oral de habla espontánea, aunque de dominio específico, es crucial para el estudio de las variaciones lingüísticas en un dialecto, puesto que cuenta con características especiales de las que los corpus de habla controlada carecen. Por ejemplo, cuando se lee un texto es menos probable que una persona modifique la pronunciación las palabras o que surjan elementos lingüísticos como las interjecciones, las pausas, las repeticiones o las autocorrecciones; mientras que cuando un hablante puede elegir sus oraciones libremente, estos fenómenos y otros aparecerán sin lugar a dudas, como vemos en el ejemplo de la Figura 6:

Usuario: Bien. Sería... ¡eh! Sería recomendable para el buen uso de la cocina para el (pausa) el buen funcio... funcionamiento

Figura 6. Ejemplo de pronunciación de oración, con pausas, repeticiones e interjecciones (Oración 35 diálogo 22).

Un problema que surge en el *Corpus DIME* es el cuidado que el hablante pone en el habla propia cuando se sabe objeto de estudio de una investigación lingüística, por lo que es necesario desarrollar estrategias que motiven la naturalidad en las oraciones producidas por los hablantes. En un corpus de habla controlada es muy difícil, pero hay estrategias útiles en los de habla espontánea. En el *Corpus DIME*, por ejemplo, los hablantes tenían que realizar tareas de diseño de una cocina, con una pantalla en la que colocaban muebles e indicaban posiciones y movimientos; esto hizo que la preocupación sobre la forma de comunicarse se fuera al segundo plano y que los hablantes se concentraran en la actividad de construcción del modelo, obteniéndose así mayor naturalidad en la comunicación oral.

En la Figura 7 tenemos un ejemplo de la interfaz gráfica del *Corpus DIME*, en la que apreciamos tres ventanas; las dos ventanas superiores indican al usuario el espacio de la cocina y la posición de los muebles en ésta y la ventana inferior muestra los muebles que están disponibles para el diseño. En el diálogo ambos hablantes están observando la pantalla y uno indica a otro el tipo de muebles que desea y la posición de éstos dentro de la cocina, mientras que el otro hablante debe dar indicaciones sobre las opciones, el tamaño de los muebles, además de indicar si la tarea que se le exige es posible y cerciorarse de que ha sido completada según las indicaciones del usuario. Como podemos notar, la atención del usuario se centra en la actividad

gráfica y no en el habla.

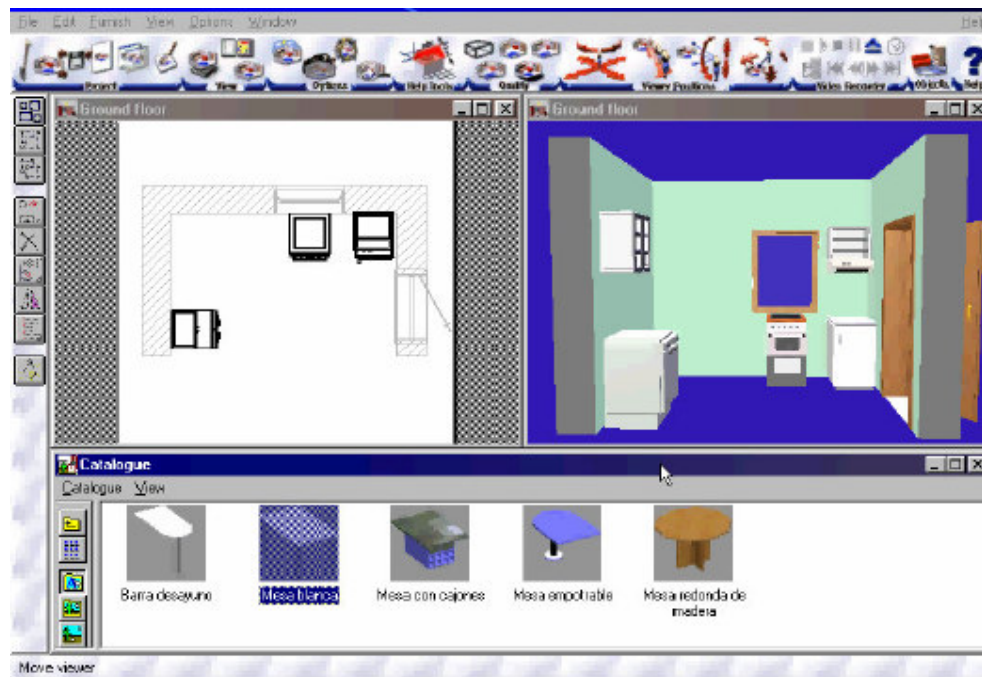


Figura 7. Interfaz gráfica utilizada para el *Corpus DIME*. Panel de diseño de cocina.

Así, se obtuvieron expresiones espontáneas, silencios en que los hablantes pensaban las siguientes acciones, interjecciones de sorpresa, de disgusto, de inconformidad, etc., además de formas distintas de asentir y de negar.

Más adelante en el capítulo cuatro, se demostrará la utilidad que tuvo este corpus en la obtención de fenómenos fonéticos. Por ahora, basta decir que las tecnologías del habla comprenden una rama computacional y una rama lingüística, por lo que deben estudiarse ambas disciplinas si se pretende hacer aplicaciones tecnológicas del lenguaje.

En este capítulo expliqué ampliamente la parte computacional e hice hincapié en la importancia de la participación del lingüista en este tipo de investigaciones; ahora, ya que este estudio pretende mostrar además un estado actual de las modificaciones fonéticas del español de México, el siguiente capítulo abordará la teoría sobre las elisiones en la lengua.

3. Fenómenos de elisión en el español

Las lenguas, como centro de la comunicación entre los seres humanos, son susceptibles de cambio. El uso constante de las formas tiende a desgastarlas y a provocar que se modifiquen. Las lenguas permanecen más o menos estables durante siglos porque son capaces de mantener su estructura básica y adaptar los cambios a ella, esto debido a que es necesario que aseguren su conservación y continuidad. Los cambios en las lenguas son lentos porque las formas que se enfrentan deben coexistir hasta que una desplace a la otra en la mayoría de los hablantes y, así, se generalice; muchas variaciones lingüísticas pueden permanecer por años tratando de desplazar a otra forma y al final desaparecer sin provocar ningún cambio. Los cambios dependen, además, de distintas circunstancias sociales, culturales, temporales y espaciales. Sin embargo, hay pequeñas modificaciones que pueden generar reajustes significativos en las estructuras con el paso del tiempo.

Son dos las tendencias que puede tomar una lengua en cuanto a sus modificaciones: “1) tendencia al debilitamiento por rapidez y práctica comunicativa; y 2) tendencia al reforzamiento para conservar estructuras comunicativas que nos permitan interpretar lo que oímos” (Company, 2001:28). Es decir, que las lenguas van a sufrir modificaciones que las lleven a cambiar sus estructuras al mismo tiempo que buscarán estrategias que las hagan mantener las bases estructurales fijas para lograr la comunicación óptima entre los hablantes. Es posible afirmar que el límite del cambio en las lenguas está determinado por la comprensión entre los hablantes.

Las causas de los cambios son naturales y tienen que ver con la relación que se establece entre los sonidos del habla y los estados de ésta. Una de las teorías que explican mejor las causas de mutaciones en el habla es la que expone Malmberg (1974:172) como la ley del menor esfuerzo. Él afirma que “quand l’homme parle, il essaie, tout comme dans l’exécution de n’importe quel travail, d’obtenir un maximum de résultat avec un minimum d’effort”¹ y es por eso que en el habla, la tarea más llevada a cabo a lo largo del día, se pretende hacer el menor esfuerzo posible, siempre y cuando los mensajes puedan interpretarse correctamente.

Como acabamos de ver, los mensajes son interpretables a pesar de los cambios porque las lenguas reaccionan ante el cambio con el reforzamiento de estructuras. El mismo Malmberg (1974:178) señala que “quand une tendance assimilatrice menace d’éliminer une importante distinction phonologique, la langue réagit en accentuant pour ainsi dire encore la différence entre les deux”², como ejemplo del reforzamiento de las lenguas.

Gili Gaya (1975:180) explica esta compensación afirmando que, si bien las lenguas son un ente cuya naturaleza tiende a la mutabilidad, la variedad de hablas individuales provocaría un verdadero problema en la conversación, por lo que las modificaciones deber ser mínimas y la lengua tiene “que ajustarse a límites estrictos para que los hombres puedan entenderse; las tendencias personales se suman unas a otras para que los signos verbales sean inteligibles, eliminando diferencias y acentuando semejanzas, de modo que las transformaciones tomen un rumbo

¹ “Cuando el hombre habla, trata de obtener, de la misma forma que en la realización de cualquier trabajo, el mayor resultado con el menor esfuerzo”.

² “Cuando una tendencia asimilatoria amenaza con eliminar una distinción fonológica importante, la lengua reacciona acentuando más, por así decir, la diferencia entre las dos”.

colectivo”; esto es, el habla individual puede presentar modificaciones, pero es raro que un cambio lingüístico ocurra puesto que las mutaciones tendrían que seguir un rumbo común, lo cual es bastante improbable, por eso las lenguas varían poco y pocos cambios trascienden, aunque cada momento haya modificaciones casuales.

Por ejemplo, en el dialecto mexicano ha existido durante mucho tiempo la alternancia entre las formas *haya* y *haiga* para el subjuntivo del verbo *haber*; los hablantes han decidido conservar la forma canónica como la más común, pero la forma alternativa nunca ha desaparecido totalmente ni ha disminuido su frecuencia. En cuanto a cambios fonéticos, como vimos anteriormente (§ 2.1.3), la palabra *pues* puede pronunciarse de distintas formas y aun así no suplantar su pronunciación canónica, claro que esto puede tener motivaciones académicas prescriptivas.

Los cambios en las palabras pueden explicarse también debido a la cadena fónica que se forma en el discurso; como apunta la ley del menor esfuerzo, en una conversación las palabras comienzan a perder fuerza -de pronunciación- a medida que se repiten. Bybee (2001) expone varias causas de los cambios en las lenguas y cita a Fowler y Housum (1987), específicamente su prueba de que la segunda vez que una palabra se pronuncia en el discurso suele ser más corta que la primera y presentar relajamiento. La explicación que da a esto es que “for the second occurrence of a word, the speaker can be less explicit phonetically because the context and the fact that the word is already primed make easier for the hearer to access it” (Bybee, 2001:58)³; además de que el oyente es capaz de percibir que una palabra se ha repetido y la entiende a

³ En la segunda ocurrencia de una palabra, el hablante puede ser fonéticamente menos explícito porque el contexto y el hecho de que la palabra fue ya introducida hace más fácil para el oyente el acceder a ella.

pesar de que no sea pronunciada totalmente o de que presente algunas modificaciones fonéticas.

Otra teoría defendida por Bybee, entre otros autores, postula que los cambios fonéticos tales como la pérdida, están muy relacionados con la frecuencia de las palabras o de los fonemas de una palabra en la lengua. Esta teoría, a modo de analogía, advierte que las formas más utilizadas son las que tienden más al desgaste. Una palabra con alta frecuencia de uso es más propensa a presentar modificaciones que aquella que permanece sin usarse por mucho tiempo o que se utiliza sólo en contextos reducidos; en palabras de Bybee (2001:58), “high-frequency words will undergo more reduction”⁴.

A continuación se mencionan los tipos de cambios que sufren las lenguas en general.

3.1 Tipología general de cambios de sonido⁵

Aunque cada lengua encuentra sus vías de cambio y circunstancias, cinco son los cambios generales que puede sufrir cualquier lengua: asimilación, disimilación, pérdida, inserción y metátesis (Lass, 1984).

Las asimilaciones se dan cuando dos sonidos están en contacto directo y en el proceso uno de los dos va a contagiar a otro de sus rasgos acústicos o articulatorios y en algunos casos forzarlo a la modificación total o a la pérdida, que puede darse cuando dos sonidos se han vuelto muy similares y el hablante decide omitir alguno

⁴ Las palabras de alta frecuencia de uso experimentarán más reducción.

⁵ A pesar de que la definición de algunos conceptos básicos pueda parecer innecesaria, incluí dichas explicaciones puesto que el estudio está dirigido también a otras disciplinas, como la computación.

de ellos, este es el llamado fenómeno de elisión o pérdida. Las asimilaciones y las pérdidas se abordarán ampliamente y a lo largo del Capítulo 4 se buscará acentuar la relación entre estos dos fenómenos que fueron, además, los que presentaron mayor frecuencia de aparición en el corpus aquí estudiado.

Sin embargo, a pesar de que las asimilaciones en muchos casos son la causa de la pérdida de sonidos, esta pérdida no siempre ocurre porque un sonido influye en otro; las pérdidas pueden tener causas diversas que a veces pueden incluso no ser fonéticas, sino más bien estar relacionadas con otros niveles de la lengua, tales como las circunstancias en que se lleve a cabo una conversación; esto es, un fonema puede perderse cuando el hablante tiene una pronunciación descuidada o veloz y no articula todos los fonemas óptimamente.

3.1.1 La elisión como fenómeno fonético productivo

Las elisiones ocurren como resultado del relajamiento en la pronunciación de algunos fonemas. Un fonema se relaja, según Martinet (1995), cuando por alguna razón en la lengua comienza a utilizarse más, como en morfología por ejemplo, y los hablantes tienden a hacer que la complejidad del segmento se reduzca. Él afirma que “les changements phonétiques résulteraient d’une tendance à réduire la complexité du phonème au degré qui lui revient du fait d’une fréquence accrue”⁶. Esta teoría responde a la de la ley del menor esfuerzo, en cuanto que las formas más utilizadas se vuelven articulatoriamente más simples para hacer más ágil el discurso.

⁶ Los cambios fonéticos resultarían de una tendencia a reducir la complejidad del fonema al grado que éste proviene del aumento de una frecuencia.

En la evolución de una lengua las elisiones son productivas puesto que facilitan la comunicación entre los hablantes simplificando las formas que dan menos información y reajustando aquellas que vale la pena conservar. Para explicar lo anterior, es necesario revisar lo dicho por Martinet en cuanto a la cantidad de información que posee un fonema y la relación de esta información con su conservación en la lengua. Martinet (1995) pone como ejemplo el caso de dos frases incompletas -para el francés- una con el segmento *act-* y otra con el segmento *at-*. Para la primera señala que las posibilidades de elección se reducen a dos palabras, mientras que para la segunda se extienden hasta treinta; lo que intenta demostrar aquí es que el primer segmento es muy rico en información y que, a pesar de su dificultad articulatoria “la distinction supplémentaire apportée par /k-/ vaut bien le surcroît de travail musculaire qu’il implique”⁷ (Martinet, 1995:140). Así veremos a lo largo del capítulo 4 que algunos contextos fonéticos pueden explicarse por la tendencia a la facilidad articulatoria y otros de alta complejidad tienden a conservarse. Esto se debe a que son los hablantes los que deciden cómo usan la lengua, por lo que a pesar de que un cambio que facilite la articulación parezca lógico, en algunas ocasiones el hablante decide conservar articulaciones complejas. Por ejemplo, en español, algunos de los llamados grupos cultos se pierden naturalmente como en /ekstríkto/ → [estríkto] (§4.2.3.2) pero muchos son conservados sin problemas como en /rápto/.

De los cambios que determinaron el paso del latín al español y que han hecho que éste se modifique a lo largo del tiempo, destaca el de la pérdida. Muchas de estas pérdidas provenían de relajamientos ocasionados por la asimilación o por el desgaste. Los tipos más comunes de elisión fueron las apócope que modificaron la

⁷ La distinción suplementaria aportada por /k-/ vale el esfuerzo del trabajo muscular que implica.

morfología, como la pérdida de la /m/ final en los sustantivos /kastél:um/ > /kastíl:o/; o las sínkopas provocadas por asimilaciones como la sonorización, como en /tépidu/ > /tíbio/.

Se puede concluir, entonces, que la pérdida es un proceso productivo porque modifica los contextos para comodidad articulatoria, por lo que se requiere un menor esfuerzo para producir el habla. Las elisiones adaptan los contextos complejos para comodidad del hablante y es así como el español ha realizado muchos de sus cambios, tanto a nivel diacrónico, como sincrónico.

3.2 Tipos de elisión

Las elisiones se dividen en tres categorías de acuerdo a la posición en la palabra en que se pierden los fonemas. Cuando la pérdida de un fonema o un segmento ocurren al inicio de palabra, la elisión se denomina aféresis. Cuando las pérdidas ocurren al interior de una palabra o de una cadena fónica, la pérdida se denomina síncope. Por último, cuando una pérdida ocurre al final de palabra se le considera apócope.

El habla es un *continuum* acústico en el que las palabras no se pronuncian de forma aislada sino que forman parte de cadenas de fonemas. De tal suerte, la clasificación de aféresis, síncope y apócope puede variar si se analizan las palabras de forma aislada o en una cadena fónica. Por ejemplo, podemos decir que la pérdida de /d/ en /berdád/ > [berdá] es una apócope si la vemos de forma aislada; pero si colocamos la palabra dentro de una cadena fónica como en /de verdad bino/ > [deberdábino]

podemos decir que la pérdida es una síncope.

Todos los sonidos de una lengua son susceptibles de relajamiento y pérdida; las condiciones dependerán de las predilecciones de una lengua en particular. Los fenómenos pueden ser sistemáticos o pueden ocurrir por casualidad. Si bien vimos en § 3.1.1, la elisión ha sido un fenómeno productivo en la historia del español y lo es actualmente también, no hay que perder de vista que sólo unos cuantos procesos llegan a consolidarse en cambio lingüístico; la mayoría de las lenguas pueden sufrir cambios casuales o variación por mucho tiempo antes de que se consideren cambios y los fenómenos pueden perderse antes de esta consolidación. Por ejemplo, en algunos dialectos del español ha prevalecido la alternancia entre el cero fonético y la /s/ final en la conjugación de segunda persona singular en el pretérito: /bíte/ vs. [bístes].

En el presente estudio no se pretende dar evidencia de mutaciones o de variaciones que indiquen algún cambio fundamental pero sí encontrar aquellas modificaciones o pronunciaciones diversas que ayuden a hacer una caracterización general del fenómeno de pérdida en el español de la Ciudad de México, para localizar esas pequeñas variaciones sistemáticas que podrían ser relevantes para mejorar los procesos de síntesis y reconocimiento de habla que son sustentadas por la frecuencia.

