

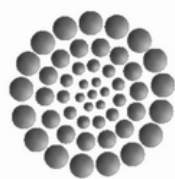


Análisis prosódico comparativo del español oral

[*Vergleichende prosodische Analyse des gesprochenen Spanisch*]

DISSERTATION ZUR ERLANGUNG DES
DOKTORGRADES
DES FACHBEREICHS PHILOSOPHIE UND GEISTESWISSENSCHAFTEN
AM INSTITUT FÜR ROMANISCHE PHILOLOGIE
DER FREIEN UNIVERSITÄT BERLIN

Eduardo Patricio Velázquez Patiño
Mai, 2008



CONACYT

Zusammensetzung der Promotionskommission:

Herr Prof. Dr. Thomas Kotschi (Erstgutachter)
Herr Prof. Dr. Guido Mensching (Zweitgutachter)
Herr Prof. Dr. Uli Reich
Frau Prof. Dr. Gabriele Knauer
Frau Dr. Dolors Font Rotchés

Datum der Disputation: 18. Juli 2008

A Luis Arturo y Concepción, quienes me dieron el ejemplo de cómo luchar por alcanzar las metas sin olvidar los propios valores humanos.

A Julia, por estar todo este tiempo a mi lado, motivándome y ayudándome.

A Carolina y Rafael, por su apoyo incondicional.

An Prof. Dr. Thomas Kotschi, für seine Bereitschaft mich zu betreuen und für seine wichtigen Ratschläge und Kommentare.

A todos los colegas y amigos que me dieron alguna idea o me hicieron algún comentario a lo largo del camino y que enriquecieron alguna parte de este texto.



Esta tesis doctoral fue escrita mientras el autor fue becario del Consejo Nacional de Ciencia y Tecnología.

Índice de Contenidos

0. Introducción	1
0.1 Preguntas de investigación	1
0.2 Objetivos	2
0.3 Hipótesis	2
0.4 Justificación	3
1. Marco Teórico	5
1.1 Prosodia	6
1.1.1 Rasgos prosódicos y unidades prosódicas	8
1.1.1.1 Acento	10
1.1.1.2 Patrón entonativo	11
1.1.1.3 Cantidad	12
1.1.1.4 Juntura	13
1.1.2 Funciones comunicativas de la prosodia	13
1.1.2.1 Caracterización del esquema foco/trasfondo	14
1.1.2.2 Caracterización del tipo de oración	16
1.1.2.3 Establecimiento de correferencia y coherencia	17
1.1.2.4 Caracterización de la coordinación y elipsis	18
1.1.2.5 Caracterización de la estructura oracional	19
1.1.2.6 Solución de ambigüedades léxicas	20
1.1.3 Constituyentes prosódicos	22
1.1.3.1 Sílaba	22
1.1.3.2 Pie	24
1.1.3.3 Palabra fonológica	25
1.1.3.4 Grupo clítico	25
1.1.3.5 Frase fonológica	26
1.1.3.6 Frase entonativa	26
1.1.3.7 Enunciado fonológico	27
1.1.3.8 Crítica	27
1.2 Entonación	29
1.2.1 Modelos teóricos de la entonación	31
1.2.1.1 Estructura interna de los contornos melódicos	33
1.2.1.2 Declinación y escalonamiento descendente	34
1.2.1.3 Rango tonal	35
1.2.2 La escuela británica: análisis por configuraciones	36
1.2.2.1 Componentes básicos del modelo	37
1.2.2.2 Las unidades entonativas y sus componentes	37
1.2.2.3 Aplicación del análisis por configuraciones al español	41
1.2.2.4 Componentes de la unidad melódica y sus significados	41

1.2.3 La escuela americana: análisis por niveles	44
1.2.3.1 Componentes básicos del modelo	44
1.2.3.2 Aplicación del análisis por niveles al español	46
1.2.4 Teoría métrica autosegmental	48
1.2.4.1 Tones and Break Indices (ToBI y Sp-ToBI)	52
1.2.4.2 Entonación en el discurso	57
1.2.4.3 Tonemas del español	58
1.2.5 La escuela holandesa: El modelo IPO	60
1.2.5.1 Componentes básicos del modelo	61
1.2.5.2 Movimientos melódicos	61
1.2.5.3 Configuraciones	62
1.2.5.4 Unidades prosódicas	63
1.2.5.5 Ventajas y desventajas	64
1.2.5.6 Aplicación al español: Patrones melódicos del español en habla espontánea	65
1.2.5.7 Tonemas y sus variantes tipo	69
1.2.6 Modelo de Aix-en-Provence	69
1.2.6.1 Componentes básicos del modelo	70
1.2.6.2 Ventajas y desventajas	72
1.2.6.3 Aplicaciones del modelo de Aix-en-Provence al español	73
1.3 Ritmo y tempo	73
1.3.1 El ritmo según la psicología y la antropología lingüística	74
1.3.2 El ritmo y el tempo como señales de contextualización	75
1.3.3 Interdependencia del tempo y el ritmo	77
1.3.4 Ritmos abstractos	78
1.4 Intensidad	81
1.5 Calidad de la voz	82
1.5.1 Prosodia de los estados emocionales	84
1.6 Resumen	85
2. Metodología	87
2.1 Transcripción de la lengua oral	88
2.1.1 Sistemas de transcripción	89
2.1.2 La computadora como medio de apoyo	92
2.2 XML (Extensible Markup Language)	93
2.2.1 Conceptos fundamentales	95
2.2.2 Etiquetado	98
2.2.3 DTD (Document Type Definition)	102
2.2.4 Corrección y validez	108
2.2.5 XSL / XSLT	109
2.2.6 Resumen	112

2.3 Transcripción prosódica de la lengua oral en XML	115
2.3.1 Antecedentes	115
2.3.1.1 <i>EXMARaLDA</i>	116
2.3.1.2 <i>TEI</i>	119
2.3.1.3 <i>Resumen</i>	127
2.3.2 Propuesta de estructura jerárquica prosódica para documentos de XML	127
2.3.3 Recolección de datos en Praat y transformación en XML	139
2.3.3.1 <i>Transcripción</i>	143
2.3.3.2 <i>Documentos básicos de XML</i>	146
2.3.3.3 <i>Hoja de estilo para documentos básicos de XML</i>	153
2.3.3.4 <i>Análisis detallado en Praat</i>	157
2.3.3.5 <i>Documentos resultantes de XML</i>	165
2.3.3.6 <i>Hojas de estilo para cada necesidad</i>	167
2.3.4 Resumen	172
3. Análisis	173
3.1 Descripción de la muestra	174
3.2 Corpus de Madrid	178
3.2.1 Sílabas y enunciados	178
3.2.2 Duración	180
3.2.3 Intensidad	185
3.2.4 Frecuencia	189
3.2.5 Entonación	195
3.2.6 Ritmo	203
3.3 Corpus de la Ciudad de México	210
3.3.1 Sílabas y enunciados	210
3.3.2 Duración	212
3.3.3 Intensidad	216
3.3.4 Frecuencia	220
3.3.5 Entonación	226
3.3.6 Ritmo	233
3.4 Análisis comparativo	240
3.4.1 Sílabas, unidades entonativas y enunciados	240
3.4.2 Duración	241
3.4.3 Intensidad	245
3.4.4 Frecuencia	249
3.4.5 Entonación	255
3.4.6 Ritmo	271
3.5 Resumen	276

4. Conclusiones	279
5. Bibliografía	283
6. Anexos	291
Anexo 1: Normas de transcripción en rejillas de texto (<i>text grids</i>) de Praat	291
Anexo 2: Definición de Tipo de Documento PROSODY_XML.DTD	292
Anexo 3: Definición de Tipo de Documento PROSODY_XML_SET.DTD	294
Anexo 4: Glosario de elementos de PROSODY_XML.DTD y PROSODY_XML_SET.DTD	296
Anexo 5: Subprograma ExtractTextGrid+Sound+XML.praat	299
Anexo 6: Subprograma Create_XML(U+).praat	317
Anexo 7: Subprograma ReviewTextGrids.praat	335
Anexo 8: Subprograma RecalculateF0var&F0std.praat	337
Anexo 9: Subprograma ExtractSound&TextGrid.praat	341
Anexo 10: Hoja de estilo Prosody.xsl	342
Anexo 11: Hoja de estilo Prosody-set.xsl	362
Anexo 12: Hoja de estilo Prosody-set-table.xsl	373
Anexo 13: Subprograma en JavaScript dynamic.js	375
Anexo 14: Contenido del DVD	379
Anexo 15: Deutsche Zusammenfassung (Resumen en alemán)	380

Índice de Ilustraciones

Ilustración 1-1: Rasgos prosódicos y unidades prosódicas	9
Ilustración 1-2: Declinación	11
Ilustración 1-3: Ejemplo de una jerarquía prosódica de constituyentes	22
Ilustración 1-4: Constituyentes de la sílaba	24
Ilustración 1-5: Ejemplos de estructuras de pies en español	25
Ilustración 1-6: Ejemplos de grupos clíticos	26
Ilustración 1-7: Ejemplo de la jerarquía prosódica de constituyentes según Fox (2000)	28
Ilustración 1-8: Segmentos moraicos y no moraicos	28
Ilustración 1-9: Movimientos de los tonos nucleares simples y compuestos	38
Ilustración 1-10: Movimientos de las sílabas posnucleares (colas)	39
Ilustración 1-11: Movimientos de las cabezas simples	39
Ilustración 1-12: Movimientos de las cabezas complejas	40
Ilustración 1-13: Tonemas según Navarro Tomás (1948)	42
Ilustración 1-14: Tonos prenucleares y sus contornos	55
Ilustración 1-15: Tonos de frontera inicial y sus contornos	56
Ilustración 1-16: Tonos nucleares en combinación con tonos de frontera final	56
Ilustración 1-17: Patrón melódico de la entonación neutra	66
Ilustración 1-18: Patrones melódicos de la entonación interrogativa	66
Ilustración 1-19: Patrones melódicos de la entonación suspendida	67
Ilustración 1-20: Patrones melódicos de la entonación enfática	68
Ilustración 1-21: Tonos absolutos y negativos según el sistema de anotación INTSINT	71
Ilustración 1-22: Ejemplo de parrilla métrica	79
Ilustración 1-23: Ejemplo de árbol métrico	79
Ilustración 1-24: Combinación de la parrilla métrica y el árbol métrico	80
Ilustración 2-1: Dimensiones para la comparación entre sistemas de transcripción	91
Ilustración 2-2: Estructura arbórea de los elementos básicos de una transcripción	96
Ilustración 2-3: Análisis fonético del archivo de audio <code>Ejemplo.wav</code> en Praat	100
Ilustración 2-4: Estructura jerárquica enriquecida con atributos	106
Ilustración 2-5: Exportación de la información de Praat a EXMARaLDA	118
Ilustración 2-6: Elemento <code><Transcript></code>	128
Ilustración 2-7: Elemento <code><Header></code>	128
Ilustración 2-8: Elementos <code><Text></code> y <code><Turn></code>	132
Ilustración 2-9: Elemento <code><U></code>	133
Ilustración 2-10: Elementos <code><IU></code> y <code><F></code>	134
Ilustración 2-11: Elemento <code><S></code>	136
Ilustración 2-12: Estructura jerárquica de <code>PROSODY_XML.DTD</code>	138
Ilustración 2-13: Estructura jerárquica de <code>PROSODY_XML_SET.DTD</code>	139
Ilustración 2-14: Fragmento de la ventana Praat objects con LongSound	140
Ilustración 2-15: Ventana emergente para la creación de un objeto TextGrid	141
Ilustración 2-16: Fragmento de la ventana Praat objects con LongSound y TextGrid	141
Ilustración 2-17: Ventana de preferencias para el espectrograma	142
Ilustración 2-18: Fragmento de la ventana de edición de subprogramas (<i>scripts</i>)	142
Ilustración 2-19: Adición de subprograma al menú al editor de TextGrids	142
Ilustración 2-20: Fragmento del TextGrid Editor con nueva opción en el menú File	143
Ilustración 2-21: Transcripción en el editor de TextGrids	146
Ilustración 2-22: Adición del subprograma al menú dinámico	147

Ilustración 2-23: Formulario del subprograma <code>ExtractTextGrid+Sound+XML.praat</code>	148
Ilustración 2-24: Imagen creada en la ventana Praat picture	150
Ilustración 2-25: Contenido compatible con XML en la ventana Info	152
Ilustración 2-26: Encabezado de la grabación en HTML	153
Ilustración 2-27: Resumen de fenómenos relevantes en el documento	155
Ilustración 2-28: Convenciones de los signos utilizados en la transcripción	155
Ilustración 2-29: Documento básico de XML transformado en HTML	156
Ilustración 2-30: Lectura de la F_0 a partir del espectrograma y los formantes	158
Ilustración 2-31: Formulario de <code>Create_XML(U+).praat</code>	159
Ilustración 2-32: Segmentación de sílabas	160
Ilustración 2-33: Cálculo del porcentaje de variación de F_0 y su valor estándar	161
Ilustración 2-34: Archivo de TextGrid con análisis completo	162
Ilustración 2-35: Menú emergente en el documento transformado en HTML	168
Ilustración 2-36: Despliegue de datos fonéticos por cada sílaba	170
Ilustración 2-37: Resultado en HTML de la hoja de estilo <code>Prosody-set-table.xsl</code>	171
Ilustración 2-38: Formato automático en Excel	171

Índice de Tablas

Tabla 1-1: Escala Universal de Sonoridad	23
Tabla 1-2: Gramática combinatoria del modelo métrico autosegmental	34
Tabla 1-3: Funciones discursivas de la entonación	58
Tabla 1-4: Combinaciones posibles de acentos y tonos de frontera en español	59
Tabla 1-5: Inventario de movimientos melódicos del holandés y su caracterización en rasgos distintivos	62
Tabla 1-6: Inventario de configuraciones para el holandés	63
Tabla 1-7: Estructuras silábicas posibles dentro de un grupo fónico	81
Tabla 3-1: Hablantes de Madrid	174
Tabla 3-2: Hablantes de la Ciudad de México	176
Tabla 3-3: Promedio de sílabas por enunciado (mujeres)	178
Tabla 3-4: Promedio de sílabas por enunciado (hombres)	178
Tabla 3-5: Promedio de sílabas por tipo de enunciado	179
Tabla 3-6: Número de enunciados de cada tipo por persona (mujeres)	179
Tabla 3-7: Número de enunciados de cada tipo por persona (hombres)	180
Tabla 3-8: Duración total de sílabas por persona (en segundos y minutos)	180
Tabla 3-9: Duración silábica máxima, mínima y promedio (segundos)	181
Tabla 3-10: Duración silábica máxima, mínima y promedio por tipo de enunciado (segundos)	181
Tabla 3-11: Número máximo, mínimo y promedio de sílabas por segundo	182
Tabla 3-12: Número máximo, mínimo y promedio de sílabas por segundo y por tipo de enunciado	182
Tabla 3-13: Variación máxima y mínima promedio de la duración silábica	183
Tabla 3-14: Variación máxima y mínima promedio de la duración silábica por tipo de enunciado	183
Tabla 3-15: Desviación estándar máxima, mínima y promedio de la duración silábica	184
Tabla 3-16: Desv. est. máxima, mínima y promedio de la duración silábica por tipo de enunciado	184
Tabla 3-17: Intensidad silábica máxima, mínima y promedio (decibelios)	185
Tabla 3-18: Intensidad silábica máxima, mínima y promedio por tipo de enunciado (dB)	186
Tabla 3-19: Variación máxima y mínima de la intensidad silábica	186
Tabla 3-20: Variación máxima y mínima de la intensidad silábica por tipo de enunciado	186
Tabla 3-21: Desviación estándar máxima, mínima y promedio de la intensidad silábica	187
Tabla 3-22: Desviación estándar máxima, mínima y promedio de la intensidad silábica por tipo de enunciado	187
Tabla 3-23: Intensidad silábica estandarizada máxima, mínima y promedio (dB inicial = 60)	188
Tabla 3-24: Intensidad silábica estandarizada máxima, mínima y promedio por tipo de enunciado (dB inicial = 60)	188
Tabla 3-25: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada	188
Tabla 3-26: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada por tipo de enunciado	189
Tabla 3-27: Frecuencia fundamental máxima, mínima y promedio (hertz)	190
Tabla 3-28: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (Hz)	190

Tabla 3-29: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (mujeres) (Hz)	191
Tabla 3-30: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (hombres) (Hz)	191
Tabla 3-31: Variación máxima y mínima promedio de la frecuencia fundamental	191
Tabla 3-32: Variación máxima y mínima promedio de la frecuencia fundamental por tipo de enunciado	192
Tabla 3-33: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental	192
Tabla 3-34: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado	192
Tabla 3-35: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (mujeres)	193
Tabla 3-36: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (hombres)	193
Tabla 3-37: Frecuencia estandarizada máxima, mínima y promedio (F_0 inicial = 100)	193
Tabla 3-38: Frecuencia estandarizada máxima, mínima y promedio por tipo de enunciado (F_0 inicial = 100)	194
Tabla 3-39: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada	194
Tabla 3-40: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada por tipo de enunciado	195
Tabla 3-41: Número y porcentaje de cada tipo de tono inicial	196
Tabla 3-42: Número y porcentaje de cada tipo de tono prenuclear	196
Tabla 3-43: Número y porcentaje de cada tipo de tono nuclear	196
Tabla 3-44: Número y porcentaje de cada tipo de tono final	197
Tabla 3-45: Número de tonos iniciales por tipo de enunciado	197
Tabla 3-46: Porcentaje de tonos iniciales por tipo de enunciado	197
Tabla 3-47: Número de tonos nucleares por tipo de enunciado	197
Tabla 3-48: Porcentaje de tonos nucleares por tipo de enunciado	198
Tabla 3-49: Número de tonos finales por tipo de enunciado	198
Tabla 3-50: Porcentaje de tonos finales por tipo de enunciado	198
Tabla 3-51: Número de tonos nucleares combinados con cada tipo de tono final	199
Tabla 3-52: Porcentaje de tonos nucleares combinados con cada tipo de tono final	199
Tabla 3-53: Número de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos	199
Tabla 3-54: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos	200
Tabla 3-55: Número de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos	200
Tabla 3-56: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos	200
Tabla 3-57: Número de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos	200
Tabla 3-58: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos	201
Tabla 3-59: Número de tonos nucleares combinados con cada tipo de tono final en enunciados neutros	201
Tabla 3-60: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados neutros	201

Tabla 3-61: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres	202
Tabla 3-62: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres	202
Tabla 3-63: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres	202
Tabla 3-64: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres	203
Tabla 3-65: Número de grupos rítmicos en enunciados producidos por mujeres	204
Tabla 3-66: Porcentaje de grupos rítmicos en enunciados producidos por mujeres	205
Tabla 3-67: Número de grupos rítmicos en enunciados producidos por hombres	206
Tabla 3-68: Porcentaje de grupos rítmicos en enunciados producidos por hombres	206
Tabla 3-69: Número total y porcentaje de grupos rítmicos	207
Tabla 3-70: Número total y porcentaje de grupos rítmicos por tipo de enunciado	208
Tabla 3-71: Número de grupos rítmicos idénticos	209
Tabla 3-72: Promedio de sílabas por enunciado (mujeres)	210
Tabla 3-73: Promedio de sílabas por enunciado (hombres)	211
Tabla 3-74: Promedio de sílabas por tipo de enunciado	211
Tabla 3-75: Número de enunciados de cada tipo por persona (mujeres)	211
Tabla 3-76: Número de enunciados de cada tipo por persona (hombres)	212
Tabla 3-77: Duración total de sílabas por persona (en segundos y minutos)	212
Tabla 3-78: Duración silábica máxima, mínima y promedio (segundos)	213
Tabla 3-79: Duración silábica máxima, mínima y promedio por tipo de enunciado (segundos)	213
Tabla 3-80: Número máximo, mínimo y promedio de sílabas por segundo	214
Tabla 3-81: Número máximo, mínimo y promedio de sílabas por segundo y por tipo de enunciado	214
Tabla 3-82: Variación máxima y mínima promedio de la duración silábica	215
Tabla 3-83: Variación máxima y mínima promedio de la duración silábica por tipo de enunciado	215
Tabla 3-84: Desviación estándar máxima, mínima y promedio de la duración silábica	215
Tabla 3-85: Desv. est. máxima, mínima y promedio de la duración silábica por tipo de enunciado	216
Tabla 3-86: Intensidad silábica máxima, mínima y promedio (decibeles)	216
Tabla 3-87: Intensidad silábica máxima, mínima y promedio por tipo de enunciado (dB)	217
Tabla 3-88: Variación máxima y mínima de la intensidad silábica	217
Tabla 3-89: Variación máxima y mínima de la intensidad silábica por tipo de enunciado	217
Tabla 3-90: Desviación estándar máxima, mínima y promedio de la intensidad silábica	218
Tabla 3-91: Desviación estándar máxima, mínima y promedio de la intensidad silábica por tipo de enunciado	218
Tabla 3-92: Intensidad silábica estandarizada máxima, mínima y promedio (dB inicial = 60)	219
Tabla 3-93: Intensidad silábica estandarizada máxima, mínima y promedio por tipo de enunciado (dB inicial = 60)	219
Tabla 3-94: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada	219
Tabla 3-95: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada por tipo de enunciado	220
Tabla 3-96: Frecuencia fundamental máxima, mínima y promedio (hertz)	221

Tabla 3-97: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (Hz)	221
Tabla 3-98: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (mujeres) (Hz)	221
Tabla 3-99: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (hombres) (Hz)	222
Tabla 3-100: Variación máxima y mínima promedio de la frecuencia fundamental	222
Tabla 3-101: Variación máxima y mínima promedio de la frecuencia fundamental por tipo de enunciado	223
Tabla 3-102: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental	223
Tabla 3-103: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado	223
Tabla 3-104: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (mujeres)	224
Tabla 3-105: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (hombres)	224
Tabla 3-106: Frecuencia estandarizada máxima, mínima y promedio (F_0 inicial = 100)	225
Tabla 3-107: Frecuencia estandarizada máxima, mínima y promedio por tipo de enunciado (F_0 inicial = 100)	225
Tabla 3-108: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada	225
Tabla 3-109: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada por tipo de enunciado	226
Tabla 3-110: Número y porcentaje de cada tipo de tono inicial	226
Tabla 3-111: Número y porcentaje de cada tipo de tono prenuclear	227
Tabla 3-112: Número y porcentaje de cada tipo de tono nuclear	227
Tabla 3-113: Número y porcentaje de cada tipo de tono final	227
Tabla 3-114: Número de tonos iniciales por tipo de enunciado	227
Tabla 3-115: Porcentaje de tonos iniciales por tipo de enunciado	228
Tabla 3-116: Número de tonos nucleares por tipo de enunciado	228
Tabla 3-117: Porcentaje de tonos nucleares por tipo de enunciado	228
Tabla 3-118: Número de tonos finales por tipo de enunciado	229
Tabla 3-119: Porcentaje de tonos finales por tipo de enunciado	229
Tabla 3-120: Número de tonos nucleares combinados con cada tipo de tono final	229
Tabla 3-121: Porcentaje de tonos nucleares combinados con cada tipo de tono final	229
Tabla 3-122: Número de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos	230
Tabla 3-123: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos	230
Tabla 3-124: Número de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos	230
Tabla 3-125: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos	231
Tabla 3-126: Número de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos	231
Tabla 3-127: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos	231
Tabla 3-128: Número de tonos nucleares combinados con cada tipo de tono final en enunciados neutros	231

Tabla 3-129: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados neutros	232
Tabla 3-130: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres	232
Tabla 3-131: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres	232
Tabla 3-132: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres	233
Tabla 3-133: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres	233
Tabla 3-134: Número de grupos rítmicos en enunciados producidos por mujeres	234
Tabla 3-135: Porcentaje de grupos rítmicos en enunciados producidos por mujeres	235
Tabla 3-136: Número de grupos rítmicos en enunciados producidos por hombres	236
Tabla 3-137: Porcentaje de grupos rítmicos en enunciados producidos por hombres	236
Tabla 3-138: Número total y porcentaje de grupos rítmicos	237
Tabla 3-139: Número total y porcentaje de grupos rítmicos por tipo de enunciado	238
Tabla 3-140: Número de grupos rítmicos idénticos	239

Índice de Gráficas

Gráfica 3-1: Duración silábica	242
Gráfica 3-2: Número de sílabas por segundo	243
Gráfica 3-3: Rango promedio de variación y desviación estándar de la duración	244
Gráfica 3-4: Intensidad silábica	246
Gráfica 3-5: Intensidad silábica estandarizada	247
Gráfica 3-6: Rango promedio de variación y desviación estándar de la intensidad	248
Gráfica 3-7: Frecuencia fundamental	249
Gráfica 3-8: Frecuencia fundamental (mujeres)	250
Gráfica 3-9: Frecuencia fundamental (hombres)	251
Gráfica 3-10: Frecuencia fundamental estandarizada	252
Gráfica 3-11: Desviación estándar de la frecuencia fundamental	253
Gráfica 3-12: Rango promedio y desviación estándar de la F_0 estandarizada	254
Gráfica 3-13: Tonos iniciales por tipo de enunciado	255
Gráfica 3-14: Tonos prenucleares y nucleares	256
Gráfica 3-15: Tonos prenucleares y nucleares agrupados	257
Gráfica 3-16: Tonos nucleares por tipo de enunciado	258
Gráfica 3-17: Tonos nucleares agrupados por tipo de enunciado	258
Gráfica 3-18: Tonos finales por tipo de enunciado	260
Gráfica 3-19: Tonos nucleares por cada tono final	261
Gráfica 3-20: Tonos nucleares agrupados por cada tono final	261
Gráfica 3-21: Tonos nucleares por cada tono final en enunciados interrogativos	262
Gráfica 3-22: Tonos nucleares agrupados por cada tono final en enunciados interrogativos	263
Gráfica 3-23: Tonos nucleares por cada tono final en enunciados exclamativos	264
Gráfica 3-24: Tonos nucleares agrupados por cada tono final en enunciados exclamativos	264
Gráfica 3-25: Tonos nucleares por cada tono final en enunciados suspendidos	265
Gráfica 3-26: Tonos nucleares agrupados por cada tono final en enunciados suspendidos	266
Gráfica 3-27: Tonos nucleares por cada tono final en enunciados neutros	266
Gráfica 3-28: Tonos nucleares agrupados por cada tono final en enunciados neutros	267
Gráfica 3-29: Tonos nucleares por cada tono final en enunciados producidos por mujeres	268
Gráfica 3-30: Tonos nucleares agrupados por cada tono final en enunciados producidos por mujeres	268
Gráfica 3-31: Tonos nucleares por cada tono final en enunciados producidos por hombres	269
Gráfica 3-32: Tonos nucleares agrupados por cada tono final en enunciados producidos por hombres	270
Gráfica 3-33: Grupos rítmicos por género	272
Gráfica 3-34: Grupos rítmicos por tipo de enunciado	273
Gráfica 3-35: Grupos rítmicos idénticos incluyendo el grupo 0	274
Gráfica 3-36: Grupos rítmicos idénticos	275

0. Introducción

En un comedor universitario se encuentra una persona sentada sola en una mesa leyendo una revista en español. Alguien se acerca y le pregunta si puede sentarse en la misma mesa y, al hacerlo, se da cuenta de que la revista está escrita en su lengua materna. Comienzan a conversar. Tan sólo necesitan intercambiar unas palabras para darse cuenta de que uno de ellos es, por ejemplo, mexicano y el otro español.

Esta situación tan trivial nos permite suponer que los participantes de una conversación no únicamente interpretan el contenido de los enunciados producidos por su interlocutor, sino que también, basados en su experiencia personal, intuyen actitudes y emociones en el habla de los demás participantes de la conversación. Además, cuando un oyente percibe diferencias a nivel fónico (y prosódico), morfológico, sintáctico o semántico con respecto a su manera propia de utilizar la misma lengua, intenta identificar los rasgos o unidades diferentes, clasificarlos por categorías, y compararlos con su realización personal y con otros sistemas conocidos. Al final, cada interlocutor irá comprobando sus hipótesis sobre el origen social y geográfico de los demás participantes de la conversación haciendo uso de cada nueva pista.

Este proceso de identificación del sociolecto o regiolecto del interlocutor, por lo general, suele durar tan sólo unos segundos, dependiendo de la familiaridad de un oyente con el sistema lingüístico del hablante, y, la mayoría de las veces, se da de manera inconsciente.

Sin embargo, cuando los hablantes imitan el habla típica de una determinada región o grupo social determinado distintos de los suyos, la adaptación de las características de su propia variedad de lengua a lo que considera inherente a la variedad imitada es la mejor prueba de que, además de las diferencias que presenta cada variedad en el plano segmental, hay rasgos prosódicos que pueden ser percibidos, abstraídos e incluso reproducidos. Por lo tanto, estos rasgos característicos del habla deben tener correlatos acústicos fácilmente observables y cuantificables.

0.1 Preguntas de investigación

Para estudiar científicamente el proceso de identificación descrito anteriormente es necesario plantearse las siguientes preguntas de investigación:

P₁: ¿Qué factores influyen en la percepción de las diferencias entre las variedades lingüísticas del hablante y del oyente, en especial con respecto al nivel prosódico?

- P₂: ¿Qué enfoques teóricos ofrecen los modelos de representación fonológica de la prosodia más fácilmente aplicables al estudio de la lengua oral y a su transcripción en medios electrónicos?
- P₃: ¿Qué factores y modelos teóricos permiten hacer una mejor comparación de las características prosódicas entre las variedades regionales de una lengua?
- P₄: ¿Qué características debe tener un sistema de transcripción para que, según los enfoques teóricos que serán utilizados en la investigación, permita la representación y lectura de los distintos fenómenos prosódicos y sus alteraciones lingüísticamente marcadas, y que sea extensible para que en el futuro sea capaz de incluir otros enfoques teóricos, otras variedades del español o incluso otras lenguas?
- P₅: ¿Cuáles pasos del análisis es necesario que sean automatizados para permitir el uso de *corpora* lingüísticos y bases de datos lo suficientemente extensos, y hasta qué grado es pertinente que se automaticen?

0.2 Objetivos

Las cinco preguntas anteriores dan origen a los objetivos que se persiguen en esta investigación:

- O₁: Establecer un conjunto de factores prosódicos que den cuenta de las diferencias que los oyentes perciben en la producción lingüística del hablante.
- O₂: Presentar y comparar los modelos de representación de la prosodia de la lengua oral, en especial los que ya han sido aplicados al español.
- O₃: Aplicar los modelos más aptos al estudio comparativo de la prosodia de las variedades castellana y mexicana del español.
- O₄: Crear un sistema de transcripción electrónica que permita codificar y decodificar distintos fenómenos prosódicos y que sea fácilmente extensible.
- O₅: Automatizar la mayor cantidad de pasos del proceso de análisis computacional que no requieran del criterio del investigador.

0.3 Hipótesis

Del mismo modo, partiendo de las preguntas y objetivos de la investigación, se originan las hipótesis que serán el eje central del presente trabajo.

- H₁: Los factores prosódicos que influyen en la comunicación son la entonación, el ritmo y el tempo, la intensidad y la calidad de la voz.

- H₂: Los modelos teóricos que mejor se aplican al estudio y representación electrónica de la entonación del español son el métrico autosegmental, Sp-ToBI, y el de los contornos melódicos con frecuencia fundamental estandarizada; para el ritmo, el modelo más apto es el de los grupos rítmicos con base sintagmática; la mejor manera de analizar el tempo y la intensidad es por medio de la variación porcentual de una sílaba con respecto a la anterior; la calidad de la voz no es fácilmente analizable en grabaciones espontáneas, al menos de manera significativa.
- H₃: Dichos modelos son suficientes para dar cuenta de las similitudes funcionales y las diferencias regionales de la prosodia del español oral de la Ciudad de México y Madrid.
- H₄: Haciendo uso del lenguaje XML, es posible representar jerárquicamente los distintos fenómenos prosódicos que aparecen en el habla espontánea y especificar por medio de variables sus características acústicas.
- H₅: El programa Praat es el más adaptado para realizar los análisis acústicos y crear automáticamente documentos de XML con la información necesaria para permitir análisis detallados de la entonación, el ritmo, el tempo y la intensidad de las producciones lingüísticas de los hablantes castellanos y mexicanos.

0.4 Justificación

El tema de la prosodia es relativamente reciente, sobre todo si se considera únicamente los estudios sobre el español, y muy pocos investigadores se han ocupado de la transcripción sistemática de los fenómenos prosódicos que componen una conversación espontánea. En este sentido, la investigación es de carácter exploratorio, y más adelante, el enfoque será más bien descriptivo y correlacional.

Se ha considerado trascendente realizar esta investigación porque, por una parte, en la actualidad sigue habiendo discrepancias entre los distintos enfoques teóricos sobre la prosodia (habla de laboratorio vs. habla espontánea, configuraciones vs. categorías discretas, etc.) y, por otra, aunque se han creado sistemas de transcripción jerárquicos y temporales, incluso haciendo uso del lenguaje XML, en ninguno de ellos se ha profundizado en la representación de la prosodia a nivel silábico.

Por lo tanto, se consideró necesario, primeramente, iniciar una búsqueda de los modelos de representación fonológica de la prosodia más fácilmente aplicables al estudio del español, del habla espontánea y cuya representación en medios electrónicos fuera lo más simple posible.

Además, puesto que en la actualidad el uso de las computadoras es imprescindible para los análisis acústicos, era necesario pensar en una solución global completamente computarizada,

desde la grabación digital de las conversaciones, pasando por la transcripción y la representación jerárquica de los fenómenos lingüísticos, hasta la visualización de los datos transformados por medio de hojas de estilo y su utilización como bases de datos para distintos tipos de análisis estadísticos.

Se intentó utilizar tecnologías al alcance de cualquier investigador, como son los lenguajes XML, XSLT, JavaScript, HTML, así como el programa Praat, para asegurar que los contenidos e ideas generadas en este trabajo sean fácilmente aplicables y reproducibles.