3.3 Relajamiento y pérdida en el español de la Ciudad de México

Para localizar los fenómenos es necesario hacer una revisión teórica sobre los

fenómenos de que diversos autores han dado fe en sus investigaciones sobre el español de México.

El español de México, y particularmente el de la Ciudad de México, han sido revisado por muchos investigadores como Henríquez Ureña (1938), Boyd-Bowman (1952), Lope Blanch (1991a), entre otros; de este dialecto se ha logrado hacer una caracterización más o menos uniforme y una clasificación de sus tendencias. En cuanto al carácter fonético, los autores coinciden en que es un dialecto que tiende a la pérdida de vocales y a la conservación de consonantes, aunque ambas pérdidas presentan fenómenos según el contexto en el que se encuentran. Canfield (1981:76) señala que “el resto de los hablantes reconoce a un mexicano por su entonación, por su tendencia a alargar la articulación de /s/ y por una preferencia general por la consonante sobre la vocal”. La predilección de pérdida de vocales sobre consonantes se da con alta frecuencia en los hablantes del altiplano, cuyo nivel de relajamiento vocálico puede provocar una simple elisión o modificaciones sintácticas como las sinalefas; por ejemplo la frase /la eskuéla/ puede transformarse en [leskwéla].

Moreno de Alba (1994) expone, para el español del altiplano mexicano, la teoría sobre que los fenómenos fonéticos están relacionados o conllevan una compensación articulatoria. Esta teoría supone que los dialectos que presenten abundancia de un fenómeno fonético en contextos particulares, mostrarán estabilidad en otros contextos para lograr la eficiencia en la comunicación y el equilibrio de la lengua. Así, para el español del altiplano, caracterizado por el relajamiento de vocales, se espera que muestre consonantes fuertes y poco proclives a la pérdida por relajamiento.

Martín Butragueño (2002) se opone parcialmente a la teoría compensatoria, advirtiendo que la división regional que presenta Moreno de Alba en sus mapas fonéticos no es exacta, pues si se superponen unos a otros es posible percatarse de que algunas zonas que son conocidas por su tendencia al relajamiento vocálico, también presentan relajamiento consonántico. Esto nos indica que tal vez la teoría compensatoria funciona en parte, y que los fenómenos pueden tener más de una causa. En esta investigación se buscarán contextos de pérdida en general, sin esperar que un contexto de conservación conlleve otro de pérdida o viceversa.

En general, los autores concuerdan en que son pocos los casos en los que un fonema se pierde completamente, pero todos los fonemas atraviesan por una serie de procesos de relajamiento previos en el habla. Moreno de Alba (1994:33) aclara, en cuanto a la frecuencia de pérdida, que “son menos los hablantes en los que el fenómeno es relativamente acusado (42.2%) que aquellos en los que sólo se da ocasionalmente o con poca frecuencia (57.8%)”.

3.3.1 Vocales

Sobre el comportamiento vocálico del español, Navarro Tomás (1918) señala que el relajamiento vocálico no depende de la posición silábica con respecto al acento; sin embargo, distingue entre las posiciones en las que generalmente una vocal es susceptible de relajamiento, las vocales débiles al final de grupo y ante pausa son muy relajadas; los ejemplos que señala son los siguientes: “espuma-**espúmã**, nueve-**nwéBë**, lunes-**lúnës**, hermanos-**ermánös**” (p.45); indica además que la vocal

penúltima de las palabras esdrújulas suele ser relajada también. Veremos más adelante (§4.2.1 y 4.2.2.1.) que de la muestra que se analizó, sólo uno o dos fenómenos relacionados con la pérdida de vocales mostraron un patrón sistemático en estos contextos; la pérdida de vocales no fue tan extensa como la de las consonantes y sí mucho menos sistemática. Por lo tanto, se ve apoyada la teoría de Navarro (1918) que afirma que el relajamiento o pérdida de una vocal depende del entorno consonántico que la envuelva y no de su posición en la palabra.

En cuanto a la pérdida de vocales en el español de México particularmente, en un principio se consideraba que estaba relacionada con la aparición de /s/ y que casi estaba reducida a este contexto de /s/ + vocal sin importar el orden (Quilis, 1979, *apúd* Boyd-Bowman,1952); sin embargo Canellada y Zamora Vicente (1965) sostienen que son muchos más los contextos y que la tonicidad no es indispensable porque se pierden tanto vocales átonas como tónicas.

Moreno de Alba (1994) afirma que la vocal más relajada y, por consecuencia, la que tiende más a la pérdida es la /e/, seguida de la /o/; la vocal /e/ es además la más frecuente⁸ en el español según diversos estudios de frecuencias (ver §2.3.1). Canfield (1981) propone que la /e/ pronunciada en México es bastante distinta a la peninsular en cuanto a que la primera es más cerrada que la segunda y a eso puede deberse su relajamiento.

En cuanto a la posición, también se puede decir que la vocal postónica es la que tiende más al relajamiento y pérdida que las posiciones nucleares o iniciales. En un

⁸ La frecuencia de los fonemas en el español de México está basada en los estudios de Cuétara, 2004.

estudio extenso sobre las vocales del español de México, Lope Blanch (1991a) distinguió cuatro niveles de relajamiento que van desde uno muy ligero hasta la pérdida total, que se da en sólo pocos casos. Lope Blanch coincide además en que la /e/ es la vocal que más se relaja por ser la más frecuente y señala que es importante tener en cuenta estos cuatro niveles de relajamiento para no cometer el error de afirmar que una vocal se ha perdido cuando en realidad sólo ha sufrido una disminución acústica. Asegura que todos los relajamientos y pérdidas -que son las menos frecuentes- están determinados por el contexto consonántico. Canfield (1981) menciona que la vocal /e/ es la más inestable y la que presenta más modificaciones. Henríquez Ureña (1938) señala que las vocales tienden al timbre cerrado y, junto con el resto de los autores, afirma que la posición en la sílaba y en la palabra influye directamente en el grado de relajamiento; el carácter átono de las vocales facilita su debilitamiento y pérdida, puesto que se hará más énfasis en pronunciar una vocal tónica y el esfuerzo disminuirá en las átonas. Por ejemplo, en la palabra *rápido* la vocal /a/ será pronunciada plenamente y el esfuerzo se irá reduciendo en /i/ y en /o/, haciendo probable que esta última se relaje considerablemente.

Como vemos, los autores no coinciden en las causas del relajamiento vocálico, puesto que se debate siempre si la causa principal es la tonicidad, la posición en la palabra o el entorno consonántico que rodea a las vocales; en los § 4.2.2.1 y 4.2.3.4 notaremos que las causas pueden ser varias y que, de las anteriormente mencionadas, más de una puede ocurrir a la vez.

3.3.2 Consonantes

En cuanto a las consonantes, Perissinotto (1975) afirma que las pérdidas de algunas de éstas tienen que ver con la predilección del español por la sílaba abierta. Así las consonantes en posición final de sílaba y de palabra tienen más posibilidades de relajarse y desaparecer que aquellas en posición inicial. El resto de las modificaciones consonánticas sistemáticas están relacionadas con procesos de asimilación, ya sea parcial, como en el caso de una nasalización o una velarización, o total, como en casos en los que una de las dos consonantes en contacto desaparece. La pérdida de consonantes en los llamados grupos cultos es una constante en la lengua (Henríquez Ureña, 1938), por ejemplo en la palabra *substancia* que se redujo a [sustánsja] como norma.

Muchas de las pérdidas en el habla tienen que ver con el límite silábico, para consonantes y vocales por igual. Sin embargo, para la pérdida de consonantes, los cambios fonéticos -y en especial las elisiones- suelen ser síncopas; es decir, pérdidas al interior de palabra. Las síncopas, como se sabe, están precedidas por procesos de relajamiento y asimilaciones parciales. Ya que el fenómeno de elisión determinó, en buena medida, el desarrollo de la lengua española es, por tanto, un proceso altamente recurrente en esta lengua. Las causas son muchas y no necesariamente fonéticas; en algunos casos la pérdida está condicionada por la rapidez o por un idiolecto, pero en la presente investigación se mostrarán las sistematizaciones que mostraron los procesos de síncopa encontrados en el *Corpus DIME*.

Para el español de México, las posiciones que más arrojan cambios son la intervocálica, la implosiva y la final absoluta. Las intervocálicas, como dice Navarro Tomás (1918), son contextos altamente asimilables puesto que están rodeadas de

sonoridad y además pueden aparecer varias de éstas en una sola palabra, lo cual facilita que después de la primera emisión las pronunciaciones cambien y se vuelvan más relajadas. Por ejemplo, en la palabra *abogado* aparecen tres intervocálicas rodeadas de vocales fuertes, lo cual puede provocar que alguna de estas consonantes se relaje considerablemente al recibir una gran cantidad de sonoridad.

De entre los fenómenos de pérdida de consonante intervocálica, se encuentra aquella de la terminación *-ado*, que se convierte en *-ao*, cuyo uso se ha extendido en algunos países y se utiliza más que la forma convencional (Navarro Tomás, 1918). Boyd-Bowman, hablando de la pérdida de /d/ intervocálica pone como ejemplos las palabras *todo* y *nada*, así como algunas frases construidas con estas palabras y argumenta que la consonante /d/ sólo se reduce a cero cuando se encuentra en palabras tan comunes como éstas o en las expresiones que forman, porque el uso es altamente frecuente y el esfuerzo perceptivo es menor; él da como ejemplo las formas ‘nada más > *namás*, todo el día > *toil día*’. Con esto sustenta que es raro que se pierdan sílabas enteras y que lo más común es que se relaje o se pierda la consonante intermedia quedando dos vocales juntas que se pronuncian, incluso si son la misma: “las formas más usadas son *na^da*, *to^do*, *naa*, *too*, y las enfáticas *nada*, *todo*, con /d/ interdental larga y fuertemente rehilada” (Boyd-Bowman, 1960:47).

Henríquez Ureña (1938) reduce la pérdida de intervocálicas a la pronunciación rápida de la terminación *-ado*, y ocasionalmente de /r/, /n/, /g/.

Otros contextos que suelen ser proclives a la pérdida de consonantes son la posición implosiva y la posición final. Este fenómeno puede estar relacionado con la

formación de sílabas en español y en las lenguas, en general. Tanto Perissinotto (1975) como Quilis (1979) señalan la predilección del español por la formación de sílabas abiertas, lo cual explica el por qué la posición implosiva o final de las consonantes es articulatoriamente más difícil que las iniciales. Según las cifras de Quilis, el contraste entre sílaba abierta y sílaba trabada es 55% contra 21% respectivamente.

En cuanto a las consonantes que individualmente tienden a la mutación, tenemos que la /s/ es terreno fértil de modificaciones. La /s/ es la consonante más frecuente del español y tiene una pronunciación particular en el español de México. Canfield (1981) señala que el alargamiento de la /s/ es una de las características más notables del dialecto. La /s/ es una alveolar fricativa sorda que tiene un timbre muy agudo y que provoca cambios tanto en consonantes como en vocales y que es susceptible de mutar en algunos contextos.

Algunos autores consideran que la pérdida de vocales está ampliamente relacionada con la presencia de /s/ y esto puede deberse a sus características acústicas. Esta consonante posee una resonancia especial y un timbre agudo, pero sus formantes no son claros; puede, sin embargo, ser núcleo silábico en lenguas similares al español, como italiano o francés, y en algunos casos -como veremos más adelante (§ 4.2.1)- se le puede ver realizando la misma función en español.

Otro contexto de pérdida de consonantes es el de los grupos cultos y los grupos de consonantes, tales como *ns*, *kt*, etc., que son de articulación compleja pero que en algunos casos se conservan bastante bien, ya que son segmentos son ricos en

información para la distinción de las palabras, como expliqué anteriormente en §3.1.1. Esto tiene, además, un origen histórico más que fonético: es decir, la inclusión de estos grupos cultos se dio de forma artificial y no siguió la evolución natural de la lengua. En español, los grupos cultos ingresaron cuando se pretendía tener una escritura más latinizante y después se volvieron norma. La realidad es que la mayoría de los grupos se ha conservado bastante bien a pesar de su dificultad, pero algunos han buscado su forma natural y se han simplificado; en el dialecto peninsular, incluso, son admitidas ambas pronunciaciones como norma, por ejemplo, la alternancia de [eksáкто] y [esáкто].

He presentado aquí los distintos tipos de elisión presentes en el español. En el siguiente capítulo compararé la teoría aquí expuesta con los resultados obtenidos del análisis de datos.

4. Fenómenos fonéticos del español de México en el *Corpus DIME*

Como se mencionó en la Introducción, el objetivo de esta investigación es hacer una clasificación general de los fenómenos de elisión en el español de la Ciudad de México para su posible aplicación tecnológica en sintetizadores y reconocedores de habla. Para esto fue necesario realizar un análisis lingüístico a una muestra de habla, proceso cuyos resultados se explicarán ampliamente a continuación.

La muestra de habla que sirvió de base para el análisis fue extraída del *Corpus DIME* y consta de cinco diálogos¹ correspondientes a siete hablantes, cinco hombres y dos mujeres; uno de los hablantes participó en cuatro de los cinco diálogos, por lo que de él se tiene una mayor cantidad de ejemplos. Cada dialogo contenía entre 92 y 210 oraciones de extensión variable.

A pesar de que esta muestra parece limitada por el número de hablantes, debe tomarse en cuenta que, para la construcción del *Corpus DIME*, los hablantes seleccionados cubrían características que les permitían ser representativos del español de México y tuvieron el perfil del usuario final del producto; algunas características que se tomaron en cuenta fueron edad, género y grado de instrucción, entre otros. En cuanto al género, hay un balance entre el número de hombres y de mujeres; y en cuanto al grado de instrucción, los hablantes eran en todos estudiantes de licenciatura o de posgrado.

¹ Los diálogos corresponden a los archivos: d08, d10, d20, d22 y d26.

4.1 Resultados del análisis primario

Para el análisis del corpus obtenido se partió de la premisa de que los fenómenos más frecuentes se transformarían en reglas fonéticas que fueran aplicables a sistemas informáticos; con esto en mente, se estudió el corpus sin buscar fenómenos en particular, pero conforme se avanzó en la clasificación de los fenómenos que aparecían, se fueron creando las siguientes categorías: apertura y cierre vocálicos, pérdidas, silabificación, sonorización y ensordecimiento. Los cambios fueron examinados de acuerdo con su frecuencia y con los hablantes. Los resultados, al concluir dicha parte del análisis, se presentan en la Figura 8.

Fenómeno	Número de apariciones	Porcentaje
Elisión	314/451	70%
Cierre vocálico	36/451	8%
Apertura vocálica	29/451	6%
Monoptongación (sin elisión)	18/451	4%
Otros, como metátesis	54/451	12%
Total	451	100%

Figura 8. Tabla de porcentajes de fenómenos fonéticos en el *Corpus DIME*.

Es importante señalar que, para obtener las reglas fonéticas que se pretendía, era necesario no sólo que los fenómenos fueran de una frecuencia considerable sino que también presentaran sistematización, por lo que, a pesar de que muchos fueron los fenómenos que se encontraron, fue necesario descartar aquellos cuyo

comportamiento fue difícil de generalizar, debido a que ocurrían en pocas ocasiones o no compartían patrones de comportamiento.

Así, los fenómenos de elisión y silabificación eran los únicos que tenían, además de alta frecuencia, una gran variedad de hablantes, por lo que sólo consideré relevante extender el análisis en estos dos fenómenos.

Al estudiar más profundamente los ejemplos de silabificación, fue posible percatarse de que la mayoría de éstos mostraban que la silabificación en una cadena de palabras se veía modificada por procesos de elisión, por lo que ambas categorías se fusionaron y la silabificación se convirtió en unas de las categorías de elisión. En este estudio se aborda ligeramente el proceso de silabificación (§4.2.3.5), sólo para mostrar las primeras etapas del análisis y demostrar su relación con los fenómenos de pérdida; sin embargo, no se muestran porcentajes ni se ahonda en esta categoría, ya que una investigación a fondo sobre la silabificación está en proceso². Así, en la presente investigación explicaré sólo los procesos de elisión que no están relacionados con silabificación.

Al finalizar el conteo y la clasificación, decidí trabajar exclusivamente con los fenómenos de pérdida encontrados; una revisión nueva del corpus resultó en la eliminación de algunas oraciones que, por su extensión o contenido, no aportaban contextos de pérdida ni contextos proclives a dicho fenómeno, por lo que al final el

² Alejandra Espinoza realiza actualmente una investigación sobre el fenómeno de silabificación en el *Corpus DIME* como tesis de la licenciatura en lengua y literaturas hispánicas, de la Facultad de Filosofía y Letras de la UNAM (2007). Los resultados de dicha tesis, al igual que este trabajo, podrían aplicarse en el *Proyecto DIME*.

corpus se redujo a 671 oraciones, de las cuales sólo 451 presentaron una o más pérdidas y el resto se utilizó para establecer un contraste.

Aunque se esperaba encontrar pérdidas de fonemas vocálicos y consonánticos por igual, en el corpus estudiado las vocales tienden a la conservación en la mayoría de los casos, excepto en silabificación, y son las consonantes las que tienen mayor incidencia. Si recordamos lo dicho por Perissinoto (1975) sobre el español de México, observamos que sus conclusiones mostraban que las consonantes eran muy estables incluso en final de sílaba y que, por el contrario, eran las vocales las que se debilitaban comúnmente; sin embargo, el corpus no reflejó este comportamiento. Por supuesto, no se puede concluir que los fenómenos de pérdida hayan cambiado radicalmente, pero la muestra da cuenta de que algunos fonemas consonánticos no son tan estables en el dialecto mexicano, como se mostrará a lo largo de este capítulo, cuando se aborden los contextos específicos de pérdida de consonantes.

4.2 Fenómenos de elisión en el Corpus DIME

Al llegar a este punto de la investigación realicé un conteo que incluía las pérdidas y también los contextos idénticos que no sufrieron modificaciones para así obtener el índice porcentual de los fenómenos. Asimismo, los fenómenos fueron clasificados teniendo como base la tipología general de cambios de sonido y en el proceso nuevas categorías fueron surgiendo dentro de los grupos. Las tres categorías principales son aféresis, síncope y apócope y los porcentajes de frecuencia se encuentran en la Figura 9.