El fin de la investigación es ofrecer nuevas posibilidades de tratamiento de la información prosódica y promover el uso de la tecnología en este fascinante campo de estudio.

1. Marco teórico

El campo de la prosodia ha sido estudiado a lo largo de las últimas décadas desde muy variados enfoques y con distintas finalidades. Esta variedad de posturas, lejos de entorpecer el avance en la búsqueda de conocimiento, ha fomentado aún más teorías y conceptos.

En esta primera sección se hará una revisión de la literatura sobre la prosodia y se dividirá en cinco grandes apartados:

En el primero de ellos se hará una introducción general a la prosodia, tomando en cuenta temas como los rasgos prosódicos y las unidades prosódicas: acento, patrón entonativo, cantidad y juntura. Más adelante se analizarán y ejemplificarán las funciones comunicativas de la prosodia. Finalmente, se hará un repaso y una crítica de la clasificación de los constituyentes prosódicos: sílaba, pies, palabra fonológica, grupo clítico, frase fonológica, frase entonativa y enunciado fonológico.

El segundo apartado, cuyas dimensiones hacen suponer que es el tema principal del capítulo, tratará el tema de la entonación, que es el aspecto de la prosodia más estudiado en la actualidad. En dicha sección se hará un repaso de distintos modelos teóricos de la entonación que han tenido mayor o menor influencia en los estudios de la entonación del español: la escuela británica, la escuela americana, la teoría métrica autosegmental, la escuela holandesa y el modelo de Aix-en-Provence.

En el tercer apartado se hará una revisión de los distintos enfoques desde los cuáles se ha estudiado el ritmo y el tempo, con la finalidad de sentar las bases para llevar a cabo un estudio de tales aspectos en el español oral.

El cuarto y quinto apartado tratarán de la intensidad y la calidad de la voz, respectivamente. En este último apartado se revisará también brevemente la prosodia de las emociones.

Los contenidos teóricos en este capítulo serán evaluados, así como las aplicaciones de los distintos modelos al estudio de la prosodia del español, para justificar la elección de algunos de los enfoques que serán utilizados más adelante en esta investigación.

1.1 Prosodia

El concepto **prosodia** se deriva de la palabra griega *προσῳδία* (*prosodía*), un término musical que parece haber significado ‘canto para acompañar la música’, implicando también que la prosodia era el acompañamiento musical de las palabras mismas. Más recientemente se ha utilizado el término para referirse a los principios de versificación, cubriendo aspectos tales como patrones rítmicos, esquemas de rimas y estructuras de versos. Sin embargo, en contextos lingüísticos se le puede encontrar más frecuentemente con un significado distinto, haciendo referencia a características de enunciados tales como la prominencia acentual y la entonación, entre otras, y sobre todo, aplicados más a la prosa que al verso. No obstante, hacer una caracterización precisa de los alcances y el significado lingüístico actual de la prosodia, así como su delimitación con respecto a otros términos relacionados, es una tarea muy difícil.

Según Fox (2000: 1-2), tradicionalmente, la descripción del habla se ha hecho mayoritariamente en términos de rasgos segmentales, es decir, a base de vocales y consonantes, así como sus atributos. Tal vez es comprensible, ya que son estos los elementos que se representan en la ortografía alfabética y son los responsables de diferenciar las palabras entre sí. Los rasgos prosódicos, a primera vista más intangibles y menos significativos, fueron fácilmente ignorados y su contribución subestimada. Las escuelas estructuralistas clásicas de la fonología basaron sus teorías en el *fonema*, la unidad segmental distintiva básica. Por lo tanto, si es que las características tales como la prominencia y la entonación se estudiaron, fue en términos de una fonología a base de fonemas, o sea, segmental. En la fonología estructuralista americana, estos rasgos se estudiaron más bien como tipos especiales de fonemas: *fonemas suprasegmentales*, que, según esta escuela, se encontraban ‘por encima’ de los fonemas segmentales. Los fonólogos de la Escuela de Praga fueron, por lo general, menos rigurosos y tuvieron una concepción más flexible de la estructura del habla que les permitió tratar los rasgos prosódicos por separado de los segmentos, pero no ofrecieron un marco teórico coherente para la descripción de dichos rasgos. La fonología generativa clásica adoptó la concepción de los rasgos suprasegmentales como atributos de unidades segmentales, y aunque este modelo le dedicó una atención considerable a la formalización de los procesos de asignación de prominencia, hizo pocos avances en el estudio de la naturaleza o rol real de los rasgos prosódicos. En los modelos recientes, llamados colectivamente ‘no lineales’, se ha llegado más lejos, ya que han adoptado explícitamente marcos teóricos más complejos y multidimensionales que van más acordes con la naturaleza de la prosodia.

A pesar de que los términos ‘suprasegmental’ y ‘prosódico’ coinciden en gran medida, es útil y deseable diferenciarlos. Para comenzar, una simple dicotomía ‘segmental’ vs. ‘suprasegmental’ no le hace justicia a la riqueza de estructura fonológica ‘por encima’ del segmento. Como veremos más adelante, esta estructura es bastante compleja, ya que involucra varias dimensiones distintas, y por lo tanto, los rasgos prosódicos no pueden ser vistos únicamente como rasgos que se superponen a los segmentos. Aún más importante, se puede hacer una distinción entre ‘suprasegmental’ como modo de descripción por un lado y ‘prosódico’ como un tipo de rasgo por el otro. Por ejemplo, algunos rasgos tales como la nasalidad o la sonoridad pueden ser estudiados suprasegmentalmente cuando se extienden más allá de los límites de un solo segmento. Tales rasgos no son prosódicos, y no obstante, pueden ser analizados suprasegmentalmente. (Fox 2000: 2-3)

Se mencionó que la lengua oral es fonológicamente compleja, debido a las varias dimensiones simultáneas. La estructura prosódica de los enunciados se representa preferentemente en distintos niveles, no solamente como estratos independientes, sincronizados de algún modo unos con otros, sino de manera mucho más interdependiente, formando juntos una estructura jerárquica integral. Si es que los hay, son pocos los rasgos prosódicos simples y unidimensionales. Se puede decir que la entonación es el resultado de un único parámetro, en este caso la altura tonal de la voz, mientras que el ritmo, por ejemplo, es un rasgo complejo que resulta de la interacción del acento, la longitud segmental y el tempo o velocidad del habla, además de que el acento por sí mismo ni siquiera se puede definir como un único parámetro. En realidad, incluso los rasgos aparentemente simples, como la entonación, involucran más de una dimensión, ya que no se puede separar claramente del acento y la calidad de la voz. Por lo tanto, no se puede determinar cuántos rasgos prosódicos individuales deberían de ser tomados en cuenta, y por eso, normalmente se parte de un pequeño número de parámetros acústicos, tales como la longitud segmental (o silábica), la frecuencia fundamental y la intensidad.

Otra dificultad más es que los rasgos prosódicos tienen una clara afinidad con los rasgos paralingüísticos y es difícil establecer las fronteras entre ellos. El ‘tono de voz’ se expresa tanto por medio de rasgos prosódicos, como la entonación, como por medio de rasgos paralingüísticos, como la calidad de la voz, y de hecho, tanto la altura como la calidad de la voz pueden tener funciones tanto lingüísticas como no lingüísticas. (Fox 2000: 9-11)

En las siguientes secciones se hará un repaso de los conceptos más importantes para el estudio de la prosodia. Primero se diferenciará entre rasgos y unidades prosódicas. Más adelante se hará un recuento de las funciones lingüísticas más importantes que cumple la prosodia.

También se analizará el orden jerárquico de los constituyentes prosódicos. Enseguida se hará un acercamiento más profundo al tema de la entonación, que por sí solo representa el tema más importante de esta investigación. Finalmente, se revisará el estado actual de las investigaciones sobre el ritmo y el tempo por una parte, y por la otra, se verán otros aspectos menos estudiados, como el volumen y la calidad de la voz o timbre.

1.1.1 Rasgos prosódicos y unidades prosódicas

Para sistematizar los numerosos fenómenos prosódicos, Rabanus (2001: 4-9) distingue entre *rasgos* prosódicos y *unidades* prosódicas. Los rasgos prosódicos son fenómenos fonéticos que en un modelo de tres niveles se pueden describir como rasgos auditivos, acústicos y articulatorios. Las unidades prosódicas, en cambio, son fenómenos fonológicos. Se habla de unidades prosódicas cuando estos fenómenos satisfacen el criterio de distintividad fonológica, es decir, cuando distinguen funciones lingüísticas (sintácticas, semánticas y pragmáticas). En otras palabras, las unidades prosódicas son rasgos prosódicos con función lingüística. En cambio, como se verá más adelante, entre las unidades prosódicas y los rasgos acústicos, no hay una relación tan clara y directa (véase la Ilustración 1-1).

Como se mencionaba antes, en fonética se distingue básicamente tres aspectos del habla: el aspecto articulatorio o producción del habla; el aspecto acústico o rasgos mensurables, es decir, la sustancia del habla, y el aspecto auditivo o percepción del habla. Los rasgos prosódicos que sirven para la descripción del nivel fonético de la prosodia deberán estudiarse de tal manera que se abarque con ellos estas tres facetas del habla (Kehrein 2002: 79).

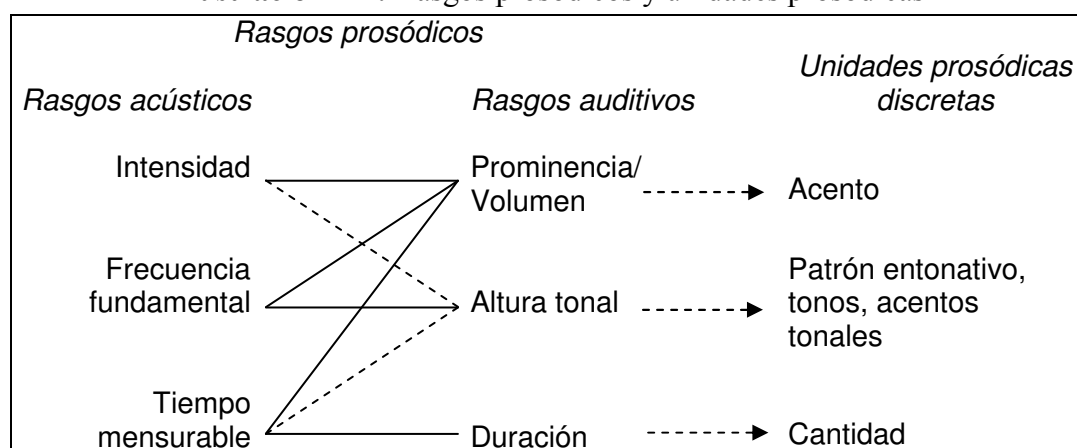
A nivel fonológico o funcional define Kehrein (2002: 82) las unidades prosódicas como magnitudes lingüísticas apoyadas en rasgos prosódicos que a nivel léxico tienen una función distintiva y a nivel oracional tienen una variedad de funciones y significados lingüístico-comunicativos básicos. Por funciones y significados básicos este autor entiende los significados de las unidades prosódicas que permanecen constantes en la mayoría de los contextos. Conjuntamente con significados básicos de unidades de otros sistemas de comunicación (lenguaje verbal, kinésico, proxémico, etc.) en el contexto lingüístico y situacional dan lugar a un significado contextual perceptible de una oración. Este es el principio de la percepción integral en la interacción comunicativa. En este sentido, las unidades prosódicas se entienden como elementos integrales del sistema lingüístico y por lo tanto son potencialmente sensibles al contexto.

Kehrein (2002: 83-84) propone una división de las unidades prosódicas según tres rasgos: 1) su extensión, 2) la manera en la que se coordinan su forma y función, y 3) su correlato auditivo primario.

- 1) *Extensión*: Las unidades prosódicas pueden extenderse sobre cadenas de segmentos de distintas longitudes. Son locales cuando se relacionan con sílabas aisladas y globales cuando abarcan varias sílabas o cadenas de segmentos.
- 2) *Coordinación de forma y función*: La coordinación de las unidades puede ser discreta o gradual/continua. Discreto significa que a partir de cierto grado de variación del rasgo se percibe una función o no, o se percibe ya sea como función A o como función B. La coordinación continua hace que la mayor o menor expresión de un rasgo modifique el significado básico, como la elevación o disminución de la velocidad del habla.
- 3) *Correlato auditivo primario*: Mientras que entre los rasgos prosódicos acústicos y auditivos existen relaciones complejas, para las unidades prosódicas se puede asignar de manera relativamente segura correlatos auditivos primarios. Por lo tanto, se puede diferenciar entre unidades prosódicas de la altura tonal (patrón entonativo, tonos, acentos tonales), de la prominencia (acento léxico, acento oracional, acento enfático) y de la duración segmental (corta vs. larga).

En la siguiente ilustración se resumen las relaciones entre rasgos y unidades prosódicas:

Ilustración 1-1: Rasgos prosódicos y unidades prosódicas



Fuente: Diseño propio, basado en Kehrein (2002: 99)

La ilustración contiene todos los aspectos importantes que deben ser tomados en cuenta para el análisis de la prosodia:

- Las unidades prosódicas del lado funcional y los rasgos prosódicos acústicos y auditivos como dos niveles de descripción del lado formal deben mantenerse separados (también terminológicamente).

- Entre los rasgos auditivos perceptibles y los rasgos acústicos mensurables no hay correlaciones directas, sino relaciones parcialmente complejas.
- Las unidades prosódicas se identifican exclusivamente de manera auditiva según su correlato auditivo primario: acento – prominencia/volumen; patrón entonativo (así como tonos y acentos tonales) – altura tonal; cantidad – duración de unidades segmentales.
- En el caso de análisis acústicos complementarios de las unidades prosódicas debe tomarse en cuenta que al correlato auditivo de una unidad prosódica le corresponde un complejo de varios rasgos acústicos. (Kehrein 2002: 99-100)

A continuación se tratará en detalle cada una de las unidades prosódicas de manera individual.

1.1.1.1 Acento

El acento se conceptualiza de maneras variadas en la literatura, pero la distinción más importante es la que se da entre acento léxico y acento tonal. Acento es el énfasis de una o más unidades lingüísticas (sílabas, palabras o constituyentes) por medios fonéticos. El rasgo auditivo correspondiente al acento es la prominencia. (Rabanus 2001:8)

El estudio de la prominencia léxica arroja luz sobre la ubicación de las sílabas prominentes dentro de las palabras, es decir, sobre los factores rítmicos, posicionales, cuantitativos y morfológicos que rigen los patrones de prominencia silábica. Aunque la realidad mental de la prominencia sea indudable, aún no se ha descubierto un correlato fonético evidente. Las sílabas prominentes son potencialmente capaces de conllevar movimientos tonales con una fuerte carga perceptiva, además de que suelen tener una mayor duración e intensidad, aunque estos dos últimos factores normalmente se subordinan a la altura tonal.

Kager (1995: 367-368) menciona cinco características del acento. Primero, el acento es culminativo, es decir, en las lenguas con sistemas acentuales, cada palabra (o núcleo de sintagma) tiene al menos una sílaba prominente. Segundo, el acento es jerárquico, ya que se puede dar una jerarquía de prominencias entre múltiples acentos. Tercero, el acento funge como delimitador en los sistemas en los que se le utiliza para marcar los finales de palabra. Cuarto, el acento es rítmico en los sistemas en los que se alternan las sílabas acentuadas e inacentuadas, y donde se evitan los choques de acentos adyacentes. Naturalmente, el acento no se asimila a las sílabas adyacentes, pues esto produciría choques acentuales. Quinto, los contrastes acentuales tienden a ampliarse a nivel segmental: las sílabas prominentes pueden reforzarse por medio del alargamiento vocálico o geminación, mientras que las sílabas sin prominencia pueden debilitarse por medio de la reducción vocálica.

1.1.1.2 Patrón entonativo

El patrón entonativo o entonación es el uso funcional de la curva melódica y se relaciona con el rasgo auditivo de la altura tonal y con el fenómeno articulatorio de la vibración de las cuerdas vocales en la laringe. De los rasgos acústicos con los que se relaciona la entonación, la frecuencia fundamental (F_0) es el más importante.

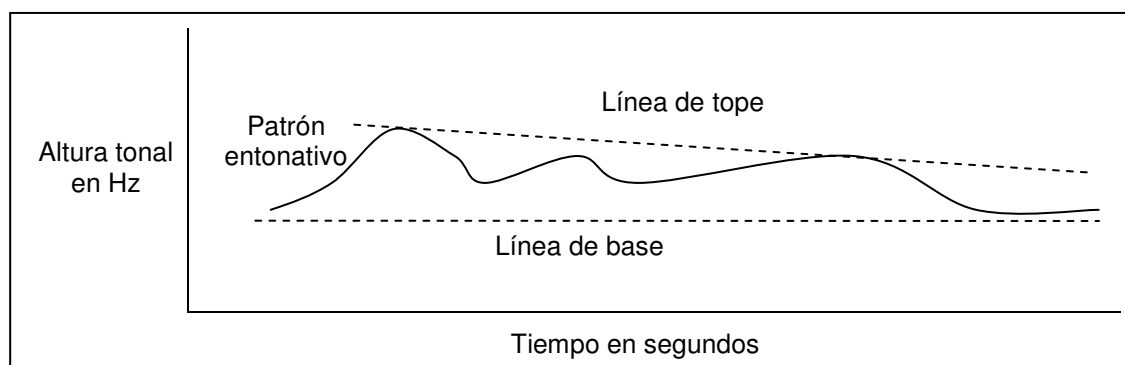
Como se verá más adelante, por medio de los tonos se pueden distinguir funciones sintácticas, semántico-discursivas y contextuales (véase 1.1.2). Dichas funciones se estudian en la literatura (Rabanus 2001:139-143) no por medio de tonos o contornos aislados, sino por medio de secuencias entonativas.

Para la percepción de la señal tonal es determinante la frecuencia fundamental, nombre con el que se designa el tono básico dentro de un espectro sonoro. Incluso si otros eventos sonoros también son físicamente analizables, los tonos parciales y las proporciones de los formantes solamente influyen en la percepción del timbre (véase 1.4).

En ciertas áreas de frecuencia, la diferenciación tonal es mucho más sensible, sobre todo en el rango entre los 100 y 250 hercios (Hz), pues es el intervalo dentro del cual se encuentra la altura media de la voz humana. La representación del patrón entonativo en hercios ofrece una orientación mucho más precisa que otro tipo de notaciones musicales, como los semitonos.

Un fenómeno intrínseco de la entonación es la declinación. A grandes rasgos, el patrón entonativo se podría representar como una línea ondulada. Si se unen todos los picos más altos de un enunciado, se obtiene una línea de tope (*topline*), mientras que al unir todos los valles más profundos se obtiene una línea de base (*baseline*). Estas dos líneas se encontrarán colocadas de manera casi paralela, pues las dos irán descendiendo y al final del enunciado tenderán a acercarse:

Ilustración 1-2: Declinación



Fuente: Diseño propio

Los primeros análisis de la frecuencia fundamental se realizaron de manera auditiva, a veces con ayuda de instrumentos de afinación. Sin embargo, en las últimas décadas para el análisis

instrumental de la frecuencia el uso de las computadoras se ha vuelto indispensable. No obstante, la complementación de procedimientos objetivos (acústicos) con métodos subjetivos (auditivos) no debe de descartarse. (Kranich 2003: 19-23)

Según esta idea, para la presente investigación se creó la siguiente fórmula de estandarización de los valores puramente fonéticos de la frecuencia fundamental, que:

$$\overline{f0}_{std} = \frac{\sum f0_i}{n} * 100 / X_1 \quad (1)$$

Con esta fórmula, que será aplicada en la sección 3, se propone asignar al valor de la primera sílaba de cada enunciado (X_1) un valor preestablecido de 100 Hz y ajustar los valores de las siguientes sílabas según las diferencias relativas y con respecto al valor inicial (100 Hz) para así obtener el promedio de estos valores estandarizados.

1.1.1.3 Cantidad

La cantidad designa las diferencias abstractas de duración, que se establecen en una lengua con la finalidad de distinguir significados. La función gramatical de esta unidad se basa en la distinción léxica. El alemán, por ejemplo, muestra una oposición binaria de cantidad entre vocales acentuadas cortas y largas. En italiano la oposición se da entre consonantes, y en lenguas como el finlandés, el estonio y el húngaro se dan oposiciones tanto vocálicas como consonánticas. Sin embargo, en español, a pesar de influir en la calidad de voz, esta unidad carece de importancia pues la lengua no posee oposiciones funcionales de cantidad segmental. A pesar de que la cantidad parezca un rasgo fonológico segmental, pertenece a las unidades prosódicas, ya que se necesita una valoración relativa para determinar la cantidad de un segmento y el estudio del segmento por sí solo no es suficiente. (Günther 1999:69)

La duración de sonidos, sílabas, palabras, enunciados y pausas es un importante indicador de estructuras discursivas y expresivas, por lo que puede alterar el tempo o velocidad del habla (véase 1.3 y 1.3.4). La duración puede estudiarse explícitamente (por ejemplo, duración segmental o silábica) o transformada en parámetros de velocidad. Un parámetro usual es el basado en intervalos articulatorios como sonidos por segundo, sílabas por segundo o palabras por segundo (o por minuto), sin tomar en cuenta la duración de las pausas, ya que debe tratarse como un parámetro independiente. (Paeschke 2003: 28)

Las pausas por sí solas poseen un gran significado con respecto a la comprensión textual, ya que le conceden a los rasgos más informatividad. Se considerará como pausa toda interrupción intrasegmental auditivamente perceptible del flujo de habla que se representa por medio de silencio, muletillas o una mezcla de ambos componentes. (Kranich 2003: 15)

1.1.1.4 Juntura

Además de las tres unidades prosódicas discutidas hasta ahora, acento, entonación y cantidad, es necesario incluir la juntura. Bajo esta categoría prosódica abstracta se resumen todos los rasgos que sirven para delimitar las fronteras o junturas. Puesto que en la cadena hablada se marcan una gama de fronteras por medios muy distintos, que además interactúan con las realizaciones de otras unidades prosódicas, como la entonación o la cantidad, el área que cubre el fenómeno de las señales de juntura resulta ser muy heterogéneo. Debido a la realización entonativa de las fronteras por medio de tonos de juntura, se debe tomar en cuenta la relación interna entre la juntura y la entonación (Günther 1999: 70-71).

La función gramatical de la juntura consiste en señalar fronteras estructurales, lo que sirve para dividir la cadena hablada de manera que apoye la comprensión y evite las ambigüedades. Con esta finalidad se señalan las fronteras de los constituyentes prosódicos como sílabas, palabras o frases prosódicas, de unidades semánticas y de segmentos discursivos.

Las principales opciones acústico-articulatorias para la delimitación de fronteras son la colocación de pausas; la modulación de la frecuencia fundamental, la duración segmental y el ritmo; y la realización específica de segmentos colocados al inicio o final de sílabas.

1.1.2 Funciones comunicativas de la prosodia

La lengua es un sistema que le sirve al ser humano para expresar y transmitir sus ideas, pensamientos e informaciones. Esta funcionalidad de la lengua dirigida hacia la comunicación evoca sobre todo un análisis funcional como un medio para describir y explicar estructuras y fenómenos lingüísticos. La lengua hablada, vista físicamente como la realización acústica de estructuras lingüísticas abstractas, junto con la lengua escrita y la gética, es una manifestación de estructuras ontológicas que representan el núcleo de un acto lingüístico. En este proceso, la lengua hablada no es solamente la realización sonora monótona de palabras aisladas de una oración. Los sonidos se someten a una modulación global controlada, en la cual se modifican los parámetros de frecuencia, duración e intensidad mediante el control de los sistemas respiratorio y articulatorio. Estas modificaciones y realizaciones específicas de los sonidos son el objeto de los análisis prosódicos para descubrir los determinantes funcionales y estructurales de los fenómenos prosódicos de la lengua hablada (Günther 1999: 21-22).

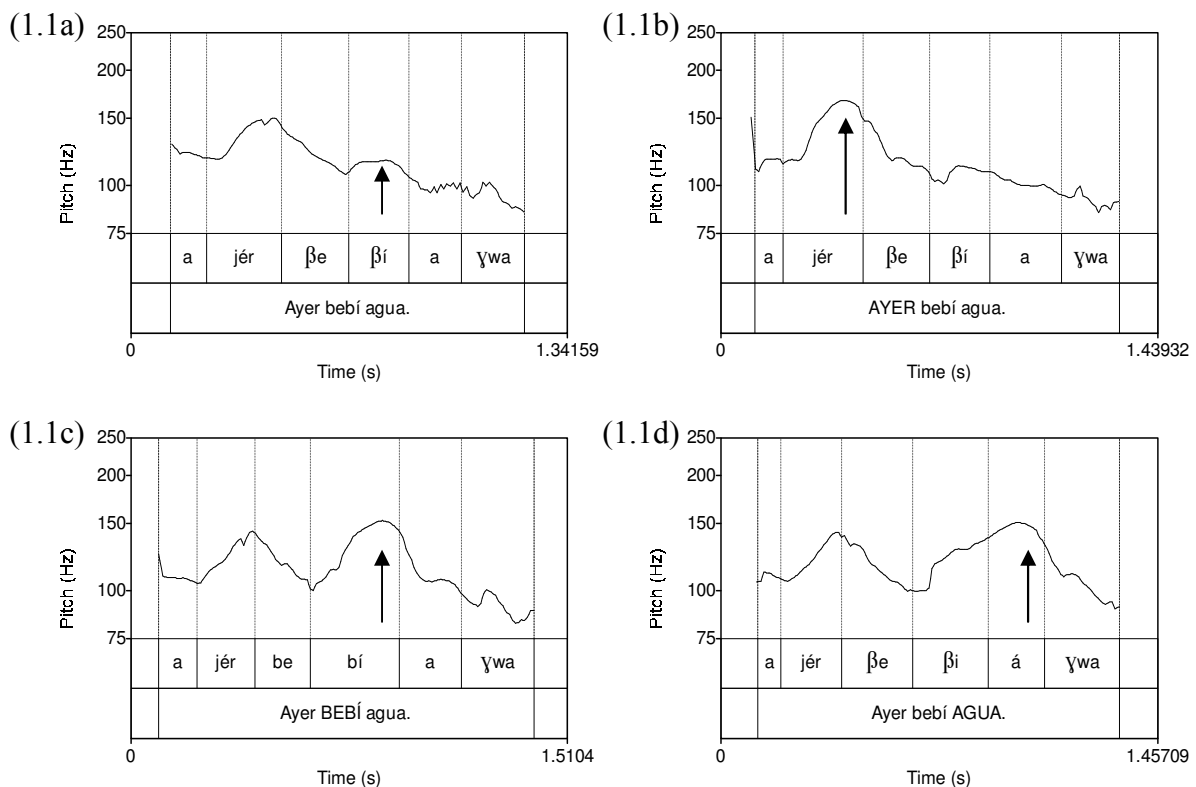
Teniendo en cuenta la función comunicativa primaria de la lengua hablada, los objetivos centrales de la comunicación (la transmisión de información, la expresión de opiniones y la satisfacción de expectativas) están especialmente sujetos al sistema prosódico.

A continuación se presentarán una serie de funciones gramaticales, es decir, semánticas, sintácticas o léxicas que se pueden asignar a los fenómenos prosódicos.

1.1.2.1 Caracterización del esquema foco/trasfondo

Una característica de los enunciados en la lengua hablada es que ciertas partes se realizan de manera más prominente que otras. Esta prominencia fonética se puede atribuir, por medio de rasgos y estructuras gramaticales, a sistemas conceptuales que se relacionan con el esquema informativo del enunciado proyectado. Los elementos del enunciado que se realizan de manera más prominente son las porciones que deben transmitirse como información relevante en el discurso, es decir, información que el emisor supone que el receptor aún no conoce. La diferenciación conceptual entre elementos informativos y no-informativos del enunciado se refleja en el nivel semántico de representación como *esquema foco/trasfondo*, que a su vez origina una realización prosódica distinta de estas particiones. Entre el esquema informativo y la acentuación existe una correlación directa y sistemática. (Günther 1999: 32-33)

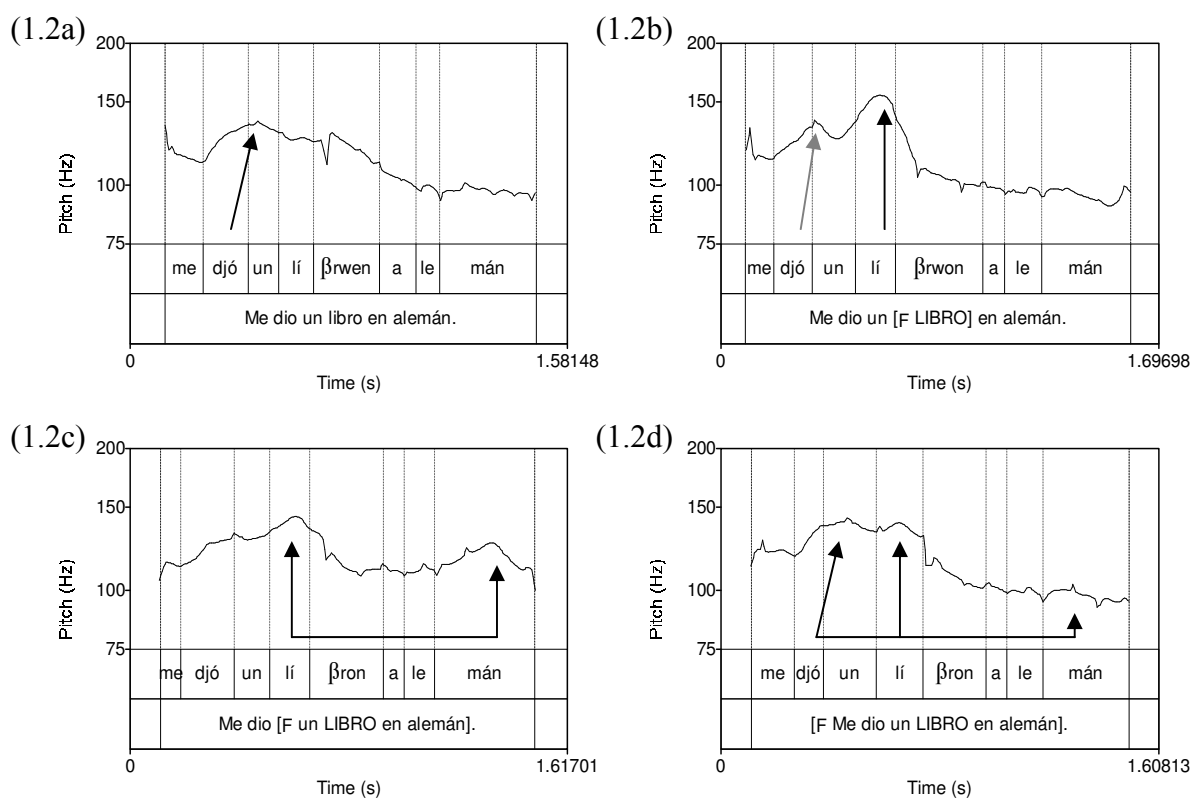
La simple variación de las relaciones de prominencia en dos oraciones con la misma disposición lineal de los mismos lexemas, la misma estructura sintáctica y el mismo valor de veracidad en un análisis semántico-oracional puede evocar distintas presuposiciones por medio de las cuales se podría interpretar la oración teniendo como trasfondo un número determinado de contextos.



1.1 Prosodia

La oración (1.1a) es un enunciado declarativo neutro. En la frecuencia fundamental o F_0 del ejemplo, pronunciado por un individuo masculino de origen mexicano, pueden verse tres prominencias, aunque la tercera (*agua*) se neutraliza parcialmente por la proximidad con la segunda (*bebí*). A grandes rasgos, el contorno se asemeja al de la Ilustración 1-2. El enunciado (1.1b) podría ser la respuesta a la pregunta de retroalimentación ‘¿Cuándo bebiste agua?’ y esta información focalizada se reflejará en un pico aún más alto que en (1.1a) sobre la primera prominencia (*ayer*). Por su parte, (1.1c) sería la respuesta a ‘¿Qué hiciste ayer?’, y (1.1d) a ‘¿Qué bebiste ayer?’. En estos dos últimos ejemplos vemos cómo el primer pico del enunciado es semejante al de (1.1a), seguido de un pico aún mayor en la sílaba prominente de la palabra con la información focalizada. Si la respuesta a estas últimas preguntas hubiera sido (1.1a), daría como resultado una respuesta correcta pero mal formulada.

Sin embargo, el foco por sí solo no siempre se puede equiparar con los elementos oracionales percibidos como prominentes. Por ejemplo, la oración (1.2a), con el acento principal en *dio*, es una respuesta bien formulada para las preguntas ‘¿Qué es lo que te dio?’, ‘¿Qué tipo de libro te dio?’ y ‘¿Por qué estaba ella aquí?’, aunque los elementos focalizados en cada uno de los ejemplos (1.2b, c y d, respectivamente) hacen que cada uno de los enunciados sea una respuesta más apropiada a cada una de las preguntas.

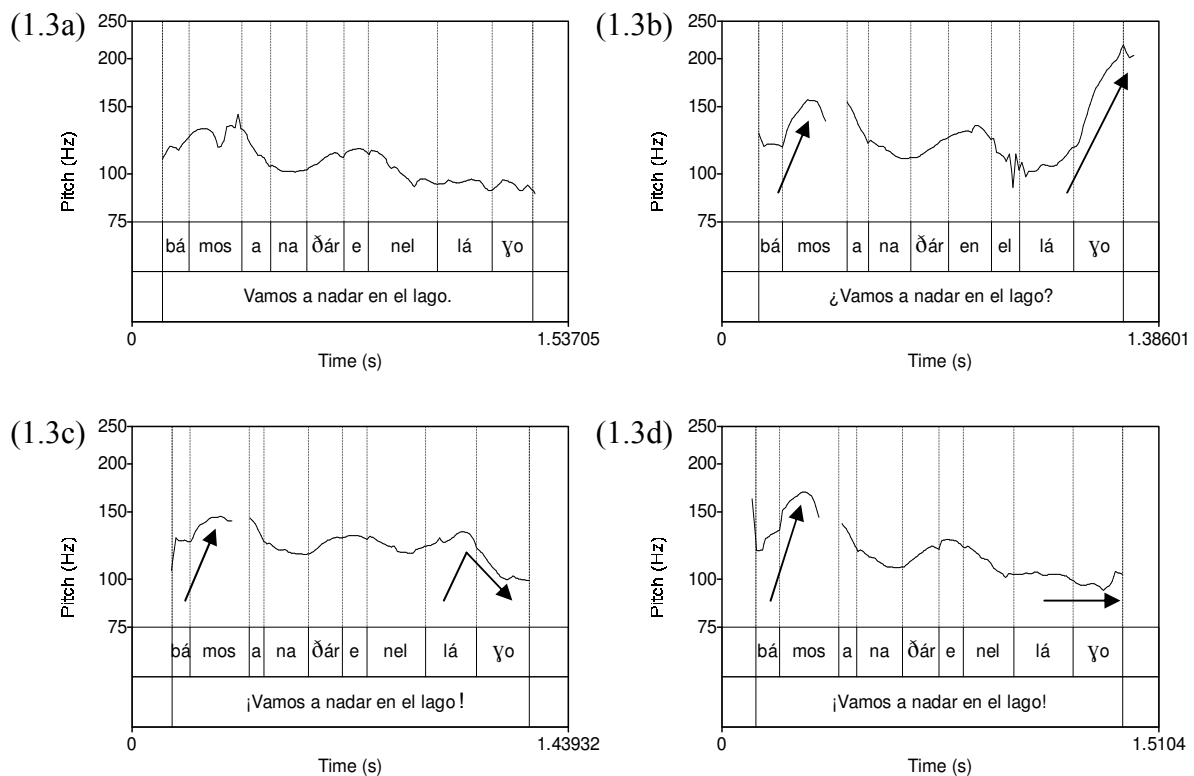


Las estructuras foco/trasfondo de los ejemplos (1.2b-d) se diferencian únicamente con respecto a sus dominios focales (representado por los corchetes indicados con una F). En

(1.2b) se da un foco *estrecho*, en (1.2c) un foco *amplio* y en (1.2d) un foco *neutral*. Por lo tanto, se puede sostener que dentro de un dominio focal solamente la sílaba tónica de un lexema se realiza de manera especialmente prominente, en estos tres ejemplos, el lexema *LIBRO* (cfr. Günther 1999: 33-35).

1.1.2.2 Caracterización del tipo de oración

El término *tipo de oración* es la relación sistemática entre los tipos formales de oraciones establecidos por medio de rasgos gramaticales y sus correspondientes tipos funcionales. Por lo tanto, los tipos de oración en español son el declarativo, interrogativo, imperativo y exclamativo, pues el tipo de función describe el significado de la estructura sintáctica autosuficiente con valor de oración. Es decir, la función de una determinada forma gramatical consiste en contribuir, en el nivel conceptual, al establecimiento de la orientación básica con respecto al contenido proposicional de una oración o del tipo de acto lingüístico. En español, el tipo de la estructura con valor de oración se establece según una serie de rasgos estructurales: contenido categórico (utilización de pronombres interrogativos y exclamativos o partículas modales), rasgos morfológicos del modo verbal (imperativo vs. indicativo) y prosodia (tipo de acentos de frase y contorno entonativo) (cfr. Günther 1999: 37).



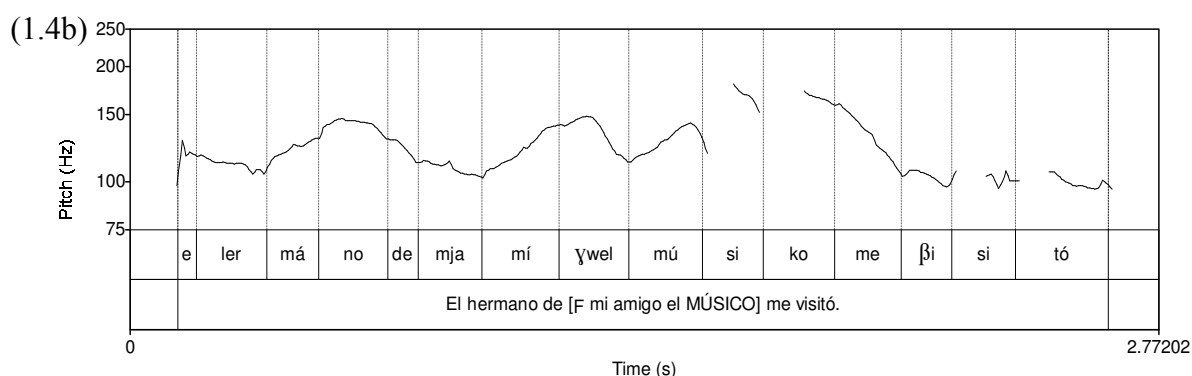
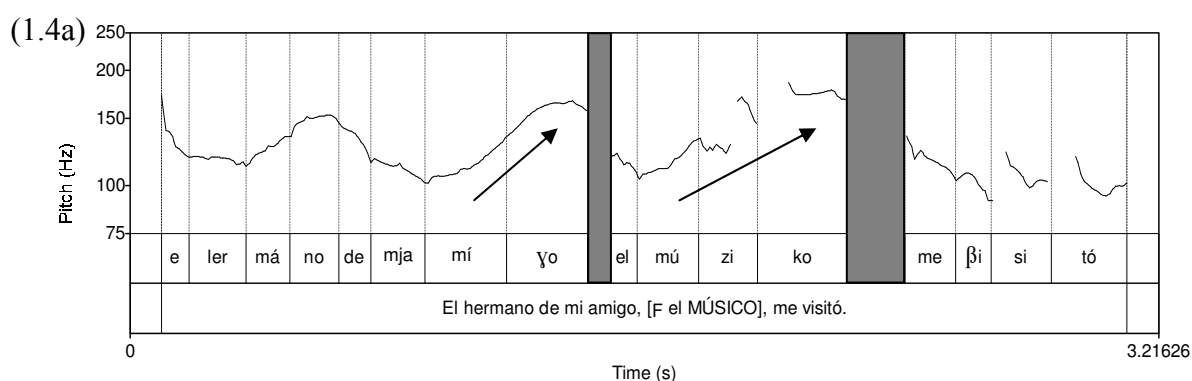
La propiedad de los rasgos prosódicos para la constitución de tipos de oración se muestra, por ejemplo, en los pares mínimos entonativos (véase 1.3a-d), cuyo contenido léxico y otros rasgos del tipo de oración gramatical no presentan diferencias entre sí.

1.1 Prosodia

A diferencia, por ejemplo, del alemán o el inglés, que exigen que el verbo se coloque en primera posición en oraciones exclamativas e imperativas, en español se puede prescindir del pronombre personal. Las cuatro oraciones son idénticas excepto a nivel prosódico, lo que permite distinguir entre una oración declarativa (1.3a) con un contorno descendente al final de la oración; una interrogativa (1.3b) con contorno ascendente final; o una imperativa (1.3c) y una exclamativa (1.3d), ambas con tono inicial alto, pero con contorno final descendente en el primer caso y suspendido en el segundo.

1.1.2.3 Establecimiento de correferencia y coherencia

La realización prominente de partes de la oración sirve no sólo para el señalamiento de la estructura del foco, sino que también expresa la función prosódica de fomentar la coherencia y así permitir una interpretación semánticamente correcta en la comunicación oral. Esta cohesión prosódica basada en la acentuación, como expresión formal de la creación de coherencia, se da, por ejemplo, en la producción de relaciones de referencia inequívoca a frases relativas extrapuestas (cfr. Günther 1999: 40).



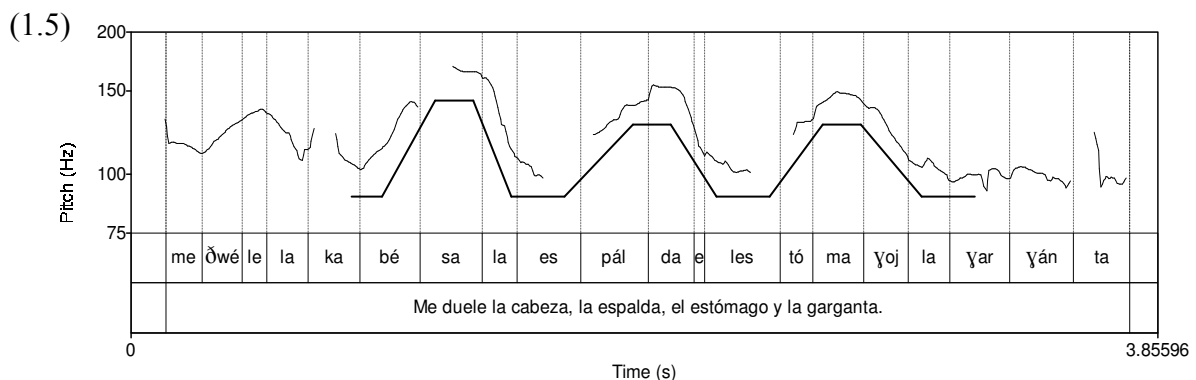
En el par de oraciones (1.4a) y (1.4b) se hace referencia a dos individuos masculinos. Solamente por medio de la acentuación y las pausas se puede entender a quién de los dos corresponde la información adicional acerca de la profesión. En (1.4a) el músico es el hermano y en (1.4b) es el amigo.

Además, en (1.4a) se puede ver cómo la información adicional es independiente del resto del enunciado, pues, si se eliminara el fragmento entre las dos barras grises, el contorno de la frecuencia fundamental aparecería inalterado.

Este fenómeno de la producción de correferencia, en especial desde el punto de vista del proceso, es interesante para la producción lingüística, ya que desde la realización articulatoria del fragmento *mi amigo*, el proceso de planeación conceptual y semántica debe estar tan avanzado, que ya se sabe que enseguida sucederá una modificación.

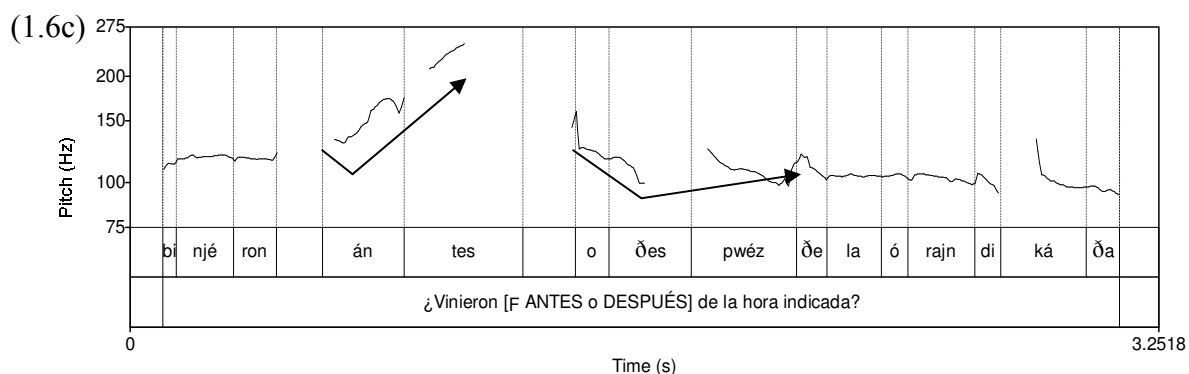
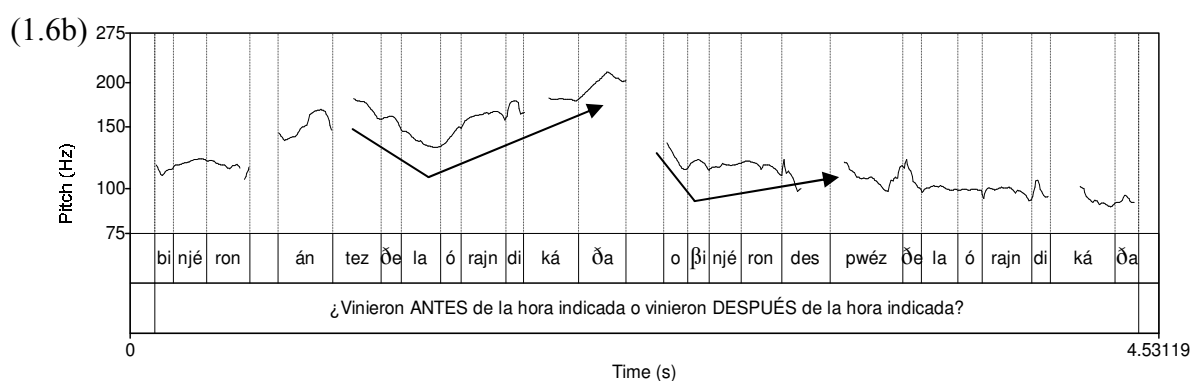
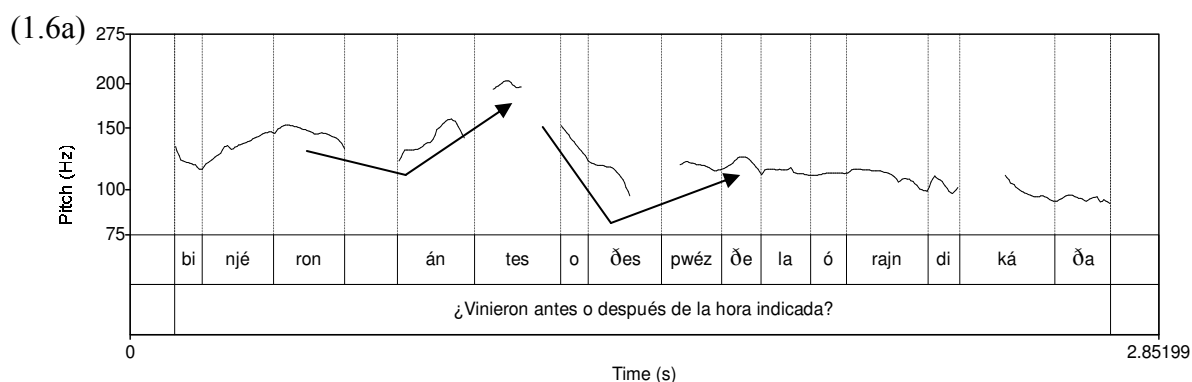
1.1.2.4 Caracterización de la coordinación y elipsis

Hay cierto tipo de contornos entonativos característicos para la realización prosódica de nexos coordinativos. De manera más evidente se da en el caso de la entonación que se utiliza para el señalamiento prosódico de las enumeraciones.



En (1.5), cada uno de los elementos unidos de manera asintética *la cabeza, la espalda, el estómago* presenta una característica secuencia de tonos altos y bajos. El final de la coordinación, señalado de manera sintáctica por medio de una construcción sintética: *el estómago y la garganta*, se señala prosódicamente mediante la continuación del tono alto previo sobre *estómago* y un tono bajo final. Este contorno entonativo terminal también es llamado, debido a su forma, contorno de sombrero.

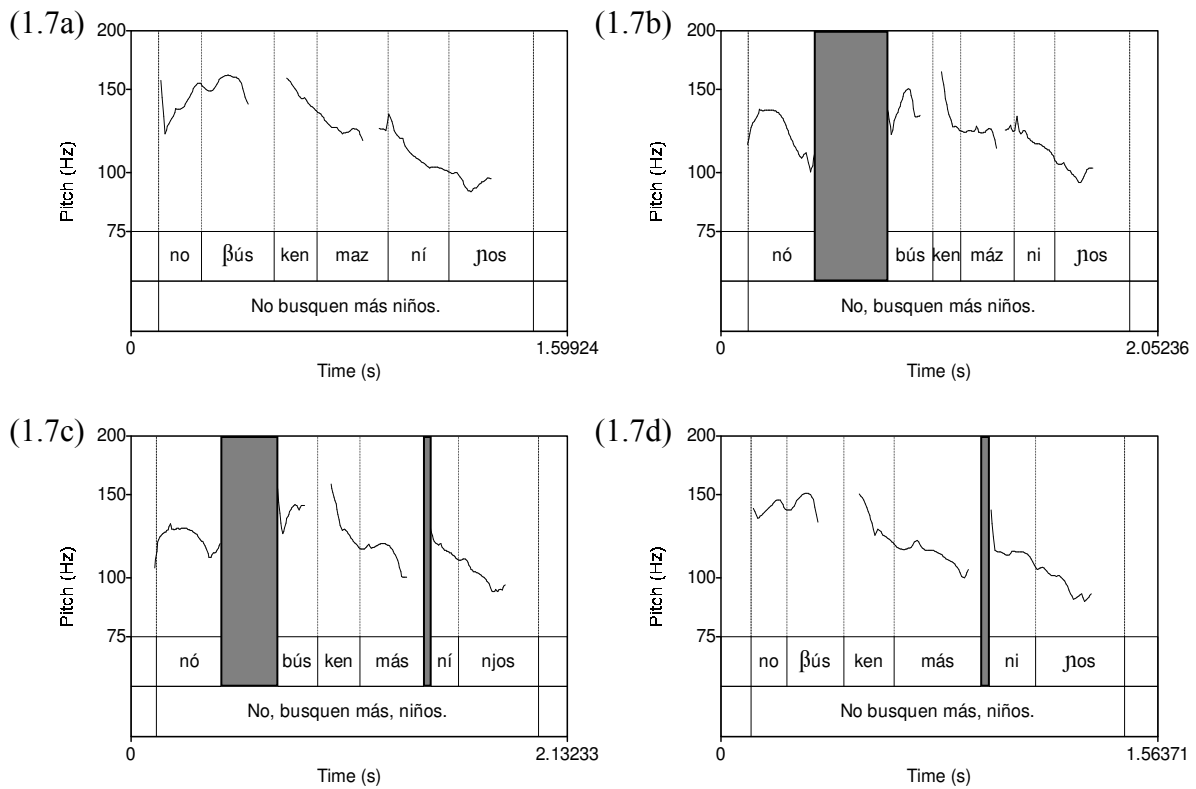
Por otra parte, las coordinaciones y las elipsis pueden ser enfatizadas por medio de contrastes a nivel prosódico. El enunciado (1.6a) se caracteriza por un pico frecuencial muy alto en la palabra *antes* y un pico de menor intensidad sobre la palabra *después*. Esta diferenciación por medio de la entonación (la duración y la intensidad de las sílabas prominentes, aunque esta última no se muestre explícitamente en los ejemplos, es aproximadamente igual) en la versión elíptica alargada (1.6b) o la corta (1.6c) da lugar a un enfoque contrastivo de ambas preposiciones, se manifiesta por medio del contorno entonativo (cfr. Günther 1999: 41-42).



1.1.2.5 Caracterización de la estructura oracional

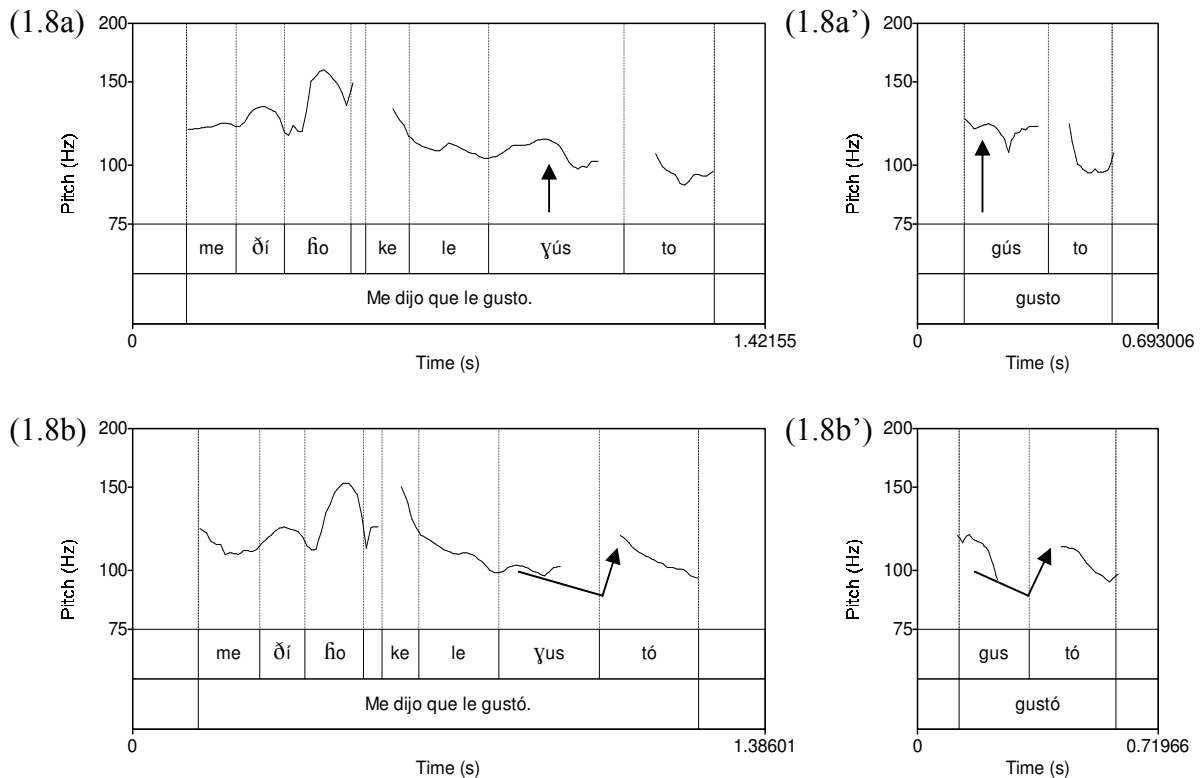
Con la ayuda de medios prosódicos como la colocación de pausas, la selección de la secuencia de tonos en las juntas entre enunciados, del movimiento y registro tonal dentro de los enunciados y del tempo, el esquema oracional se marca como un apoyo, a veces obligatorio, para que el receptor analice la oración de manera sintácticamente correcta.

En los ejemplos (1.7a-d) se da una muestra de la utilización de las pausas (marcadas con barras de color gris) que pueden modificar ligera o drásticamente el significado de la oración, sobre todo al permitir que el interlocutor relacione el verbo *buscar*, conjugado en tercera persona de plural, con el sujeto implícito, *ustedes*, o explícito, *niños*, que también puede funcionar como sintagma nominal acusativo o vocativo. De este modo se puede crear al menos cuatro enunciados distintos partiendo de la misma cadena léxica.



1.1.2.6 Solución de ambigüedades léxicas

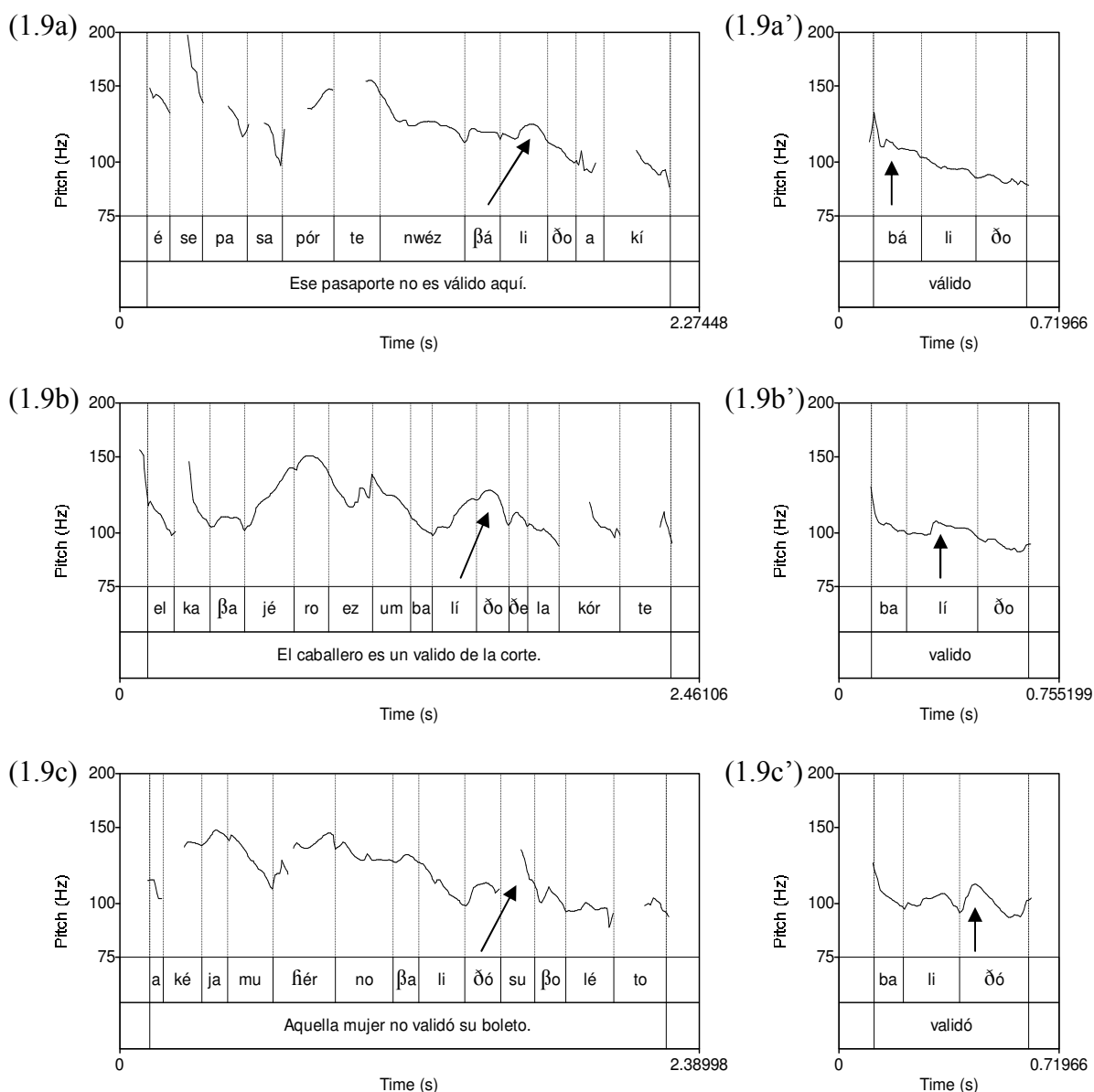
En español, la única unidad prosódica que se utiliza para evitar ambigüedades léxicas es el acento, es decir, la prominencia interna de cada palabra. De esta manera, la localización del acento léxico da información sobre la función gramatical de las palabras.



1.1 Prosodia

El ejemplo (1.8a) contiene la conjugación del verbo *gustar* en primera persona de singular en presente de indicativo, y (1.8b) la conjugación en la tercera persona del singular en pretérito perfecto simple. En este caso, la acentuación evita la confusión de formas verbales. En los ejemplos (1.8a') y (1.8b'), para los cuales se grabó cada una de las palabras de manera aislada, queda demostrado que en algunas ocasiones no sólo la altura tonal es un rasgo distintivo, sino también la duración y la intensidad, aunque esta última no se incluya en el análisis.

Por otra parte, en el caso de (1.9a) – adjetivo, (1.9b) – sustantivo, y (1.9c) – forma verbal conjugada, la posibilidad de confusión es casi nula, debido a que tales palabras nunca aparecen en el mismo contexto. En los ejemplos (1.9a'), (1.9b') y (1.9c') se produjo las mismas palabras de manera aislada para demostrar los mecanismos de contraste de la altura tonal y la duración con respecto a las sílabas átonas.



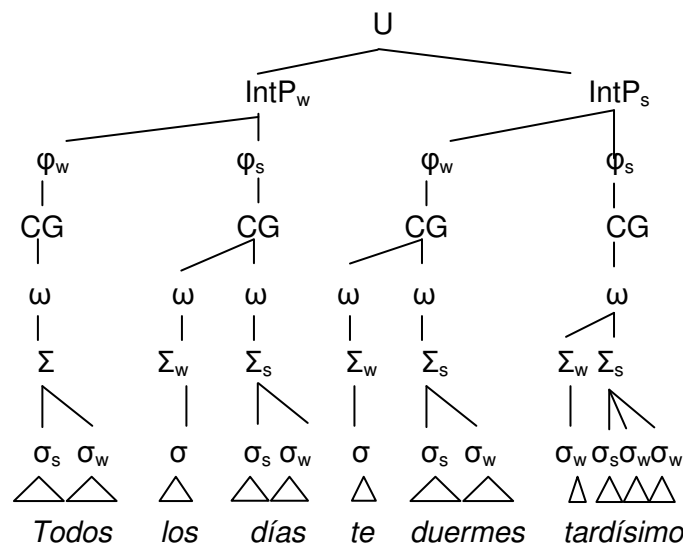
Para dar cuenta de la estructura jerárquica que rige la distribución e interrelaciones de las distintas unidades prosódicas, en la siguiente sección se abordará el tema de los constituyentes prosódicos, abarcando desde las partes de la sílaba, categoría central de análisis, hasta el enunciado fonológico.

1.1.3 Constituyentes prosódicos

La fonología prosódica recomienda que los constituyentes prosódicos se ordenen jerárquicamente para lograr la representación abstracta de la estructura prosódica de un enunciado. A continuación se discutirá cada uno de los constituyentes prosódicos propuestos por Nespor y Vogel (1986: 11): sílaba (σ), pie (Σ), palabra fonológica (ω), grupo clítico (CG), frase fonológica (φ), frase entonativa (IntP) y enunciado fonológico (U).

La Ilustración 1-3 muestra la jerarquía estructural de una oración, cuyos elementos se describirán en las siguientes secciones:

Ilustración 1-3: Ejemplo de una jerarquía prosódica de constituyentes



Fuente: Diseño propio, basado en Günther (1999: 76)

1.1.3.1 Sílaba

Como unidad fonética, la sílaba (σ) es problemática, ya que las fronteras entre sílabas no tienen correlatos exactos desde el punto de vista acústico o articulatorio (Rabanus 2001: 10). Sin embargo, en el nivel abstracto fonológico, la sílaba forma el constituyente prosódico más pequeño y permite una representación estructurada de la información segmental de un enunciado. La especial importancia de la sílaba como unidad estructural del sistema lingüístico consiste, primero, en que la sílaba es una unidad lingüística que se comporta de

manera uniforme, pues la división de un enunciado en sílabas es una tarea más sencilla que la división en segmentos o morfemas; segundo, la sílaba forma una unidad fonotáctica, dado que las restricciones de cada lengua sobre la secuencia segmental sólo pueden ser constatadas dentro de las sílabas; tercero, la sílaba funge como dominio de procesos fonológicos tales como la aspiración en coda silábica, y cuarto, la sílaba sirve como unidad para la representación de rasgos suprasegmentales como el acento. Además, la sílaba es una realidad mental que se expresa en el hecho de que, incluso para los hablantes maternos sin conocimientos lingüísticos, la división de un enunciado en sílabas no representa ninguna dificultad. Por lo tanto, la sílaba también es una unidad perceptivamente relevante (Günther 1999: 77, Rabanus 2001: 9)

Según Núñez Cedeño y Morales-Front (1999: 170), dentro de la sílaba, los segmentos se organizan de acuerdo con una escala universal de sonoridad (Tabla 1-1, Ilustración 1-4), de modo que el segmento con mayor sonoridad ocupa el lugar central en la sílaba (el núcleo) y otros segmentos a su izquierda o su derecha han de descender progresivamente en sonoridad.

Tabla 1-1: Escala Universal de Sonoridad

Vocales bajas	/a/	6
Vocales medias	/e/, /o/	5
Vocoides altos, incluyendo semivocales/semiconsonantes	/i/, /u/ /j/, /w/	4
Líquidas	/r/, /R/, /l/, /ʎ/	3
Nasales	/m/, /n/, /ɲ/	2
Obstruyentes y fricativas	/p/, /t/, /k/, /b/, /d/, /g/, /tʃ/, /f/, /θ/, /s/, /x/	1

Fuente: Diseño propio, basado en Núñez Cedeño y Morales-Front 1999: 181

Si asignamos un número a los diferentes segmentos según su posición en la escala de sonoridad, podemos observar cómo el valor de los segmentos en esta escala disminuye progresivamente según nos alejamos del núcleo. Por lo tanto, la sílaba consiste en un conjunto de segmentos agrupados alrededor de una cumbre o pico de sonoridad que nunca es menor que el de los más alejados.

Para referirse a posiciones o grupos de segmentos dentro de la sílaba se utilizan los términos núcleo, ataque (o arranque), coda y rima. Como ya se mencionó, el núcleo es el centro de la sílaba y el elemento de mayor sonoridad dentro de ella. En español todas las vocales constituyen núcleos silábicos y todos los núcleos silábicos contienen una vocal. Tal identidad no se da en lenguas como el inglés o el alemán, que tienen consonantes silábicas.

El ataque es la consonante o grupo de consonantes que preceden al núcleo dentro de las sílabas. Algunas de las restricciones en español son */dl-/ y obstruyente+nasal. El grupo /tl-/

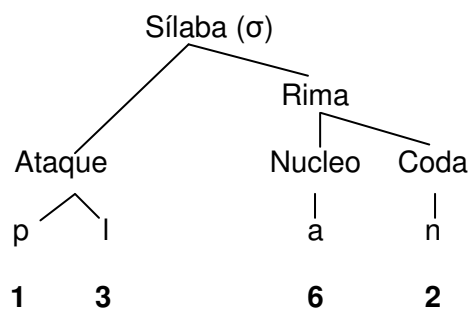
solamente aparece de manera dialectal en México en topónimos y préstamos del náhuatl: *Tlaxcala, tlalpalería*.