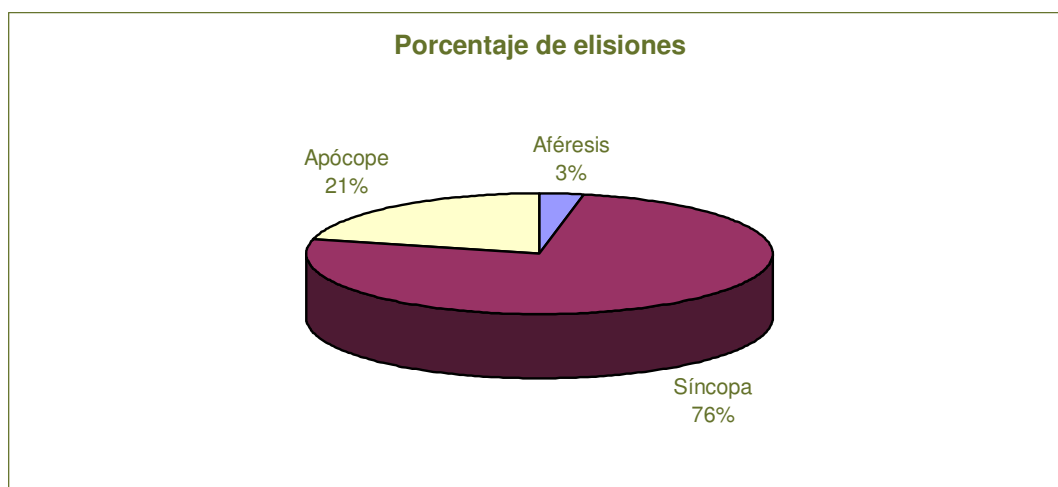


Figura 9. Gráfica de porcentajes de elisiones en el *Corpus DIME*.

Cabe mencionar que, al descartar el fenómeno de silabificación, los criterios de elisiones cambiaron de pérdidas en cadenas fónicas a pérdidas en palabras; esto es, consideraremos que aféresis es la omisión de un fonema inicial absoluto, síncopa es sólo al interior de palabra y apócope se considera igual si es final absoluto o de palabra, siempre y cuando no tenga que ver con fenómenos de reajuste silábico.

En la gráfica anterior podemos percatarnos de que el fenómeno de aféresis es mínimo, al menos en la muestra estudiada y que, por el contrario, el apócope y la síncopa tiene una aparición más frecuente, siendo la última la más abundante.

A continuación se abordará cada una de las categorías para mostrar cuáles son más frecuentes y constantes y cuáles son los motivos por los que algunos contextos sufren elisiones frecuentes. Ya que esta es una investigación sobre pérdida de fonemas, las imágenes que se muestren serán aquellas en que los sonidos se hayan perdido, pero el

Apéndice 1 muestra la imagen clara de los cada uno de los fonemas cuya pérdida se expondrá a continuación.

4.2.1 Aféresis

Las aféresis aparecieron solamente en poco más de 3% del corpus. La posición inicial es generalmente estable en los fonemas y es raro que algún fonema en esta posición se relaje y, mucho más raro, que se pierda. Sin embargo, algunos contextos permiten la reducción inicial. La vocal /e/ fue la que presentó mayor pérdida, sobre todo cuando aparecía en la preposición *en* seguida de artículo, como observamos en la Figura 10. Se perdió en 80% de los casos, lo cuál puede indicarnos que algo está ocurriendo en ese contexto; las causas pueden ser fonéticas o léxicas, pues la /e/ es la vocal más frecuente en español,³ y la preposición /en/ es, de igual forma, altamente frecuente en la lengua. La teoría fonética supone que la /n/ asimila a la /e/ y que la desaparece porque /n/ es capaz de mantenerse como núcleo silábico, por su calidad de nasal; en este sentido Gil Fernández (1990) menciona que las nasales y las líquidas poseen características vocálicas y esto les permite ser núcleo en algunas lenguas. Además, como teoría léxica puede considerarse que, además de ser asimilada, la palabra en la que aparecen es la misma y es una palabra de uso altamente frecuente en la lengua, por lo que su desgaste es mayor y el esfuerzo por pronunciarla es menor (Bybee, 2001).

³ Datos basados en el recuento que hace Cuétara (2004) de los diversos estudios de frecuencia aplicados al español de México y en Emilio Pérez Hernán (2003) para el español en general.

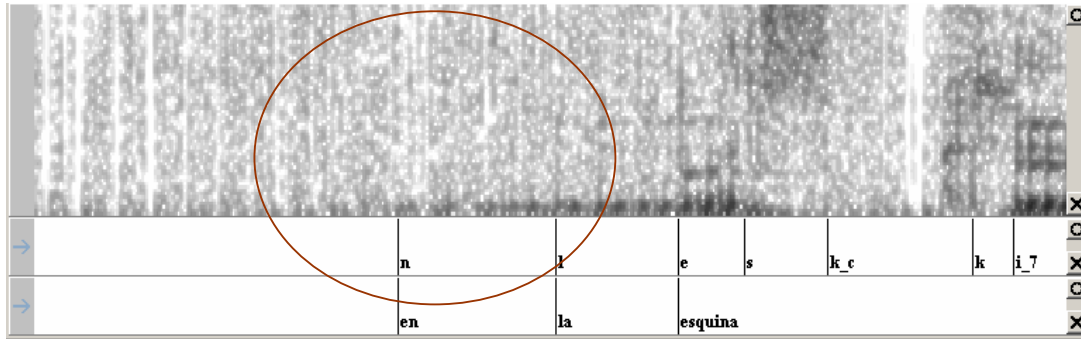


Figura 10. Aféresis de /e/ ante /n/ en la frase *en la esquina* (Diálogo 10, oración 42, hablante 2).

La /e/ sufre aféresis también ante /s/, un fonema con características alveolares y fricativas. La /s/, como vimos anteriormente (§2.3.2) tiene un timbre muy agudo y fácilmente puede funcionar como núcleo silábico. El contexto de /s/ inicial más consonante no es raro en otras lenguas, emparentadas o no, con el español directamente. En la Figura 11 podemos observar que la oración inicia con las sombras características de /s/ y no es posible localizar la presencia de la /e/ inicial.

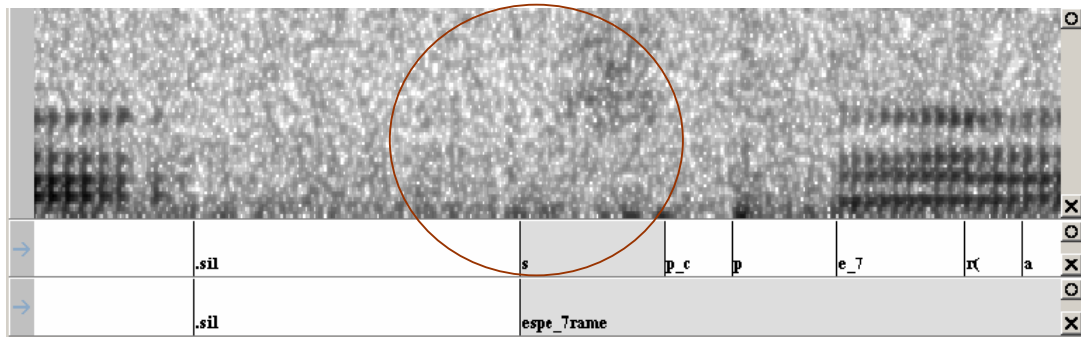


Figura 11. Aféresis de /e/ ante /s/ en la palabra *espérame* (Diálogo 8, oración 151, Mago 1).

4.2.2 Apócope

El fenómeno de apócope tuvo mayor incidencia en las muestras del corpus de esta

investigación, y se dio lo mismo en vocales que en consonantes.

4.2.2.1 Apócope de vocales

En cuanto a las vocales encontramos que fueron las medias las que sufrieron apócope; la que más sufre pérdida es la /o/ en contextos postónicos. Como vemos en el ejemplo de la Figura 12.

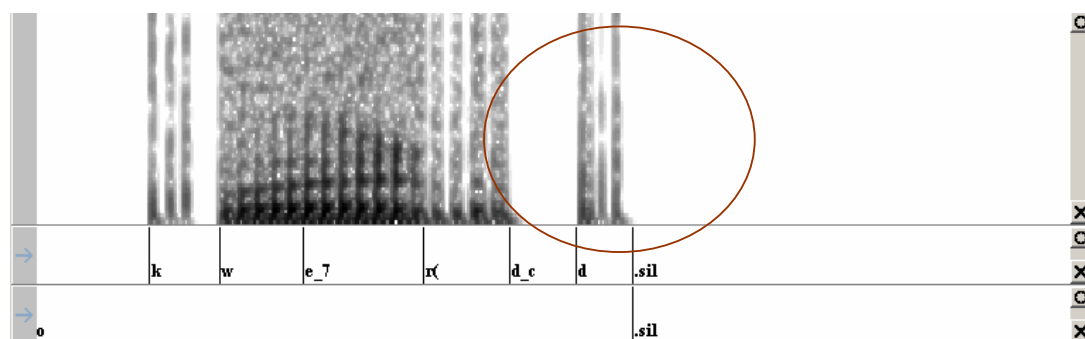


Figura 12. Apócope de /o/ en la palabra *acuerdo*
(Diálogo 20, oración 3, Mago 2).

La /o/ postónica se elimina cuando la frase va perdiendo fuerza hacia el final. En la Figura 12 notamos que la frase concluye con la explosión de la /d/ oclusiva y que el silencio se prolonga por varios milisegundos, por lo que notamos claramente que la /o/ final ha sido omitida.

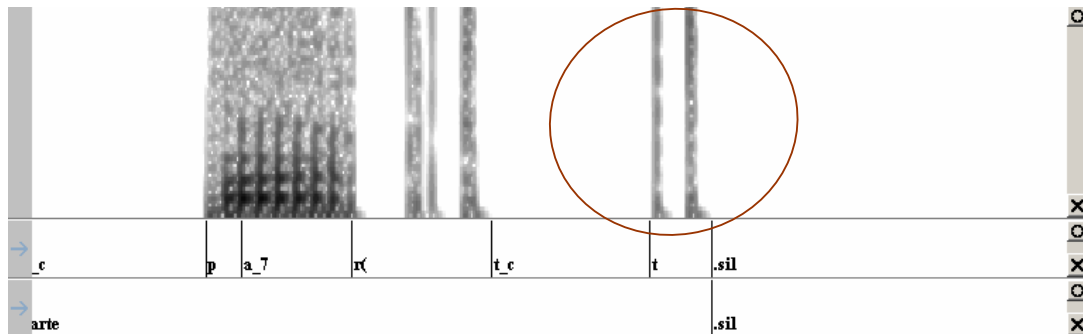


Figura 13. Apócope de /e/ en la palabra *parte*
(Diálogo 20, oración 76, Mago 2).

De la misma manera, la /e/ sufre apócope cuando es postónica y la entonación de la oración va decreciendo. En la Figura 13, al igual que ocurrió en el ejemplo de la Figura 12, la frase termina con la explosión de la consonante oclusiva y no hay evidencia de la presencia de alguna vocal final.

Cualquier pérdida de vocales entonces, está relacionada a la tonicidad y a su posición en la palabra, según nos muestran los ejemplos de aféresis y de apócope de vocales, aunque, como señala Lope Blanch (1991a), a pesar de que las condiciones tónicas estén dadas, una vocal sólo se pierde si su entorno consonántico lo permite.

Otra característica que puede influir en la pérdida de las vocales es la velocidad del habla; Llisterri (1991:81) señala: “a velocidades rápidas de elocución, las vocales no llegan a alcanzar las características propias de una vocal «plena»”.

4.2.2.2 Apócope de consonantes

La pérdida de consonantes finales puede tener los mismos motivos que el proceso de

silabificación que se mencionará más adelante (§3.2.3.5); las consonantes a final de sílaba suelen ser articulatoriamente más complejas y presentar variaciones tales como ensordecimiento o sonorización. Las consonantes que presentaron apócope fueron /d/, /n/, /r/, y /s/, consonantes, además, de alta frecuencia en la lengua, por lo que este cambio puede vincularse con la teoría del desgaste de formas y de menor esfuerzo (Martinet, 1995).

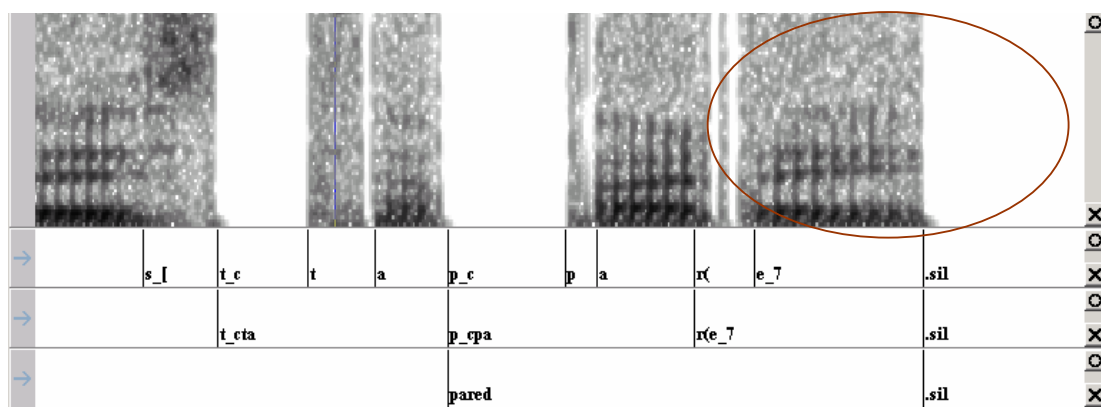


Figura 14. Síncopa de /d/ final de palabra en la palabra *pared* (Diálogo 20, oración 88, Mago 2).

En la Figura 14 encontramos que la /d/ de la palabra *pared* se pierde incluso en contexto tónico. Al final de la imagen espectrográfica notamos la /e/ de manera clara y que muestra mayor duración por ser tónica pero que va seguida de un silencio (.sil) con el que se concluye la frase; la /d/ final se perdió en 21% de los casos.

Al respecto de la pérdida de /d/ en México, Matluck (1951) señala que hay diversos comportamientos para este fonema en final absoluto, pero que el más común es la elisión o debilitamiento casi total. Dice, además, que esto suele ocurrir en palabras muy comunes como son: *verdad*, *usted*, *ciudad*, etc. La pérdida de este fonema se

debe a que naturalmente es de una pronunciación complicada; en palabras de Navarro Tomás (1918:103):

La *d* final absoluta, seguida de pausa, se pronuncia particularmente débil y relajada: la punta de la lengua toca perezosamente el borde de los incisivos superiores. Las vibraciones laríngeas cesan casi al mismo tiempo que se forma el contacto linguodental, y además, la corriente aspirada, preparando la pausa siguiente, suele ser tan tenue, que de hecho la articulación resulta casi muda.

Navarro Tomás (1918) señala que para esa época ya era común en toda España la pérdida de /d/ final absoluta, incluso en las personas letradas o ilustres; es decir, que no era sólo un rasgo particular de un sociolecto, y, como vemos, este fenómeno aparece también actualmente en el dialecto mexicano.

La /n/ sufrió se apocopa en el 11.5% de los casos en que este contexto se dio. La /n/ puede perderse al nasalizar la vocal que acompaña. En algunos casos, la /n/ puede aparecer mínimamente, pero en la Figura 15 notamos que los formantes de la /e/ se expanden hasta el final de la oración y que posteriormente hay un silencio largo que evidencia la pérdida de la /n/; esto ocurre también en contexto tónico de palabra y de oración.

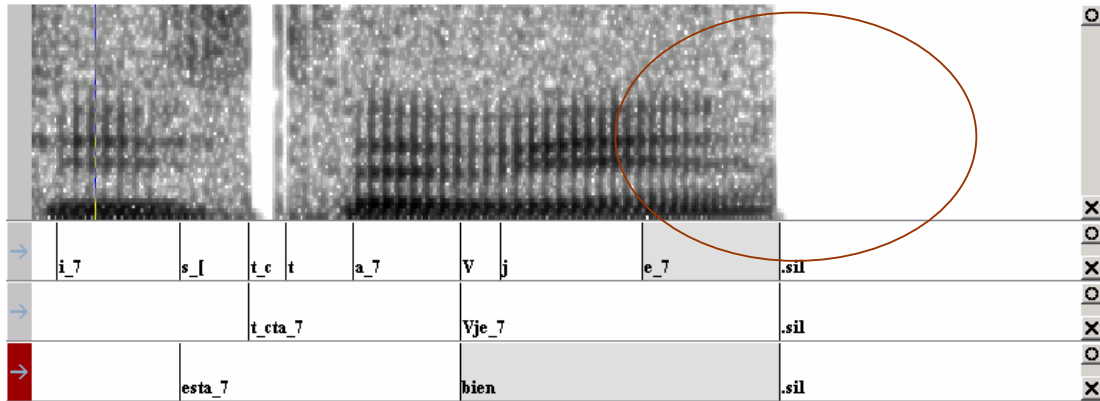


Figura 15. Apócope de /n/ en la palabra *bien*
(Diálogo 22, oración 155, hablante 4).

Navarro Tomás (1918) sostiene que la /n/ final ante pausa suele tener una pronunciación muy debilitada porque la lengua queda pegada a los alvéolos durante la presión de aire y la vibración de las vocales, lo cual la ensordece hasta hacerla casi imperceptible.

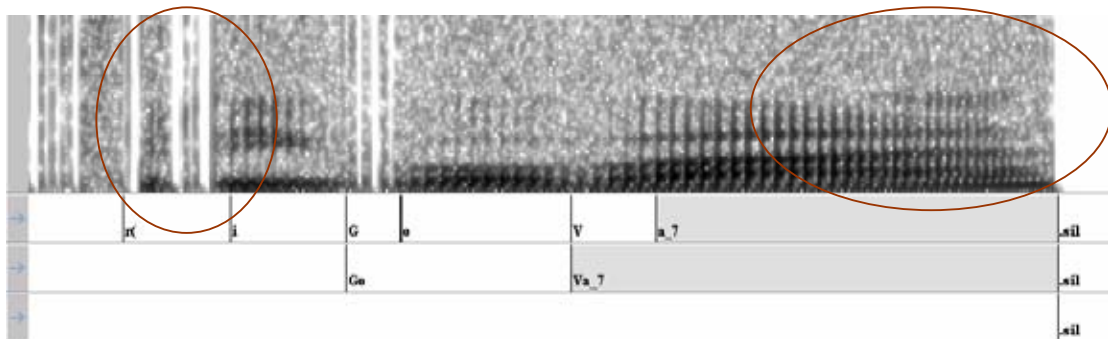


Figura 16. Apócope de /r/ en la palabra *frigobar*
(Diálogo 20, oración 17, Mago 2).