La coda es lo que sigue al núcleo en la sílaba. Tras el núcleo podemos encontrar una deslizada (*boi.na, flau.ta*) o una consonante (*pren.sa, ac.to*). Podemos tener también una deslizada seguida por consonante (*claus.tro, vein.te*), o, menos frecuentemente, un grupo de dos consonantes (*trans.por.te, pers.pec.ti.va*) de las cuales la segunda siempre es /s/. Sin embargo, no todas las consonantes son igualmente comunes en posición de coda y, además, se debe distinguir entre codas finales y codas interiores. Cualquier consonante puede aparecer como coda (al menos en el habla cuidada), excepto las palatales ($/tʃ/$, $/j/$, $/ʎ/$). Las nasales y laterales están sujetas a restricciones más generales que determinan su neutralización y asimilación en posición de coda. Otras neutralizaciones, como la pérdida de la distinción entre sordas y sonoras, afectan también a otras consonantes (con diferencias entre dialectos), lo que hace que el número efectivo de oposiciones entre consonantes que encontramos en posición de coda sea bastante menor que el que tenemos en posición de ataque.

Finalmente, núcleo y coda se agrupan en una unidad superior denominada rima. La rima en español no puede contener más de tres segmentos en total (**auns.tral, *muens.tra*). (Núñez Cedeño y Morales-Front 1999: 170-173)

Los términos ataque, núcleo, coda y rima definen, pues, estructuras como la representada en la siguiente Ilustración:

Ilustración 1-4: Constituyentes de la sílaba



Fuente: Diseño propio, basado en Günther (1999:77), Núñez Cedeño y Morales-Front (1999: 173)

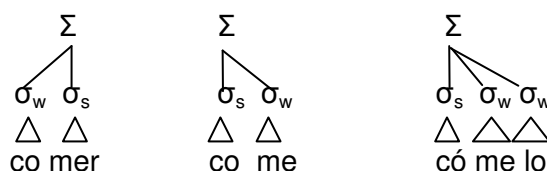
1.1.3.2 Pie

El constituyente prosódico jerárquicamente por encima de la sílaba es el pie (Σ) que juega un papel importante en la determinación métrica de las sílabas fuertes y débiles. Por consiguiente, la estructura de los pies de una palabra tiene una gran influencia sobre la distribución de prominencias y el ritmo. Además, hay una serie de reglas fonológicas que se relacionan con este constituyente. Por lo general, el pie domina directamente una sílaba

1.1 Prosodia

métricamente fuerte y una o más sílabas débiles. Las limitaciones específicas de cada lengua tienen que ver con el número de sílabas que acepta cada pie (si permite pies binarios o pies libres, con número libre de sílabas) y la posición dentro del pie de la sílaba métricamente fuerte (inicial o final) (Günther 1999: 79). Por ejemplo:

Ilustración 1-5: Ejemplos de estructuras de pies en español



Fuente: Diseño propio, basado en Günther (1999:80)

1.1.3.3 Palabra fonológica

Por encima del pie en la jerarquía viene la palabra fonológica (ω). Este constituyente es una unidad fonológica interna pero que establece una relación con los componentes morfológicos de la gramática. Sin embargo, no hay isomorfismo entre la palabra fonológica y el término utilizado en la morfología. Por lo general, por medio de reglas fonológicas se puede derivar las palabras fonológicas de la estructura morfosintáctica (Günther 1999: 80-81). En español, el uso de afijos, en especial los sufijos, origina en algunos casos que las fronteras entre sílabas no coincidan con las fronteras morfológicas. La palabra fonológica es la categoría que domina directamente al pie. Todos los pies de una secuencia deben agruparse en palabras fonológicas y ninguna otra categoría se agrupa de ese modo, es decir, nunca se da el caso de que un solo pie pertenezca a dos palabras fonológicas (Nespor y Vogel 1986: 109).

La palabra fonológica conforma un dominio en el que la asignación del acento léxico se restringe a una de las tres últimas sílabas. Esto es lo que explica el deslizamiento de la prominencia en el ejemplo (1.10):

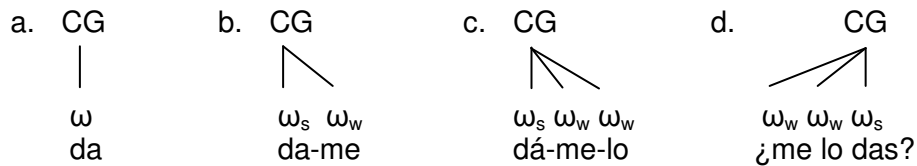
(1.10) **tarde** → **tardí**simo, **tardec**ito

1.1.3.4 Grupo clítico

Los clíticos son híbridos por naturaleza, dado que tienen una posición intermedia entre los afijos y las palabras, y así es como se refleja en el componente fonológico de la gramática, en la que el grupo clítico se coloca entre la palabra fonológica (que agrupa los afijos con sus raíces) y la frase fonológica (que agrupa unas palabras con otras) (Nespor y Vogel 1986:149). Algunas combinaciones de palabra más clítico se comportan como si fueran dos palabras independientes, mientras que otras se comportan como si fueran una sola palabra, es decir, en algunos casos los clíticos se comportan como palabras independientes y en otros como afijos,

dependiendo de las reglas específicas, como el tipo de oración y la forma verbal (Nespor y Vogel 1986:145-146). A manera de ejemplo, véase la siguiente Ilustración:

Ilustración 1-6: Ejemplos de grupos clíticos



Fuente: Diseño propio

En español, los clíticos carecen de acento y se acoplan a unidades portadoras de acento. En la Ilustración 1-6 vemos cómo en el imperativo los pronombres clíticos se comportan como afijos (b y c), pero en indicativo se comportan como palabras aisladas (d).

1.1.3.5 Frase fonológica

El constituyente prosódico que domina directamente al grupo clítico en el orden jerárquico prosódico es la frase fonológica (ϕ), para cuya constitución se refiere en mucho mayor medida a la estructura sintáctica que en el caso del grupo clítico (Günther 1999:83-85). La regla para la formación de la frase fonológica hace referencia a nociones generales como frase sintáctica y núcleo oracional, así como al parámetro que establece la dirección en la que los enunciados se alinean en cada lengua con respecto a los núcleos oracionales, sin tomar en cuenta constituyentes específicos. La frase fonológica resultante es otro ejemplo del no isomorfismo entre las jerarquías sintácticas y prosódicas (Nespor y Vogel 1986: 184-185).

Junto a esa frase fonológica determinada sintácticamente hay una serie de propuestas alternativas para constituyentes entre la palabra fonológica y la frase entonativa. Para esta investigación, es de especial importancia mencionar al menos una de estas propuestas, la de Pierrehumbert y Hirschberg (1990: 227), quienes la denominan frase intermedia (*intermediate phrase*), caracterizada por contornos tonales especiales (véase 1.2.4).

1.1.3.6 Frase entonativa

El siguiente constituyente prosódico en la jerarquía es la frase entonativa (IntP), para cuya constitución se basa no tanto en información sintáctica, sino semántica. Estas unidades significativas se determinan, sobre todo, por medio del esquema foco/trasfondo (véase 1.1.2.1). La característica prosódica de las frases entonativas es, como su nombre lo indica, que el dominio especificado presenta un contorno entonativo cargado de significado. El contorno entonativo dentro del enfoque de la secuencia de tonos se puede describir de manera

abstracta por medio de acentos tonales, de frase y de juntura. Las fronteras de las frases entonativas las marcan las pausas y otras características de juntura, como el alargamiento de segmentos. Los acentos tonales, de los cuales al menos debe haber uno en la frase entonativa, se relacionan con el esquema informativo del enunciado. Los tonos de juntura, por el contrario, señalan principalmente el tipo de oración (véase 1.1.2.2) (Günther 1999:85-86).

1.1.3.7 Enunciado fonológico

El constituyente prosódico más alto que se analiza en la fonología prosódica es el enunciado fonológico (U). El enunciado fonológico es importante para el análisis de fenómenos fonológicos cuyos dominios de aplicación no pueden ser formulados estrictamente en términos de la estructura de constituyentes ofrecida por la sintaxis. Al igual que otros constituyentes prosódicos, aunque el enunciado fonológico utilice información sintáctica en su definición, no presenta isomorfismo con ningún constituyente sintáctico. Sin embargo, lo que hace más interesante trabajar a un nivel de enunciados fonológicos es que no sólo depende de factores sintácticos sino también de factores de naturaleza lógico-semántica. Por lo tanto, en el nivel más alto de análisis fonológico encontramos una interacción entre varios componentes de la gramática, una interacción que no sólo tiene implicaciones para la organización de la fonología, sino también para la organización de la gramática en general (Nespor y Vogel 1986: 221).

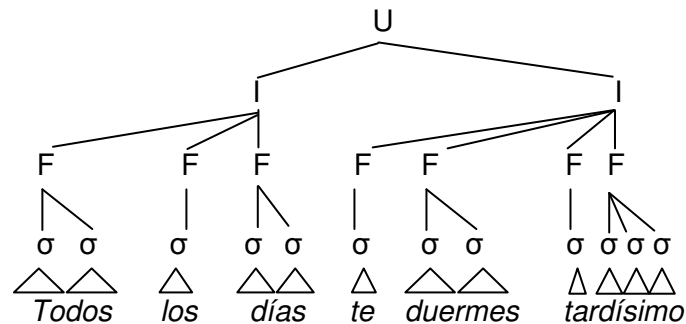
Un ejemplo de las ventajas de analizar el discurso al nivel global de unidades oracionales es que los segmentos muestran una realización prosódica específica, ya que al principio se habla más rápido y con una frecuencia fundamental más alta y el final de un segmento discursivo a menudo se señala con una caída de la frecuencia fundamental hacia la línea de base (véase 1.1.1.2) del hablante, es decir, una caída más profunda que al final de las estructuras oracionales (Günther 1999: 86-87).

1.1.3.8 Crítica

Para evitar la creación de unidades de análisis superfluas, Fox (2000: 330-365, 355), tras hacer un repaso de los distintos modelos de estructuración prosódica, hace una selección de las unidades principales: sílaba (σ), pie (F), unidad entonativa (I) y enunciado (U), eliminando varias unidades prosódicas por debajo, por encima o en medio de estas cuatro. Las principales dificultades las encuentra él en el aspecto terminológico, ya que se da distintos nombres a las mismas unidades y además hay diferencias tipológicas entre las lenguas que hacen que el conjunto de unidades varíe.

Tomando en cuenta este nuevo repertorio en el que la relación binaria de unidades estructurales fuertes y débiles no es un criterio imprescindible, se puede reformular el ejemplo en la Ilustración 1-7 (cfr. Ilustración 1-3), que nos daría el siguiente resultado:

Ilustración 1-7: Ejemplo de la jerarquía prosódica de constituyentes según Fox (2000)



Fuente: Diseño propio

Por otra parte, Nespor y Vogel (1986) no mencionan el concepto de mora (Hyman 2003 [1985]: 5-17, McCarthy y Prince 1999 [1986]: 244-278, Hayes 1999 [1989]: 353-368, Broselow 1995: 188-203, Perlmutter 1995: 310-316, Cantero 2002: 63-66) que nos permite asignar un valor relativo a la duración de un sonido cuando se trata de un rasgo distintivo. En general, se puede decir que una vocal larga tiene dos moras y que una vocal breve tiene una sola mora. Sin embargo, en español no hay diferencias cuantitativas relevantes entre los sonidos, aparte del fenómeno de la geminación y, por tanto, la cantidad, en sí, no es relevante. En el caso de los diptongos, en el modelo moraico tanto las vocales altas /i/, /u/, como las deslizadas /j/, /w/ tienen los mismos rasgos segmentales y difieren en que /i/, /u/ tienen una mora asociada, mientras que /j/, /w/ no la tienen. Por ejemplo *d[je]n.te* y *cl[ie]n.te* tienen diferente estructura moraica (y, consecuentemente, silábica) para hablantes con este contraste léxico:

Ilustración 1-8: Segmentos moraicos y no moraicos



Fuente: Diseño propio basado en Núñez Cedeño y Morales-Front 1999: 185

Un factor de complicación es que las vocales altas pueden realizarse como deslizadas a nivel postléxico. Así, mientras que en *Jua.n[i].Ma.rí.a* tenemos una vocal, en *Pe.dro[j].Ma.rí.a* tenemos una semivocal. Compárese también por ejemplo *t[u].pa.dre* y *t[w]a.bue.lo*. Es decir, una vocal puede perder su mora asociada en ciertos contextos.

Otra característica del español es la existencia de procesos de resilabificación a través de fronteras de palabras. Una consonante final de palabra suele resilabificarse como un ataque cuando va seguida por otra palabra que empieza por vocal. Además de este proceso de resilabificación de secuencias CV, tenemos también resilabificación en secuencias vocálicas (dos vocales seguidas pueden formar una sola sílaba aunque se encuentren en palabras diferentes) en el discurso seguido. Este fenómeno se conoce como sinalefa.

La sinalefa es especialmente frecuente cuando ninguna de las dos vocales afectadas lleva acento. Si las vocales tienen diferente altura, la más alta de las dos se convierte en semivocal. Si las dos vocales son de la misma altura, la primera de las dos es la que pierde su silabicidad. Finalmente, si las dos vocales son idénticas pueden reducirse a la duración de una sola vocal: *m[ja].mi.go*, *t[wa].mi.go*, *bi.c[ju].sa.da*, *es.t[eo].so*, *l[oe]s.pe.ro*, *t[e]s.pe.ro*, *m[i].de.a*. En una sílaba el grado de sonoridad asciende hasta el núcleo y desciende a partir de él: *cul.t[aeu].ro.pa*, *si.t[iou]m.bro.so*, *es.t[aoa].que.lla*, *sie.t[euo].cho*.

Una restricción importante en español es que en general no encontramos secuencias de dos consonantes idénticas dentro del dominio de silabificación (excepto: *perenne* y *obvio*, que no contrastan). (Núñez Cedeño y Morales-Front 1999: 185-188)

A continuación, se hará un repaso de algunos de los distintos enfoques bajo los que se han venido estudiando la entonación, el ritmo y el tempo, así como el volumen y la calidad de la voz o timbre. Por cuestiones de espacio, únicamente se eligió los enfoques de mayor relevancia para los fines de esta investigación, pues algunas aplicaciones específicas de estos enfoques conformarán el sustento teórico del presente trabajo. En cada una de las secciones se resumirá las aportaciones de las distintas escuelas, se hará algunos comentarios y críticas, y en caso de haberlas, se presentará sus aplicaciones al español.

1.2 Entonación

La lengua hablada, según Prieto (2003: 13), es mucho más que la articulación de sonidos para la formación de palabras de manera consecutiva, es también la asignación de melodías a los enunciados. Todo hablante de una lengua conoce el sistema particular de fórmulas melódicas por medio de las cuales puede producir una serie de efectos semánticos que se comparten entre toda la comunidad de hablantes de la misma lengua. De la misma manera, los hablantes somos capaces de identificar los contornos ajenos al repertorio de nuestra lengua nativa. En los últimos años se ha ido tomando cada vez más seriamente el hecho de que la entonación es

un fenómeno lingüístico que forma una parte muy importante del sistema fonológico de la lengua.

A diferencia de las llamadas **lenguas tonales** que utilizan las variaciones tonales para expresar oposiciones de tipo léxico o morfológico, en **lenguas entonativas**, como el español, las variaciones en la melodía no se utilizan para distinguir palabras sino para expresar diferentes sentidos pragmáticos que afectan a todo el enunciado. El carácter lingüístico de la entonación se justifica con el hecho de que los patrones melódicos son modelos bien definidos que se utilizan para expresar las intenciones comunicativas del hablante (véase 1.1.2).

Más allá del sentido referencial del mensaje (también llamado *dictum*), el hablante, mediante la entonación, manifiesta su actitud subjetiva respecto de ese contenido: el *modus*. La expresión del *modus* es esencial en los procesos de interacción comunicativa, ya que muy a menudo el oyente está menos interesado en saber el contenido literal del mensaje que en interpretar la manera en la que se ha dicho.

Además de la llamada **función expresiva**, la entonación también tiene una **función focalizadora** en la lengua: el hablante selecciona la información central del mensaje y le confiere relevancia y prominencia entonativas. Finalmente, la **función demarcativa** de la entonación es una función que indica la manera en la que se segmenta y organiza el discurso: el emisor divide el discurso en unidades tonales para que el oyente pueda segmentarlo e interpretarlo con mayor facilidad (cfr. 1.1.2).

La entonación es un fenómeno lingüístico complejo cuyo tratamiento sistemático precisa de la combinación de tres niveles de análisis complementarios: el **eje físico** (la evolución del parámetro físico de frecuencia fundamental a lo largo del tiempo de la emisión del enunciado) el **eje fonológico** (las unidades melódicas con importancia significativa en una lengua) y el **eje semántico** (los efectos significativos que producen esas variaciones melódicas). Tal como ocurre en el plano segmental, presumiblemente los oyentes sólo se percatan de una clase restringida de movimientos tonales que son los que producen contrastes lingüísticos en una lengua determinada, y no de todas las modificaciones fonéticas de la curva melódica. El **análisis lingüístico** de la entonación pretende poner en relación los ejes físico y funcional de ésta y ‘descubrir’ las unidades tonales capaces de generar oposiciones distintivas o producir diferencias de significado. (Prieto 2003: 13-14)

Como base del análisis fonológico, no es suficiente tomar únicamente la representación gráfica producida por los programas digitales de extracción de la frecuencia fundamental pues hay que descartar todavía los efectos microprosódicos producidos por los mismos segmentos. Los rasgos prosódicos pueden clasificarse ya sea en microprosódicos o en macroprosódicos.

Por macroprosodia se entiende el control consciente de los parámetros prosódicos para producir fenómenos perceptibles como las melodías de oraciones y acentos tonales, mientras que la microprosodia, por el contrario, no puede ser controlada conscientemente, ya que está condicionada por el segmento, la coarticulación y el hablante. (Sosa 2003: 205-206, Paeschke 2003: 29)

1.2.1 Modelos teóricos de la entonación

Los primeros modelos lingüísticos de la entonación, enfocados en la descripción de la prosodia del inglés, surgieron de dos escuelas que iniciaron sus actividades a principios del siglo XX: la **escuela británica** (véase 1.2.2) y la **escuela americana** (véase 1.2.3). Una de las diferencias más notables entre estas dos corrientes radica en el tipo de elementos fonológicos subyacentes que proponen: mientras la escuela británica analiza los contornos melódicos como secuencias de patrones o ‘configuraciones’ expresados mediante movimientos tonales, la escuela americana los analiza mediante una serie de niveles tonales estáticos. De ahí que las propuestas de esas dos corrientes se hayan denominado **análisis por configuraciones** y **análisis por niveles**, respectivamente. Asimismo, ambas escuelas se distinguen por el tratamiento que dan a la organización interna de los contornos: la escuela británica separa los contornos en unidades funcionales independientes (cabeza, núcleo y cola), mientras que la escuela americana considera que el contorno central no presenta estructura interna alguna. Otro aspecto que distingue a las dos escuelas es la orientación especialmente fonemicista que caracteriza a la escuela americana. Esta escuela consideró prioritaria la representación formal de los contornos, es decir, la tarea de definir un repertorio de elementos fonemáticos capaces de dar cuenta de todos los contrastes melódicos en una lengua y, a diferencia del análisis por configuraciones, dejó de lado la caracterización semántica y fonética de los contornos.

Tanto el análisis por niveles como el análisis por configuraciones tienen como objetivo encontrar las unidades mínimas de análisis entonativo y estudiar los contrastes fonémicos que éstas generan. Ninguna de las dos escuelas se preocupa por hacer explícita la relación entre la representación fonológica subyacente y la forma melódica final. Actualmente se considera que todo modelo lingüístico de la entonación debe incluir al menos los dos componentes siguientes:

- 1) un **componente fonológico** que caracteriza las curvas melódicas mediante un inventario de unidades contrastivas, y
- 2) un **componente fonético** que describa de forma explícita el vínculo existente entre la forma subyacente de las curvas y el *continuum* melódico. (Prieto 2003: 15-17)

Hasta cierto punto, se puede considerar que el viejo debate entre niveles y configuraciones continúa vigente en la actualidad, puesto que diferentes modelos actuales han optado por uno u otro sistema de representación. La **escuela holandesa** (véase 't Hart *et al.* 1990, Garrido 2003), por ejemplo, considera que las unidades básicas de análisis tonal son los movimientos tonales, no los niveles (véase 1.2.5). En este sistema, la gramática tonal del holandés consta de diez clases diferentes de movimientos (cinco de tipo ascendente y cinco de tipo descendente) que constituyen las unidades básicas del análisis melódico. En cambio para el **modelo de Aix-en-Provence** (véase Hirst y Di Cristo 1998, Alcoba y Murillo 1998, Baqué y Estruch 2003), las unidades básicas de un contorno son los niveles siguientes: T (*Top*), o altura tonal máxima del locutor; B (*Bottom*), o altura tonal mínima del locutor; y M (*Mid*), valor medio del locutor (véase 1.2.6). Igualmente, el modelo **métrico autosegmental** (véase 1.2.4) propone una versión radical del análisis por niveles y defiende que los contornos se pueden representar adecuadamente utilizando sólo dos niveles tonales, el alto (H – *high*) y el bajo (L – *low*). Esta solución permite minimizar el problema de la sobregeneración de contornos tan criticado al análisis por niveles tradicional. La utilización de únicamente dos niveles es posible técnicamente por dos razones: por un lado, la versión de Pierrehumbert, presentada en su tesis doctoral de 1980, incorpora una regla de escalonamiento descendente que genera la declinación de los picos a lo largo de la frase; por otro lado, la variación en el campo tonal de las patrones melódicos se atribuye a variaciones graduales (no fonológicas) que reflejan el nivel de énfasis del enunciado. (Prieto 2003: 20)

Una de las aportaciones más innovadoras del modelo métrico autosegmental propuesto por Pierrehumbert es el reconocimiento del estrecho vínculo que existe entre la acentuación y la entonación, y del papel de la estructura métrica como columna vertebral de los movimientos melódicos. Es decir, las posiciones métricamente fuertes actúan como puntos de anclaje para los movimientos melódicos relevantes del contorno, lo cual permite ‘predecir’ la aparente multiplicidad de formas de un mismo patrón en diferentes contextos. (Prieto 2003: 18)

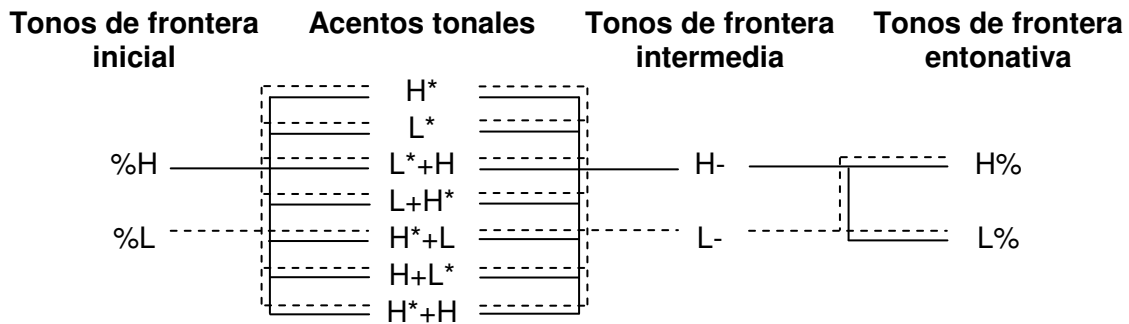
El modelo métrico autosegmental toma como punto de partida la existencia de acentos tonales que, si bien hacen referencia a los niveles originales, también indirectamente describen el tipo de ‘contorno’ que producen esas unidades. En esencia, no se trata de saber exclusivamente qué ‘nivel’ recibe una determinada sílaba, sino cómo se alinean los descensos y ascensos con la sílaba acentuada, lo cual puede producir un contraste fonológico. Precisamente uno de los inconvenientes que presenta el análisis por niveles tradicional es la falta de información relativa a la alineación de los niveles tonales con el texto. (Prieto 2003: 21-22)

1.2.1.1 Estructura interna de los contornos melódicos

Según el **modelo británico**, las curvas melódicas se componen de los siguientes elementos o **configuraciones**: el **núcleo** (*nucleus*), que es el único componente obligatorio de los contornos y se define como la sílaba más prominente del enunciado, que puede ir opcionalmente precedido por la **cabeza** (*head*) y la **precabeza** (*prehead*), y seguido de una **cola** (*tail*) que tiene la función de continuar y completar el movimiento tonal iniciado por el núcleo. La precabeza comprende la porción del contorno de las primeras sílabas átonas y la cabeza, la porción que va de la primera sílaba tónica hasta el inicio del núcleo. A su vez, estos elementos se suelen agrupar en **contorno** (o configuración) **prenuclear** (formado por la precabeza y la cabeza) y el **contorno nuclear** (formado por el núcleo más la cola), que es el pilar alrededor del cual se organiza la melodía del enunciado (véase 1.2.2.1). (Prieto 2003: 23)

Un aspecto que diferencia el **modelo americano** del británico es su interpretación de los contornos entonativos como unidades sin estructura interna. Este modelo considera que el inventario de unidades fonológicas es el mismo en posición nuclear que prenuclear, aunque sí se admite la impotencia entonativa que tiene la sílaba más prominente del enunciado, es decir, el núcleo. Todo patrón entonativo consta como mínimo de dos **niveles tonales** y una **juntura terminal** (véase 1.2.3.1). Igualmente, el **modelo métrico autosegmental** concibe las curvas melódicas como una concatenación lineal de dos clases de unidades fonológicas: los **acentos tonales** (*pitch accents*, asociados con las sílabas acentuadas) y los **tonos de frontera** (*boundary tones*, asociados con límites prosódicos), organización parcialmente heredada del modelo americano tradicional. La gramática combinatoria (véase Tabla 1-2) resultante es capaz de generar todos los contornos bien formados del inglés. Cualquier contorno entonativo puede empezar con un tono de frontera inicial opcional (%H o %L) y consta obligatoriamente de uno o más de los siete acentos melódicos siguientes de la gramática del inglés (H*+L, L*+H, etc.). En principio, la gramática admite su libre combinación, es decir, el hablante puede ir colocando tantos acentos tonales como quiera en el orden que le plazca. Finalmente, el contorno debe acabar obligatoriamente en un tono de frontera intermedia (H- o L-) seguido de un tono de frontera entonativa (H% o L%). Los movimientos intermedios se generan mediante reglas de interpolación del componente fonético (véase 1.2.4). (Prieto 2003: 23-24)

Tabla 1-2: Gramática combinatoria del modelo métrico autosegmental



Fuente: Diseño propio, basado en Prieto (2003: 24)

A continuación trataremos otros dos aspectos para los cuales cada una de las tres escuelas tiene propuestas alternativas: la declinación, el escalonamiento descendente y el rango tonal.

1.2.1.2 Declinación y escalonamiento descendente

En la mayoría de las lenguas se constata que la curva tonal típicamente desciende a lo largo del enunciado. Este fenómeno de declive, en el cual tanto los límites superior como inferior del campo tonal van ‘declinando’ progresivamente, se conoce con el nombre genérico de **declinación** (*downdrift*, *declination*) (véase 1.1.1.2).

Durante mucho tiempo se creyó que todo declive tonal respondía a un mecanismo fisiológico y semiautomático de producción del habla atribuible al descenso gradual de la presión subglotal, hasta el punto que algunos lingüistas lo han considerado un universal de la producción entonativa. Recientemente se ha demostrado que hay un tipo de descenso tonal en varias lenguas que no proviene únicamente de un fenómeno universal de producción del habla, sino de un mecanismo que los hablantes utilizan con fines claramente lingüísticos. Sobre todo en muchas lenguas tonales africanas se distingue entre dos tipos de descenso tonal: 1) un descenso producido por un proceso fonológico condicionado por la presencia de tonos bajos – también conocido con el nombre de **escalonamiento descendente** (*downstep*) –, y 2) un descenso tonal de menor magnitud producido por un efecto temporal llamado **declinación**. (Prieto 2003: 25-26)

En este punto es importante introducir una característica más que distingue a las teorías actuales de la entonación, según la concepción que tienen de la organización de la estructura entonativa. Los modelos de análisis de la entonación se pueden clasificar en **modelos globales** (jerárquicos, o de **interacción de contornos**) y **modelos secuenciales** (discretos). Por un lado, los modelos globales defienden la existencia de dos niveles de representación tonal independientes: un componente ‘local’, que contiene una serie de unidades fonológicas; y un componente ‘global’, que contiene rasgos tonales como la declinación y que afectan a la

frase entera (cfr. 1.1.1). En los modelos globales, el contorno se genera mediante la información combinada de diferentes dominios prosódicos independientes que se van sobreponiendo uno a otro. Por otra parte, los modelos secuenciales generan los contornos entonativos exclusivamente por medio de la concatenación lineal de unidades fonológicas, sin referirse a otro dominio prosódico superior. La teoría métrica autosegmental, un modelo secuencial por excelencia, genera los contornos mediante la suma de elementos tonales subyacentes, asumiendo que su forma global se deriva de la aplicación de reglas locales. Así, la declinación que se produce a lo largo del enunciado se obtiene mediante la aplicación de una regla fonológica local de escalonamiento descendente que va bajando el nivel de altura tonal de los acentos melódicos uno a uno. (Prieto 2003:18-19)

En general, se puede decir que la adscripción a uno u otro marco teórico a menudo se ha visto condicionada por el foco y el ámbito de aplicación del estudio: los modelos globales se han utilizado más a menudo en aplicaciones de síntesis de habla y sus defensores se han interesado por aspectos de modelización fonética, mientras que los promotores de modelos secuenciales se han centrado en el análisis lingüístico y fonológico de la entonación. Esta relativa ‘especialización’ de los modelos explica el hecho de que actualmente contemos con modelos secuenciales con un componente fonológico explícito, pero con un componente fonético poco desarrollado, y al revés, con modelos globales que incluyen precisos algoritmos de generación fonética de contornos pero no lo suficientemente versátiles como para dar cuenta de la diversidad de contornos melódicos de una lengua. (Prieto 2003: 19-20)

La investigación sobre el comportamiento de la declinación es crucial para poder valorar la certeza de las predicciones de los modelos globales y locales. como El modelo métrico autosegmental, que es estrictamente local, predice que el descenso de un pico tonal no se va a modificar en función de la distancia temporal al acento anterior o en función de aspectos de tipo oracional como la modalidad. Pero si se demuestra que el factor temporal influye significativamente en el descenso tonal, entonces habrá indicios para suponer que el hablante tiene un control ‘global’ de la declinación y hará falta referirse a un dominio independiente referido al dominio de frase. (Prieto 2003: 27)

1.2.1.3 Rango tonal

El rango o campo tonal (*accent range*, *pitch range*) de un patrón entonativo o inflexión se define como el intervalo existente entre el valle y el pico de una inflexión ascendente o entre el pico y el valle de una inflexión descendente (véase la Ilustración 1-2). Por ejemplo, una de las diferencias más obvias que existe entre la entonación declarativa y la exclamativa suele ser

el rango tonal que ocupan los patrones tonales. Asimismo, en español, el primer pico de la oración suele ser más alto en oraciones interrogativas e imperativas que en oraciones declarativas.

Una de las propuestas más radicales del modelo métrico autosegmental consiste en considerar que el rango tonal de los acentos melódicos se amplía o se reduce en el componente fonético en función de la implicación del hablante en la emisión del enunciado, de tal manera que cuanto más enfático sea un enunciado más aumenta el rango tonal de la inflexión. Pierrehumbert (1980) argumenta que el campo tonal tiene un uso fundamentalmente expresivo y que no precisa de representación en la forma fonológica. Sin embargo, existe una serie de fenómenos que pone de manifiesto que en algunos casos el incremento del campo tonal de un acento tonal aporta una interpretación distinta del enunciado. (Prieto 2003: 27-28)

A continuación haremos un repaso más detallado de las propuestas de la escuela británica tanto en los modelos de análisis del inglés como en su aplicación al español.

1.2.2 La escuela británica: análisis por configuraciones

El **análisis por configuraciones** o contornos melódicos (*tunes*) debe su nombre al hecho de que reconoce que el contorno global de un grupo tiene función semántica. Este contorno resulta de la entonación que se aplica a cada uno de sus componentes. El hablante elige el tono o melodía que aplicará al acento principal y también el tratamiento melódico de las otras sílabas tónicas y átonas. Su combinación resultará en una determinada configuración global, con significado propio, pues el significado depende fundamentalmente de la elección del tono principal del grupo. El resto de los eventos entonativos matizan y cualifican dicho significado, y en ocasiones pueden incluso llegar a determinarlo. (García-Lecumberri 2003: 36-37)

El análisis por configuraciones de la entonación del inglés ha tenido gran repercusión no sólo en Gran Bretaña sino también en otros países. Este es el que ha sido el empleado mayoritariamente por la escuela británica y el que más ha trascendido en el mundo del estudio del inglés como lengua extranjera. García-Lecumberri (2003: 44-48 y 48-56) describe detalladamente el más conocido dentro de estos modelos, el de O'Connor y Arnold (1973), así como el modelo de Navarro Tomás (1948) para el español. En esta investigación nos limitaremos a resumir los componentes básicos y las características generales que comparten todos los modelos adscritos a la escuela británica. Más adelante se hará un breve repaso del modelo de Navarro Tomás (véase 1.2.2.3).

1.2.2.1 Componentes básicos del modelo

En primer lugar, se tiene en cuenta el potencial de una sílaba para recibir prominencias (*word stress*), es decir, la posición de los acentos que viene marcada en la entrada léxica de las palabras (descontextualizadas) y que por ello se denomina a menudo **acento léxico** o ‘potencial’ (véase 1.1.1.1). Por ejemplo, en las siguientes palabras, el acento léxico recae respectivamente en la última, penúltima y antepenúltima sílaba, partiendo de la misma cadena segmental: ‘numeró’, ‘numero’, ‘número’. (Gracia-Lecumberri 2003: 38)

Según García-Lecumberri (2003: 38-39), los primeros autores de la escuela británica (Sweet, Jones y Armstrong, así como Ward) asociaban la prominencia silábica con la intensidad (*loudness*), es decir, con la fuerza del aire expirado. Al mismo tiempo se reconocía que hay otras formas de realzar una sílaba, tales como su duración, su sonoridad y su entonación. Así nace una distinción que ha sido mayoritariamente aceptada en la escuela británica: la intensidad – u otros mecanismos como la duración – realzan una sílaba, son responsables de su **prominencia** (*stress*), pero algunas sílabas se destacan aún más por medio de un movimiento de la entonación y se consideran acentuadas por medio de un **acento tonal** (*accent*). Otros miembros de la escuela británica (Abercrombie 1965, 1967; Halliday 1970; Lieberman 1979; Ladd 1980; Couper-Kuhlen 1986) consideran que más que la intensidad, el factor que hace prominente a una sílaba es su posición en la cadena rítmica.