En palabras como *frigobar* y *refrigerador* (esta última de uso muy frecuente en el *Corpus DIME*, por el tipo de tema), la /r/ final se perdió en sólo el 8.1% de los casos.

Como lo comprobaremos más adelante (§ 4.2.3.1 y § 4.2.3.2), la /r/ es una consonante muy estable que tiende poco hacia el debilitamiento; este fonema enfrenta sus propios fenómenos de asibilación o reforzamiento. En la Figura 16, en el círculo de la derecha, notamos que hacia el final de la imagen los formantes de la /a/ son muy marcados y no es posible apreciar rastros de /r/. De nuevo, nos encontramos ante un silencio provocado por apócope en contexto tónico. Esto queda más claro si miramos el círculo pequeño (izquierda) que señala la aparición de una /r/ (transcrita como /r(/); al final de palabra no es posible encontrar una /r/.

Otra consonante que es muy estable en los hablantes del corpus estudiado es la /s/. No presentó muchos contextos de pérdida; sólo se perdió en posición final de palabra o de oración en 3% de los casos.

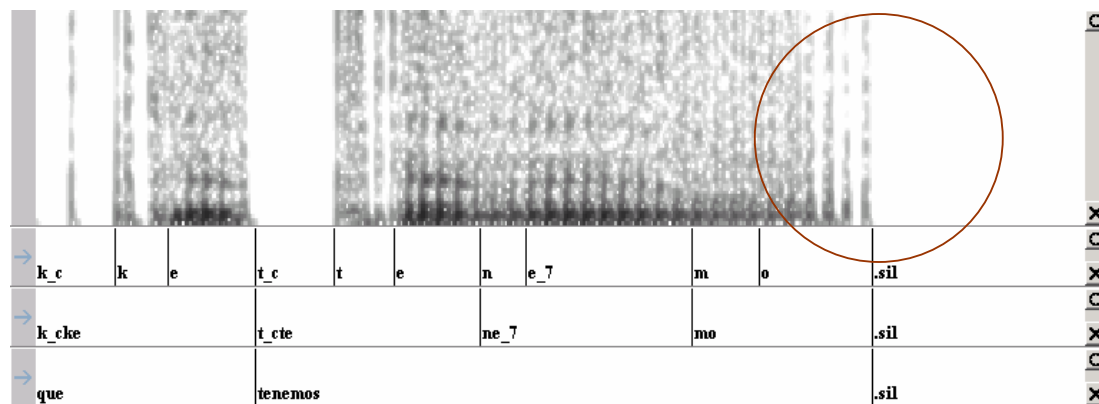


Figura 17. Apócope de /s/ en la palabra *tenemos* (Diálogo 26, oración 36, Mago 1).

En la Figura 17 notamos que se ha perdido en un contexto postónico, lo cual además nos hace pensar que probablemente la tonicidad es irrelevante en los apócopies y que las causas pueden ser la entonación o a la articulación compleja de los fonemas consonánticos finales, aunque, cabe decir, la /s/ tiene un papel morfológico principal

al presentarse en todos los plurales de español, por lo que podemos pensar que, en este caso, puede ser que estas modificaciones sean sólo casuales debido a su baja frecuencia. Por lo tanto, este fenómeno no podría ser considerado para la formulación de reglas fonéticas para las tecnologías del habla.

Una mención especial requiere la palabra inglesa *okay* (transcrita como *okey* aquí) que es de las palabras más utilizadas en el corpus. Esta palabra sufre reducción de diptongo en 21% de los casos y se le ha englobado en esta categoría puesto que el proceso asimilativo de la /e/ sobre la /i/ en el diptongo /ei/ siempre se resuelve en la pérdida de la última vocal. El ejemplo puede observarse en la Figura 18

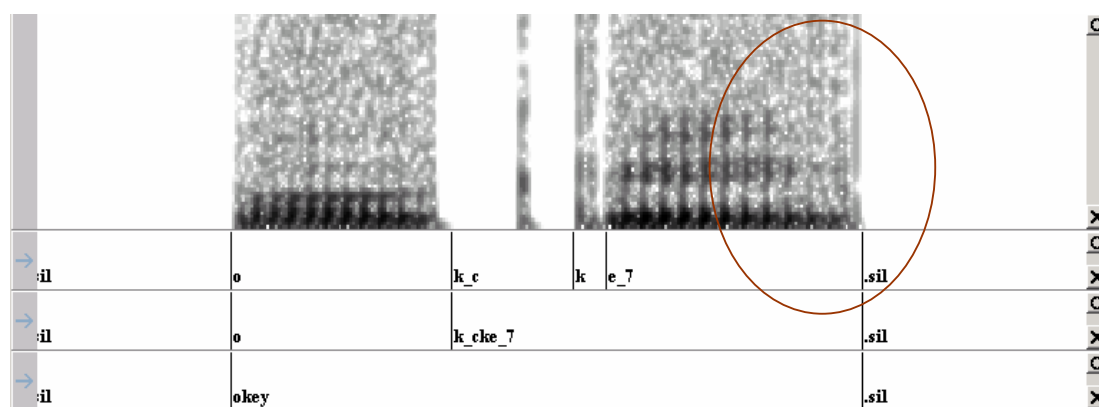


Figura 18. Reducción vocálica en el diptongo [ei] en la palabra *okey* (Diálogo 26, oración 170, hablante 5).

Podemos concluir, tratando de encontrar regularidades, a partir de los fenómenos más frecuentes, que las apócope tienen que ver con el comportamiento entonativo del español, que suele ocasionar que los elementos átonos en las cadenas fónicas pierdan fuerza, se debiliten y se pierdan totalmente en algunas posiciones limítrofes. El proceso ocurrió en vocales y consonantes. Las vocales finales mostraron diversos procesos de debilitamiento y sólo algunos de pérdida; los resultados del corpus

coinciden con lo dicho por Lope Blanch (1991a) en cuanto a que la pérdida se da sólo en algunas ocasiones, pero que las vocales átonas suelen pasar hasta 4 grados de debilitamiento, pues de igual forma, en nuestro corpus se encontró que la pérdida fue poca, pero que ocurría más en vocales postónicas, principalmente en las medias. Para las consonantes, la pérdida está condicionada también por la posición, pero no necesariamente por la tonicidad, pues en algunos casos la consonante se pierde aún si la sílaba final es tónica. Esto tiene que ver, más bien, con la construcción de las sílabas en español. Como lo plantea Perissinoto (1975:68), “es un hecho bastante conocido que el español en general tiene una predilección por las sílabas abiertas”.

4.2.3 Síncopa

El fenómeno de síncopa fue el más prolífico de los tres fenómenos de pérdida. En general, está relacionado con procesos de asimilación total; Bybee (2001:77) afirma con respecto a la relación entre la pérdida y la asimilación que “regular sound change is entirely reductive or assimilation, with the result that articulatory changes all move toward the compression and reduction of articulatory gestures”⁴. En este estudio obtuvimos resultados variados para la síncopa; algunos contextos presentaron una frecuencia de pérdida muy elevada, lo cuál puede, además de ayudar a los procesos de síntesis y reconocimiento de habla, aportar información relevante sobre el comportamiento lingüístico general actual del español de la Ciudad de México.

Los procesos de síncopa del Corpus *DIME* fueron mayormente de consonantes y ocurrieron en dos contextos: 1) en posición intervocálica y 2) en posición implosiva.

⁴ Los cambios de sonido regulares tienden a la reducción o a la asimilación, así que todo cambio articulatorio tiende a la compresión y reducción de los gestos articulatorios.

Las consonantes que presentaron pérdidas frecuentes fueron: /b/, /r/, /d/, /n/, /s/ y /k/. La mayoría de éstas mostraron pérdidas en ambos contextos -excepto la /k/. Para vocales, tenemos que ocurrieron sólo en reducción de diptongo; asimismo, se presentó un tercer contexto que incluye la pérdida de sílabas en algunas palabras (§ 4.2.3.3).

4.2.3.1 Síncopa de consonantes intervocálicas

Las pérdidas de consonantes en posición intervocálica tienen como causa principal la sonoridad que las vocales contagian a la consonante interior. Como se sabe, muchas de las modificaciones que han ocurrido en el español tienen que ver con la sonoridad que reciben las consonantes que les provoca sonorización, si son sordas, y pérdida, si son sonoras. Las consonantes que presentaron el fenómeno en el corpus fueron /b/, /r/ y /d/, todas sonoras y con pronunciaciones variables dependiendo de su posición en la palabra. Consideré para esta categoría los contextos en que /b/ va acompañada de la líquida /l/ y en los que /r/ va acompañando a otra consonante, como en: *amable* y *trabajo*, respectivamente.

La /b/ intervocálica se perdió en 20 % de los casos, la mayoría de las veces entre las vocales /a/ y /o/, como vemos en la Figura 19.

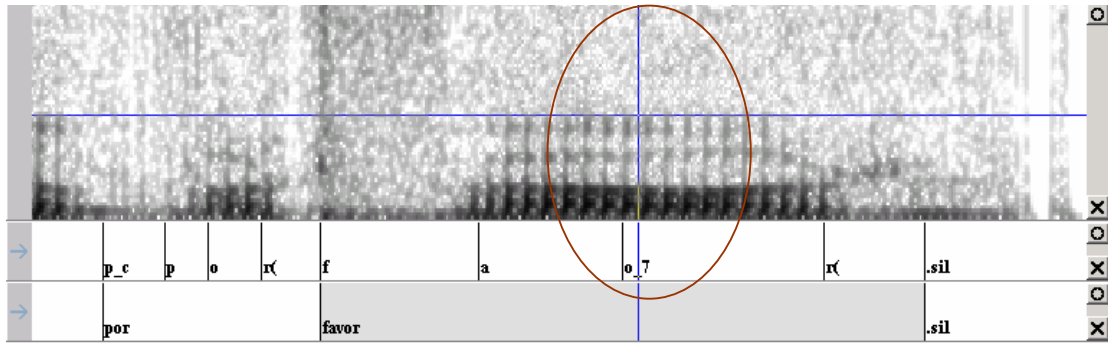


Figura 19. Síncopa de /b/ intervocálica la frase *por favor* (Diálogo 8, oración 20, Mago 1).

Si miramos la imagen notamos que hay un cambio de formantes que van de /a/ a /o/, pero no hay ningún rastro intermedio de consonante. Casos como éste ocurrieron varias veces con la combinación de distintas vocales, incluyendo los contextos en que /b/ forma parte del segmento /bl/ intervocálico.

El segmento /bl/ se perdió en el 18% de los casos con vocales distintas también y siendo casi siempre tónico, como vemos en la Figura 20, en la que es posible observar que entre la /o/ y la /e/ de la palabra *problema*, hay una clara manifestación de /l/ pero no de /b/.

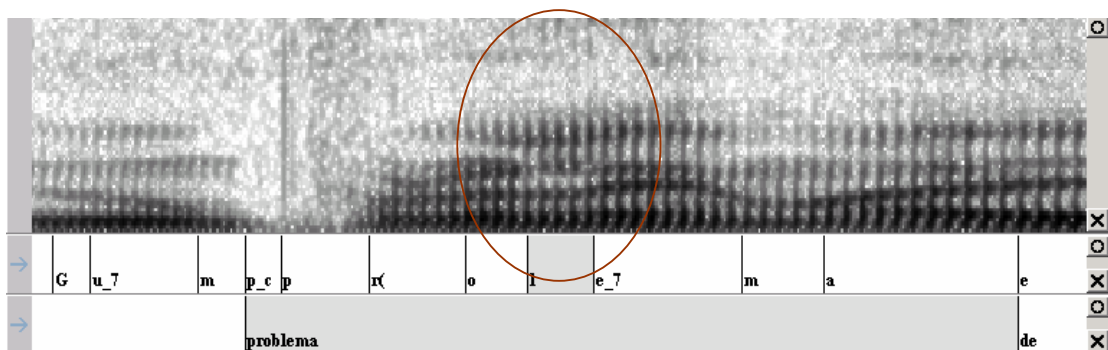


Figura 20. Reducción del segmento /bl/ a [l] entre vocales en la palabra *problema* (Diálogo 22, oración 162, hablante 4).

Para ambos casos, las palabras en que se perdió la /b/ intervocálica son usadas en

varias ocasiones por los mismos hablantes, como son: *problema, mueble* o *por favor*, por lo que una vez más notamos que la frecuencia determina las condiciones de pérdida, además de la asimilación que ya mencionamos (§ 3.1) que sufren las consonantes en esta posición.

La /r/ es una consonante bastante estable. Se perdió en posición intervocálica en sólo el 6% de los casos y generalmente en palabras derivadas de los verbos *querer* y *ser* y en la palabra *refrigerador*. En la Figura 21, es posible observar un cambio de formantes de [e] a [j] con ausencia de /r/, lo cual se refleja en la transición inmediata de una vocal a otra.

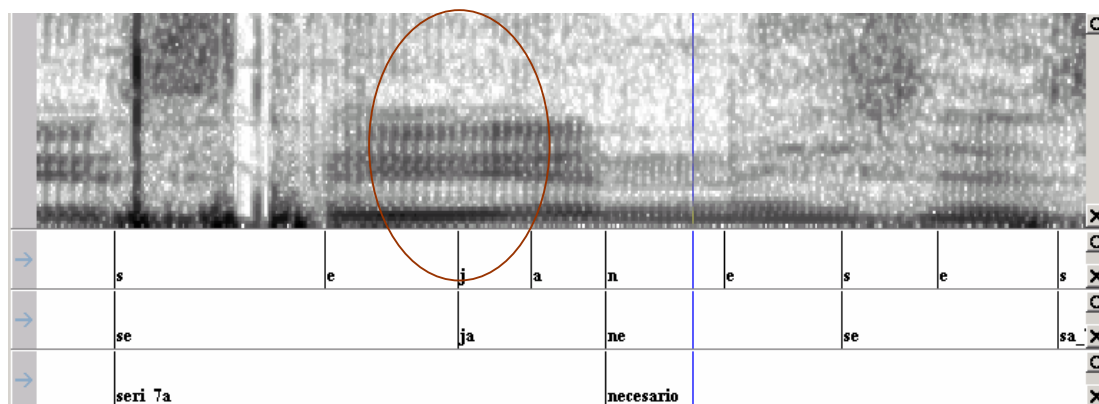


Figura 21. Síncopa de /r/ intervocálica en la palabra *sería* (Diálogo 22, oración 32, hablante 4).

La consonante /r/ se pierde más abundantemente cuando acompaña a una consonante como /t/, /f/ o /d/. En estos contextos, se perdió en 10.5%, 20.2% y 25% de los casos, respectivamente. Lope Blanch (1991a) mencionaba que el contexto /tr/ provoca que la /r/ se asibile o se ensordezca casi hasta volverse imperceptible, más en palabras de uso frecuente. En la Figura 22 observamos la ausencia de /r/ entre la explosión de /t/ y los formantes de /e/. El verbo *mostrar* y sus conjugaciones es de uso frecuente en el

Corpus DIME, por lo cual podemos decir que la pérdida de la consonante puede deberse a la frecuencia también, pero no se puede negar que la combinación de una consonante más una líquida CCV es articulatoriamente más compleja que la formación CV, además de que las palabras no tienen pares con los que podrían confundirse al perder la /r/, por lo que el esfuerzo no es indispensable.

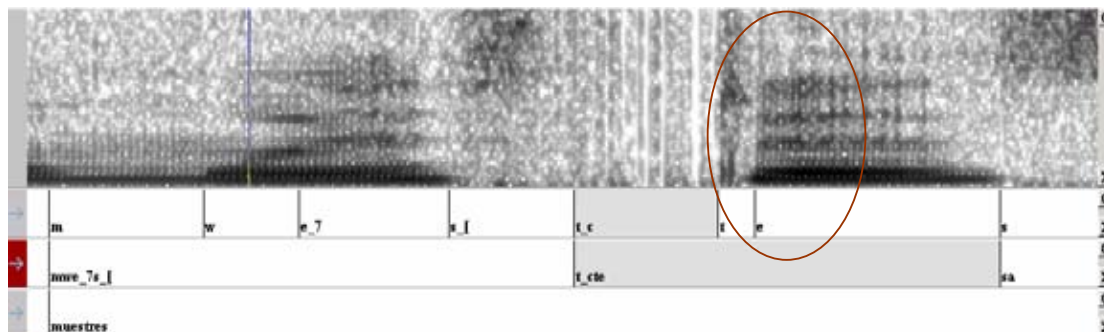


Figura 22. Síncopa de *r* en el segmento /tr/ en la palabra *muestras* (Diálogo 26, oración 33, hablante 5).

Si recordamos lo dicho por Martinet (§3.1.1), notamos que en algunos casos los fonemas van a conservarse cuando su pérdida signifique provocar una confusión; es decir, que los segmentos más ricos en información van a mantenerse para facilitar la comunicación. En los casos que aquí se presentan, no hay riesgo de confusión de palabras por la pérdida de fonemas por lo que el hablante obedece a la ley del menor esfuerzo.

En cuanto a la /d/ encontramos que se pierde en 18% de los casos y con muchas combinaciones de vocales. En la Figura 23 vemos que la /d/ se pierde entre dos vocales idénticas.

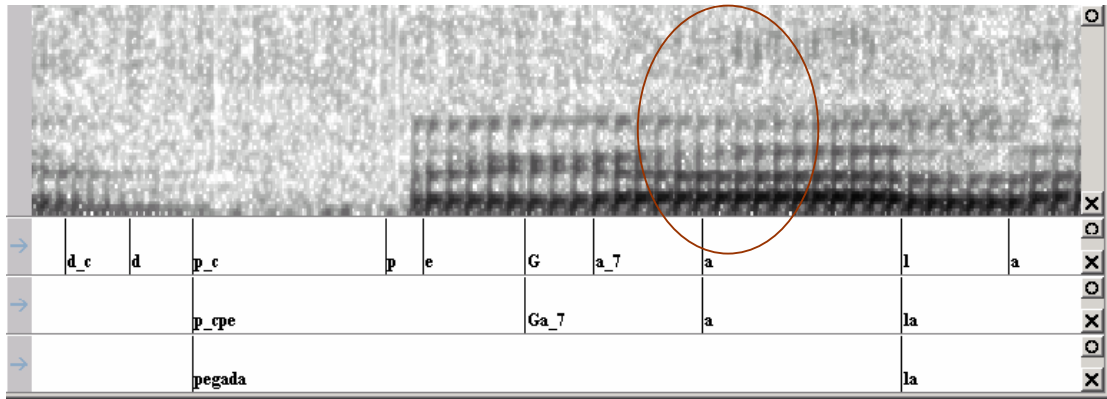


Figura 23. Síncopa de /d/ intervocálica en la palabra *pegada* (Diálogo 10, oración 92, hablante 2).

El fenómeno de pérdida de /d/ intervocálica ha sido estudiado por muchos autores, mencionamos ya algunos en §2.2; Alarcos (1981) señala que la /d/ sólo se pierde en el segmento *-ado*, cambio que se presenta en todo el español. En este corpus encontramos que éste no es el único contexto y tampoco el de mayor frecuencia. De hecho, pocos fueron los casos en los que la terminación *-ado* perdió la intervocálica.

4.2.3.2 Síncopa de consonantes en posición implosiva

La posición implosiva de las consonantes es también susceptible a la pérdida por la tendencia, mencionada anteriormente, del español por la sílaba abierta (§3.3.2). Si una consonante se encuentra en posición de límite silábico el hablante probablemente preferirá eliminarla o modificarla. Quilis (1997) advierte sobre las consonantes en posición de límite silábico que pueden tener muchas articulaciones que van desde la pronunciación canónica, hasta la pérdida de la consonante, pasando por varios niveles de debilitamiento, ensordecimiento, etc. Como sabemos, dos consonantes en contacto tienden a enfrentar procesos de asimilación y los casos aquí presentados son de consonantes -la mayoría- que comparten rasgos articulatorios y que, por tanto, se

simplifican con la elisión de la consonante implosiva, facilitando así la pronunciación de las palabras o de cadenas fónicas. En su estudio sobre fonética instrumental, Herrera (2001) realiza un análisis similar al que aquí se presenta y concluye que la posición implosiva, sobre todo de las oclusivas, genera un cambio en la mayoría de los casos y que estos cambios suelen ser ensordecimientos, elisiones y otros procesos de asimilación relacionados con el punto de articulación. Las consonantes que en la posición implosiva presentaron pérdida en el corpus son: /n/, /r/, /s/ y /k/.

La /n/ se perdió en 20% de los casos en que apareció en esta posición. Se perdió principalmente ante /s/ y ante /l/, ambas consonantes alveolares como la /n/ misma. Se perdió ante /s/ con un porcentaje de 35%, y este caso es típico y sigue funcionando en el español en el que incluso se están haciendo modificaciones ortográficas que admiten la reducción del segmento /ns/ a /s/, sobre todo en el prefijo *trans-*; Matluck (1951) afirmaba, ya en su época, que el segmento /ns/ que pertenecían a prefijos como *trans-* se conservaba totalmente, aunque la consonante /n/ estaba debilitada y a veces desaparecía después de contagiar su nasalidad a la vocal que la antecedía.

En la Figura 24 notamos que la consonante /n/ se ha perdido cuando antecede a /s/, lo cual se comprueba cuando observamos que la transición de los formantes de /e/ a las sombras de /s/ no tiene ningún espacio en que pudiera localizarse la /n/; Por esta razón podemos pensar que la /n/ se ha perdido al ser asimilada por otra consonante alveolar.

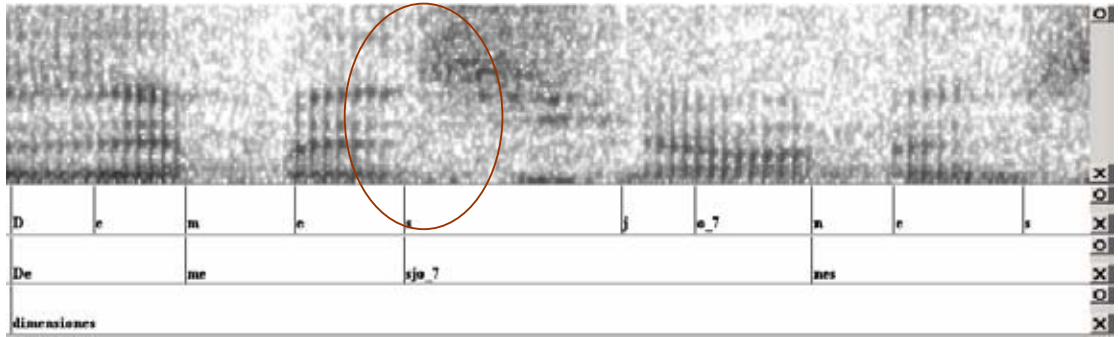


Figura 24. Reducción del segmento /ns/ en posición implosiva a [s] en *dimensiones* (Diálogo 8, oración 3, Mago 1).

Navarro Tomás (1918) advierte que la /n/ se modifica cuando antecede a todas las consonantes, excepto a las alveolares, pero que en algunos prefijos, como los anteriormente mencionados, la /n/ no se pronuncia; Quilis (1997) afirma que en particular el segmento /ns/ suele reducirse a /s/ en España en la mayoría de los casos y que también es una práctica general del español.

La consonante /n/ se pierde también ante /l/ con un porcentaje de 14%, y este cambio, igual que el anterior, puede explicarse por la similitud de las características de las tres consonantes, las cuales comparten el punto de articulación, aunque el resto de las características sean distintas. Cuando se está haciendo segmentación fonética, los fonemas /nl/ suelen ser de los más complicados de separar, puesto que son pocas las características físicas que los diferencian, además de ser acústicamente muy semejantes. En la Figura 25 notamos que la /l/ ha asimilado a la /n/ completamente y que esto puede apreciarse también por la extensión del segmento.

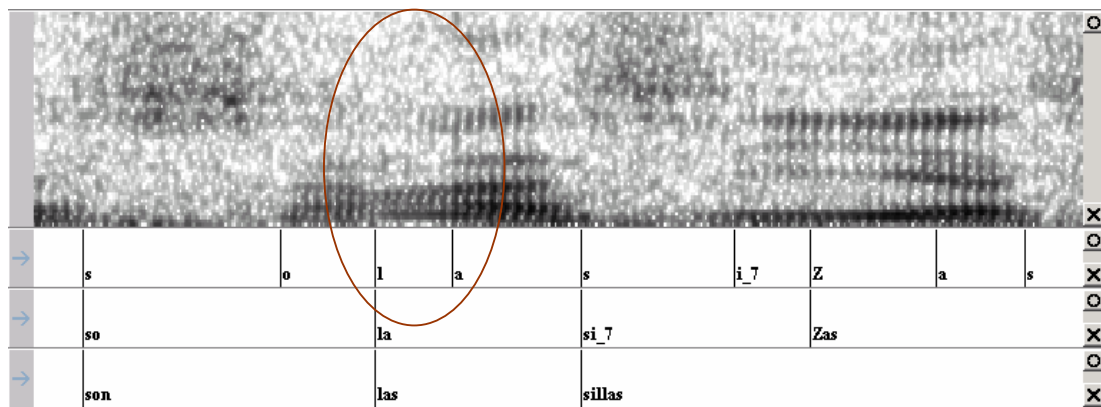


Figura 25. Síncopa de /n/ ante /l/ en la oración *son las sillas* (Diálogo 8, oración 107, Mago 1).

La /r/ en posición implosiva, como en el resto de las posiciones, tiene poco índice de pérdida. Se omitió en 5% de los casos ante /l/ y en 22% ante /f/, aunque este último caso ocurrió sólo en la fórmula *por favor*, utilizada ampliamente en el corpus; esto no es indicio de que la /r/ suela perderse en posición implosiva. Ni siquiera sería posible afirmar que se pierde ante /f/ pero, dado el porcentaje considerable de pérdida y que ésta es una palabra de uso muy frecuente en los diálogos prácticos, se consideró necesario trabajar la expresión *por favor* como una modificación de tipo léxico. En la Figura 26, observamos que el tránsito de /o/ a /f/ se hace inmediatamente y no hay algún segmento que pudiera indicar la presencia de /r/.

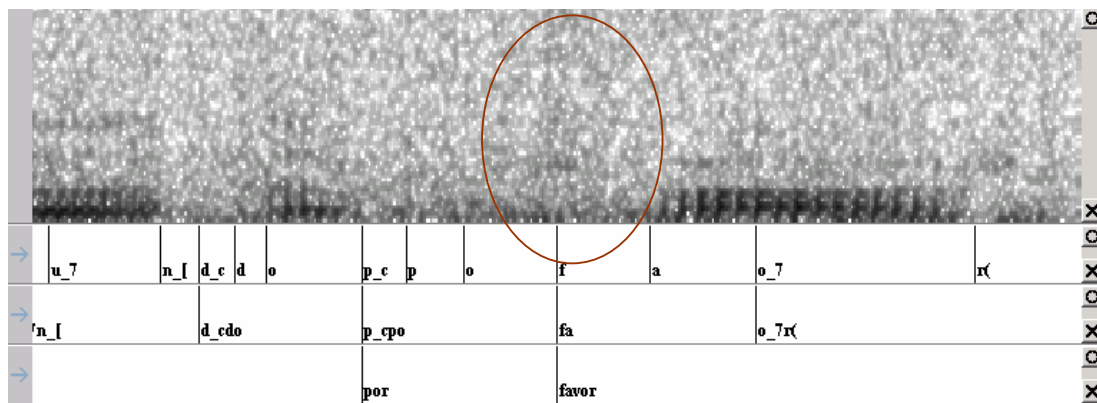


Figura 26. Síncopa de /r/ final de palabra ante /f/ en la frase *por favor* (Diálogo 26, oración 179, hablante 5).

Sobre el comportamiento general de /r/, Lope Blanch (1991:84-85) advierte que su pérdida es común en final de sílaba ante consonante, sobre todo en la construcción infinitivo + artículo iniciado por /l/, como en *ver las estufas*, o cuando un clítico de objeto forma parte de la morfología verbal, como en *hacerlo*; estos contextos aparecieron varias veces en el corpus y presentaron pérdida de la vibrante simple.

La /s/ que se encuentra en posición implosiva presentó un porcentaje muy elevado en dos de sus contextos. Se perdió en 44% de los casos en que antecedía a una /l/; estas dos consonantes son alveolares y, como se mencionó en el presente apartado, en el caso de pérdida de /n/ ante /s/ (Figura 24), la consonante lateral pudo haber asimilado a la /s/ hasta desaparecerla.

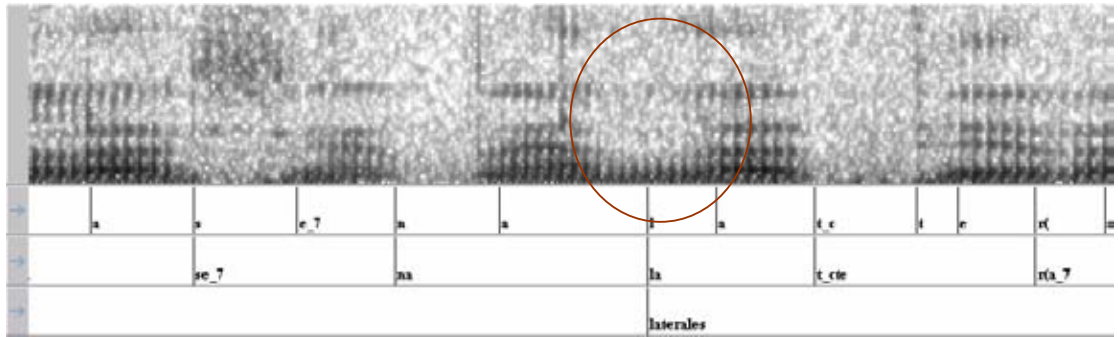


Figura 27. Síncopa de /s/ final de palabra ante /l/ en la frase *alacenas laterales* (Diálogo 8, oración 28, Mago 1).

La /s/ se perdió con mayor incidencia aun cuando antecedía a una /r/, de nuevo una consonante alveolar. La consonante /s/ se omitió en 50% de los casos. Este contexto, además, presentó varios casos de asibilación y de homogeneidad de los dos fonemas. La /s/, como el resto de las fricativas sordas, tiene una apariencia espectrográfica muy peculiar, “se caracterizan por tener ruido turbulento; es decir, inarmónicos” (Martínez 1998:69), por lo que es fácil percatarse de su desaparición. En las figuras 27 y 28 vemos que su omisión es obvia, al no observar las sombras características de la articulación de /s/ (ver Apéndice 1).

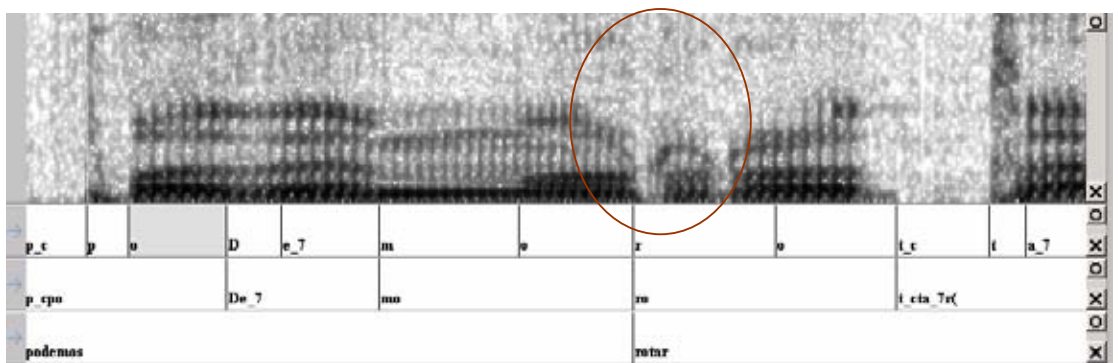


Figura 28. Síncopa de /s/ final de palabra ante /r/ en la oración *podemos rotar* (Diálogo 20, oración 147, hablante 3).

El fenómeno que sin lugar a dudas presentó el mayor porcentaje de pérdida es el de

/k/ implosiva que antecede a la alveolar /s/ en el segmento /ks/. La pérdida se dio en 62.5% en todas las palabras que presentaron este contexto. Esta pérdida es la que presenta mayor sistematización, lo cual puede entenderse fácilmente debido al gran esfuerzo articulatorio que implica mantener dos consonantes sordas juntas a veces ambas en posición de cierre silábico. En la Figura 29, tenemos un ejemplo de la pérdida de la /k/ que se evidencia con la ausencia de oclusión de /k/.

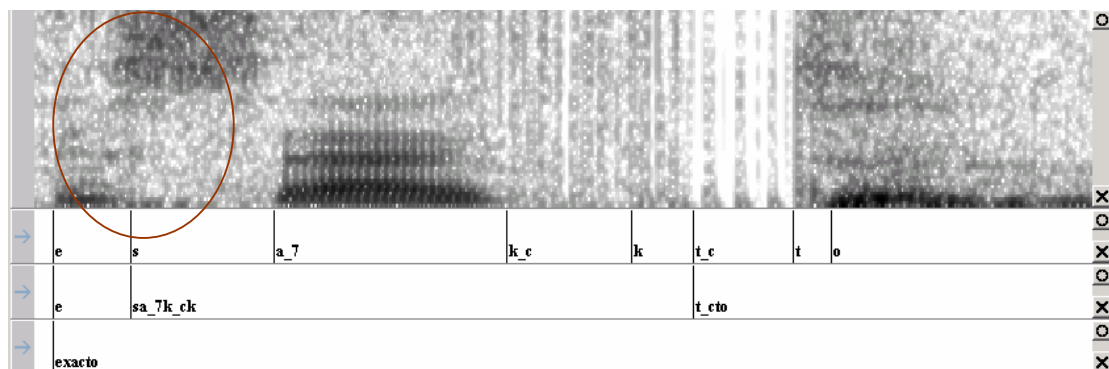


Figura 29. Reducción del segmento /ks/ intervocálico a [s] en la palabra *exacto* (Diálogo 26, oración 141, hablante 5).

El segmento se redujo, además, en cuatro de los siete hablantes en más de una ocasión. Este fenómeno tiene que ver con el comportamiento de los llamados grupos cultos en el español. En el dialecto español por ejemplo, no son pocos los casos en que es completamente admisible pronunciar [esáкто], pero según Matluck (1951) México suele caracterizarse por poseer un dialecto que conserva generalmente los grupos cultos. Sin embargo, los resultados del estudio confirman que se tiende a la simplificación, por lo menos de este segmento.

4.2.3.3 Síncopa en palabras específicas

Algunos de los fenómenos que se habían englobado en categorías distintas fueron exclusivos de algunas palabras, que son, además, de alta frecuencia en el corpus; por lo que decidí darles un tratamiento léxico. Algunas palabras adoptaban la misma forma siempre y otras más tenían varias formas de transformarse. La siguiente tabla explica las palabras que se analizaron aisladamente y su transformación más común.

Palabra	Pronunciación	Frecuencia
<i>puedes</i>	<i>pues</i> [pwés]	52%
<i>quieres</i>	<i>quies</i> [kjés]	28%
<i>necesito</i>	<i>nesito</i> [nesíto]	18%

La palabra *puedes* presentó, además, pérdida de /d/ intervocálica en muchos casos, lo cual puede indicar que tal vez la reducción se deba a la pérdida de /d/ y a la homologación de las dos /e/ en contacto, ya que hubo muy pocos casos en que un alargamiento de /e/ fue visible. En la Figura 30 notamos que la duración de la palabra *puedes* se ha reducido y que no es posible observar una /d/, ni tampoco un alargamiento de la vocal /e/, por lo que la palabra se reduce a la pronunciación [pwés]. }

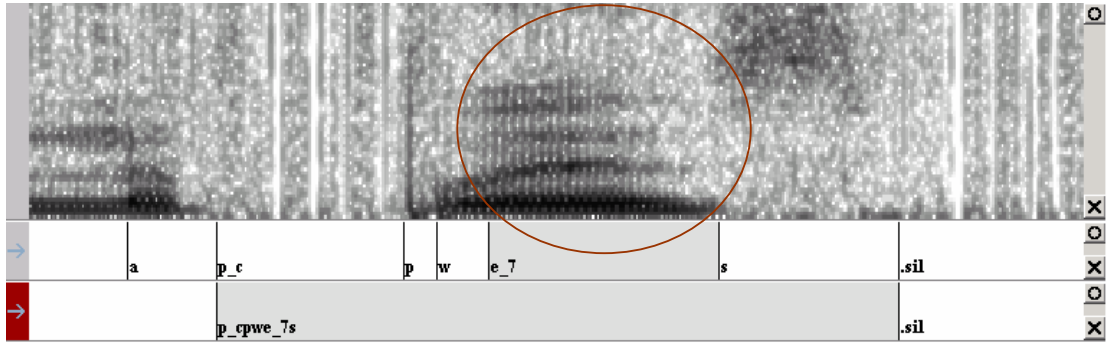


Figura 30. Reducción silábica de la palabra *puedes* a [pwés]
(Diálogo 10, oración 45, hablante 2).