Al igual que los primeros fonetistas ingleses, los españoles (Navarro Tomás 1948; Alarcos Llorach 1950; Gili Gaya 1950; Quilis y Fernández 1985) consideraban que la prominencia silábica se debe principalmente a la intensidad (‘acento de intensidad’), si bien en la cadena hablada estas prominencias de intensidad suelen ir acompañadas por un contorno entonativo específico. Más adelante (Contreras 1963; Quilis 1981; Enríquez *et al.* 1989) se comprobó que, como en inglés, también en castellano las variaciones de entonación ejercen un papel fundamental para realzar las sílabas. Finalmente, Gil Fernández (1988) reconoce la diferencia entre sílabas resaltadas por un conjunto de parámetros o correlatos (‘prominencia’ o ‘acento en general’) y sílabas destacadas principalmente por la entonación (‘acento tónico’).

1.2.2.2 Las unidades entonativas y sus componentes

Todos los estudiosos de la entonación utilizan algún tipo de grupo melódico que se conoce por diversos nombres: grupo entonativo, grupo tonal, unidad melódica, etc. (*intonation group, tone group, tonic unit, word group*). Esta unidad presenta un patrón entonativo completo, y puede delimitarse atendiendo a varios criterios: algunas veces hay pausas reales que ayudan a establecer los límites; en otros casos se pueden determinar las fronteras atendiendo a indicios

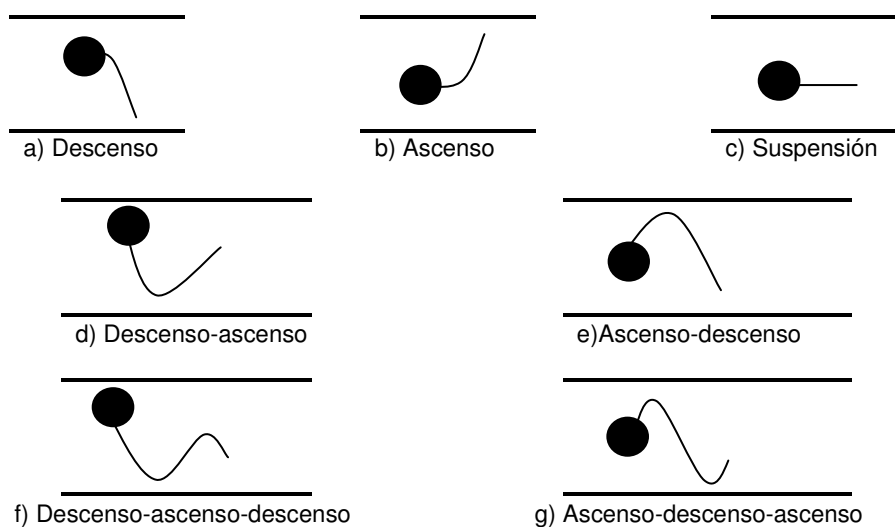
fonéticos como la declinación de la entonación a lo largo del grupo, la anacrusis (aceleración) de las sílabas átonas iniciales o el restablecimiento del nivel tonal al comienzo de un nuevo grupo. También se emplean criterios fonológicos que atañen a la estructura del propio sistema, como, por ejemplo, la presencia de un núcleo del grupo entonativo. (García-Lecumberri 2003: 39)

El grupo entonativo, según los modelos británicos de análisis por configuraciones, se compone de los siguientes elementos:

Núcleo: El único componente esencial es el **núcleo**, que es, por definición, el último acento del grupo entonativo, y que se llama, por consiguiente, **acento nuclear**.

El **tono nuclear** es el tratamiento entonativo que el hablante aplica al núcleo del grupo. Los tonos nucleares se pueden dividir en dos grupos: simples y complejos (véase la Ilustración 1-9). Los tonos simples son unidireccionales: a) descenso, b) ascenso y c) suspensión (*falling, rising y level*) y también se contemplan subdivisiones de estos dependiendo del punto de arranque del movimiento entonativo. Por otra parte, los tonos nucleares complejos pueden ser bidireccionales: d) descenso-ascenso y e) ascenso-descenso; o multidireccionales: f) ascendente-descendente-ascendente o g) descendente-ascendente-descendente.

Ilustración 1-9: Movimientos de los tonos nucleares simples y compuestos

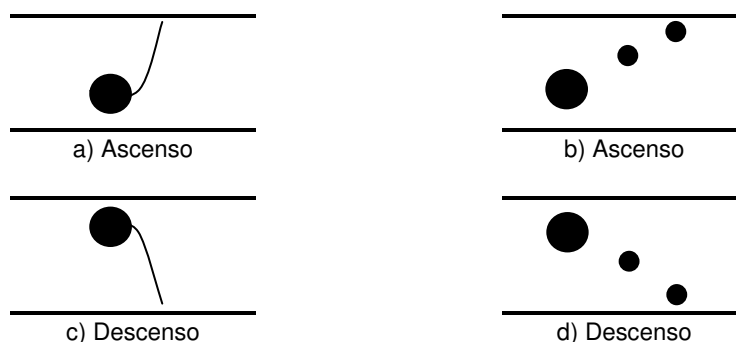


Fuente: Diseño propio, basado en García-Lecumberri 2003: 41-42

Cola: La sílaba nuclear, aquella donde recae el acento nuclear, puede ir seguida de sílabas átonas. Estas sílabas constituyen la **cola** (*tail*) del grupo entonativo. La cola sigue el movimiento iniciado por la sílaba nuclear a menos que se trate de un tono complejo (véase la Ilustración 1-9, d-g). En los tonos ascendentes (véase la Ilustración 1-10), la subida puede realizarse a) con movimiento en la sílaba nuclear o b) en las sílabas de la cola mientras que la

sílaba nuclear permanece estática en el punto de arranque del ascenso. En los tonos descendentes, el movimiento puede realizarse c) en la sílaba nuclear o d) como un salto desde ésta a la sílaba o sílabas posnucleares.

Ilustración 1-10: Movimientos de las sílabas posnucleares (colas)



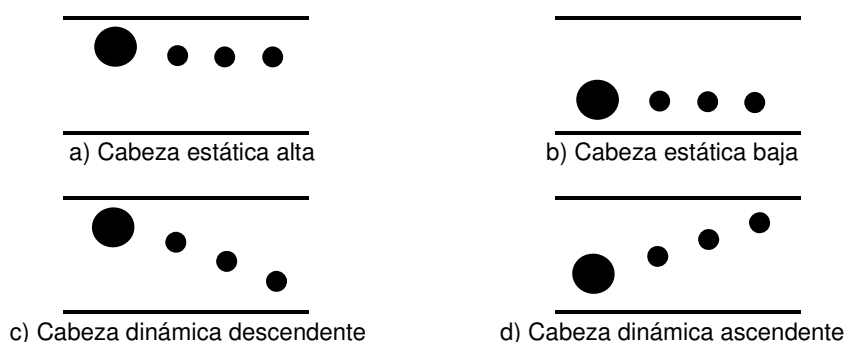
Fuente: Diseño propio, basado en García-Lecumberri 2003: 42

Cabeza: La sílaba nuclear puede ir precedida de sílabas acentuadas, prominentes o átonas. La primera sílaba acentuada del grupo señala el comienzo de la **cabeza** o **cuerpo** (*head* o *body*). La cabeza abarca e incluye desde esta primera sílaba acentuada hasta la inmediatamente anterior.

Los acentos tonales que aparecen en la cabeza del grupo y que, por lo tanto, son previos al núcleo se conocen como acentos prenucleares. Éstos son fonética y fonológicamente distintos de los acentos nucleares. Sin embargo, algunas configuraciones prenucleares se asemejan fonéticamente a las nucleares y viceversa.

El primer acento pre nuclear puede iniciar una línea entonativa que, si no hay otros acentos prenucleares, se extenderá hasta la sílaba nuclear (sin incluirla). Este movimiento puede ser alto o bajo estático (véase la Ilustración 1-11, a y b) o dinámico (c y d). Éstas son las llamadas ‘cabezas simples’ (*simple heads*).

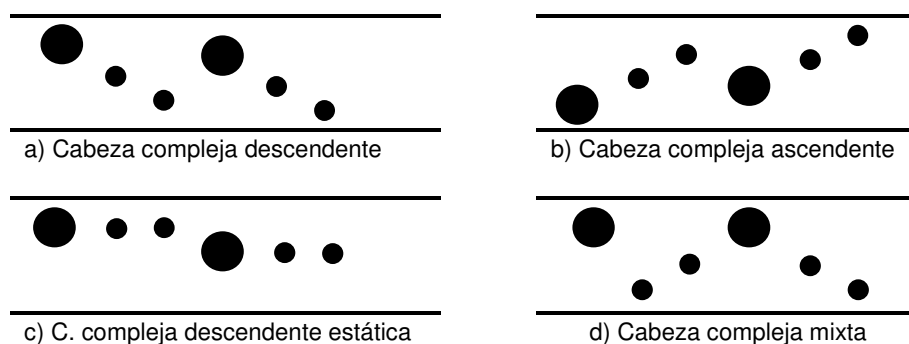
Ilustración 1-11: Movimientos de las cabezas simples



Fuente: Diseño propio, basado en García-Lecumberri 2003: 43

Las cabezas simples también pueden repetirse dentro de un mismo grupo. Se les denomina ‘cabezas complejas’ (*complex heads*). La condición es que debe haber al menos otro acento prenuclear. Cada uno de estos acentos sucesivos puede iniciar una línea entonativa que abarque hasta el siguiente acento. Los resultados son secuencias de movimientos descendentes, ascendentes, de tonos estáticos progresivamente descendentes, o incluso, combinaciones de los movimientos simples:

Ilustración 1-12: Movimientos de las cabezas complejas



Fuente: Diseño propio, basado en García-Lecumberri 2003: 43

Precabeza: Las sílabas átonas que puedan preceder a la cabeza (o al núcleo, si no hay cabeza) se denominan **precabeza** (*pre-head*).

Así, un grupo entonativo que presente todos los elementos posibles, aunque no imprescindibles, ya que sólo el núcleo es necesario, constaría de precabeza, cabeza, núcleo y cola en este orden. (García-Lecumberri 2003: 39-44)

Algunos modelos británicos han sido criticados por su excesivo ‘detalle’ en el plano semántico. Paralelamente, a veces se ha censurado la falta de datos acústicos verificables, ya que se basan en análisis auditivos, e incluso subjetivos en algunos casos. Sin embargo, hay que tomar en cuenta que en la época de apogeo de estos modelos el análisis acústico de la entonación resultaba mucho más costoso. De cualquier modo, se ha comprobado que muchos análisis auditivos han resultado ser bastante fiables al cotejarlos con datos instrumentales.

Además, el modelo de Crystal (1969) sí enfatiza lo fonético y analiza los componentes de la unidad tonal y otros rasgos suprasegmentales con una riqueza y minuciosidad de detalle. A estos últimos los sistematiza de manera innovadora: establece una categorización en la que distingue rasgos prosódicos (tono, rango entonativo, pausas, intensidad, etc.) y rasgos paralingüísticos (tensión articulatoria o calidad de la voz).

Resumiendo lo que hasta ahora se ha dicho, los análisis de la escuela británica pretenden ser fonológicos: su propósito es clasificar los elementos distintivos, sin embargo, como en

muchas otras escuelas, no se llega a un total acuerdo en cuanto a cuáles y cuántos son estos elementos. Unos distinguen dos tonos nucleares complejos, otros tres o cuatro. Unos dan importancia a la dirección del movimiento, otros al punto de partida o al de llegada. Estas diferencias se deben a lo que para unos es una variación ‘alotónica’ y no significativa, pero para otros es un contraste fonológico. No obstante, estas divergencias no son tan graves, ya que los significados a los que refieren son muchas veces pragmáticos o paralingüísticos y no gramaticales. Los diferentes inventarios reflejan los distintos puntos de vista e intenciones de cada autor. (García-Lecumberri 2003: 56-57)

1.2.2.3 Aplicación del análisis por configuraciones al español

También para el castellano existen análisis comparables a los británicos. En concreto, el modelo de Navarro Tomás (1948) fue el de mayor influencia durante gran parte del siglo XX y en el que se basaron muchos estudios posteriores.

Para Navarro Tomás, el tono principal o tonema es el responsable fundamental del significado de la entonación, pero no adjudica significados diferenciados a las distintas configuraciones de la unidad melódica, a las cuales llama ‘sintonemas’ y las equipara a los ‘*tunes*’ británicos. Sin embargo, su repertorio de sintonemas y su denominación son iguales a los tonemas. Por eso, a pesar de que menciona las configuraciones, el modelo de Navarro Tomás se clasificaría mejor dentro de los análisis por tonos en vez de los de configuraciones.

1.2.2.4 Componentes de la unidad melódica y sus significados

El grupo entonativo, denominado ‘unidad melódica’, se compone de los siguientes elementos: **inflexión inicial** (correspondiente a la precabeza de los modelos ingleses), **cuerpo** (cabeza) e **inflexión final** (núcleo + cola).

Inflexión final: el último acento es el enclave entonativo más importante del grupo, ya que dentro de él se localiza el tono nuclear o ‘tonema’. Sin embargo, para Navarro Tomás no existe un elemento separado como ‘cola’, sino que la inflexión final comprende el último acento y las subsiguientes sílabas átonas.

Cuerpo: su definición es igual a la británica de cabeza. Se extiende desde la primera sílaba acentuada hasta la inmediatamente anterior al último acento.

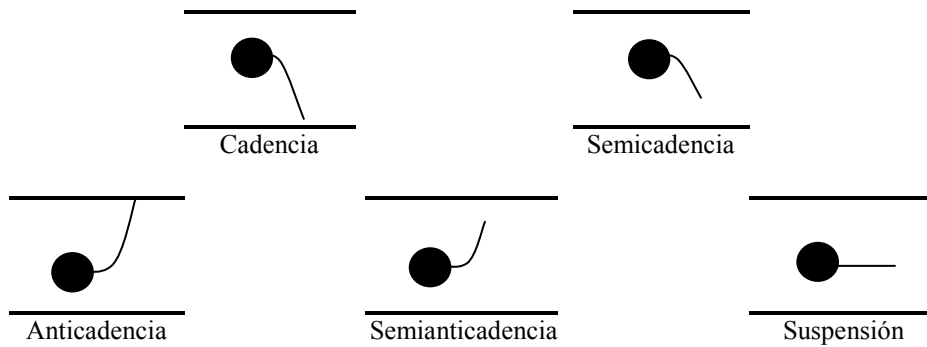
Inflexión inicial: equivale a la precabeza, pues la constituyen las sílabas previas al primer acento de la unidad melódica. (García-Lecumberri 2003: 48-50)

Los tonemas se pueden clasificar, al igual que los británicos, según la dirección del movimiento entonativo. Así encontramos tonemas ascendentes, descendentes o estáticos que a

su vez se subclasifican. Sin embargo, esta división no depende del punto de inicio del movimiento sino del punto en que el movimiento ascendente o descendente termina.

El inventario (véase la Ilustración 1-13) cuenta con cinco tonemas: cadencia, semicadencia, anticadencia, semianticadencia y suspensión:

Ilustración 1-13: Tonemas según Navarro Tomás (1948)



Fuente: Diseño propio, basado en García-Lecumberri 2003: 41-42

Para describir el uso de los tonemas principales, Navarro Tomás explica la función de cada uno de ellos dentro de las oraciones enunciativas, y lo hace basándose en el grado de contraposición que cada tonema marca entre el grupo entonativo en el que se ubica y el siguiente. Para esto recurre a la diferenciación clásica entre ‘tema’ y ‘rema’ que corresponden en la frase respectivamente a la ‘rama tensiva’ y ‘distensiva’. La rama tensiva (tema) termina en uno de los tonemas ascendentes (anticadencia o semianticadencia) mientras que la rama distensiva (rema) finaliza con un tonema descendente (cadencia o anticadencia).

Cadencia: es el tonema con el que terminan las declarativas en grupos finales absolutos, el rema del enunciado; también se emplea (con variaciones de rango tonal) para muchas interrogativas (por ejemplo, las pronominales), exclamativas e imperativas.

Anticadencia: indica continuidad, información incompleta; por esta razón se emplea en la rama tensiva que señala el tema, precediendo típicamente al grupo que termina en cadencia; también se encuentra en algunas interrogativas (reiterativas) y exclamativas.

Semicadencia: en grupos enunciativos no finales absolutos, por ejemplo, en las enumeraciones, o en cierto tipo de interrogativas (aseverativas).

Semianticadencia: separa grupos con menos diferenciación que los señalados por la anticadencia.

Suspensión: se emplea entre grupos estrechamente relacionados, por ejemplo, antecediendo a una aposición o un vocativo. (García-Lecumberri 2003, 50-52)

En el modelo de Navarro Tomás existen diferentes tipos de cabeza que funcionan distintivamente, ya que él mismo señala que es la cabeza o cuerpo el elemento que distingue, por ejemplo, a las enunciativas de las interrogativas.

La realización habitual tiene un nivel medio sostenido. Esta es una diferencia con respecto a los modelos británicos en los que la cabeza habitual es de un nivel alto sostenido.

Cabeza media: Tono medio y uniforme (con variaciones alotónicas). Es característica de las declarativas. En este tipo de oraciones la inflexión inicial y el cuerpo no tienen valor fonológico, sino fundamentalmente idiomático (caracterización del hablante) y presenta muchas menos variaciones entonativas que el tonema final.

Cabeza descendente: Empleada distintivamente en las interrogativas, por ejemplo, en las preguntas absolutas, aseverativas, restrictivas y pronominales. También se utiliza en las exclamativas que indican decepción, queja, resignación, amenaza o admiración (dependiendo del rango del descenso).

Cabeza alta: Frente al nivel medio sostenido de la cabeza de las declarativas, y el descenso de muchas interrogativas, también podemos encontrar un tipo de cabeza uniforme, pero más elevada que en las declarativas. Se emplea en las preguntas relativas, reiterativas y exclamativas.

Cabeza ascendente: Se encuentra en las preguntas hipotéticas, preguntas reiterativas y preguntas exclamativas. Las cabezas ascendentes también son propias de las exclamaciones con sentido de extrañeza, sorpresa, rectificación, protesta o desaprobación.

No se menciona la posibilidad de cabezas complejas. Sin embargo, Navarro Tomás señala que la línea tonal recta presenta ‘ondulaciones’ que reflejan las actitudes y emociones del hablante.

La inflexión inicial o precabeza no se subclasifica explícitamente según sus características entonativas o significado. Su forma habitual corresponde a un nivel entonativo bajo y a menudo asciende hasta el nivel del primer acento. La inflexión inicial puede presentar variaciones en cuanto a la altura y movimiento de las sílabas, que reflejan el matiz emocional que el hablante otorga a su enunciado. (García-Lecumberri 2003: 53-55)

Al igual que en el estudio del inglés, los modelos de análisis entonativo del español no sólo se identificaron con las propuestas británicas. Por tanto, es necesario que hagamos un repaso detallado de las propuestas generales de la escuela americana.

1.2.3 La escuela americana: análisis por niveles

Martínez Celdrán (2003: 63-64) comenta que cualquier aspecto de la lingüística americana tiene siempre sus raíces en Bloomfield (1933), quien establece una primera división entre sonidos básicos y modificaciones. Es decir, plantea claramente la diferencia entre segmentos y suprasegmentos, ya que lo que él llama modificaciones hace referencia a las cuestiones de duración, acento y tono. Bloomfield deja claro que el tono responde a la frecuencia de vibración de la voz y, quizá, la primera diferencia que merece la pena ser destacada respecto del resto de lingüistas de la época es que él consideró el tono y cualquiera de las modificaciones como ‘fonemas secundarios’. Únicamente este autor y algunos de sus seguidores han considerado a los elementos suprasegmentales como fonemas y los clasifican como ‘secundarios’ para distinguirlos de los segmentos que se constituyen en fonemas primarios.

El uso del tono como fonema secundario (a diferencia de su uso como fonema primario en las lenguas tonales) está referido a lo que se ha considerado la entonación que, como Bloomfield reconoce, funciona principalmente al final de las oraciones. Es lo que más adelante se conocerá con el nombre de **tonema**. Distingue, además, un total de cinco terminaciones fonológicas: declarativa, interrogativa absoluta, relativa, exclamativa y pausa.

Partiendo de esta base, sus seguidores fueron dando forma a la teoría de la entonación de la escuela americana. Wells (1945) habla de cuatro fonemas tonales para el inglés, mientras Pike (1945) es uno de los primeros que hablan de ‘niveles’ y establece, al igual que Wells, cuatro niveles tonales para el análisis del inglés. Además, Pike indica que estos cuatro niveles son relativos, es decir, que al analizarlos forzosamente hay que ponerlos siempre en relación entre sí. El nivel más alto será /1/ y el más bajo /4/. También pone gran énfasis en la estrecha relación entre acentos y tonos.

Pike menciona tres diferencias básicas en la entonación según el significado:

- las que no afectan al significado (no implican una actitud especial del hablante),
- las que reflejan un estado de ánimo, y
- las que modifican el significado oracional.

1.2.3.1 Componentes básicos del modelo

Las teorías de Pike fueron importantes para el establecimiento de las bases, pero fueron Trager y Smith (1951) los que elaboraron la teoría de la entonación de la escuela americana de forma más detallada. Ellos son quienes realizaron el estudio de los elementos prosódicos

1.2 Entonación

más influyente en las generaciones posteriores. Siguiendo la tradición de dicha escuela, también hablan de fonemas y establecieron los siguientes elementos prosódicos:

- a) 4 tipos de acentos: / ' ^ ` ~ /, también llamados primario, secundario, terciario y débil.
- b) 1 juntura interna: /+/, para las llamadas transiciones abruptas, distintas de las normales.
- c) 4 tonos o niveles tonales: / ¹ ² ³ ⁴ /, también llamados bajo, semibajo, semialto y alto.
- d) 3 junturas terminales: / | || # /, también llamadas sostenida, ascendente y descendente.

Los acentos

Los acentos se manifiestan básicamente sobre las vocales. Cuando la vocal se manifiesta con un máximo de intensidad, entonces se dice que posee un acento primario /'/, incluso las palabras monosílabas. El acento débil /~/ indica la ausencia de acento primario o de otros acentos. También son posibles otros acentos intermedios en intensidad, los terciarios /'/. En inglés sólo un acento débil puede preceder a un acento primario al principio de una palabra. Tras un acento primario u otros acentos fuertes, el patrón más común es que les sigan uno o más acentos débiles. El acento secundario aparece en secuencias complejas como la siguiente: *cónfèrènce+côördīnàtōr* ('coordinador de conferencias'), donde la primera sílaba de la segunda unidad, simbolizada como *ô*, tiene un acento intermedio entre el fuerte y el que se ha denominado terciario.

La juntura interna abierta ('plus-juncture')

Se distinguen dos tipos de transición entre los componentes silábicos de las palabras: la transición normal y la interrumpida o abrupta. Si comparamos *nitrate* y *night-rate*, veremos que a pesar de tener los mismos elementos fonemáticos, la **transición** es diferente: normal en /náytrèt/ y abrupta en /náyt+rèt/. En el primer caso el fonema /t/ forma parte de la segunda sílaba, y en el segundo se une a la primera.

Los tonos

Trager y Smith llegan a la conclusión de que en inglés hay cuatro fonemas tonales, numerados como / ¹ ² ³ ⁴ / yendo del nivel más bajo al más alto. Pero en su afán de establecer el hecho fonológico perfectamente delimitado frente al fonético, establecen variantes alofónicas dentro de cada uno de los niveles. En principio, son cuatro las variaciones normales:

- ~ 'lowest': representa la variante más grave dentro de cualquiera de los niveles, así pues [¹ ~] significa, dentro del nivel 1, la posibilidad más baja y [⁴ ~] indicará que es el alófono más bajo dentro del tono más agudo.

- ◦ ‘*next higher*’: indica que el tono marcado con este diacrítico es un poco más agudo que el anterior. Se podría clasificar como semibajo, por ejemplo [◦¹].
- ˘ ‘*still higher*’: señala un tono un poco más alto que el anterior, es decir, un tono semialto dentro del mismo nivel, por ejemplo [˘¹].
- ˉ ‘*highest*’: indica el tono más elevado dentro del mismo nivel, por ejemplo [ˉ¹].

Por lo tanto, por cada fonema tonal habrá cuatro alófonos, ya que cada nivel puede ser bajo, semibajo, semialto y alto, por ejemplo [˘² ◦² ˘² ˉ²]. En total habrá dieciséis alófonos y cuatro fonemas tonales.

Las junturas terminales

Hay tres tipos de fenómenos terminales en inglés (aunque no coincide obligatoriamente con el final de una oración). Las junturas terminales sostenidas, // (fonéticamente [ˉ]), indican que el tono se mantiene inalterable al final de esa secuencia. Por otra parte, las junturas ascendentes, /|| (fonéticamente [+]), son típicas de las interrogativas absolutas. Finalmente, las junturas descendentes, /#/ (fonéticamente [-]), son las más comunes en oraciones declarativas y en las interrogativas pronominales. Estos tres tipos de junturas pueden encontrarse después de cualquiera de los cuatro niveles tonales.

Gracias a este modelo de análisis prosódico, los integrantes de la escuela americana logran formalizar la sustancia prosódica al crear unidades discretas en los puntos relevantes de la melodía eliminando todo lo redundante de la transcripción fonética, es decir, someten todos los aspectos de la sustancia fónica a criterios fonológicos. (Martínez Celdrán 2003: 64-69)

1.2.3.2 Aplicación del análisis por niveles al español

Al igual que en el análisis entonativo del inglés, en el castellano también se dio la división entre los análisis por niveles y análisis por configuraciones: Navarro Tomás y sus seguidores, como ya hemos visto, adoptan un modelo con semejanzas a los análisis de tonos británicos, mientras que otros autores, como Stockwell, Bowen y Silva-Fuenzalida (1956) y Quilis (1985), estudiaron la entonación del español utilizando ‘niveles’. Las diferencias entre ambos tipos de análisis son muchas, desde la naturaleza y grados de la prominencia silábica hasta los significados prosódicos, pasando por los componentes de la unidad entonativa. (García-Lecumberri 2003: 58)

Para un recuento más detallado de las distintas aplicaciones del análisis por niveles al español, se recomienda consultar el artículo de Martínez Celdrán (2003: 69-84), donde hace una

revisión crítica de los modelos propuestos por Stockwell *et al.* (1956), Silva-Fuenzalida (1956-57), Hockett (1971) [1958] y Quilis (1985, 1993).

Este último autor introduce una innovación en el estudio de la entonación española basada en niveles que consiste en el establecimiento de los niveles partiendo de las curvas melódicas, como posteriormente sería lo habitual en escuelas como la métrica autosegmental. Sin embargo, según Martínez Celdrán (2003: 79-82), Quilis dice partir de un análisis de configuraciones para después establecer los niveles pertinentes de la frase a estudiar. Lo que él entiende por configuración de un grupo entonativo es la curva melódica tal cual. No obstante, el procedimiento metodológico desde el punto de vista de la teoría de niveles es novedoso, ya que en los años cincuenta los análisis se basaban sobre todo en el oído del investigador. En cuanto a los niveles tonales, fonemas acentuales y junturas terminales, Quilis adopta las propuestas de Stockwell *et al.* (1956).

Además, Quilis (1993) divide las funciones de la entonación según se presenten en el nivel lingüístico, en el nivel expresivo y en el nivel sociolingüístico. En el nivel lingüístico incluye la función integradora y la distintiva que incluye la declarativa, interrogativa absoluta y la pronominal, también expone curvas de las funciones demarcativas. En el nivel expresivo estudia varias entonaciones enfáticas: tanto de tipo declarativo como interrogativo de varias clases: enfática, de cortesía, reiterativa, relativa, confirmativa, etc. Por último, en el nivel sociolingüístico incluye diversas variedades dialectales.

Por su parte, Martínez Celdrán (2003: 84-94) hace su propia propuesta de análisis prosódico por niveles y acepta la mayor parte de las premisas de la teoría de los niveles:

- 1) La importancia de las junturas terminales, marcadas por la ralentización del tempo de la última sílaba antes de que la voz se desvanezca al comenzar la pausa, además de su dirección tonal descendente, ascendente o mantenida. (La juntura interna abierta no desempeña ningún papel desde el punto de vista entonativo, sino que es esencialmente una cuestión de silabeo.)
- 2) La existencia de tres niveles tonales: bajo /1/ y medio /2/ para los tonos neutros, y /3/ para marcar los elementos enfáticos.
- 3) La sílaba más prominente, debido a su duración e intensidad, es el punto de anclaje de los niveles tonales.

Martínez Celdrán concluye que la renovación fundamental que necesita la teoría consiste en partir de curvas reales y establecer el nivel fonológico a partir del análisis fonético.

1.2.4 Teoría métrica autosegmental

Ahora que hemos estudiado los planteamientos teóricos de las escuelas británica y americana para el análisis de la entonación, pasemos a revisar las propuestas del modelo métrico autosegmental (*autosegmental-metrical*, AM), que es el más generalizado en los estudios actuales de la entonación.

El modelo métrico autosegmental tiene mucho en común con sus predecesores, ya que, por ejemplo, contempla niveles tonales (únicamente dos: alto – *high*, H – y bajo – *low*, L) y tonos de juntura, como el análisis americano por niveles. Con respecto a la escuela británica, el modelo métrico autosegmental no abandona los tonos y configuraciones, sino que realiza un análisis de la estructura interna de dichas unidades desde otro punto de vista. En este modelo se analizan todos los acentos con el mismo inventario de acentos tonales sin tomar tanto en cuenta si pertenecen a la cabeza o al tono nuclear según los términos británicos. Sin embargo, se considera que el último de estos acentos es el más importante (cfr. ‘núcleo’) y también es el que más movimiento entonativo puede manifestar, por lo que es necesario emplear otros mecanismos además de los acentos tonales para poder describir la complejidad de las terminaciones. A pesar de las grandes diferencias entre el análisis por configuraciones y el métrico autosegmental, la continuidad y la compatibilidad entre los dos se hacen más aparentes al observar interpretaciones autosegmentales de las configuraciones británicas y también de los componentes del grupo entonativo (véase 1.2.1.1) (García-Lecumberri 2003: 58-59), por ejemplo:

Cabeza alta + descenso bajo	H+L* L-L%
Cabeza alta + descenso-ascenso	H+L* L-H%
Cabeza baja + ascenso alto	L+H* H-H%
Cabeza – núcleo	Dominio inicial – Dominio nuclear

Según Hualde (2003: 155-157), en muchos aspectos podemos hallar antecedentes directos de este modelo en los principios para el análisis de las lenguas tonales que se conocen como teoría fonológica autosegmental (Goldsmith 1976) y otras ideas relacionadas que se propusieron durante los mismos años para el análisis de la entonación del inglés (Libermann 1975) y del sueco (Bruce 1977).

En la teoría fonológica autosegmental se asume que la melodía o modulación tonal de los enunciados constituye un nivel separado y en cierto modo independiente de los demás rasgos fonológicos. Los tonos que caracterizan un enunciado (melodía) son autosegmentos asociados con el nivel segmental (texto) por medio de reglas en parte universales y en parte específicas para cada lengua.

La teoría segmental se aplicó a lenguas tonales con unidades tonales contrastivas mínimas que pueden dar lugar a contornos complejos mediante su combinación. Entre otros aspectos, la independencia del nivel tonal con respecto al segmental y la concepción del aspecto melódico de los enunciados como el resultado de la combinación a nivel fonológico de una serie de especificaciones tonales contrastivas como H y L, también se encuentran en el modelo métrico autosegmental.

El término ‘métrico’ proviene del hecho de que en algunas lenguas la sílaba tiene un tono léxicamente especificado. En lenguas como el inglés o el español no encontramos esta característica, ya que en estas dos lenguas los tonos tienen funciones pragmáticas, es decir, constituyen la melodía de enunciados, no de palabras, y se asocian de manera distintiva sólo con ciertas sílabas. En su estudio sobre la entonación del inglés, Liberman (1975) propuso que en esta lengua la asociación de la melodía con el texto, es decir, el hecho de que ciertos tonos aparezcan asociados con ciertas sílabas en un enunciado concreto, depende de las relaciones de prominencia que se establecen de manera independiente tanto entre las sílabas de cada palabra como entre palabras en la oración. En otras palabras, en muchas lenguas los elementos con los que se asocian los tonos no son todas y cada una de las sílabas, sino sólo aquellas sílabas que tienen prominencia a nivel léxico o discursivo.

La teoría métrica autosegmental parte de la premisa de que el contorno entonativo de un enunciado es el resultado de la interpolación fonética entre eventos tonales fonológicamente especificados y asociados con determinadas sílabas. Básicamente, a nivel fonológico en lenguas como el inglés o el español hay **acentos tonales**, cuando un tono o secuencia de tonos se asocia con sílabas con acento léxico, y **tonos de juntura o de frontera**, cuando se asocian fonológicamente con el límite de una frase.

Tomando esta premisa como punto de partida, Pierrehumbert (1980) y Pierrehumbert y Hirschberg (1990: 272-278) afirman que hay tres rasgos decisivos al describir los patrones entonativos: **acentuación** (*stress*), **melodía** (*tune*) y **fraseo** (*phrasing*). La acentuación se refiere al patrón rítmico y la relativa prominencia de las sílabas de un enunciado. La melodía es la fuente abstracta del patrón de frecuencia fundamental. El fraseo permite la división de un enunciado complejo en distintas frases. Estos rasgos y sus relaciones se explican a continuación.