La palabra *quieres* (así como otras conjugaciones del verbo *querer*) se redujeron a [kjés] al perderse la /r/ intervocálica, de la misma forma que sucedió con *puedes*; las causas pueden ser las misma, que la pérdida de /r/ creara una homologación vocálica de los elementos restantes. Vemos un ejemplo de esta reducción en la Figura 31.

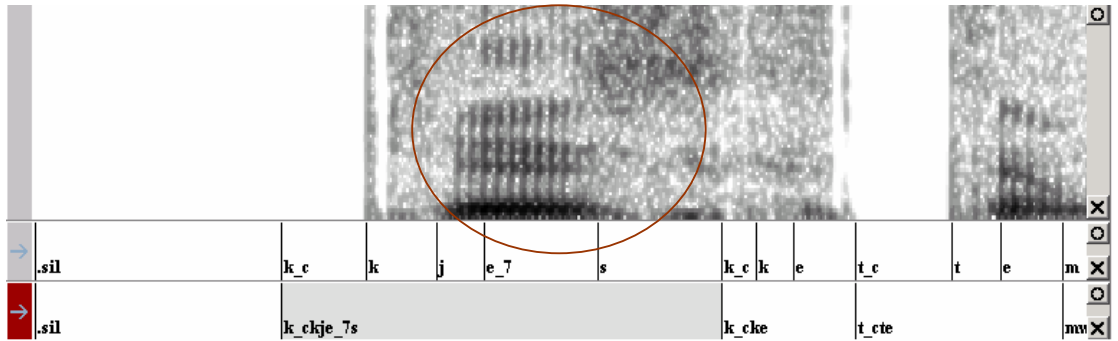


Figura 31. Reducción silábica de la palabra *quieres* a [kjés]
(Diálogo 22, oración 110, Mago 1).

La palabra *necesito* se redujo a [nesító] en muchos casos, y tuvo algunos cambios intermedios en que ocurre una síncope de la vocal /i/, lo que provoca un alargamiento notable de las dos /s/ en contacto, como es posible apreciar en la Figura 32.

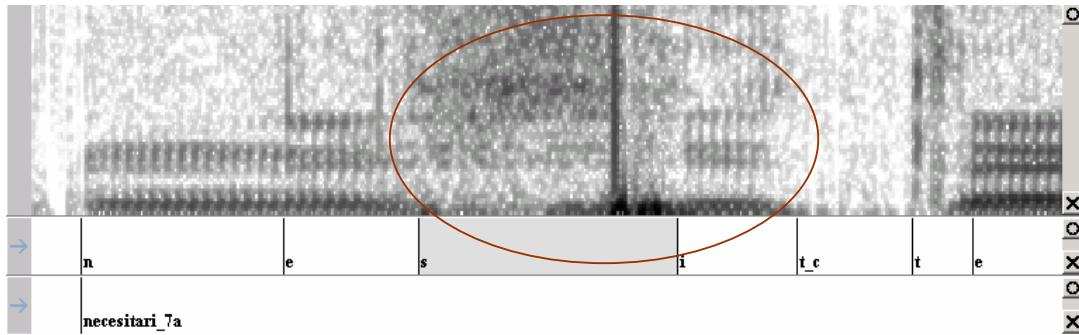


Figura 32. Reducción silábica de la palabra *necesitaria* a [nesitaría]
(Diálogo 22, oración 86, hablante 4).

Estas tres palabras, *puedes*, *quieres* y *necesito* (junto con algunas otras de menor incidencia) no se considerarán para reglas de contextos fonéticos sino que sus pronunciaci3nes más frecuentes serán adheridas a la lista de pronunciaci3nes a nivel léxico con las que actualmente cuenta el diccionario de pronunciaci3n del reconocedor de habla del *Proyecto DIME*. La adhesi3n de estas palabras se debe a que son muy frecuentes en el corpus y en el contexto en el que se pretende utilizar estos sistemas de reconocimiento y síntesis de habla.

4.2.3.4 Síncopa de vocales por reducci3n de diptongos

El último contexto que encontré es el de la monoptongaci3n o reducci3n de dos elementos vocálicos a uno solo. Este caso sólo se di3 con el diptongo /je/ y ocurri3 en 17% de los casos. La mayor parte de las ocasiones el diptongo suele reducirse a /i/; en este proceso de asimilaci3n, las dos siendo vocales palatales se reducen a una siendo ésta átona. En la Figura 33 podemos observar que no hay cambio de formantes en el espacio que correspondía al diptongo [je], además de que la duraci3n del segmento está reducida considerablemente.

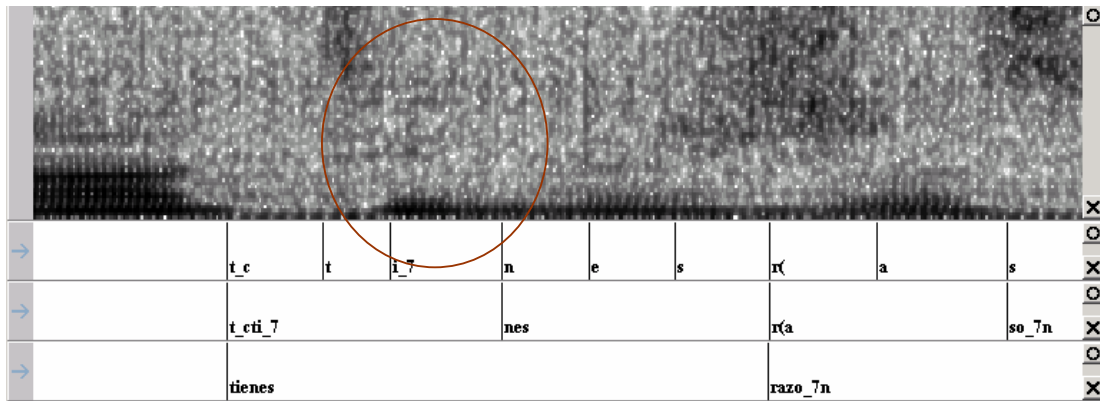


Figura 33. Reducción del diptongo [je] a [i] en la palabra *tienes* (Diálogo 8, oración 130, hablante 1).

El diptongo se convierte en [e] en muy pocos casos y sólo con la palabra *bien*. En este caso la vocal tónica es la que permanece. Podemos notar, en la Figura 34, que los formantes son idénticos, lo cuál indica que no hay dos vocales en contacto; lo que sí notamos es el alargamiento del segmento, que de la misma forma que en el ejemplo de la Figura 33, es tónico.

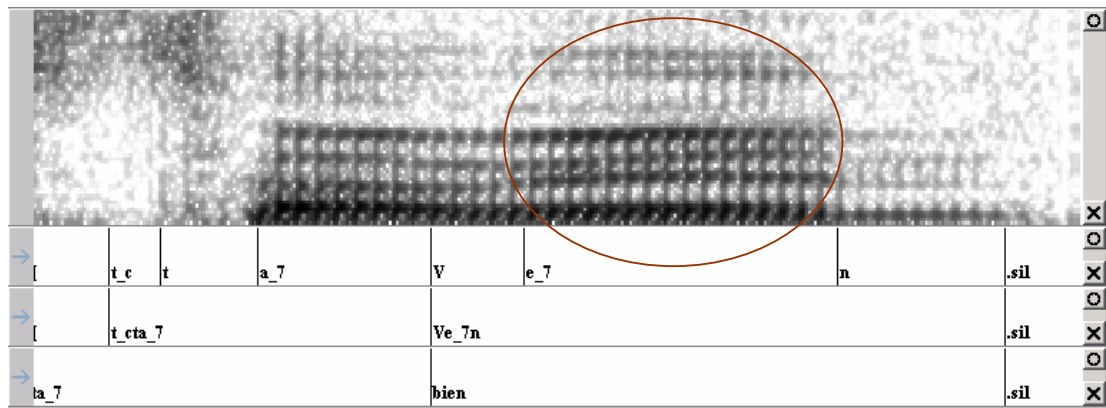


Figura 34. Reducción del diptongo [je] a [e] en la oración *está bien*. (Diálogo 10, oración 58, hablante 2)

Sobre las vocales en contacto, se sabe que “nuestra pronunciación tiende, preferentemente, a convertir, siempre que es posible, todo conjunto de vocales en un grupo monosilábico” (Navarro Tomás, 1918:148), aunque en este corpus sólo encontramos un contexto que se reduce en interior de palabra, hubo mucho ejemplos de reducción vocálica que produjeron silabificación.

4.2.3.5 Silabificación

Este fenómeno ha sido abordado por diversos teóricos, tales como Navarro Tomás (1918) y Quilis (1997). Es un fenómeno difícil de detectar aunque algunas marcas de pronunciación lo indica; en general, somos capaces de dividir las palabras en sílabas sin ningún problema incluso de forma artificial, “We can tell which part comes first because we know how to make these sounds. We can say syllables slowly and work out how they are made” (Ladefoged, 2001:175)⁵. Pero, a pesar de que hacer una división abstracta de las palabras no entorpece nuestra pronunciación, cuando se pretende que los sistemas informáticos produzcan el habla de la forma más parecida a la humana, el proceso de silabificación real se vuelve imprescindible.

El fenómeno de silabificación es interesante porque la variación de la silabificación fonológica y la fonética afectan el comportamiento de los sistemas de síntesis de habla, y muchos fueron los ejemplos que se encontraron en estos cinco diálogos. La silabificación consistió generalmente en el apareamiento de consonantes y vocales

⁵ Podemos decir qué parte va primero porque sabemos cómo hacer estos sonidos. Podemos decir sílabas lentamente y determinar cómo están hechas.

para la formación de sílabas abiertas, es decir, aquellas que terminan con vocal C + V. Un segmento fónico como CVVCVCVCV que se divide de la siguiente forma: CVV-CVC-V-CV, en el habla, muchas veces puede transformarse a CVV-CV-CV-CV, como en el siguiente ejemplo:

Cadena fónica	Silabificación fonológica	Silabificación fonética
La quieres aquí	La quie-res a-quí	La-quie-re-sa-quí

Otro caso frecuente de silabificación es aquél en el que algún segmento que marca frontera entre palabras se pierde y provoca un reajuste en la estructura fónica, como en el siguiente ejemplo:

Cadena fónica	Silabificación fonológica	Silabificación fonética
Puedes subir	Pue-des su-bir	Pue-de-su-bir

En el caso anterior las dos eses se vuelven un solo segmento y se silabica de forma similar al primer ejemplo.

Estos ejemplos de silabificación fueron altamente frecuentes en el corpus puesto que en el habla se van enlazando los sonidos sin hacer pausas en cada palabra, lo cual provoca modificaciones en los límites de éstas. Este fenómeno está siendo ampliamente analizado en una investigación de tesis en la UNAM, por lo que baste con señalar aquí algunos ejemplos y mencionar su alta frecuencia en el habla espontánea.

3.3 Reglas fonéticas

Los criterios para formular reglas fonéticas son la frecuencia, la variedad de hablantes y la relevancia de las palabras en los sistemas en que piensan aplicarse. En cuanto a frecuencia, se consideró como mínimo el 25% de pérdida en un contexto específico, para poder pensar que es relevante crear entradas con elisión para todas las palabras que se encuentran en el diccionario de pronunciación de un reconocedor de habla. Por último, las palabras más utilizadas no sólo en el contexto en el que se desarrolla la muestra sino en la resolución de tareas en general se considerarán para el tratamiento léxico y se incluirán las nuevas pronunciaciones en los sistemas existentes.

A continuación, se muestran las reglas que se obtuvieron del proceso de análisis. La expresión de las reglas está basada en los símbolos convencionales para la formulación de reglas en las tecnologías del habla (Jurafsky, 2000). Al inicio, se coloca el fonema que sufre la pérdida; la flecha (→) indica la transformación del segmento; a continuación se muestra un cero (0) que representa la omisión; la barra o diagonal (/) significa: “en el contexto”; el guión bajo () indica la posición del segmento que se pierde; los corchetes ({ }) incluyen más de un contexto; cuando una palabra o letras están entre barras o diagonales (/palabra/) se representa su transcripción fonológica y cuando está entre corchetes ([]) la fonética; y por último las letras V y C significan “vocal” y “consonante” respectivamente. A partir de esta codificación se presentan las siguientes reglas:

Regla	Explicación
$n \rightarrow 0 / _s [C \text{ o } V]$	La /n/ se reduce a cero cuando precede al segmento /s/ más consonante o vocal.
$s \rightarrow 0 / \{ _l, _r \}$	La /s/ se reduce a cero cuando precede al segmento /l/ o /r/ más vocal.
$k \rightarrow 0 / _s [C \text{ o } V]$	La /k/ se reduce a cero cuando precede a /s/ seguida de vocal o consonante.
$VfrV \rightarrow VfV$	La /r/ se pierde cuando forma parte del segmento /fr/ intervocálico.
$VbV \rightarrow VV$	La /b/ se pierde cuando se encuentra en posición intervocálica.

Para el nivel léxico, se propone que se incluyan las siguientes realizaciones para las palabras:

Palabra	Realización fonética alternativa	Explicación
<i>puedes</i>	[pués]	La palabra <i>puedes</i> pierde la sílaba /de/ y se reduce a [pwés].
<i>quieres</i>	[kjés]	La palabra <i>quieres</i> pierde la sílaba /re/ y se reduce a [kjés].
<i>Okey</i>	[oké]	La palabra <i>okey</i> pierde la última vocal del diptongo y se reduce a [oké].
<i>Favor</i>	[fabó]	La palabra <i>favor</i> pierde la /r/ final y se reduce a [fabó].
<i>Pared</i>	[paré]	La palabra <i>pared</i> pierde la /d/ final y se reduce a [paré].
Por favor	[pofabór]	La /r/ se pierde antecediendo a la /f/ en la frase [<i>porfabór</i>].

Las reglas aquí presentadas serán analizadas para evaluar si es posible aplicarlas al diccionario de pronunciación del *Proyecto DIME*; el proceso sobre la evaluación y la aplicación de las reglas se presenta a continuación.

4.4 Aplicación de las reglas al diccionario de pronunciación

Las reglas que se obtuvieron del análisis anterior pueden ser aplicadas tanto a la síntesis como al reconocimiento de habla. Por lo que respecta de este estudio, las reglas formuladas serán aplicadas a un diccionario de pronunciación perteneciente a uno de los reconocedores del *Proyecto DIME*. Para esto, es necesario explicar brevemente la capacidad del diccionario y el impacto que las reglas fonéticas tienen sobre los diccionarios en general.

El reconocedor de habla *DIMEx200x* del *Proyecto DIME* está compuesto por veintidós modelos acústicos, un modelo del lenguaje y un diccionario de pronunciación (así como uno de relleno⁶). Este diccionario contiene aproximadamente 4300 palabras, las cuales muestran la pronunciación canónica y otras pronunciaciones fonéticas. Es necesario explicar los tipos de diccionarios de pronunciación, su utilidad y el proceso para su obtención.

Los diccionarios de pronunciación pueden ser canónicos; es decir, que pueden contener una única pronunciación para cada palabra basados en la forma tradicional o académica; o tener varias pronunciaciones para cada palabra. El diccionario que se basa en la pronunciación canónica puede obtenerse automáticamente, a partir de una base léxica contenida en un corpus. Así, contendría una correspondencia fonética para cada palabra existente en el corpus. Con esto se limita mucho el proceso de reconocimiento puesto que cualquier modificación afectaría la capacidad de interpretación del sistema. Cuando a un diccionario se le aplican reglas fonéticas de contextos específicos, el diccionario aumenta las pronunciaciones de cada palabra

⁶

creando hipótesis de posibles modificaciones; por ejemplo, si el diccionario contiene las palabras *transmitir*, *transparente* y *transportar*, y se le aplica una regla que indique que la /n/ se pierde ante /s/ implosiva, es decir, $n \rightarrow 0/ _sC\$,$ además de la pronunciación canónica de las tres palabras, el diccionario adquiriría las siguientes: [trasmitír], [trasparénte] y [trasportár]; si además de esto se le aplica una regla que afirme que la /r/ puede perderse cuando forma parte del segmento /tr/, es decir, $VtrV \rightarrow VtV$ (suponiendo que se creara esta regla fonética), además se le agregaría a cada palabra las siguientes pronunciaciones: [tasmitír], [tansmitír], [tasparénte], [tansparénte], [tasportár] y [tansportár]. Así, tendríamos, al menos, cuatro pronunciaciones distintas para cada palabra.

Ahora, hablando de las reglas aquí obtenidas, al aplicar cuatro de contextos fonéticos específicos, el diccionario canónico aumentó de 4300 a 4637 pronunciaciones. Así, vemos que el impacto de la aplicación de sólo cuatro reglas en un diccionario canónico es considerable, pues, suponiendo que se crearan bastantes reglas, se podría contar con la mayor cantidad de pronunciaciones posibles.

Existen también los diccionarios robustos, que son aquéllos que contienen, además de la pronunciación canónica, otras posibles pronunciaciones de las mismas palabras. Estas realizaciones fonéticas se obtienen con base en estudios de habla, semejantes al que aquí se presentó. El diccionario de pronunciación al que aplicaremos las reglas es ya bastante robusto, puesto que se han dedicado varias horas al análisis manual de muestras de habla que indican las variaciones de las palabras en el habla natural y en el habla controlada; sin embargo, algún impacto se obtuvo para las reglas cuyos contextos fueron los más frecuentes. En cuanto al léxico, se logró agregar una

pronunciación de cuatro palabras de alta frecuencia en el corpus y en la lengua en general, que son: [pwes] para *puedes*, [kjés] para *quieres*, [oké] para *okay* y [fabó] para *favor*.

Observamos que el impacto de reglas fonéticas de contextos específicos no es tan grande en los diccionarios de pronunciación robustos, puesto que estos ya contienen pronunciaciones variadas para la misma palabra; sin embargo, el impacto en los diccionarios canónicos es importante ya que, si se tuvieran suficientes reglas, podrían construirse pronunciaciones alternativas automáticamente al ser aplicadas dichas reglas fonéticas; lo cuál además de robustecer el diccionario de pronunciación, ahorraría muchas horas de trabajo manual.