El **patrón acentual** (*stress pattern*) de un enunciado es el patrón de relativa prominencia de las sílabas (véase 1.1.1.1). El **acento léxico** (*word stress*) se asigna por medio de reglas léxico-fonológicas. El **acento de frase** (*phrasal stress*), por su parte, se afecta al considerar la estructura de la información.

El acento se manifiesta en la duración, amplitud y características espectrales de los segmentos del habla. Por lo general, las sílabas acentuadas se articulan de manera más completa que las sílabas átonas. El patrón acentual es independiente de la melodía, en el sentido de que una melodía dada puede aplicarse a materiales con distintos patrones acentuales y un patrón acentual dado puede producirse con distintas melodías.

En el sistema de descripción entonacional de Pierrehumbert, las melodías se describen como secuencias de tonos bajos y altos que determinan la forma del contorno de la frecuencia fundamental. Algunos de estos tonos, los que participan en los **acentos tonales** (*pitch accents*), se dan en sílabas acentuadas. Si el patrón acentual de un enunciado se altera, el número y ubicación de los acentos tonales se alterará del mismo modo. Otros tonos, los **tonos de frase** (*phrasal tones*), delimitan las frases fonológicas. Si se modifica la manera en que un enunciado se divide en frases, el número y ubicación de los tonos de frase cambiará.

Los acentos tonales marcan la unidad léxica con la que se asocian como prominente. Hay seis diferentes tipos de acentos tonales en inglés: dos tonos simples – H y L – y cuatro tonos complejos (véase la Tabla 1-2). El tono alto, que es el utilizado con mayor frecuencia, aparece como un pico sobre la sílaba acentuada y se representa como H*. La ‘H’ indica un tono alto y el asterisco significa que el tono se alinea con una sílaba acentuada. Los acentos L* ocurren más abajo en el **rango tonal** (*pitch range*) que H* y se realizan fonéticamente como mínimos locales de la frecuencia fundamental o F₀. Los otros acentos tonales en inglés tienen dos tonos, de los cuales uno se selecciona para alinearlos con el acento por medio del diacrítico ‘*’. Estos acentos pueden representarse como L*+H, L+H*, H*+L o H+L*. Según Pierrehumbert y Hirschberg (1990), los acentos con dos tonos iguales no existen en inglés.

La sílaba tónica sirve de punto de anclaje para los eventos tonales que contribuyen a dar prominencia a esta sílaba sobre las otras de la palabra o grupo rítmico, pero el tipo de contorno tonal que se asocia con la sílaba acentuada depende del tipo de enunciado y de la posición y relevancia pragmática de la palabra dentro del mismo. La definición de puntos de anclaje para la melodía es un elemento esencial del modelo métrico autosegmental y que lo distingue de otros modelos de la entonación. (Hualde 2003: 159)

Pierrehumbert y Hirschberg (1990) proponen que para el inglés hay dos niveles de **fraseo** (*phrasing*) que se involucran en la especificación de la melodía: la **frase intermedia** (*intermediate phrase*) y la **frase entonativa** (*intonational phrase*). Una frase intermedia consta de uno o más acentos tonales, más un tono simple alto o bajo (ya sea H o L, anteriormente representados como H⁻ y L⁻), que indica el final de la frase. Utilizando un término obsoleto, Pierrehumbert se refiere a este tono como **acento de frase** (*phrase accent*),

que debería llamarse más bien tono de frase. Una importante propiedad fonética del acento de frase es que controla la F_0 entre el último acento tonal y el inicio de la siguiente frase intermedia o el final de un enunciado. Las frases intermedias y entonativas se identifican por medio de pausas o por alargamiento de la sílaba al final de la frase.

Las frases entonativas se componen de una o más frases intermedias. El final de una frase entonativa se marca con un tono H o L adicional llamado **tono de frontera** o **marginal** (*boundary tone*) y se indica por medio del diacrítico '%'. Este tono recae exactamente en el final de la frase. Ya que el final de cada frase entonativa es también el final de una frase intermedia, hay en total cuatro maneras en las que la melodía podría aparecer después del último acento tonal: L L%, H L%, L H% o H H%.

La melodía se define por su particular secuencia de acento tonal, acento de frase y tono de frontera. Por lo tanto, un patrón declarativo ordinario con un valle final se representa como H* L L%, es decir, una melodía a base de un acento tonal H*, un acento de frase L y un tono de frontera L%. Un contorno interrogativo típico se representa como L* H H%.

Ladd (1996: 112), a quien se debe la clasificación bajo el término 'métrico autosegmental' del modelo de Pierrehumbert y de todos los modelos posteriores que parten de planteamientos teóricos similares, identificó cuatro principios básicos que hacen del modelo de Pierrehumbert una descripción efectiva de la entonación: primero, su estructura lineal; segundo, la idea de que el acento es un fenómeno entonacional que se asocia con ciertas sílabas acentuadas, y no con la esencia fonética de la acentuación; tercero, el análisis de los acentos en los componentes de niveles tonales H y L; y, por último, la especificación local de tendencias globales de F_0 . Además, Ladd (1996: 42) identificó dos omisiones específicas en el análisis específico que Pierrehumbert hace del inglés – el acento de frase y el análisis del escalonamiento descendente (*downstep*). Para resolver este problema, Ladd introduce el concepto de frase intermedia, que como hemos visto más arriba, fue implementado en las nuevas versiones del modelo de Pierrehumbert. Por otro lado, el escalonamiento descendente, que se produce cuando hay secuencias de tonos H cuyo descenso en la F_0 no se debe únicamente a la declinación natural, sino que el hablante baja voluntariamente la altura tonal de manera escalonada (véase 1.2.1.2). Pierrehumbert y Hirschberg (1990: 280-283) los analizan como una secuencia de H*+L...H* y Ladd (1996: 90-92) prefiere considerarlo como una posibilidad lingüística independiente expresada por medio del diacrítico '!', es decir H*...!H*.

El número y tipo de acentos tonales puede variar de una lengua a otra. Sin embargo, para todas las lenguas que se han analizado hasta ahora empleando este modelo, hay acuerdo casi

unánime en que basta con especificar cuando mucho dos tonos como constituyentes del acento tonal. En principio, el sistema permite distinguir los siguientes acentos monotonaes y bitonaes:

H*	Pico en la sílaba tónica.
L*	Valle en la tónica.
L+H*	Pico en la tónica precedido por un valle (subida de la pretónica a la tónica).
L*+H	Valle en la tónica seguido por un pico (subida de la tónica a la postónica).
H+L*	Valle en la tónica precedido por un pico (bajada de la pretónica a la tónica).
H*+L	Pico en la tónica seguido por un valle (bajada desde la tónica).

El análisis de los contornos entonativos de una lengua determinada es lo que nos llevará a postular la existencia de unos u otros acentos contrastivos en esa lengua. Además de estas combinaciones, existen otras posibilidades como H+H* (en algunos análisis del inglés) y acentos tritonaes como L+H*+L, aunque éstos últimos aún no se hayan utilizado en ninguna lengua analizada bajo este modelo.

En cuanto a los tonos de frontera, en algunos análisis se ha empleado M%, además de L% y H%, para indicar un tono medio final que contrasta tanto con una subida final como con una bajada. Este es el caso, por ejemplo, del sistema notacional *Spanish Tones and Break Indices* (Sp-ToBI) que veremos a continuación. (Hualde 2003: 164-165)

1.2.4.1 Tones and Break Indices (ToBI y Sp-ToBI)

Una versión modificada y simplificada del modelo de Pierrehumbert se incorporó al sistema de transcripción ToBI (*Tones and Break Indices*), que se propuso como el estándar para etiquetar rasgos prosódicos de grabaciones digitales de voz en inglés. Una transcripción completa de ToBI contiene varios niveles o estratos que incluyen la transcripción ortográfica y otros peldaños para comentarios. Los peldaños más importantes son los que indican los **tonos** (*tones* - To) en el contorno de la frecuencia fundamental y los **índices de disyunción** (*break indices* - BI) que señalan la fuerza de cada límite léxico, y son estos dos niveles los que le dan el nombre al sistema. (Beckman y Ayers 1997, Beckman *et al.* 2002)

El sistema ToBI no es como un alfabeto fonético internacional que pueda ser utilizado para transcribir todas las lenguas. Cada idioma necesita un conjunto especial de signos. Siguiendo la costumbre de añadir un prefijo denotador de la lengua en cuestión al nombre genérico ToBI, un equipo de trabajo de la Ohio State University, encabezados por Beckman, propuso la creación del Sp-ToBI (*Spanish Tones and Break Indices*). La notación prosódica que presentan pretende ser panhispánica y fonémica. Esto significa que eligieron un conjunto inicial de etiquetas para capturar los fenómenos principales que aparecen en todos los dialectos y después se incrementará el sistema para que pueda capturar todas las distinciones

significativas en cada dialecto que se vaya investigando a mayor profundidad. Sin embargo, capturar tanto los paralelos funcionales como las diferencias de superficie implica que algunas de estas etiquetas parecerán alofónicas a los hablantes de algunos dialectos (Beckman *et al.* 2002: 10-11).

El propósito del modelo notacional Sp-ToBI es ofrecer transcripciones sistemáticas de la prosodia de las múltiples variedades del español, que puedan ser utilizadas consistentemente para la notación de *corpora* de habla de diversos tipos y orígenes.

Una típica transcripción Sp-ToBI consiste en el oscilograma junto con el trazado de la frecuencia fundamental del enunciado, en ventanas paralelas y coordinadas. Los símbolos notacionales, al igual que en el ToBI, están divididos en cinco niveles o estratos distintos: el estrato de palabras, el de sílabas, el de índices de disyunción, el de tonos y el de código. También se le puede añadir un estrato ‘miscelaneo’ para observaciones diversas. Las convenciones acordadas hasta ahora son provisionales, pues el Sp-ToBI todavía debe ser considerado como un proyecto en proceso de elaboración. Sin embargo, según la opinión de muchos investigadores, esta es una iniciativa que tendrá un impacto importante en el estudio y la regularización de la notación de la prosodia del español. (Sosa 2003: 188)

El actual conjunto provisional de convenciones del Sp-ToBI, según Beckman *et al.* (2002: 32-34) y Sosa (2003: 188-190), consta de los siguientes elementos:

Estrato de palabras

- Transcripción ortográfica ordinaria y segmentación del enunciado en palabras ortográficas.

Estrato de sílabas

- Transcripción de los segmentos sílaba por sílaba. Para que estas etiquetas sean transferibles entre distintas plataformas, la transcripción deberá ser hecha en algún conjunto de caracteres ASCII, como el alfabeto SAMPA (Wells 1989). Mientras se adopta una herramienta automática para el alineamiento de la transcripción con el archivo de audio, se recomienda utilizar este estrato únicamente para sílabas con alofonía segmental potencialmente reveladora, por ejemplo en los casos de ensordecimiento, aspiración, etc.

Estrato de índices de disyunción

- Señalamiento de impresiones subjetivas de disyunción entre pares de palabras y antes de cada pausa. Se distingue entre los siguientes tres niveles:
 - 0** En cualquier caso de reducción silábica por contacto de vocales entre palabras, como en casos de sinalefa,
 - 1** Cualquier otro caso ordinario de juntura entre palabras,

4 Frase entonativa (grupo melódico).

- Se reservan los niveles **2** y **3** para posibles unidades con claros marcadores fonológicos y un sentido de disyunción intermedia entre **4** y **1**; los posibles candidatos para estos niveles podrían ser la frase intermedia, el grupo tónico o el grupo clítico.

Estrato de tonos

- Se distingue entre los siguientes tres tipos de acentos tonales:

L*+H Acento con subida tardía, con pico después de la sílaba acentuada y valle hacia el principio de la sílaba acentuada (como en casos de acento prenuclear en el español mexicano y algunas variedades peninsulares, o acento focal en las regiones catalanoparlantes de España) o hacia la mitad de la sílaba acentuada (como el acento prenuclear en algunas variedades caribeñas).

L+H* Acento con subida temprana, con pico durante la sílaba acentuada (en variedades peninsulares) o inmediatamente después de la sílaba acentuada si ésta es intrínsecamente corta (en variedades caribeñas).

H+L* Clara caída de un tono previo más alto a un tono más bajo durante la sílaba acentuada, comenzando alrededor del inicio de la rima.

- Cuando una sílaba suena acentuada, pero es difícil clasificarla bajo uno de los tres tipos anteriores, o sus versiones escalonadas descendente o ascendente, usar temporalmente alguna de las siguientes dos etiquetas en su lugar para después volver a analizarla, cuando se haya investigado más sobre el inventario de tonos del dialecto:

H* Un claro pico menor durante la sílaba acentuada, más o menos al mismo nivel que un claro **L*+H** previo, cuando su altura no pueda ser atribuida a un escalonamiento ascendente o a alguna reducción.

* La percepción de acento en una sílaba en caso de que la forma tonal sea demasiado ambigua, incluso para **H***.

- Se señala el escalonamiento ascendente y descendente directamente en el acento tonal, como en:

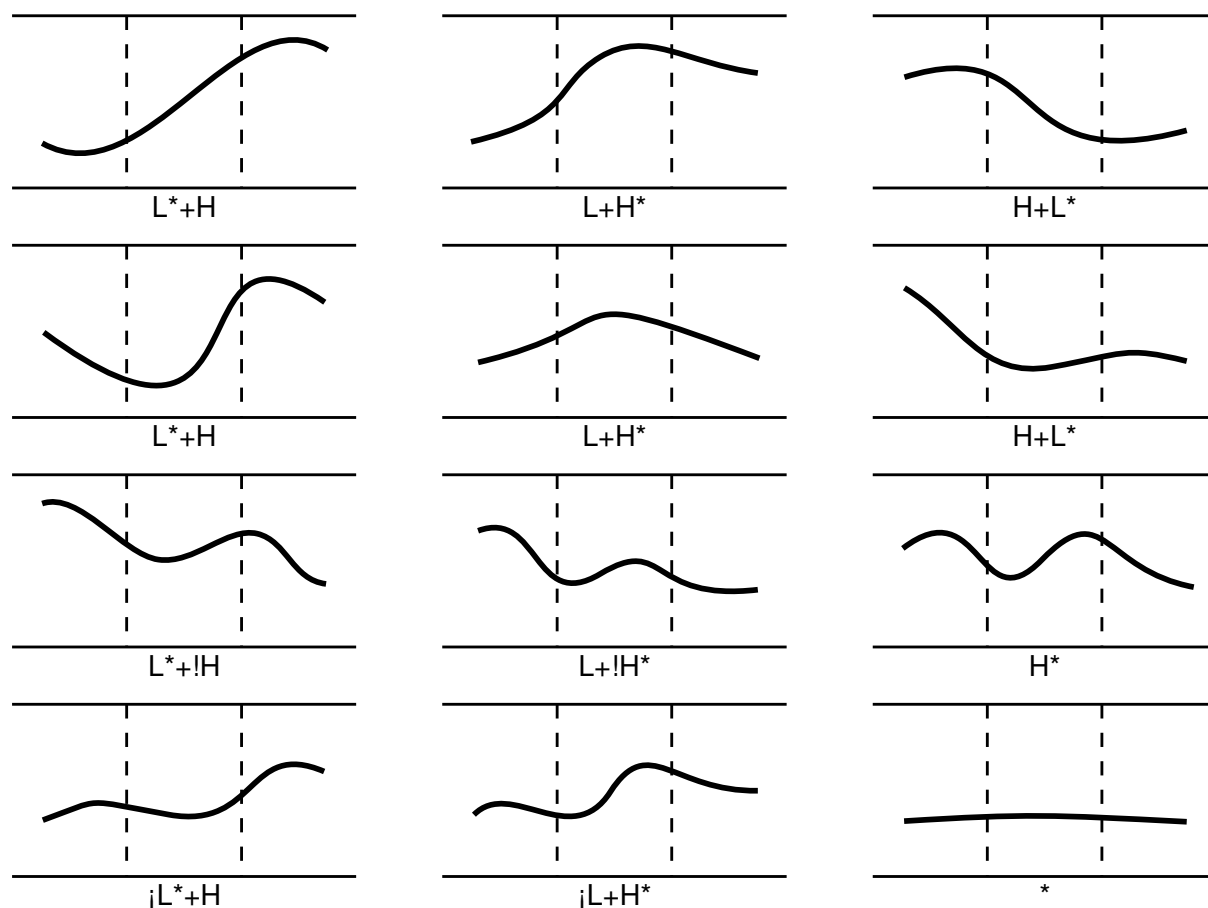
L*+!H Variante escalonada descendente de **L*+H**

L+!H* Variante escalonada descendente de **L+H***

¡L*+H Variante escalonada ascendente de **L*+H** (adición extraoficial propia)

¡L+H* Variante escalonada ascendente de **L+H***

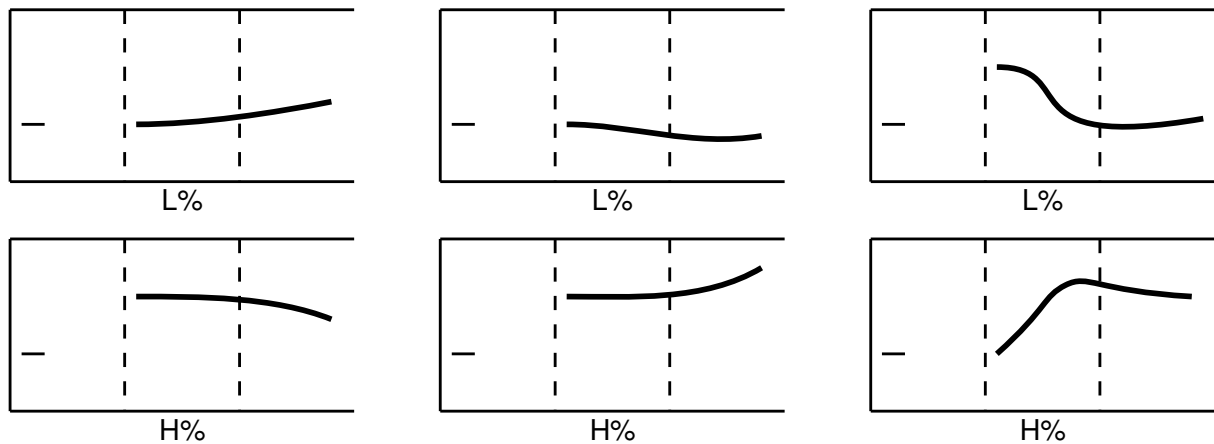
En la siguiente Ilustración (1-14) se hace una clasificación de los tonos nucleares más comunes en español y sus contornos más comunes.

Ilustración 1-14: Tonos prenucleares y sus contornos

Fuente: Diseño propio

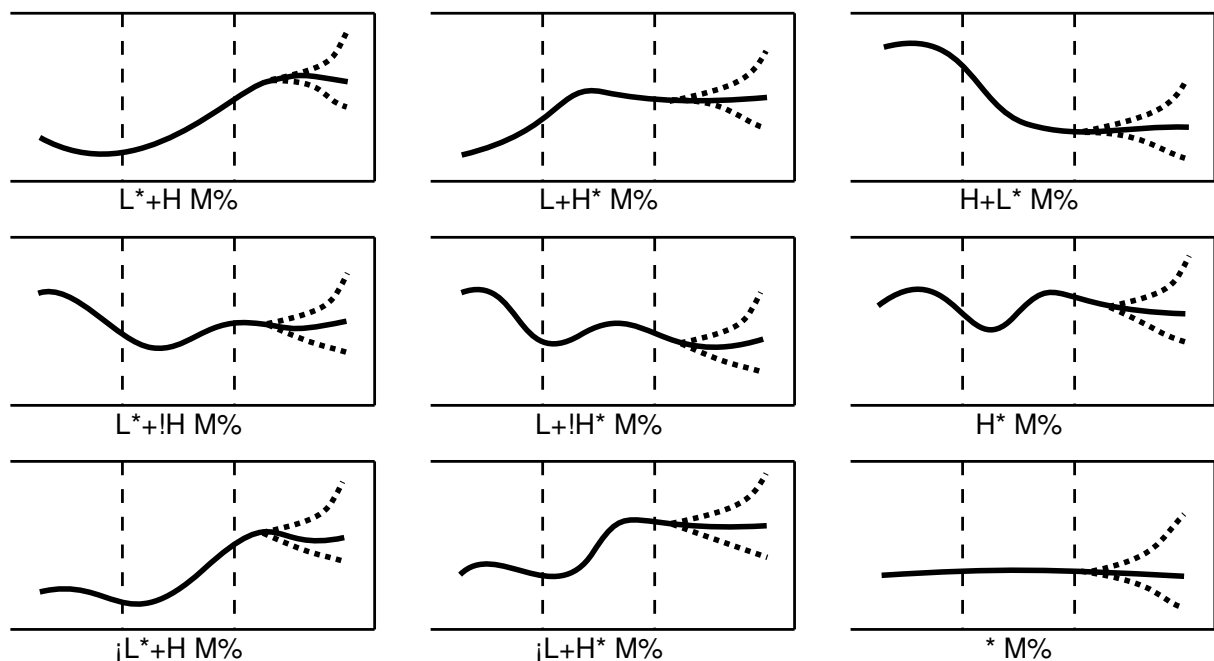
- Se distingue entre las siguientes dos formas de tonos de frontera:
 - L%** Caída a una frecuencia fundamental (F_0) menor después de $L+H^*$ u otro tono ascendente, o mantenimiento del mismo nivel de F_0 después de $H+L^*$.
 - H%** Subida a un nivel de F_0 mayor después de cualquier acento tonal.
- No optar prematuramente por un análisis a base de tonos de frontera con escalonamiento, ¡L% o !H%, para un dialecto sin un inventario de tonos previamente establecido. En su lugar, utilizar:
 - M%** Media subida o meseta tonal a nivel medio después de $L+H^*$, H^* , etc.

En esta investigación, aunque se utilizan los tres tonos como marcas de frontera final, para señalar el inicio de las frases entonativas se considera únicamente los tonos L% y H%. En la Ilustración 1-15 Podemos observar los contornos más típicos que pueden tener estos tonos de frontera en español. Dado que la designación de un contorno con L% o H% es relativa al rango tonal de cada hablante, en las imágenes se señala con un guión la altura media de voz. Como se podrá ver, no únicamente es decisiva la altura tonal de la que arranca el contorno, sino su dirección al dirigirse hacia el primer acento de la frase.

Ilustración 1-15: Tonos de frontera inicial y sus contornos

Fuente: Diseño propio

Por otra parte, en la Ilustración 1-16, podemos ver los tonos nucleares de la Ilustración 1-14 combinados con el tono M% en posición final. Con líneas punteadas se indica por encima y por debajo del contorno la forma que podría tener el contorno estilizado al combinarse con H% o con L%.

Ilustración 1-16: Tonos nucleares en combinación con tonos de frontera final

Fuente: Diseño propio

Estrato misceláneo

- Etiquetado de fenómenos tales como las pausas dubitativas, disfluencias, risas o cualquier otro que complique el análisis del patrón tonal y/o del fraseo melódico. También se utiliza este estrato para anotar las propuestas de análisis alternativos para el enunciado.

Estrato de código

- Identificación del dialecto (y sociolecto) del hablante, en caso de que se sepa. Eventualmente, se espera poder utilizar este estrato para marcar también casos de alternancia de códigos entonativos.

Más adelante (v. 3.2.5), se verá que se puede simplificar y agrupar esta multitud de tonos en sólo cuatro grupos dependiendo de la posición del pico frecuencial con respecto a la sílaba acentuada: 1) pico durante la sílaba acentuada – \acute{S} (L+H*, L+!H*, ¡L+H*, H*), 2) pico después de la sílaba acentuada – \acute{S} (L*+H, L*+!H, ¡L*+H), 3) pico antes de la sílaba acentuada – S' (H+L*), y 4) sin pico o poco perceptible durante la sílaba acentuada – S (*).

1.2.4.2 Entonación en el discurso

Para ilustrar una de las aplicaciones de la teoría métrica autosegmental al análisis del discurso, se presenta la propuesta de Wennerstrom (2001: 17), quien parte de la premisa de que los hablantes pueden manipular la altura tonal en palabras, frases e incluso constituyentes temáticos particulares para concederles significado sobre las relaciones en el discurso. La entonación no se deriva automáticamente de los patrones acentuales o la sintaxis de un enunciado; en su lugar, un hablante decide asociar patrones acentuales particulares a constituyentes particulares, dependiendo del contexto del discurso.

Por lo tanto, hay una gran interacción entre la entonación de un hablante y el significado y organización discursivos. Wennerstrom describe un sistema de significados entonativos en el que pequeñas unidades significativas se combinan para conferir una interpretación global de la entonación en el discurso. Para este fin, propone que hay cuatro categorías principales de significados entonativos. El primero es el acento tonal, que indica cuáles elementos de la estructura informativa del discurso se introducen por vez primera o que conceden una relación cohesiva con los elementos que ya se encuentran dentro del contexto de la interacción. Una segunda categoría de la entonación es el tono de frontera al final de la frase entonativa, que indica las interdependencias entre enunciados o turnos. Un tercer tipo de significado es el que confiere la elección del rango tonal al inicio de cada frase, que permite identificar la actitud o reacción del hablante con respecto al enunciado previo. Finalmente, los ‘paratonos’ (cfr. ‘*paragraphs*’, correspondencia con los párrafos escritos) son las elecciones de rango tonal asociadas con unidades organizacionales que indican la manera en la que una nueva unidad será integrada temáticamente a las unidades previas (Wennerstrom 2001: 45). En la Tabla 1-3 se presenta un resumen de las funciones discursivas de la entonación:

Esta aplicación del modelo métrico autosegmental expande las posibilidades de análisis interdisciplinarios entre especialistas de distintas áreas de la lingüística como son los fonetistas, fonólogos y analistas del discurso.

Tabla 1-3: Funciones discursivas de la entonación

Entonación	Funciones discursivas
<i>Acentos tonales</i>	<i>Rol de los elementos léxicos sobresalientes en la estructura informativa</i>
H*	Agregado: nuevo
L+H*	Agregado: contrastante
L*	No agregado: accesible, extraproposicional, proposición cuestionada
L*+H	No agregado: relevancia incierta en contraste con otro elemento
<i>Tonos de frontera</i>	<i>Grado de dependencia de una frase con la siguiente</i>
ascenso alto ↑	Anticipa el siguiente elemento (normalmente de otro hablante)
ascenso bajo ↗	Anticipa el siguiente elemento (normalmente del mismo hablante)
suspensión →	Anticipa el siguiente elemento (enumeración o pausa dubitativa)
descenso parcial ↘	Anticipa el siguiente elemento (conexión muy cercana)
descenso ↓	Independiente del siguiente elemento
sin frontera (interrupción)	Pensamiento incompleto (normalmente dentro de un mismo turno)
<i>Rangos tonales</i>	<i>Reacción o actitud al inicio de una frase con respecto a la anterior</i>
rango tonal alto ↗	contrastivo
rango tonal medio →	aditivo
rango tonal bajo ↘	reformulación de lo obvio, conclusión ya conocida
<i>Paratonos</i>	<i>Organización temática</i>
alto ↑↑	nuevo tema
bajo ↓↓	aparte del tema principal, observación parentética

Fuente: Basado en Wennerstrom (2001: 26)

1.2.4.3 Tonemas del español

Un aspecto por el cual las descripciones métrico autosegmentales de la entonación del español han divergido en relación al número e identidad de las unidades subyacentes tiene que ver con los diferentes niveles de abstracción y de mayor o menor ‘superficialidad’ de los análisis. Esta diferencia en la interpretación de lo fonológico y fonético es la principal causa de los desacuerdos entre los analistas. Mientras para algunos los acentos tonales son pocos y las variaciones se explican mayoritariamente debido al contexto fonético y a las modalidades de implementación, para otros, como Sosa (2003), la fonología tonal debe dar cuenta de manera transparente de todas las configuraciones tonales significativas.

La categoría tonema, heredada de Navarro Tomás (1948), corresponde al último acento tonal del enunciado más los tonos de juntura del modelo métrico autosegmental. Asimismo, la sílaba métricamente fuerte de la última palabra acentuada es el núcleo. Aunque no ha sido costumbre en el modelo métrico autosegmental original distinguir entre los acentos tonales que pueden estar en posición nuclear y los prenucleares, en el caso del español ha sido muy productivo distinguir entre acentos tonales tonemáticos y pretonemáticos. (Sosa 2003: 199-201)

Sosa (1999: 87, 2003: 201) menciona que la noción de acento de frase puede ser menos útil, incluso superflua, en lenguas como el español y otras lenguas de núcleo tonemático fijo. En estas lenguas, el último acento tonal no puede estar muy lejos del borde derecho de la frase, por lo que un eventual acento de frase no podría generar ningún tipo de contraste. En su modelo prescinde de la noción abstracta de acento de frase, comprobando que es posible generar todos los contornos del español sin necesidad de utilizarla.

Por lo tanto, las representaciones subyacentes de los tonemas del español que Sosa (1999: 132-133) recoge en su investigación se presentan a continuación combinando acentos y tonos de frontera (entre paréntesis aparecen las correspondencias con el Sp-ToBI):

Tabla 1-4: Combinaciones posibles de acentos y tonos de frontera en español

1. Descendentes:	{	H* L%	(; L+H* L%)
		L* L%	
		H+L* L%	
		L+H* L%	
		H+H* L%	
2. Ascendentes:	{	H* H%	
		L* H%	
		H+L* H%	
		L+H* H%	
		L*+H H%	
3. Suspensivo:	{	H*+H L%	(H* M%)

Fuente: Diseño propio, basado en Sosa (1999)

Al comparar los trabajos de Sosa (1999) y Sosa (2003) se pueden notar ligeros cambios en su repertorio de tonemas subyacentes que lo asemejan más al Sp-ToBI. Por ejemplo, ya no incluye la posibilidad de combinar tonos iguales en el acento tonal de los tipos H*+H y H+H*, ni las otras tres combinaciones posibles: la secuencia H*+L H%, que produciría un tonema descendente-ascendente (*fall-rise*); la secuencia L*+H L%, que produciría un tonema ascendente-descendente (*rise-fall*); y la secuencia H+H* H%, que produciría un efecto de escalonamiento ascendente (*upstep*). En su obra más reciente, Sosa adopta los tonos con escalonamiento ascendente y descendente según se propone en el sistema Sp-ToBI.

En mi investigación personal, que se irá presentando en los siguientes capítulos, se pretende analizar muestras del español hablado en Madrid y en la Ciudad de México para comprobar en cada uno de los dos casos la existencia de las distintas combinaciones de tonos, tanto en posición pretonemática como tonemática, y en un segundo paso, comparar paralelamente las realizaciones fonéticas de los distintos tipos de enunciados en ambas variedades para poder hablar de similitudes y diferencias dialectales a nivel prosódico. Dado que la mayoría de los investigadores trabajando actualmente con la entonación del español utilizan el modelo métrico autosegmental y considero que la compatibilidad de sistemas de notación tonal es de gran importancia. Es por eso que también tomaré como base el repertorio de unidades tonales del sistema Sp-ToBI.

1.2.5 La escuela holandesa: El modelo IPO

El rasgo que caracteriza y diferencia más claramente la aproximación del IPO (cuyas siglas corresponden al *Institute for Perception Research* con sede en Eindhoven, Holanda, donde la descripción de las curvas melódicas comenzó a desarrollarse a principios de los años 60) con respecto a otros modelos entonativos es su orientación marcadamente fonética. Su metodología parte del análisis experimental, exhaustivo y sistemático, del principal objeto fonético de la entonación: las curvas melódicas. Su objetivo es la definición de un modelo predictivo de la **forma** de las curvas melódicas, determinando sus constituyentes básicos (los **patrones melódicos**) y las reglas que determinan su combinación, como paso previo para el estudio de su función comunicativa o su representación lingüística. Por el contrario, dedica poca atención a la definición de las unidades prosódicas que subyacen a la forma fonética en la entonación, aunque éstas quedan determinadas de una forma más o menos implícita como ámbito de aplicación de los patrones definidos. (Garrido 2003: 97-98)

El modelo IPO concibe las curvas melódicas como el resultado de la superposición de una serie de patrones melódicos de ámbito reducido (varias sílabas), los **movimientos** y las **configuraciones**, a un patrón de ámbito mayor (un grupo entonativo), las llamadas ‘líneas de **declinación**’.

En este sentido, el modelo IPO puede considerarse un modelo jerárquico, ya que asume la existencia de los patrones melódicos de diferentes ámbitos, unos más globales (la declinación) y otros más locales (los movimientos y las configuraciones), que se combinan para la construcción de una curva melódica. (Garrido 2003: 98)

1.2.5.1 Componentes básicos del modelo

El modelo IPO considera que la forma global de las curvas melódicas se rige por el fenómeno de la declinación. Se denomina **declinación** a la tendencia que tienen las curvas melódicas a ir rebajando su nivel tonal a medida que evolucionan en el tiempo.

En el límite entre dos grupos entonativos, la F_0 sufre un proceso de **reajuste** (*reset*), que sitúa el nivel global de F_0 del contorno siguiente en un nivel más alto que al final del contorno anterior, para dar lugar a un nuevo proceso de declinación a lo largo del contorno siguiente. (Garrido 2003: 99)

La altura del reajuste de F_0 entre dos grupos entonativos consecutivos está controlada por la existencia de fenómenos de declinación de ámbito superior al grupo entonativo, que Garrido (2003: 100-101) denomina **supradclinación**. De la misma forma que a lo largo de un grupo entonativo se produce un descenso progresivo del nivel medio de la F_0 , entre grupos entonativos consecutivos también se observa esta tendencia. El reajuste de F_0 volvería a elevar el nivel de F_0 al inicio, pero no tanto como al inicio del grupo anterior. Es lo que se denomina **reajuste parcial** (*partial reset*) de la F_0 .