5. Conclusiones

El objetivo de este estudio fue realizar un análisis fonético de una muestra de habla del español de la Ciudad de México para obtener un estado actual de dicho dialecto en relación con el fenómeno de pérdida de fonemas. Ello con el propósito de construir reglas fonéticas que se aplicarían a sistemas computacionales de síntesis y reconocimiento de habla.

Primero, hice una revisión teórica sobre los antecedentes científicos que propiciaron el desarrollo de la inteligencia artificial, especialmente en el campo de la comunicación entre humanos y máquinas; hablé también del aspecto que conecta esta ciencia con la lingüística en los apartados sobre tecnologías de habla que son base de esta investigación (§2.1). Expliqué ampliamente las tecnologías de habla, haciendo una relación de las áreas de síntesis y reconocimiento de habla (§2.1.1, §2.1.2, §2.1.3, §2.1.4). Con esto justifiqué el papel del lingüista en el desarrollo de estos sistemas informáticos.

A continuación, revisé la teoría sobre el fenómeno de pérdida ubicándolo en la tipología de cambios de sonido, al mismo tiempo que se hizo hincapié en la importancia del proceso en la constitución primera y el desarrollo del español a lo largo de su historia (§3.1, §3.1.1). En el mismo capítulo, ahondé en la teoría sobre las pérdidas y relajamiento de fonemas en el español de la Ciudad de México (§3.4), esto para contrastar la teoría con los resultados que se obtendrían en el Capítulo 4.

Posteriormente, mostré los resultados del análisis del comportamiento del corpus (§4.1, §4.2), con lo cual podemos concluir que la lengua tiende a la ley del menor esfuerzo y las pérdidas se realizan en contextos articulatoriamente complejos -como grupos de dos consonantes- o que conllevan una pronunciación débil -como las apócope. La entonación también juega un papel importante en las pérdidas, pues los fonemas al final de oración tienden a pronunciarse muy relajadamente y a perderse en algunos casos; la posición postónica es proclive a la omisión, pero la tonicidad en las vocales no es un factor determinante para su pérdida, pues tónicas y átonas se pierden por igual.

En cuanto al tipo de pérdida, vimos que las aféresis son principalmente de /e/ ante /n/ o /s/; debido a la capacidad de estas consonantes a ocupar la posición nuclear de la sílaba.

Las apócope son de las consonantes /d/ y /n/ generalmente pero ocurren en algunas palabras muy frecuentes en el corpus y en la lengua, como son: *pared* y *bien*.

Para las síncope, la posición intervocálica provoca la pérdida de /b/ principalmente y de /r/ en el segmento /fr/. Vimos además, que la posición implosiva es la que más pérdidas aporta para las consonantes, pues es la que representa mayor dificultad articulatoria y está relacionada con la formación de las sílabas abiertas en español. De estas consonantes, se obtuvo que la /n/ y la /k/ se pierden ante /s/; la /s/ ante alveolares.

Asimismo, vimos que las palabras *puedes*, *quieres*, *okay*, *favor*, *pared* y *por favor*

presentan muchas modificaciones y se debe crear para ellas nuevas representaciones fonéticas. Por último, hablé brevemente sobre el proceso de silabificación mencionando sus ocurrencias más frecuentes.

Finalmente, mostré las reglas fonéticas que obtuve del análisis y expliqué el proceso de aplicación tecnológica de las reglas a los sistemas informáticos (§4.3, §4.4). Con esto vimos que las reglas fonéticas ayudan en gran medida a los diccionarios de pronunciación canónicos, proveyéndolos de pronunciaciones diversas de forma automática; no así a los diccionarios robustos que han sido construidos con base en análisis manual de datos, a los que sólo puede aportárseles algunas pronunciaciones y léxico distinto.

Vemos, por tanto que los análisis lingüísticos tienen importancia para el campo, puesto que muestran panoramas actuales de los fenómenos fonéticos de un dialecto o de una lengua, a la vez que tienen aplicaciones a la vida diaria. Los resultados de este estudio aportaron reglas aplicables a nivel léxico y fonético para las tecnologías del habla. Las reglas a nivel léxico se utilizan para ampliar las realizaciones fonéticas de una palabra de uso frecuente o relevante en el corpus que sirve de base para reconocedor de habla. Las reglas fonéticas son importantes porque nos permiten generar pronunciaciones a nivel de contextos fonéticos, lo cual facilita crear realizaciones automáticas para todas las palabras que contengan estos fenómenos. Con este estudio se logró que un diccionario de pronunciación ampliara su información léxica y fonética, consiguiendo así una mayor precisión en el proceso de interpretación de mensajes orales.

Queda aquí un estudio fonético que da cuenta del comportamiento del español de México que puede ser retomado con una muestra mayor y con un enfoque distinto al de las elisiones. Además, es posible utilizar el trabajo de transcripción y el análisis empírico del corpus en investigaciones diversas que pueden estar enfocadas en otras ramas de la lingüística, tales como la sintaxis y la pragmática. Queda, finalmente, un estudio de elisiones que tiene la utilidad de aplicarse a otras investigaciones prácticas en las tecnologías del habla.

6. Referencias bibliográficas

- ALARCOS LLORAH, EMILIO, 1981. *Fonología española*, Madrid: Gredos.
- BOYD-BOWMAN, PETER, 1960. *El habla de Guanajuato*, México: UNAM.
- BYBEE, JOAN, 2001. *Phonology and Language use*, Cambridge: Cambridge University Press.
- CANFIELD, LINCOLN D., 1981. *El español de América: Fonética*, Barcelona: Crítica.
- COMPANY COMPANY, CONCEPCIÓN, 2001. *Manual de fonología hispánica*, México: UNAM.
- CUÉTARA PRIEDE, JAVIER, 2004. *Fonética de la Ciudad de México. Aportaciones desde las tecnologías del habla*, tesis de Maestría inédita, México: UNAM.
- GIL FERNÁNDEZ, JUANA, 1990. *Los sonidos del lenguaje*, Madrid: Síntesis.
- GILI GAYA, SAMUEL, 1975. *Elementos de fonética general*, Madrid: Gredos.
- GORDON, E. PELTON, 1993. *Voice Processing*. McGraw-Hill, Inc.
- HENRÍQUEZ UREÑA, PEDRO, 1938. “Mutaciones articulatorias en el habla popular”, en *El español en Méjico, los Estados Unidos y la América Central*. Ed. P. Henríquez Ureña. Universidad de Buenos Aires, Buenos Aires, pp. 329-379.
- HERRERA Z. ESTHER, 2001. *Temas de fonética instrumental*, México: El Colegio de México.
- I FONTANALS, JOAQUIM RAFEL, 2003. “El procesamiento de corpus: La lingüística empírica” en *Tecnologías del lenguaje*. Martí Antonín, Ma. Antònia (Coord.). Barcelona, UOC. Cap. 2.
- INTERNATIONAL PHONETIC ASSOCIATION. 1949/1971. *The principles of the International Phonetic Association*, Londres: University College.
- _____. 1999. *Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet*, Cambridge: Cambridge University Press.
- JURAFSKI, DANIEL y James H. Martin, 2000. *Speech and Language Processing*, New Jersey: Prentice-Hall.
- LADEFOGED, PETER, 2001. *Vowels and consonants, an introduction to the study of language*, Massachusetts: Blackwell.
- LASS, ROGER. 1984/1985. *Phonology. An introduction to basic concepts*, Cambridge: Cambridge University Press.
- LLISTERRI BOIX, JOAQUIM, 1991. *Introducción a la fonética: El método experimental*, Barcelona: Anthropos.
- _____. 2003. “Las tecnologías del habla”, en *Tecnologías del lenguaje*, María Antonia Martí (Coord.), Barcelona: UOC, pp. 249-281.
- _____. 2003a “Lingüística y tecnologías del lenguaje”, *Lynx. Panorámica de Estudios Lingüísticos* (Departament de Teoria dels Llenguatges, Universitat de València) 2: 9-71.
http://liceu.uab.es/es/~joaquim/publicacions/TecnoLing_Lynx02.pdf
- LLISTERRI BOIX, JOAQUIM, LOURDES AGUILAR, JUAN M. GARRIDO, MARÍA JESÚS MACHUCA, RAFAEL MARÍN, CARMÉ 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 i Universitat Autònoma de Barcelona, pp. 449-479.
- LLISTERRI BOIX, 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 Universitat de Barcelona, pp. 13-28.
- LLISTERRI BOIX, 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.
- LOPE BLANCH, JUAN M., 1991. “La –r final del español mexicano y el sustrato nahua” en *Estudios sobre el español de México*, México: UNAM, p.79-95
- _____, 1991a. “En torno a las vocales caedizas del español mexicano” en *Estudios sobre el español de México*, México: UNAM, p.57-77
- 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: UNAM.
- MALMBERG, BERTIL, 1974. *Manuel de phonétique générale*, Paris: Editions A & PICARD.
- MARTÍ, M. ANTÒNIA JOAQUIM LLISTERRI, 2002. “Las tecnologías lingüísticas en la sociedad de la Información” en Martí, M. Antònia y Joaquim Llisteri (eds.), 2002. *Tratamiento del lenguaje natural*, Barcelona: Universitat de Barcelona.
- MARTÍ ROCA, JOSEPH, “Situación actual de las tecnologías del habla, Actas II Jornadas en Tecnología del Habla, Granada, del 16 al 18 de Diciembre de 2002
<http://lorien.die.upm.es/~lapiz/rtrth/JORNADAS/II/index.htm>, consultado 05 de abril de 2007-04-05
- MARTÍN BUTRAGUEÑO, PEDRO, 2002. *Variación lingüística y teoría fonológica*, México: Colegio de México.
- MARTINET, ANDRE, 1995. *Économie des changements phonétiques*, Benne: Éditions A. Franke S.A..
- MARTÍNEZ CELDRÁN, EUGENIO, 1994. *Fonética*, Barcelona: Teide.
- _____, 1998. *Análisis espectrográfico de los sonidos del habla*, Barcelona: Ariel.
- MATLUCK, JOSEPH, 1951. *La pronunciación en el español del valle de México*, México: UNAM.
- MORA, ELSA y MANUEL RODRÍGUEZ. *Síntesis de habla en español venezolano: usos prácticos y de investigación*
www.saber.ula.ve/.../alexandr/db/ssaber/Edocs/pubelectronicas/investigacion/inum12/avance-001.pdf. MORENO DE ALBA, JOSÉ G. 1994. *La pronunciación del español en México*, México: El Colegio de México.
- NAVARRO TOMÁS, TOMÁS, 1918/2004, *Manual de pronunciación española*, Madrid: Consejo superior de Investigaciones Científicas.
- PÉREZ PAVÓN, E. PATRICIA, 2006. *Construcción de un reconocedor de voz utilizando Sphinx y el corpus DIMEx100* tesis de licenciatura inédita, México: UNAM.
- PÉREZ, HERNÁN EMILIO. 2003 (Ms.). *Frecuencia de fonemas*, Concepción: Universidad de Concepción.
- PERISSINOTTO, GIORGIO, 1975. *Fonología del Español hablado en la Ciudad de México. Ensayo de un método sociolingüístico*, México: El Colegio de México.
- 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.

- POZA LARA, M. J.- VILLARRUBIA GRANDE, L.- SILES SÁNCHEZ, J.A. (1991) *Teoría y aplicaciones del reconocimiento automático del habla*, *Comunicaciones de Telefónica I+D 3*.
<http://www.tid.es/presencia/publicaciones/comsid/esp/articulos/vol23/habla/habla.html>
- QUILIS ANTONIO y Joseph A. Fernández, 1979. *Curso de Fonética y fonología españolas*, Madrid: C.S. I. C.
- QUILIS, ANTONIO, 1997. *Tratado de fonología y fonética españolas*, Madrid: Gredos.
- RAFEL I FONTANALS, JOAQUIM Joan Soler i Bou, 2003. “El procesamiento de corpus. La lingüística empírica” en *Tecnologías del lenguaje*, Martí Antonín, Ma. Antònia (Coord.), Barcelona: UOC.
- REVISTA DE FILOLOGÍA ESPAÑOLA. 1915. “Alfabeto fonético”, *Revista de Filología Española*, Madrid, 2, pp. 374-376.
- RODRÍGUEZ HONTORIA, HORACIO, 2003. “Las interfaces en lenguaje natural” en *Tecnologías del lenguaje*, Martí Antonín, Ma. Antònia (Coord.), Barcelona: UOC, Cap. 5.
- TAPIAS MERINO, DANIEL. 2002. “Interfaces de voz con lenguaje natural”, en *Tratamiento del lenguaje natural*, María Antonia Martí, Joaquim Llisterra Boix (Eds.), Barcelona: Edicions Universitat de Barcelona, pp. 189-208.
- 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.
- VILLARRUBIA L., “Tecnología del Habla para desarrollo de aplicaciones Multilingües, Multiservicio y Multiplataforma” Actas II Jornadas en Tecnología del Habla, Granada, del 16 al 18 de Diciembre de 2002.
- 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.

7. Apéndices

7.1 Apéndice 1. Tabla de porcentajes

Este apéndice muestra los resultados del análisis empírico de datos; indica el tipo de fenómenos que ocurrieron, el número de apariciones total contra el número de contextos con modificación; asimismo, indica el número de hablantes que presentaron el fenómeno y el porcentaje general de modificación.

CONTEXTO	NÚMERO DE APARICIONES EN EL <i>CORPUS DIME</i>	NÚMERO DE CONTEXTOS CON FENÓMENOS	NÚMERO DE HABLANTES CON FENÓMENOS	PORCENTAJE DE PÉRDIDA
<i>Puedes</i> → <i>pues</i>	27	14	4	51.85%
<i>Quieres</i> → <i>quies</i>	54	14	1	27.77%
<i>Necesito</i> → <i>nesito</i>	22	4	3	18.18%
e → 0/ #_n		8	3	3.18%
VbV → VV	70	15	5	21.42%
VblV → VIV	28	5	2	17.85%
VrV → VV	266	14	4	5.26%
VtrV → VtV	104	11	3	10.57%
VfrV → VfV	79	16	2	20.25%
VdrV → VdV	8	2	2	25%
VdV → VV	189	37	5	17.98%
n → 0/ n#	130	15	2	11.53%
n → 0/ _sV	20	7	2	35%
n → 0/ _lV	41	4	2	12.19%
r → 0/ r#	61	5	4	8.19%
r → 0/ _rf	18	4	3	22.22%
r → 0/ _rl	57	3	2	5.26%
s → 0/ _sl	16	7	3	43.75%
s → 0/ _sr	14	7	3	50%
s → 0/ s#	167	5	2	2.99%
k → 0/ _kC o _kV	24	15	4	62.50%
d → 0/ _7d#	23	5	2	21.73%
je → i, e	248	42	5	16.93%

7.2 Apéndice 2. Imágenes espectrográficas de contextos sin pérdida

Este apéndice busca auxiliar a los lectores que no están familiarizados con los espectrogramas, mostrándoles imágenes óptimas de los fonemas cuya pérdida se menciona a todo lo largo del capítulo 4.

Al contrastar estas imágenes con aquéllas que evidencian pérdida se puede comprobar la ausencia de los fonemas en cuestión.

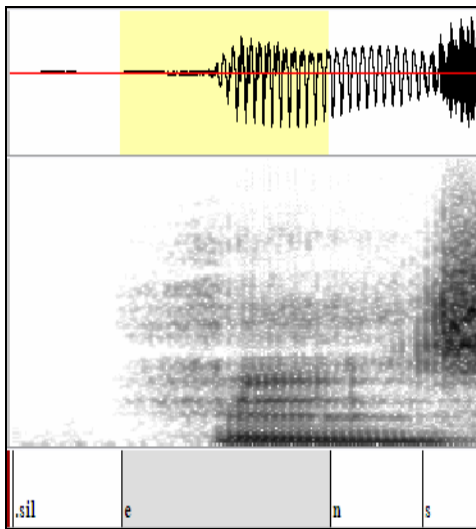


Figura 35. Espectrograma de /e/ inicial ante /n/.

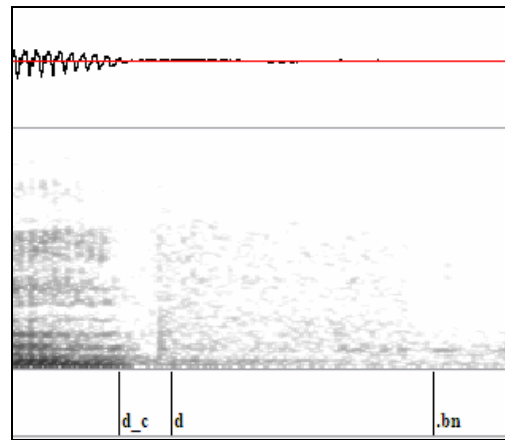


Figura 37. Espectrograma de /d/ oclusiva final.

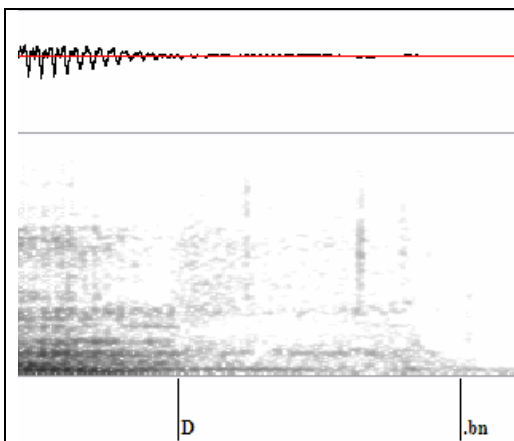


Figura 36. Espectrograma de /d/ fricativa final.

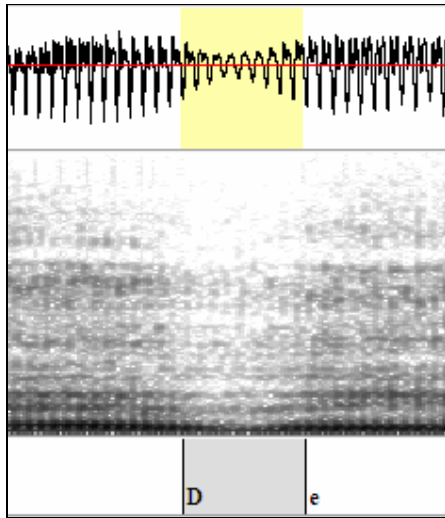


Figura 38. Espectrograma de /d/ fricativa intervocálica.