El IPO representa la declinación mediante rectas, las líneas de declinación, que marcarían los hipotéticos límites inferior (**línea de declinación baja**) y superior (**línea de declinación alta**) de los contornos melódicos en cada instante de tiempo. También considera una tercera línea, equidistante de las dos (**línea de declinación intermedia**).

Estas tres líneas de declinación definen tres niveles tonales (*Top*, *Mid* y *Bottom*) en los que se moverán los contornos melódicos. En el modelo IPO, se considera que estas tres líneas son paralelas, si se establecen a partir de una representación logarítmica de la altura tonal de la curva, en semitonos.

1.2.5.2 Movimientos melódicos

El modelo IPO asume que las curvas melódicas pueden modelarse como series de líneas rectas que, concatenadas y superpuestas a las líneas de declinación, las conforman. Cada uno de estos segmentos es en el modelo IPO un movimiento melódico (*pitch movement*). (Garrido 2003: 103)

Para cada lengua existe un número limitado y reducido de movimientos melódicos: en el caso del holandés, diez, más los movimientos ‘planos’ que siguen las líneas de declinación. Los movimientos tienen cada uno una etiqueta que los identifica (un número si es un movimiento ascendente, y una letra si es un movimiento descendente, según la notación propuesta por el IPO). Además, cada movimiento melódico se caracteriza por una matriz de rasgos distintivos

binarios, al estilo de los que se emplean para definir los fonemas. El inventario de rasgos que se propone en 't Hart, Collier y Cohen (1990: 152-153) para el holandés es el siguiente:

- 1) En función de su dirección, los movimientos se definen como [+ascendente] o [-ascendente] ([±*rise*]). Los movimientos ‘planos’, que no implican ni ascenso ni descenso de la F₀, no se consideran en el modelo del IPO propiamente movimientos, y, por tanto, no se definen en función de ningún rasgo.
- 2) En función del punto final del movimiento, establecido con respecto al inicio de la parte sonora de la sílaba, los movimientos se definen por medio de dos rasgos binarios: [±anticipado] ([±*early*]), que indica que éste finaliza cerca del inicio de la parte sonora; y [±retardado] ([±*late*]), que indica que el final del movimiento está cerca del final de la parte sonora. Si un movimiento acaba hacia la mitad de la parte sonora, se etiqueta como [-anticipado] y [-retardado].
- 3) De acuerdo con el número de sílabas que abarca el movimiento, éste puede ser [±extendido] ([±*spread*]): [+extendido] si el movimiento está asociado con 2 o más sílabas, y [-extendido] si abarca sólo una sílaba.
- 4) De acuerdo con el rango tonal del movimiento, éste puede ser [±completo] ([±*full*): [+completo] si el movimiento comprende todo el rango tonal del grupo (entre la línea superior y la línea inferior de declinación), y [-completo] si el movimiento cubre un rango menor.

Este conjunto de rasgos permite caracterizar cada movimiento y distinguirlo de todos los demás. La Tabla 1-5 Presenta el conjunto de movimientos definidos para el holandés.

Tabla 1-5: Inventario de movimientos melódicos del holandés y su caracterización en rasgos distintivos

	/1/	/2/	/3/	/4/	/5/	/A/	/B/	/C/	/D/	/E/
[ascendente]	+	+	+	+	+	-	-	-	-	-
[anticipado]	+	-	-	-	+	-	+	-	-	+
[retardado]	-	+	-	+	-	-	-	+	+	-
[extendido]	-	-	-	+	-	-	-	-	+	-
[completo]	+	+	+	+	-	+	+	+	+	-

Fuente: Garrido 2003: 104

1.2.5.3 Configuraciones

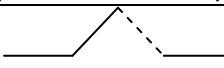
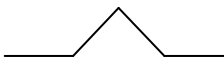
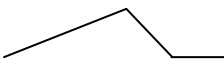
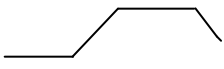
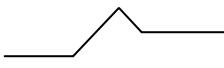
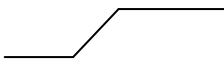

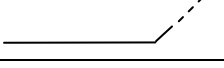
El modelo IPO considera que los movimientos son a su vez los constituyentes de otros patrones de ámbito superior, las llamadas **configuraciones**, formas recurrentes en las curvas melódicas que pueden modelarse como series de uno o más movimientos melódicos. (Garrido 2003: 104-105)

Se distinguen tres tipos de configuraciones, en función de la posición de ésta en el contorno:

- configuraciones **prefijo** (las que aparecen al inicio de un contorno melódico);
- configuraciones **raíz** (las que aparecen en posición interior de un contorno);
- configuraciones **sufijo** (las que aparecen en posición final de contorno).

En la Tabla 1-6 Se presenta el inventario de configuraciones definidas en 't Hart, Collier y Cohen (1990: 154-157) para el holandés.

Tabla 1-6: Inventario de configuraciones para el holandés

<i>Tipo</i>	<i>Etiqueta</i>	<i>Representación esquematizada</i>
Prefijo	/1B/	
Raíz	/1A/	
	/4A/	
	/3C/	
	/1E/	
	/1/	
Sufijo	/2/	
	/2/	

Fuente: Garrido 2003: 106

Una o varias configuraciones combinadas nos darían el contorno final, que es la representación estilizada de la curva melódica de todo un grupo entonativo. Las posibilidades combinatorias de las distintas configuraciones para formar los distintos contornos posibles en una lengua dada se definirían de acuerdo a una gramática como la propuesta por t'Hart *et al.* (1990): (Prefijo)ⁿ + Raíz + (Sufijo). Es decir, una configuración raíz, que es el único elemento obligatorio, puede tener, opcionalmente, una configuración sufijo y una o más configuraciones prefijo.

1.2.5.4 Unidades prosódicas

Debido a la orientación fonética del modelo IPO, la definición de las unidades que sirven de marco de aplicación a los distintos tipos de patrones del modelo no fue abordada de manera explícita en su formulación original. Sin embargo, el modelo asume que las unidades relevantes son aquellas que actúan como ámbito de los patrones definidos en el modelo. Así, el grupo entonativo es la unidad básica del modelo sobre el cual se aplica el fenómeno de la declinación.

El límite entre dos grupos entonativos (o sea, el final de un patrón de declinación determinado y el inicio de otro nuevo) viene marcado fonéticamente por la presencia de un reajuste de F_0 , que indicaría el inicio de una nueva línea de declinación. Normalmente, un reajuste de F_0 suele darse coincidiendo con una pausa y con el alargamiento del sonido final del grupo, pero no necesariamente.

Además, el modelo considera la existencia de casos de supradclinación en el ámbito del párrafo, pues tiene, al igual que el grupo entonativo, un correlato fonético en la curva melódica. Esto le da el título de entidad como unidad entonativa. (Garrido 2003: 106-107)

La metodología aplicada por el IPO para el desarrollo del modelo incluye cuatro pasos:

- a) estilización de curvas melódicas y definición de las líneas de declinación;
- b) definición de los movimientos relevantes (estandarización);
- c) definición de las configuraciones posibles y de su combinatoria;
- d) análisis de la relación entre los patrones definidos y la información lingüística que transmiten. (Garrido 2003: 109-110)

1.2.5.5 Ventajas y desventajas

El método IPO permite obtener una descripción detallada y exhaustiva del componente fonético de la entonación de una lengua: patrones melódicos y reglas de combinación que llevan a la construcción de una curva melódica.

Por otra parte, los resultados obtenidos tienen el aval del método experimental, a partir de un riguroso análisis de una gran cantidad de datos. Además, la metodología descrita es aplicable a cualquier lengua, ya que se basa en principios puramente fonéticos, no dependientes de una lengua en concreto.

Entre sus desventajas se encuentra la gran cantidad de tiempo y material fonético que requiere la aplicación de esta metodología según el proceso de estilización original. Sin embargo, los avances tecnológicos en el campo de la computación y del procesamiento de la señal van reduciendo constantemente este esfuerzo.

Por otro lado, el IPO presta poca atención a los aspectos fonológicos de la entonación. No obstante, se trata de una limitación consciente, dado que el propio procedimiento de análisis implica la descripción exhaustiva de los componentes fonéticos de la entonación como paso previo para abordar la cuestión de su función lingüística. Seguramente, completar el modelo requerirá tanto esfuerzo como el que ha supuesto la definición original del modelo. (Garrido 2003: 119-120)

1.2.5.6 Aplicación al español: Patrones melódicos del español en habla espontánea

En el marco teórico de Cantero (2005), la entonación constituye un fenómeno comunicativo diverso: un fenómeno discursivo (o **prelingüístico**, según su terminología), un fenómeno lingüístico (de tipo fonológico) y un fenómeno paralingüístico (o **expresivo**).

La entonación prelingüística (que reúne los fenómenos del acento y la entonación, en el complejo que Cantero llama **jerarquía fónica**) constituye el elemento integrador del habla que permite su inteligibilidad, es decir, el principio unificador del discurso.

La entonación lingüística se refiere al nivel de análisis mediante el cual podemos designar y estudiar las unidades fonológicas de carácter suprasegmental capaces de distinguir unidades de discurso significativas (como las frases). Dichas unidades fonológicas, llamadas **tonemas**, se caracterizan mediante los rasgos fonológicos binarios: /± interrogativo/, /± enfático/, /± suspendido/ (I, E, S, respectivamente). En Cantero (2002) se distingue un total de 8 tonemas en español:

- | | |
|---------------|---------------|
| 1. /+I +E +S/ | 5. /-I +E +S/ |
| 2. /+I +E -S/ | 6. /-I +E -S/ |
| 3. /+I -E +S/ | 7. /-I -E +S/ |
| 4. /+I -E -S/ | 8. /-I -E -S/ |

La entonación paralingüística, por su parte, se refiere a los amplios márgenes de dispersión de los tonemas, en los que caben las muy variadas realizaciones expresivas, emotivas e idiosincrásicas del hablante. La complejidad de la entonación expresiva, de hecho, constituye los márgenes de dispersión de los tonemas /+ enfáticos/.

En Cantero (2005) se denomina como **patrones melódicos** a las variantes tipo de los tonemas, es decir, las melodías típicas de cada tonema, encontradas con asiduidad en un *corpus* de grabaciones tomadas de transmisiones televisivas en España, que han permitido la síntesis melódica y que en las pruebas perceptivas han aparecido como entonaciones claramente reconocibles.

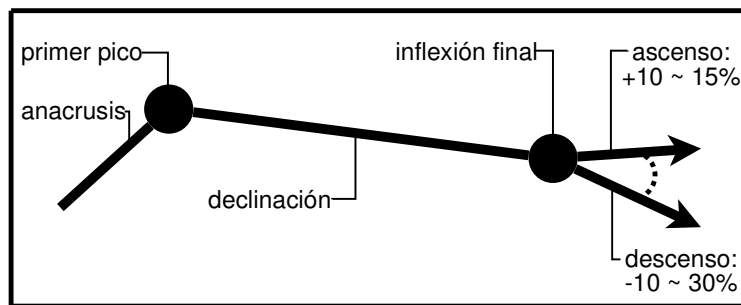
Los patrones melódicos constituyen un nivel de abstracción intermedio entre el nivel fonológico (cuya unidad es el tonema) y el nivel fonético (cuya unidad es el rasgo melódico).

El primer patrón melódico, que está caracterizado fonológicamente como una variante de tonema /-I -E -S/, es el patrón melódico de la entonación neutra, llamado así por no estar marcado por ningún rasgo fonológico positivo.

En la Ilustración 1-17 aparece el **patrón melódico I** con los distintos elementos constituyentes de la mayoría de los contornos entonativos: **anacrusis** (segmentos tonales previos al primer pico), **primer pico** (normalmente, la primera vocal tónica o la siguiente

vocal átona), la **declinación** (sucesión de segmentos tonales entre el primer pico y la inflexión final, que típicamente constituye un leve descenso) y la **inflexión final** (la parte más relevante, y, a veces, la única parte del contorno, constituida por los segmentos tonales a partir de la última vocal tónica).

Ilustración 1-17: Patrón melódico de la entonación neutra



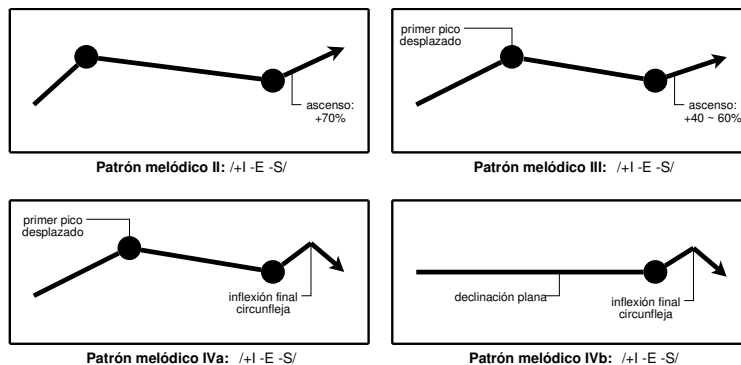
Patrón melódico I: /-I -E -S/

Fuente: Diseño propio, basado en Cantero (2005)

En el patrón melódico I, la inflexión final se caracteriza por un descenso moderado (entre un -10~30%), o bien por un leve ascenso final, no mayor a un +10~15%.

Los patrones melódicos de la entonación interrogativa, representados en la Ilustración 1-18, constituyen variantes del tonema /+I -E -S/.

Ilustración 1-18: Patrones melódicos de la entonación interrogativa



Fuente: Diseño propio, basado en Cantero (2005)

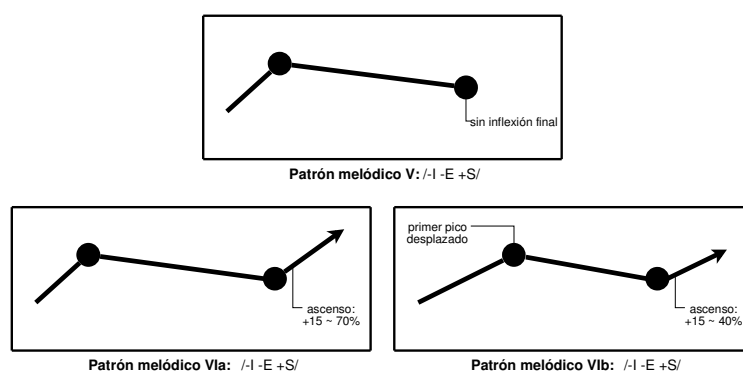
El **patrón melódico II** constituye el patrón típico de la entonación interrogativa, y se define en función de la inflexión final, cuyo ascenso debe ser superior a un +70~80% (y que, normalmente, es superior a un +100%).

El **patrón melódico III** constituye una alternativa al patrón melódico II, con un ascenso final menos pronunciado (+40~60%) que se compensa con el primer pico desplazado (desplazado a la siguiente vocal átona). Esta conjunción de rasgos (primer pico desplazado más ascenso final moderado) define el patrón melódico como /+ interrogativo/.

Los **patrones melódicos IVa** y **IVb** se definen en función de su inflexión final circunfleja ascendente-descendente. En el primer caso (IVa), conjuntamente al primer pico desplazado; en el segundo caso (IVb), conjuntamente a la declinación plana (sin pendiente, además de carecer de pico acentual y anacrusis).

La entonación suspendida, caracterizada fonológicamente como $/-I -E +S/$, aparece rara vez en contornos finales, puesto que consiste, precisamente, de melodías (Ilustración 1-19) que carecen de una inflexión final, como el **patrón melódico V**, o bien, de melodías cuya inflexión final está marcada por no ser terminativa, como el **patrón melódico VIa**, cuya inflexión final es ascendente (entre los márgenes impuestos por los patrones I y II: $+15\sim 70\%$) y el **patrón melódico VIb**, cuyo ligero ascenso final se mueve entre los márgenes impuestos por los patrones I y III (primer pico desplazado y un ascenso de $+15\sim 40\%$).

Ilustración 1-19: Patrones melódicos de la entonación suspendida



Fuente: Diseño propio, basado en Cantero (2005)

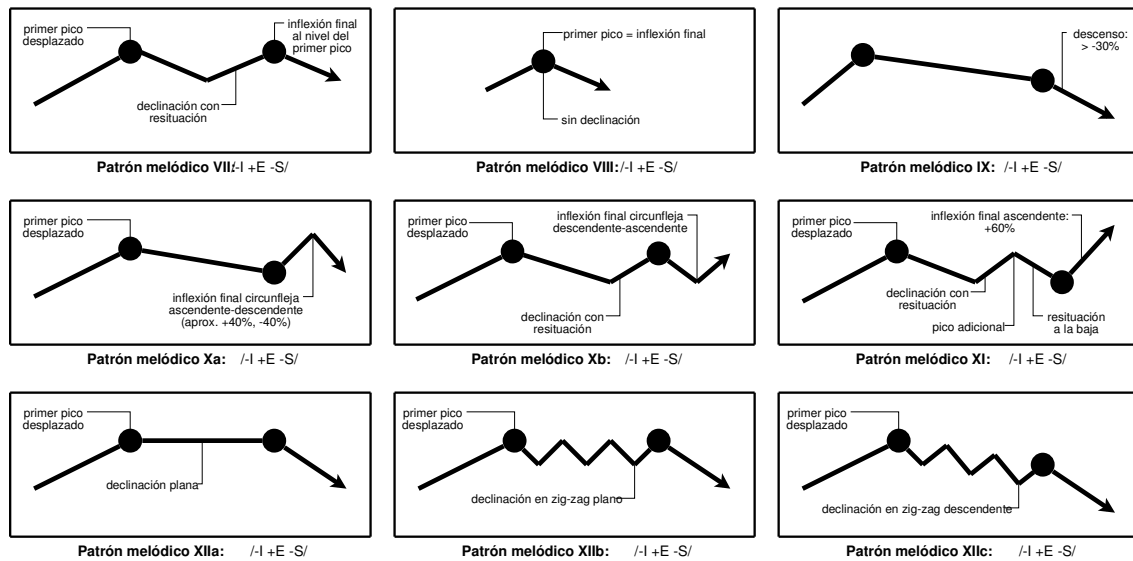
Comúnmente, los contornos suspendidos aparecen como **contornos interiores** del discurso. En el análisis de la conversación, los contornos interiores $/+S/$ permiten al hablante mantener su turno de palabra, por lo que constituyen uno de los recursos más relevantes en los procesos de interacción oral.

La entonación enfática se caracteriza fonológicamente como $/-I +E -S/$. Cantero (2005) define al énfasis como el fenómeno melódico contrastante que permite individualizar cada enunciado dentro de los amplios márgenes de dispersión del tonema $/+ \text{enfático}/$. Estos márgenes de dispersión tan amplios son los que constituyen lo esencial de la entonación paralingüística o expresividad personal.

Los patrones melódicos de énfasis (Ilustración 1-20) pueden dividirse en tres grupos:

- patrones melódicos en los que el énfasis se establece mediante la interacción del primer pico y la inflexión final (patrones VII y VIII)
- patrones melódicos definidos en función de su inflexión final (IX, Xa, Xb y XI)
- patrones melódicos definidos en función de su declinación (XIIa, XIIb y XIIc)

Ilustración 1-20. Patrones melódicos de la entonación enfática



Fuente: Diseño propio, basado en Cantero (2005)

En el **patrón melódico VII** la inflexión final se coloca al nivel del primer pico, mediante una resituación en la declinación.

En el **patrón melódico VIII** el primer pico y la inflexión constituyen el mismo segmento tonal, además de que desaparece toda declinación.

El **patrón melódico IX** es similar al patrón I, pero con un gran descenso en la inflexión final, superior al 30%.

Los **patrones melódicos Xa** y **Xb** se definen por su inflexión final circunfleja, ascendente-descendente en Xa y descendente-ascendente en Xb (en este caso, la inflexión comienza al alza mediante una resituación en la declinación). En ambos patrones, el primer pico desplazado es un rasgo conjunto a la inflexión final circunfleja (cfr. IVa).

El **patrón melódico XI** es similar a los patrones II o III, con el primer pico desplazado a la vocal átona, pero con un gran ascenso, superior a un 60%, precedido por una declinación quebrada en la que aparece un pico interior adicional (como si se tratara de un énfasis de palabra), lo cual permite marcar el énfasis con un valor fonológico /- I/.

El **patrón melódico XIIa** se define por un primer pico desplazado y por una declinación plana, sin pendiente, cuyo efecto de énfasis por contraste es muy llamativo.

Los **patrones melódicos XIIb** y **XIIc**, por su parte, se caracterizan, junto al primer pico desplazado, por una declinación en zigzag plana (XIIb) o descendente (XIIc), y constituyen el modelo más característico de las entonaciones enfáticas, como si se tratara de una sucesión de énfasis marcando cada palabra.

1.2.5.7 Tonemas y sus variantes tipo

En resumen, Cantero (2005) encuentra un total de doce patrones melódicos claramente identificables, con diversas variantes:

- un patrón melódico de entonación neutra (I)
- tres patrones melódicos de entonación interrogativa (II, III, IVa-b)
- dos patrones melódicos de entonación suspendida (V, VIa-b)
- seis patrones melódicos de entonación enfática (VII, VIII, IX, Xa-b, XI, XIIa-b-c)

De los ocho tonemas que se distinguían en Cantero (2002), cuatro de ellos se describen por medio de estos doce patrones melódicos (es decir, variantes tipo):

8. /-I -E -S/	→	I
4. /+I -E -S/	→	II , III , IVa, IVb
7. /-I -E +S/	→	V , VIa, VIb
6. /-I +E -S/	→	VII , VIII , IX, Xa, Xb , XI, XIIa, XIIb, XIIc

A partir de aquí, pueden generarse nuevas variantes melódicas añadiendo énfasis (para pasar de /-E/ a /+E/) o eliminando su inflexión final (para pasar de /-S/ a /+S/), hasta generar variantes de todos los tonemas del español.

1.2.6 Modelo de Aix-en-Provence

El modelo de Aix-en-Provence se ha venido desarrollando durante los últimos años en el laboratorio *Parole et Langage* de la Universidad de Aix-en-Provence por Daniel J. Hirst, Albert Di Cristo y Robert Espesser y con él se pretende obtener una modelización fonológica del sistema prosódico de distintas lenguas con el fin de dar cuenta de las características prosódicas de cada una de ellas y, a su vez, llevar a cabo estudios de comparatismo prosódico para extraer aquellos aspectos que son invariantes o universales y aquellos que son específicos de un sistema lingüístico determinado. (Hirst y Di Cristo 1998) (Baqué y Estruch 2003: 123)

Siguiendo la línea del generativismo y de las teorías fonológicas de la entonación, el modelo postula la existencia de cuatro niveles:

- un **nivel fonológico ‘profundo’** o ‘subyacente’ que incluye las representaciones funcionales que codifican la información necesaria para la interpretación semántica y sintáctica de la prosodia de un enunciado,
- un **nivel fonológico ‘superficial’** en el cual se hallan las categorías discretas que permiten describir los fenómenos interlingüísticos de superficie,

- un **nivel fonético** que atañe a los fenómenos variables a partir de los cuales se han puesto de manifiesto los condicionantes universales de producción y percepción prosódicas, y
- un **nivel físico** que da cuenta de las características acústicas o fisiológicas de un enunciado.

Es necesario que cada nivel intermedio se pueda interpretar en los dos niveles adyacentes, superior e inferior. Así, las funciones prosódicas se estructuran en formas prosódicas subyacentes según parámetros prosódicos abstractos específicos de cada lengua. Estas formas se estructuran a su vez en las formas superficiales, y luego en las representaciones fonéticas, para llegar al final a los correlatos físicos de los enunciados. (Baqué y Estruch 2003: 124-25)

1.2.6.1 Componentes básicos del modelo

El modelo de Aix-en-Provence plantea la recuperación de una representación simbólica a partir de la curva de F_0 y su representación parabólica a base de puntos de inflexión (*target points*), por medio del sistema MOMEL (*MOdelling MELody*), que después serán etiquetados mediante el sistema INTSINT (*International Transcription System for INTonation*). Posteriormente, se alinea la señal con la curva de F_0 , los símbolos de la codificación, las etiquetas de límite de palabra, las etiquetas de inicio de sílaba tónica y la transcripción ortográfica. Este proceso engloba los cuatro distintos niveles de análisis: el nivel físico (acústico y fisiológico), el nivel fonético, el nivel fonológico superficial y el nivel fonológico profundo.

La diferencia entre el nivel físico y el fonético se sustenta en la hipótesis de que todas las lenguas obedecen las mismas restricciones físicas en la producción y la percepción de las variaciones de la frecuencia fundamental, la intensidad y la duración. Por lo tanto, el nivel físico es aquel que podríamos denominar como universal, mientras que en el nivel fonético, estas restricciones universales ya no se tendrían en cuenta.

El objetivo del nivel fonético, teniendo en cuenta que las curvas de F_0 se modelan como la superposición de los componentes micro y maroprosódicos, es tratar de extraer automáticamente sólo la información macroprosódica relevante de la señal de habla. Esta extracción se concreta en dos fases: la estilización o modelización y la codificación simbólica. La primera consiste en la sustitución de la curva de F_0 por una función numérica sencilla que conserva la información macroprosódica original y deja de lado la información no relevante, el resultado es una curva simplificada perceptivamente equivalente a la original.

La codificación se realiza mediante un alfabeto de símbolos que convierte la curva estilizada en una secuencia de categorías discretas que, si se utilizan para regenerar la curva, aunque perceptivamente no sea idéntica a la original, lingüísticamente sí es equivalente. (Baqué y Estruch 2003: 130)

Los puntos de inflexión obtenidos con el algoritmo MOMEL son codificados mediante el sistema de anotación INTSINT. Cada uno de los puntos de la curva estilizada es codificado como un **tono absoluto** (definido globalmente en referencia al rango frecuencial del locutor) o como un **tono relativo** (definido localmente en relación a los puntos de inflexión anterior y posterior) (véase la Ilustración 1-21). Los tonos relativos se dividen en **iterativos** (definidos como puntos de inflexión intermedios en secuencias ascendentes o descendentes) y **no iterativos** (definidos como puntos de inflexión que no corresponden a tonos absolutos y que no constituyen etapas intermedias en secuencias ascendentes o descendentes). (Baqué y Estruch 2003: 132-133)

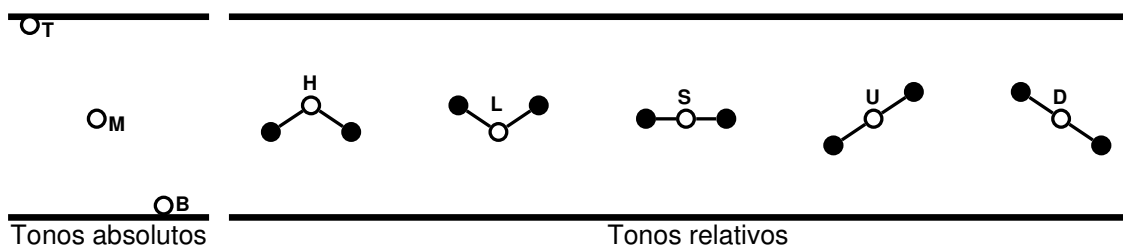
Los tonos absolutos son estos:

- T (*Top*): altura tonal máxima del locutor
- B (*Bottom*): altura tonal mínima del locutor
- M (*Mid*): etiqueta inicial correspondiente a un valor medio del locutor

Los tonos relativos son los siguientes:

- no iterativos:
 - H (*Higher*): punto más alto que los dos puntos adyacentes
 - L (*Lower*): punto más bajo que los dos puntos adyacentes
 - S (*Same*): punto igual que su precedente
- iterativos:
 - U (*Up - Upstepped*): punto en una secuencia ascendente
 - D (*Down - Downstepped*): punto en una secuencia descendente

Ilustración 1-21: Tonos absolutos y relativos según el sistema de anotación INTSINT



Fuente: Diseño propio basado en Baqué y Estruch (2003: 133)

Por último, en el nivel fonológico profundo, una serie de selecciones lingüísticamente significativas interactúan con un conjunto de parámetros prosódicos abstractos específicos de cada lengua para generar los patrones entonativos concretos. Los pasos que se siguen para determinar el patrón entonativo son, primero, la división del enunciado en unidades entonativas y éstas a su vez en unidades tonales, después, la asignación de un patrón

entonativo a cada unidad y, por último, la aplicación opcional de la regla de ajuste del escalonamiento descendente en caso de que exista un descenso progresivo de la altura de los picos tonales, así como otras reglas de ajuste. (Baqué y Estruch 2003: 134-135)

1.2.6.2 Ventajas y desventajas

Teniendo en cuenta que en este modelo se contempla la entonación como un fenómeno que se manifiesta en las variaciones de frecuencia fundamental, intensidad y duración, aún falta incorporar una representación de las características de los dos últimos parámetros. La inclusión de la unidad rítmica en el modelo y primeros estudios sobre la posible codificación de la duración y la intensidad permiten anunciar próximos desarrollos en estas áreas.

En cuanto al nivel fonético, aún debe resolverse problemas tales como la declinación, los reajustes frecuenciales, las pausas y la normalización de los datos acústicos pertenecientes a distintos locutores.

En el nivel de representación fonológica superficial aparece también un problema de normalización de los tonos absolutos (T, M y B) en producciones más extensas que un enunciado o un párrafo, pues la atribución de dichos tonos depende en el modelo actual directamente del segmento fónico analizado.

Por último, el nivel fonológico profundo aún no se ha estudiado en conjunto con los componentes semántico y sintáctico, pero sí está previsto.

Así, el modelo de Aix-en-Provence, que pretende llegar a ser un modelo completo de la prosodia, constituye actualmente una interesante teoría fonológica de la entonación, a pesar de las dificultades que aún presenta, y facilita el comparatismo lingüístico. Además, su innovador enfoque permite no sólo generar patrones acústicos de la melodía a partir de una entrada simbólica, sino también llevar a cabo el proceso inverso, es decir, obtener una codificación de la curva de F_0 a partir de la señal acústica, facilitando así el estudio de las relaciones que existen entre los niveles acústicos y fonológicos de la entonación.

Otra de las ventajas del modelo es el uso de algoritmos altamente automatizados que reduce el grado en el que el análisis depende del investigador, lo que permite llegar a resultados más fácilmente comparables y generalizables.

Si bien este modelo comparte ciertas bases teóricas con otros modelos y aún no puede considerarse del todo completo, constituye una teoría no sólo coherente en su estado actual de desarrollo, sino que también es muy original por algunos de sus enfoques. (Baqué y Estruch 2003: 148-150)

1.2.6.3 Aplicaciones del modelo de Aix-en-Provence al español

El modelo va más allá de la elaboración de una teoría que contemple aspectos universales que permitan llevar a cabo un comparatismo interlingüístico y especificar la estructura prosódica de cada lengua, pues también realiza descripciones pormenorizadas de grandes *corpora* para poder validar las distintas hipótesis que fundamentan el modelo y precisar los parámetros específicos de la lengua como objeto de estudio.

La versión actual del modelo es muy reciente, pues fue desarrollada a partir de 1995 y aún son muchos los trabajos parciales o en preparación que contribuirán a aplicarlo en distintas lenguas. En 1998, Hirst y Di Cristo editaron una publicación en la que presentan los principios básicos del modelo y su aplicación para el estudio sistematizado y estandarizado de las características prosódicas de veinte lenguas, entre las que también se encuentra el español, a cargo de Alcoba y Murillo (1998). El artículo presenta la misma estructura que los trabajos correspondientes a otras lenguas y hace una descripción general de las características prosódicas de la lengua; una descripción de los patrones entonativos –patrones no enfáticos, modo y expresividad, focalización y efectos contextuales, fraseo y organización textual, y patrones estereotipados o estilizados; así como una comparación con otros sistemas entonativos –otros dialectos y otros idiomas. Sin embargo, es necesario desarrollar investigaciones aún más sistemáticas y profundas para poder precisar con mayor certeza las hipótesis tanto del español como de otros sistemas lingüísticos, sobre todo, su aplicación al estudio del habla espontánea.

1.3 Ritmo y tempo

El ritmo, según Auer *et al.* (1999: 3), es uno de los aspectos más erráticos de la condición humana; está en el mundo a nuestro alrededor y en el mundo dentro de nosotros, en nuestros cuerpos y en nuestras mentes, en nuestra vida y nuestros pensamientos. También el lenguaje humano es profundamente rítmico. El ritmo de las interacciones verbales se sincroniza entre los participantes y al mismo tiempo se logra gracias a los participantes. Es decir, el grado y tipo de ritmicidad en el habla cotidiana puede variar: los ritmos interaccionales pueden ser compartidos o idiosincrásicos, pueden emerger o desintegrarse, volverse más o menos distintos. Lo más importante, el uso que hacen los interactantes de las estructuras rítmicas es un importante medio para que la interacción funcione.

A pesar de la relevancia del ritmo y el tempo para la interacción verbal, la lingüística se ha ocupado poco de estudiarlos. En las siguientes secciones se abordará primeramente los intentos en psicolingüística y en antropología lingüística por reintroducir las nociones de

ritmo y tempo en el estudio de la interacción verbal. Más adelante se verán los planteamientos más sobresalientes de los analistas del discurso y de los fonólogos.

1.3.1 El ritmo según la psicología y la antropología lingüística

Fuera de la lingüística, la psicología ha contribuido a la comprensión del posicionamiento en tiempo real del lenguaje en tres áreas. La primera tiene que ver con los fundamentos generales de las configuraciones psicológicas de las estructuras rítmicas y su percepción; la segunda es la investigación sobre la ‘sincronización’ que tiene una inclinación más interaccional; la tercera es la investigación psicolingüística sobre la relevancia del ritmo para la estructura y recuperación de la información. (Auer *et al.* 1999: 13)

El resultado más importante de la investigación sobre configuraciones mentales para el ritmo es el reconocimiento del papel activo que juega la percepción humana en la construcción de los ritmos. Para un conjunto dado de estímulos que se perciben como rítmicos, las duraciones concretas que conforman este conjunto son menos importantes que la estructura temporal que se les impone. Como consecuencia, el análisis del ritmo no puede comenzar con la detección de estructuras rítmicas reales en los datos acústicos; debe comenzar por la percepción y después relacionarla con las estructuras reales de duración. Pero esta relación no es nada simple. La llamada ‘ritmización subjetiva’ es la responsable del descubrimiento bien documentado de que incluso los estímulos físicamente idénticos (como el tic tac – y no tic tic – de un reloj) se escuchan como si estuvieran estructurados rítmicamente en grupos de dos o, más raramente, tres elementos. En tal estructura rítmica subjetiva, el intervalo entre dos grupos rítmicos parece más larga que el intervalo intragrupal, y el primer elemento (o a veces el último), el acento, más largo que los otros. También es importante para el estudio lingüístico del ritmo el descubrimiento de que la manipulación de estímulos a tiempos iguales con respecto a la longitud, el movimiento tonal o el volumen es la responsable de la percepción de los acentos. El resultado es una agrupación rítmica que conjunta los elementos no acentuados a la derecha (o a veces a la izquierda) del acento percibido.