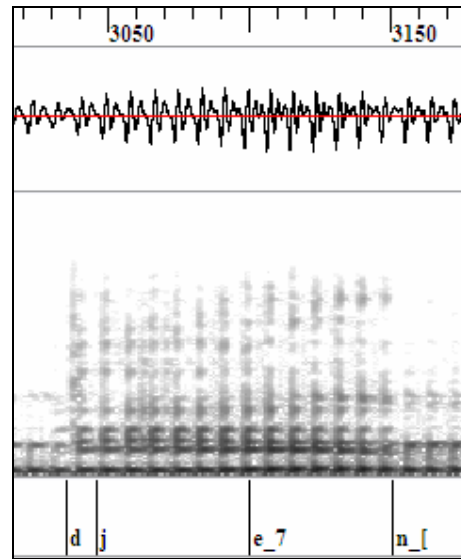


Figura 41. Espectrograma de diptongo /ie/.

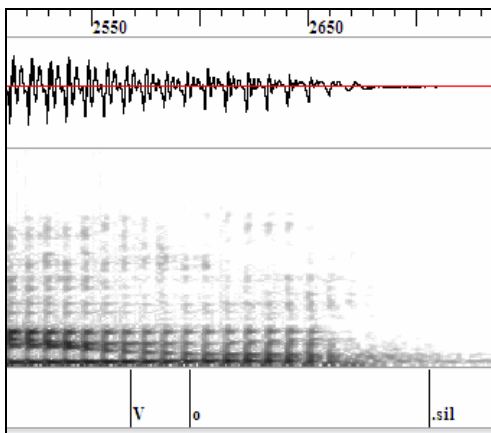


Figura 39. Espectrograma de /o/ final.

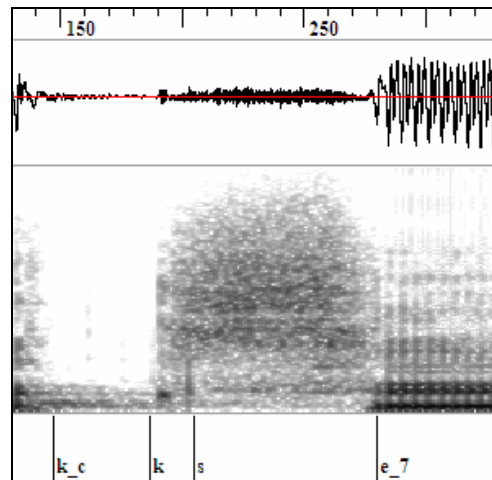


Figura 42. Espectrograma de /ks/.

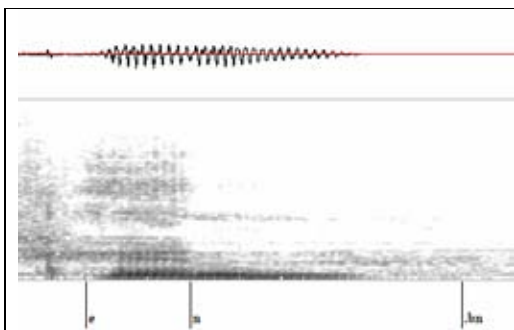


Figura 40. Espectrograma de /n/ final en sílaba /en/.

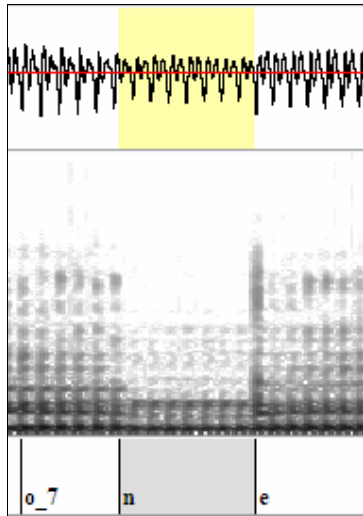


Figura 43. Espectrograma de /n/ intervocálica.

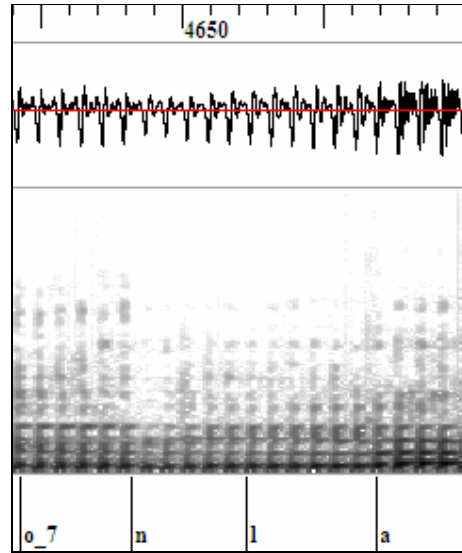


Figura 45. Espectrograma de /n/ ante consonante //l/.

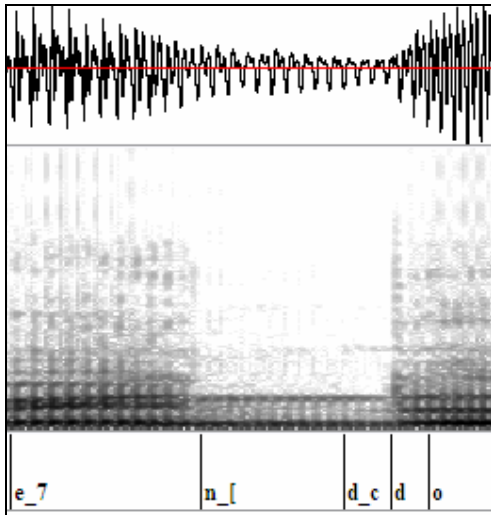


Figura 44. Espectrograma de /n/ ante consonante dental.

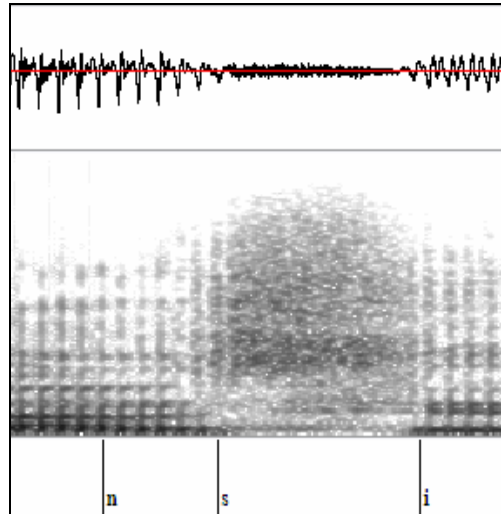


Figura 46. Espectrograma de /n/ ante /s/.

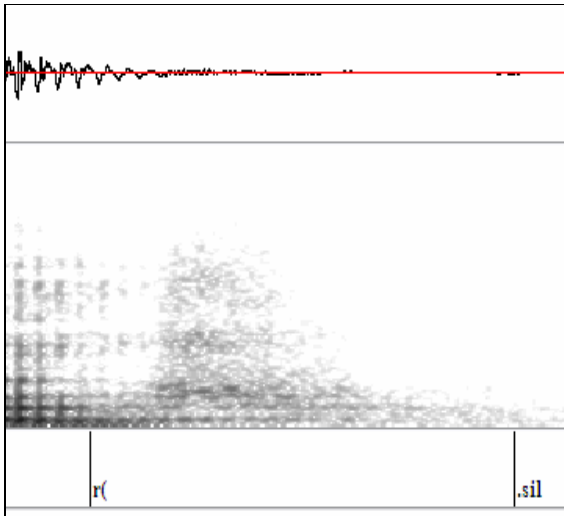


Figura 47. Espectrograma de /r/ final asibilada.

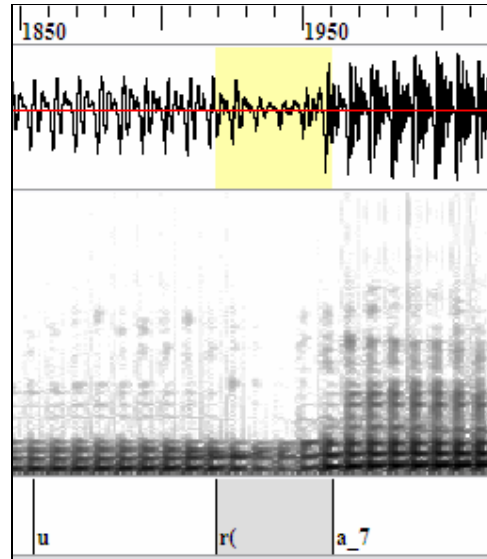


Figura 49. Espectrograma de /r/ intervocálica.

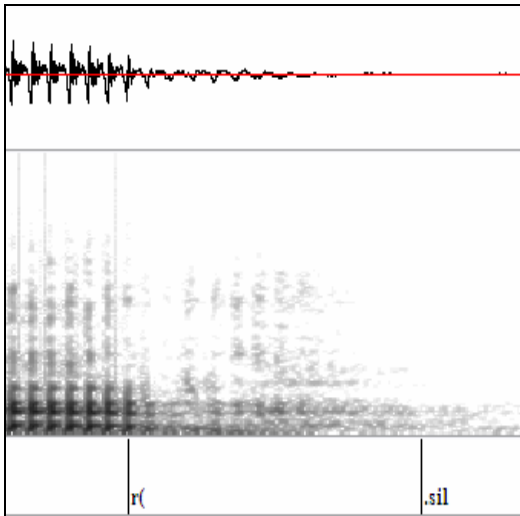


Figura 48. Espectrograma de /r/ final.

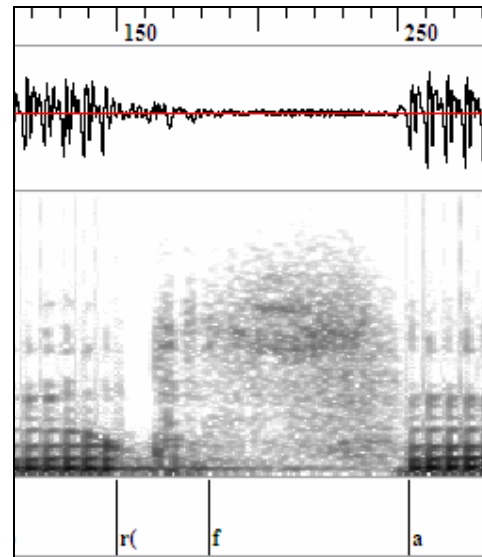


Figura 50. Espectrograma de /r/ ante /f/.

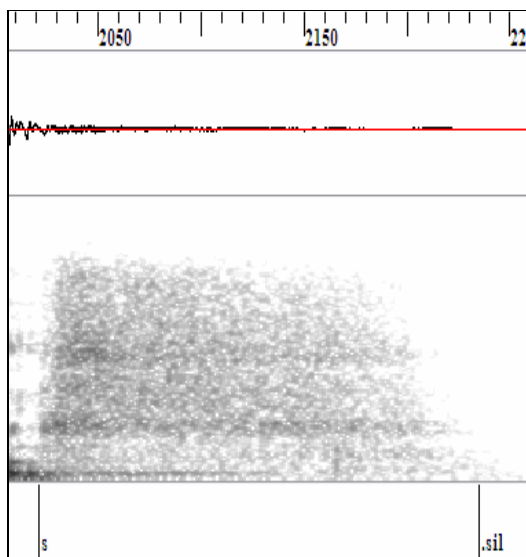


Figura 51. Espectrograma de /s/ final.

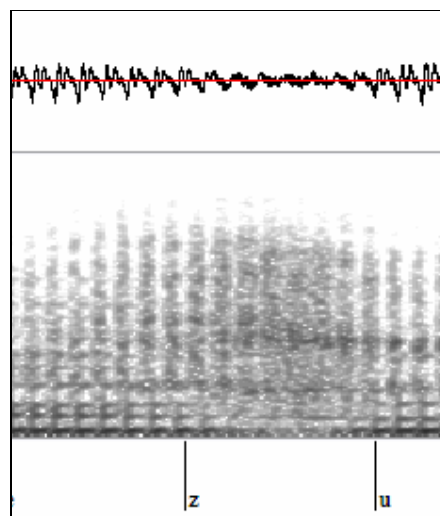


Figura 53. Espectrograma de /s/ intervocálica sonora.

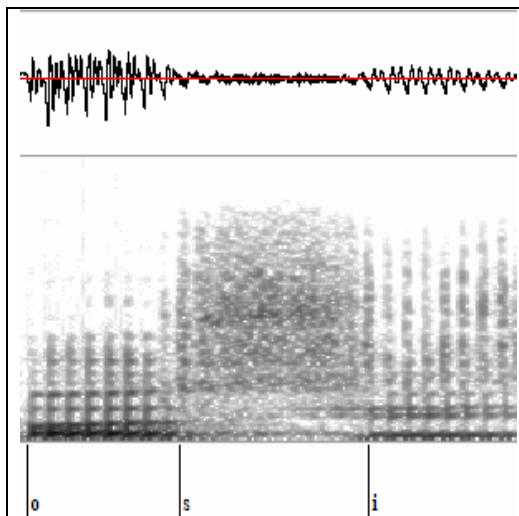


Figura 52. Espectrograma de /s/ intervocálica sorda.

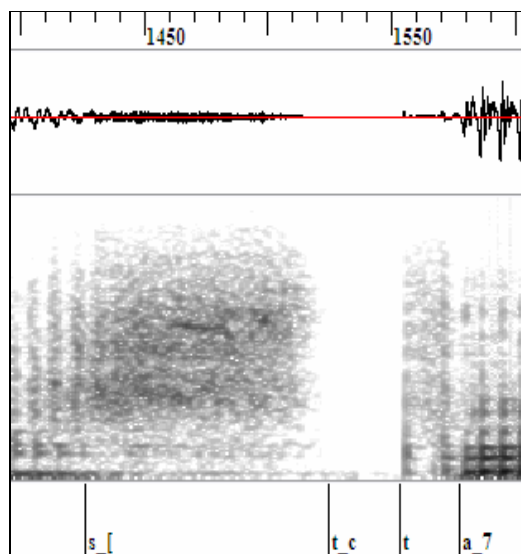


Figura 54. Espectrograma de /s/ ante consonante dental.

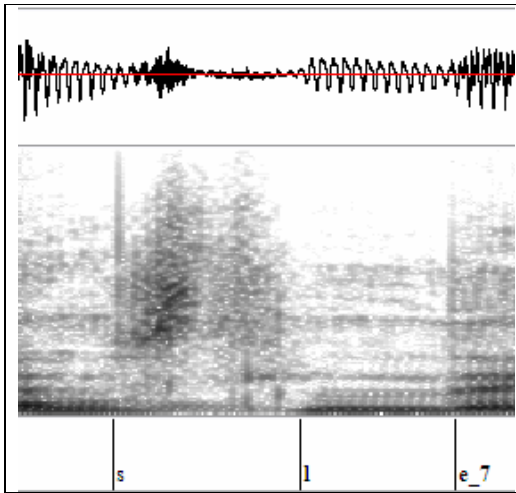


Figura 55. Espectrograma de /s/ ante /l/.

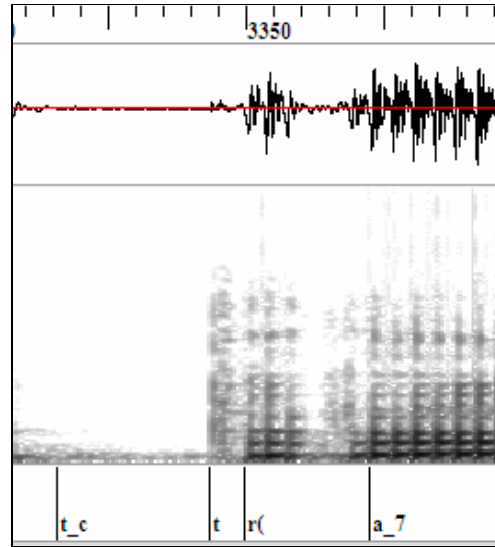


Figura 57. Espectrograma del segmento /tr/.

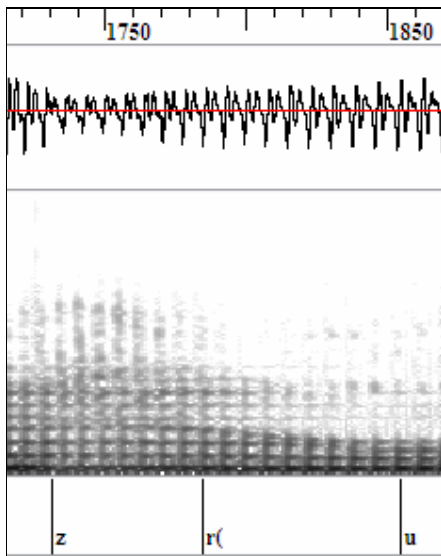


Figura 56. Espectrograma de /s/ ante /r/.

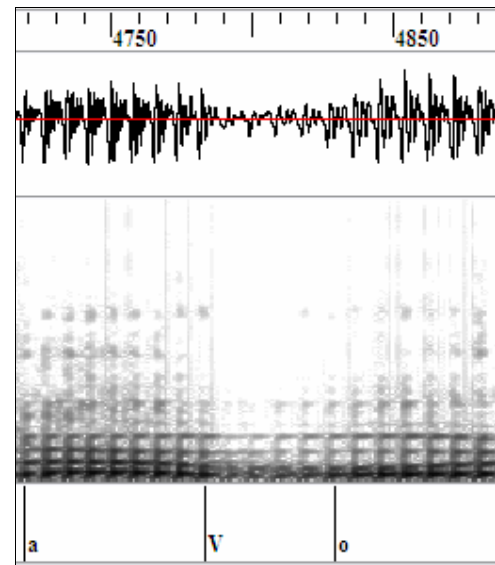


Figura 58. Espectrograma de /b/ intervocálica.

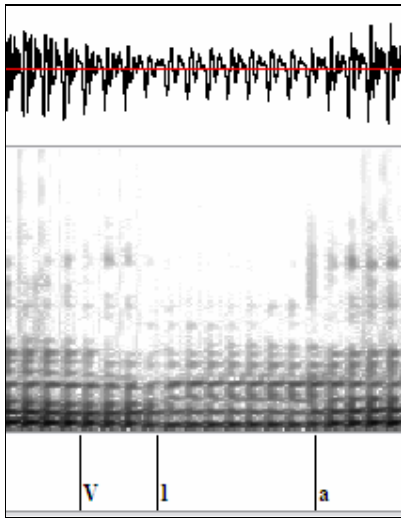


Figura 59. Espectrograma del segmento
/bl/