La segunda tradición de estudio psicológico del ritmo, que a primera vista se relaciona mucho más obviamente con el análisis de los ritmos conversacionales, se puede resumir bajo el encabezado de sincronía interaccional. Contrasta con las investigaciones previas en que esta corriente se ha interesado principalmente en los procesos adaptativos por medio de los cuales dos o más individuos sincronizan sus conductas verbales y no verbales en términos de velocidad y duración.

La tercera tradición son las investigaciones psicolingüísticas sobre la relevancia del ritmo para el procesamiento de la información. La idea básica es que el ritmo es esencial para el procesamiento de la información porque hace predecible la ocurrencia de ciertos eventos en el tiempo. Mientras que el hecho de poner atención a estos eventos implica un estado de atención perceptiva elevado, el tiempo que transcurre entre los eventos rítmicos esperables puede utilizarse para el procesamiento de los estímulos percibidos. Un ejemplo de los resultados de estos estudios es que los mensajes verbales se comunican más fácilmente cuando se organizan rítmicamente, dado que la posibilidad de predicción de la siguiente prominencia fonética guía la atención de los oyentes y les permite alternar cíclicamente entre fases de percepción y procesamiento. Además, hay bastante evidencia de que las estructuras con ritmo se recuerdan mejor que las que carecen de él. (Auer *et al.* 1999: 13-16, Paeschke 2003: 45-48)

Pero no sólo en la psicología se ha estudiado el ritmo de manera científica, también los especialistas en antropología lingüística han trabajado con la estructura rítmica del habla por varias décadas, sobre todo en el campo de la etnografía de la comunicación y más recientemente en el de la etnopoética. En esta última, sobre todo, se encontró que la estructuración rítmica en la ejecución de ciertos géneros literarios está vinculada en más de una manera con la oralidad de una cultura. El ritmo en una ejecución es una consecuencia directa o indirecta del hecho de que la estructuración rítmica facilita la memorización, aludido previamente. Por lo tanto, el ritmo contribuye de manera crucial en la transmisión del saber social a través de los géneros orales. Por otra parte, los ‘microetnógrafos’ han ido llevando a cabo un análisis técnico del significado interaccional del ritmo para el habla cotidiana, trabajando en detalle a pequeña escala las interacciones en las culturas occidentales basándose en grabaciones en cinta o película. (Auer *et al.* 1999: 16-18)

1.3.2 El ritmo y el tempo como señales de contextualización

El ritmo no es una propiedad de las señales acústicas. Los datos acústicos pueden contener rasgos fonéticos tales como picos y valles de frecuencia mínima, y cambios en amplitud o duración que pueden ocurrir a intervalos de tiempo objetivamente iguales o similares. Sin embargo, la percepción del ritmo no se relaciona ni se deriva directa o automáticamente de estos eventos físicos. Es la mente humana la que recibe ciertas señales físicas como si formaran un patrón rítmico. Más precisamente, el receptor humano de la señal acústica debe realizar numerosas tareas interpretativas para oír un ritmo en esa señal. Las variaciones acústicas de la frecuencia fundamental deben interpretarse como cambios en la entonación, las

variaciones de amplitud como ascensos o descensos del volumen y las variaciones de la duración de los segmentos como cambios de longitud. Estos rasgos tienen que combinarse en la percepción del acento, el correlato perceptivo de la prominencia fonética. Una serie de acentos puede percibirse como patrón rítmico. Por lo tanto, hablar de ritmo implica necesariamente una interpretación de los datos físicos, un proceso constructivo durante el cual los datos se vuelven parte de un esquema holístico capaz de incorporar aún más detalles de la señal que ingresa. El ritmo siempre implica algún tipo de repetición en el tiempo; en el sentido más simple, es la recurrencia de eventos que, para los fines prácticos de la construcción de ritmos, se considera que son del mismo tipo y que se perciben como que ocurren a una velocidad constante. (Auer *et al.* 1999: 23)

Para este enfoque es esencial la diferencia entre eventos rítmicos y acentos. A los primeros se les llama compases (*beats*); la duración entre dos compases es un intervalo rítmico (*rhythmic interval*). No todas las sílabas acentuadas necesitan percibirse como compás rítmico. Y al menos se necesita percibir tres sílabas acentuadas para la formación de un patrón isócrono, aunque otros acentos ocurran fuera de ritmo.

Estos dos niveles, el reconocimiento de los acentos y la identificación de formaciones rítmicas, deben ser cuidadosamente distinguidos, a pesar de que se influyan mutuamente. Nótese que no hay prioridad lógica o temporal entre ellos. Un ritmo isócrono establecido preconfigura los eventos futuros; su propiedad esencial es la de hacer predecible la sincronización de un evento posterior integrado rítmicamente. Por tanto, en cualquier punto del tiempo dentro de un ritmo establecido, el siguiente compás se anticipará. Esta expectativa facilita la percepción de cualquier evento que ocurra ‘al compás’ como prominencia fonética y, sobre todo, como un acento. (Auer *et al.* 1999: 25)

La extraordinaria atracción del enfoque de Gumperz (1982: 101-102) sobre la contextualización reside en el hecho de que permite entender el trabajo interaccional realizado por las señales prosódicas tales como la entonación, el tempo, el volumen y el ritmo (con todos sus rasgos que las componen), mostrando cómo adquieren un significado interpretativo. La prosodia no es un agregado accidental o derivativo de los datos propiamente lingüísticos, sino una parte esencial de la interacción. El significado de estas señales prosódicas es de naturaleza básicamente indéxica, es decir, que funcionan como marcadores de la identidad personal y la pertenencia a distintos grupos sociales (Crystal 1997: 173), y no puede ser capturado por medio de las técnicas tradicionales de descripción semántica desarrolladas para el significado representacional o simbólico. Las señales prosódicas únicamente ‘significan’ dentro del

contexto. Por lo tanto, analizar la labor de la prosodia en la interacción necesita una nueva teoría de la ‘pragmática fonológica’. (Auer *et al.* 1999: 28)

En el enfoque de la contextualización aplicado al ritmo, sería de poco valor preguntarse cuál es el significado interaccional de esta señal prosódica como tal, debido a que las maneras en que el ritmo contribuye a la construcción del significado en la interacción dependen esencialmente del contexto local.

1.3.3 Interdependencia del tempo y el ritmo

Según Auer *et al.* (1999: 28), el tempo es una noción fuertemente relacionada con la de ritmo. Este concepto no es sencillo, pues nociones laicas tales como habla ‘lenta’ o ‘rápida’ no parecen utilizarse de manera consistente, pues pueden referirse a diferentes tipos de fenómenos. Por ejemplo, la percepción del tempo se ve influida por la segmentación de frases en unidades entonativas, y por la duración de las pausas dentro y entre estas unidades. También se ve influida por fenómenos segmentales como la naturaleza y grado de asimilación al nivel fonológico oracional o post-léxico. Además de estos factores, la noción laica de tempo incluye al menos los siguientes dos parámetros: el número de sílabas producidas durante una unidad de tiempo dada, y el número de acentos por unidad de tiempo. Estos mismos parámetros son la base de la distinción entre **velocidad** (*speech rate*), es decir, el número de sílabas por unidad de tiempo, y **densidad** (*density*), el promedio de sílabas acentuadas e inacentuadas.

Precisamente porque la densidad se define por medio de acentos y no de compases, es necesario introducir definiciones ligeramente diferentes que permitan capturar los usos interaccionales del promedio y la densidad pero que sean parte natural de los análisis rítmicos. Así, el **tempo rítmico** (*rhythmical tempo*) es la duración de un intervalo rítmico (la distancia entre dos compases rítmicos). Debido a que puede haber sílabas acentuadas fuera de los patrones rítmicos o compases silenciosos dentro de él, el tempo rítmico y el promedio no son idénticos. La **densidad rítmica** (*rhythmical density*) se define entonces en términos del número de sílabas por intervalo rítmico, o, lo que es lo mismo, en términos del promedio de compases y sílabas fuera de compás. (Auer *et al.* 1999: 31)

Nótese que el ritmo y el tempo en esta nueva definición dependen uno del otro. Por una parte, el tempo rítmico no puede percibirse, medirse o interpretarse sin haber establecido previamente el ritmo del pasaje conversacional. Por otra parte, el tempo influye en la claridad perceptiva de una formación rítmica, produciendo un patrón rítmico distinto y claro.

1.3.4 Ritmos abstractos

La armonía rítmica, según Di Cristo (2003: 31), no constituye la finalidad última del habla, pues su organización acentual en realidad debe considerar una gama de restricciones que dependen del tamaño y la función de las unidades léxicas, de las opciones sintácticas, de las variaciones del tempo, del estilo, del señalamiento de las diversas formas de focalización, etc. Del mismo modo, según Almeida (1999: 80), la alternancia acentual puede considerarse un patrón universal de las lenguas, ya que existen no pocas evidencias de su actuación, pero su efecto puede aparecer bloqueado por tendencias particulares de cada lengua: factores sintácticos, pragmáticos, del estilo de habla, etc.

Por otro lado, Cantero (2002: 55) considera que en el habla espontánea el ritmo no parece cumplir ninguna función significativa, y no pasa de ser un fenómeno coarticulatorio. Es decir, el ritmo es un fenómeno que tiene como función exclusiva la integración de los sonidos en bloques (o grupos rítmicos) para posibilitar la comprensión auditiva del discurso. A esta función se le añade, tal vez, la de caracterizar el registro poético de la lengua.

El grupo rítmico se caracteriza como unidad gracias a su contorno melódico, en algunos casos por el alargamiento de las palabras o sílabas finales y/o por pausas; se determina semánticamente y contiene al menos una vocal acentuada. Si se desea describir el ritmo de una lengua, deberá investigarse: 1) el tamaño de los grupos rítmicos, 2) la frecuencia y duración de las pausas, 3) la frecuencia y distribución temporal y espacial de los acentos (es decir, la posición de los acentos en los grupos rítmicos), y 4) los procesos de acentuación y desacentuación. (Benkowitz 2003: 25, 43)

En la teoría métrica, las alternancias se representan por medio de pies binarios cuyos constituyentes inmediatos son uno fuerte y el otro débil (véase 1.1.3.2 y 1.1.3.8). El elemento fuerte se considera el núcleo de la estructura. Las dos principales propuestas de formalización para el ritmo, formado por una secuencia de pies, son la **parrilla métrica** (*metrical grid*) y el **árbol métrico** (*metrical tree*). (Núñez Cedeño y Morales-Front 1999, Kager 1995, Liberman y Prince 1999 [1977], Halle e Idsardi 1995)

Núñez Cedeño y Morales-Front (1999: 205-208) hacen la aplicación de ambas propuestas al español. En la parrilla métrica, cada sílaba (la unidad potencialmente portadora de acento) tiene una serie de asteriscos que marcan su nivel de prominencia. A continuación se muestra la representación en parrilla métrica de la palabra *antioxidante*:

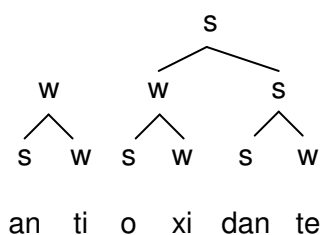
Ilustración 1-22: Ejemplo de parrilla métrica

Fuente: Diseño propio, basado en Núñez Cedeño y Morales-Front (1999: 206)

Los asteriscos en cada nivel señalan la proyección de cada sílaba en un nivel de prominencia fonológica. En este ejemplo, la penúltima sílaba es la más prominente porque se proyecta en tres niveles. Esta formulación de la prominencia acentual tiene la ventaja adicional de poder captar de una manera visual el fenómeno de los acentos que se repelen. Esta aversión entre acentos tiende a ocurrir cuando dos sílabas adyacentes tienen un mismo nivel de prominencia en la línea 1. A este nivel, dos cimas adyacentes son arrítmicas porque rompen la sucesión armónica de cimas y valles de prominencia.

Por lo tanto, la parrilla métrica no sólo permite captar niveles de prominencia, sino que también sirve para expresar la melodía rítmica (entendida como subidas y bajadas de prominencia).

Por otra parte, la representación arbórea es una propuesta que intenta incluir en la representación la existencia de constituyentes métricos, dispuestos en estructuras jerárquicas. La palabra *antioxidante* se representa en la Ilustración 1-23 (cfr. Ilustraciones 1-4 y 1-22).

Ilustración 1-23: Ejemplo de árbol métrico

Fuente: Diseño propio, basado en Núñez Cedeño y Morales-Front (1999: 206)

Lo que importa destacar de esta representación es la presencia de constituyentes binarios que pueden ser fuertes (*s – strong*) o débiles (*w – weak*). El grado de prominencia también puede calcularse a partir de la comparación del número de nodos fuertes que un determinado nodo domina. Cuantos más nodos fuertes se domine, mayor será el grado de prominencia. Este tipo de formalismo es muy conveniente para el análisis de procesos que usan de forma directa los constituyentes métricos como la armonía vocálica en lenguas como el húngaro y el turco. Una desventaja es que resulta más difícil expresar la incompatibilidad de prominencias adyacentes

y su movimiento. Pero dado que en la parrilla esto es una ventaja, se puede combinar las dos formalizaciones (véase Ilustración 1-24).

Ilustración 1-24: Combinación de la parrilla métrica y el árbol métrico

.	.	.	.	*	.	línea 2
(*)	.	*	.	*)	.	línea 1
(*)	*)	(*)	*)	(*)	*)	línea 0
an	ti	o	xi	dan	te	

Fuente: Diseño propio, basado en Núñez Cedeño y Morales-Front (1999: 207)

En cada línea de la parrilla, los paréntesis representan constituyentes y su núcleo se representa como un asterisco en el nivel de prominencia siguiente. En la línea 0 se marcan todos los elementos que podrían llevar acento. En la línea 1 se marcan los núcleos de cada constituyente en la línea 0. En la línea 2 se marcan los núcleos de los constituyentes en la línea 1.

Universalmente, el tipo de pie más común es el troqueo silábico (un pie silábico con el núcleo a la izquierda: (* ·)). Si se analiza una secuencia de sílabas en troqueos, se derivará una alternancia rítmica (Núñez Cedeño y Morales-Front 1999: 207).

Como se había mencionado al principio de este apartado, para Cantero (2002: 56, 75), hablar de isocronía acentual o silábica no es suficiente en español, pues la unidad rítmica no es ni el pie acentual (entendido como el trecho que va de acento a acento) ni la sílaba. La unidad rítmica en la que la isocronía actúa en español es la **palabra fónica** (que incluye las sílabas átonas previas y posteriores a un acento léxico o paradigmático) o **grupo rítmico**, cuyo esquema rítmico es denominado por Cantero como isocronía de palabra fónica o isocronía léxica. La cadena fónica, por lo tanto, se organiza en bloques fónicos alrededor de los acentos, que son los verdaderos núcleos del discurso: los sonidos consonánticos se organizan alrededor de los sonidos vocálicos para formar sílabas y éstas a su vez se agrupan alrededor de las vocales tónicas para constituir las palabras fónicas o grupos rítmicos; las sucesivas palabras fónicas se organizan también y forman bloques alrededor de un acento jerárquicamente superior que actúa a nivel sintagmático, por lo que Cantero los llama **acentos sintagmáticos**.

Como se verá más adelante, en el marco de esta investigación se utilizará un sistema de notación del ritmo parecido al de la parrilla métrica, con dos grandes excepciones: primero, la sílaba prominente se señalará con *s* (o con *s*, si además de ser prominente, está focalizada), las sílabas débiles se marcarán con *w* (o *w*, si se encuentran dentro de un grupo rítmico focalizado), y las sílabas aisladas se señalarán con *o*; segundo, partiendo de la idea de grupo rítmico según Cantero (2002), los pies o grupos de sílabas no se limitarán a ser binarios, sino

que pueden contener una sílaba prominente y hasta cinco sílabas subordinadas. En la Tabla 1-7 se esquematizan todas las posibles combinaciones de sílabas fuertes y débiles (no focalizadas) dentro de los grupos rítmicos en español, aunque no todas existen en la realidad.

Tabla 1-7: Estructuras silábicas posibles dentro de un grupo fónico

1 sílaba	2 sílabas	3 sílabas	4 sílabas	5 sílabas	6 sílabas
(0)	(s w)	(s w w)	(s w w w)	(s w w w w)	(s w w w w w)
	(w s)	(w s w)	(w s w w)	(w s w w w)	(w s w w w w)
		(w w s)	(w w s w)	(w w s w w)	(w w s w w w)
			(w w w s)	(w w w s w)	(w w w s w w)
				(w w w w s)	(w w w w s w)
					(w w w w w s)

Fuente: Diseño propio

La aplicación y análisis de este sistema de transcripción del ritmo se verá de manera más detallada en las secciones 2 y 3 del presente volumen.

1.4 Intensidad

El paralenguaje es la variación de la altura tonal, el tempo, el volumen y la calidad de la voz que un hablante hace por razones pragmáticas, emocionales y estilísticas. Este tipo de lenguaje se manifiesta en cada acto de habla y está disponible para los hablantes sin importar su lengua materna, aunque los contextos apropiados para el uso de ciertos tipos de efectos paralingüísticos pueden estar influidos culturalmente y bastante ritualizados. (Wennerstrom 2001: 60)

Cada acto comunicativo exitoso depende de que el hablante transmita su mensaje con una intensidad sonora adecuada para el receptor y la situación. Durante el proceso comunicativo, los participantes normalmente se adaptan recíprocamente, por ejemplo, en un ambiente muy ruidoso. Variando la posición de la cabeza y el cuerpo o de manera metacomunicativa el receptor se esfuerza por recibir el mensaje sonoro en una intensidad óptima. Condicionada por el proceso comunicativo y fuentes externas de interferencia se da una continua variación de la intensidad, por lo cual se puede concluir que no son los valores dinámicos absolutos sino la medida relativa de la variación de la intensidad la que es decisiva para el acto comunicativo. La percepción de sonido está relacionada con las fluctuaciones de presión del aire cuya amplitud se correlaciona con la percepción del volumen. Sin embargo, la *intensidad* y la percepción del volumen no tienen una relación completamente lineal, a pesar de que esta relación sea más clara que la de la frecuencia fundamental y la percepción de la altura tonal.

Desde el punto de vista de la física, la intensidad de una señal acústica representa la fuerza que se ejerce sobre una superficie, pero a nivel psíquico, la escala de intensidad de sonido es

un fenómeno complejo que está estrechamente relacionado con la altura tonal y las propiedades espectrales de los sonidos del habla. (Kranich 2003: 25-26, Neuber 2002: 76)

En comparación con la entonación, las investigaciones de las propiedades de la intensidad de la lengua hablada se han realizado mucho más esporádicamente y casi siempre sólo como descripciones globales. Esto se debe en gran medida a que los análisis fonéticos de la intensidad implican varias dificultades técnicas a la hora de la obtención de la información. En los últimos años, gracias a la posibilidad de utilizar programas de análisis computacional más apropiados, ha aumentado el interés por esta problemática. Por ejemplo, se han realizado investigaciones desde distintos enfoques teóricos para intentar identificar los contornos globales de intensidad y medir la intensidad de vocales y consonantes. Pero para poder interpretar experimentalmente los valores de la intensidad es absolutamente necesario que, por ejemplo, el hablante mantenga siempre la misma distancia con respecto al micrófono, y esto, incluso con sistemas de grabación que corrigen automáticamente este problema, arroja datos poco precisos (cfr. Kranich 2003: 27, 26), a menos que se siga el mismo procedimiento de estandarización de la frecuencia fundamental con los valores de la intensidad. Si, por ejemplo, la intensidad de la primer sílaba de un enunciado se fijara en 60 decibeles y a cada valor sucesivo se le sumara o restara el porcentaje de variación de la intensidad de la siguiente sílaba (multiplicada por 0.60), se obtendría valores estandarizados que podrían compararse indistintamente entre distintos enunciados y hablantes.

Para los fines de este estudio se creó la siguiente fórmula que será aplicada en la sección 3:

$$\overline{Int}_{std} = \frac{\sum Int_i}{n} * 60 / X_1 \quad (2)$$

Por medio de esta fórmula, similar a la presentada en 1.1.1.2, se asigna al valor de la primera sílaba en un enunciado (X_1) el nuevo valor de 60 dB y se ajusta los valores de las siguientes sílabas según las diferencias relativas y con respecto al valor de base (60 dB) para obtener el promedio de estos valores estandarizados.

Sobre todo, en esta investigación, cuyo objetivo principal era recoger muestras del español oral, se intentó realizar las grabaciones en lugares donde los participantes se sintieran cómodos y, en algunos casos, realizando otras actividades mientras hablaban, además de que se utilizaron distintos tipos de micrófonos con distintas calidades de sensibilidad. Por estas razones, el uso de valores estandarizados toma una mayor importancia. Sin embargo, tomando en cuenta que esta es una propuesta nueva, la intensidad se analizará de manera global, sabiendo de antemano que los resultados serán preliminares hasta que se lleve a cabo la misma metodología con un *corpus* grabaciones mayor y de una calidad mucho más alta.

1.5 Calidad de la voz

La calidad de la voz, correlato del ‘tono de voz’ o *timbre*, ha sido dejado varias veces fuera de la clasificación de los rasgos prosódicos. No obstante, el timbre constituye un complejo de rasgos que es modificado por la frecuencia, la amplitud y las relaciones entre los tonos parciales y el tono básico. (Kranich 2003: 33, Neuber 2002: 74)

El sonido característico de la voz de un hablante se debe a dos tipos de factores distintos: **orgánicos** y **fonéticos**. Los factores orgánicos son estructurales, mientras que los factores fonéticos tienen que ver con los patrones neuromusculares de control voluntario. Los factores orgánicos se derivan de aspectos arquitectónicos y mecánicos del aparato fonador, determinados biológicamente, mientras que los factores fonéticos son el resultado de hábitos musculares aprendidos social e idiosincrásicamente que un hablante adquiere a lo largo de toda su vida (Laver 2000: 37). Estos ajustes musculares dan como resultado un tipo de configuración general del aparato fonador que inevitablemente afectan la calidad del sonido que éste produce.

Muy pocos adultos hablan con un tipo completamente neutral de fonación. Una fonación neutral se caracteriza por la vibración de las cuerdas vocales únicamente, con una periodicidad regular, sin asperezas o fricciones y con un grado moderado de tensión en todos los sistemas musculares del aparato fonador. La variación de estas condiciones haría que la configuración fonatoria se describiera como áspera, (*harsh* – disperiodicidad, con irregularidades de frecuencia, *jitter*, o de intensidad, *shimmer*), o rechinante, (*creaky* – frecuencia baja, pulsos irregulares), o cuchicheada, (*whispery* – con fricción audible causada por el escape del aire por una glotis ligeramente abierta), o soplada, (*breathy* – con poca tensión muscular de todos los sistemas musculares del aparato fonador que origina un gasto mayor de aire que en el caso de la voz cuchicheada), o falsetto (rango tonal alto, con una mayor tensión muscular longitudinal de las cuerdas vocales), o varios tipos de combinaciones de estos modos de fonación dentro de las posibilidades de compatibilidad fisiológica o acústica. El cuchicheo y la voz rechinada son las variaciones más comunes de la fonación neutral o modal. (Laver 2000: 41, Ladefoged 2003: 169-181)

Por lo general se emplean metáforas procedentes de otros ámbitos de la realidad para describir el timbre cualitativamente. Muchas veces se han establecido grandes inventarios de metáforas pero parcialmente con conceptos o con representaciones demasiado objetivas que son difíciles de asociar con descripciones fonéticas. Con mayor frecuencia se han sugerido descripciones visuales (claro, oscuro), grados de dureza (suave, duro), metáforas de volumen y masa (lleno, vacío, delgado, robusto, plano, hueco, compacto) o grados de agudeza (agudo, metálico, estridente). A pesar de los intentos por crear un inventario unificado de sonidos, definir el timbre por medio de paráfrasis como estas o similares es difícil, pues cada persona ha desarrollado distintas ideas, suposiciones y sensaciones con respecto a dichas categorías de

análisis. De ahí la importancia de elegir mejor un reducido número de metáforas que sea suficiente para el análisis y prescindir de las más exóticas. (Kranich 2003: 35)

1.5.1 Prosodia de los estados emocionales

La voz humana es el sonido por medio del cual se transmite la lengua hablada y no sólo hace que el habla sea audible, sino que puede expresar una amplia gama de información sobre la identidad, la salud física, el estado emocional y la personalidad del hablante. Por eso, cada voz humana es única debido a los distintos factores anatómicos, fisiológicos, psicológicos, culturales, sociolingüísticos y conductual que contribuyen para el producto final.

A pesar de que las características anatómicas y las configuraciones musculoesqueléticas dan lugar a una permanencia vocal, los rasgos paralingüísticos son unos aspectos que están en constante cambio para reflejar emociones y ayudar a expresar mensajes no verbales. Un hablante emplea inconscientemente variaciones de la altura tonal, de la intensidad y de la calidad de la voz para realzar o contradecir las palabras que está produciendo. De este modo, un oyente puede intuir los verdaderos sentimientos del hablante, y el hablante puede dar indicios de las emociones cuando es difícil o socialmente inapropiado manifestarlas abiertamente. Los rasgos paralingüísticos también pueden ser manipulados conscientemente ya sea para que el hablante pueda expresar una emoción que en realidad no está sintiendo o bien para influenciar intencionalmente la conducta del oyente. La voz tiene también una importante función lingüística en la fonología segmental y suprasegmental. En su rol segmental, la voz es parte integral de los fonemas individuales en los casos en los que la capacidad de coordinación de la fonación y la articulación es esencial para la inteligibilidad. Los rasgos suprasegmentales del habla que componen la prosodia se expresan por medio de la altura tonal y la intensidad en conjunto con la duración segmental y el uso de pausas.

La emoción produce cambios en el aparato fonador que, a su vez, afectan varios parámetros vocales. El rango frecuencial es el parámetro que tiene el efecto más poderoso cuando los oyentes juzgan una emoción. Un rango frecuencial estrecho se percibe como tristeza o neutralidad, mientras que un rango frecuencial amplio expresa una alta estimulación, que puede ser positiva o negativa. (Mathieson 2000: 3-6)

Los estudios sistemáticos de las calidades acústicas de la voz en diferentes estados emocionales de un mismo hablante conllevan varios problemas. A las grabaciones de campo de enunciados emocionales por lo general les falta la alta calidad que es esencial para el análisis de los parámetros acústicos de la calidad de la voz. Por otra parte, bajo condiciones de laboratorio es muy difícil inducir exitosamente las diferentes emociones. Una posible solución a este problema es grabar en laboratorio el habla emocional producida por actores, combinado con un experimento perceptivo con oyentes ajenos a la investigación para que evalúen la

naturalidad y la capacidad de reconocimiento de las emociones. (Klasmeyer y Sendlmeier 2000: 339)

En un experimento de este tipo, realizado por Klasmeyer y Sendlmeier (2000), la calidad de la voz en los enunciados neutrales pudo clasificarse como voz modal normal en todos los casos. En los enunciados en los que se agregó la emoción de felicidad, la calidad de la voz pudo clasificarse como voz modal con volumen alto y cambios extremadamente rápidos en el contorno de la frecuencia fundamental. En los enunciados con la emoción de enojo, los hablantes utilizaron la fonación a base de gritos con una prosodia fuertemente marcada. Los enunciados con tristeza fueron producidos con fonación soplada (*breathy*). El tipo de fonación cambió a menudo en este tipo de enunciados, pues también se dieron casos de fonación rechinante (*creaky*) al inicio de los segmentos sonoros. En los enunciados producidos con susto la frecuencia fundamental en promedio es alta con poca variación para fines lingüísticos, pero, en parte, este tipo de enunciados se produce con voz soplada o cuchicheo, falsetto y aparecen irregularidades en la frecuencia fundamental. Los enunciados producidos con aburrimiento a menudo se parecen a los del habla triste, aunque no muestran irregularidades en la fonación o voz rechinante como en este último tipo de enunciados. Más bien, este tipo de fonación puede clasificarse como voz modal normal con poca dinámica en el contorno de la frecuencia fundamental.

Como se mencionó en apartados anteriores, las grabaciones sobre las que se sustenta esta investigación no cumplen con las características necesarias para realizar un estudio sistemático de la calidad de voz, pero se deja abierta la posibilidad para incluirse en versiones posteriores.

1.6 Resumen

En este primer apartado se hizo un breve repaso de la literatura sobre prosodia, sobre todo la enfocada en el estudio del español, y como este término incluye una gran variedad de subtemas (estructura y funciones prosódicas, entonación, ritmo, intensidad y calidad de la voz) y, a su vez, cada una de estas áreas de la prosodia ha sido estudiada desde distintos puntos de vista y enfoques teóricos, este repaso no deja de ser solamente un panorama general en el que se contextualiza y presenta cada uno de los elementos teóricos que se irá aplicando a lo largo de esta investigación.

En la siguiente sección se presentará la metodología de la investigación que se materializa en un modelo de transcripción, representación y análisis de grabaciones de habla espontánea recopiladas en Madrid y en la Ciudad de México.

2. Metodología

La transcripción de la lengua oral en el lenguaje XML es el tema principal de esta sección, pues aquí es donde se propone un sistema global de tratamiento de las producciones lingüísticas desde su grabación hasta la visualización de los datos en páginas *web* con distintos formatos según las necesidades de la investigación. Este capítulo se divide en tres grandes apartados: antecedentes, propuesta y metodología.

Primeramente se desarrolla el tema de la transcripción de la lengua oral en medios informáticos y sus ventajas para el lingüista. También se hace un recuento de los distintos tipos de transcripción por computadora para desembocar en los lenguajes de etiquetas, en especial, el lenguaje extendido de etiquetas o XML.

En los antecedentes, sobre todo, se hace mención de dos de los sistemas de transcripción en XML más representativos: EXMARaLDA, cuya información es ordenada según el eje temporal, y TEI, cuya organización es jerárquica. Estas dos propuestas sirven para ubicar más adelante mi propuesta personal, que, aunque inicialmente fue desarrollada evitando tener algún contacto con cualquiera de los dos sistemas y así hacer menos probable una semejanza sustancial, saca provecho de las ideas principales de temporalidad y jerarquización de los elementos que componen el texto de una conversación.

En el segundo apartado se hace una presentación paso a paso de mi propuesta personal de estructura jerárquica de la prosodia para documentos de XML, tomando como ejemplo una estructura simplificada a la que se le irán agregando cada vez más componentes hasta llegar a la presentación de la estructura completa. En este mismo capítulo se explica, por medio de la estructura simplificada, la sintaxis y composición de las definiciones de tipo de documento (DTD), de las hojas de estilo en el lenguaje extendido de transformaciones de hojas de estilo o XSLT, así como el producto final: la presentación de la información ordenada ‘gramaticalmente’ en un formato de página *web* interactiva.

Finalmente, a manera de manual de usuario, se analiza y describe cada uno de los pasos del proceso de creación de documentos detallados de análisis en XML, pasando por la transcripción y el procesamiento de las grabaciones y rejillas de texto en Praat (una herramienta programable de análisis acústico), la creación de hojas de estilo para cubrir distintas necesidades analíticas, la representación de los datos en lenguaje de etiquetas de hipertexto, HTML, y finalmente su adaptación a hojas de estilo de Excel para poder proceder al análisis cuantitativo de la información prosódica contenida en las grabaciones de los hablantes de Madrid y de la Ciudad de México.

2.1 Transcripción de la lengua oral

Para recolectar muestras de interacciones comunicativas espontáneas en situaciones naturales solamente es necesario encender una grabadora de audio o de video en el momento y lugar adecuados. Con un poco de suerte, se podrán ir acumulando horas y horas de interacciones naturales, pero la simple recolección de datos es sólo el inicio de una tarea mucho más compleja, pues las grabaciones deberán pasar por los procesos de transcripción y análisis. (MacWhinney 2000: 1)

En la actualidad hay una gran variedad de métodos que los investigadores de distintas disciplinas han desarrollado para representar la lengua oral utilizando el medio escrito (véase Dittmar 2002, Edwards y Lampert 1993, Sperberg-McQueen y Burnard 2004). Por una parte, esta variedad de sistemas es un reflejo de las dificultades que implica la elección de los fenómenos a estudiar mediante la codificación o transcripción del habla, puesto que el habla es dependiente de factores que afectan dimensiones que no tienen una contraparte directa en la escritura (por ejemplo, el tempo, el volumen o la altura tonal).

Por otra parte, las grabaciones del habla recogida en situaciones naturales de comunicación, independientemente de que la calidad de grabación no siempre sea la ideal, son transcritas de manera diferente según las necesidades específicas de investigaciones lingüísticas, acústicas, antropológicas, psicológicas, etnográficas o periodísticas, entre muchas otras. Incluso dentro del mismo campo de estudio, los intereses y perspectivas teóricas de los distintos transcritores puede llevarlos a preferir distintos niveles de detalle en la transcripción y los distintos estilos de representación visual. (Sperberg-McQueen y Burnard 2004: 10.1)

La producción y comprensión del habla están íntimamente ligadas a la situación en la cual ocurre el evento comunicativo, mucho más que en el caso de los textos escritos. Además, los problemas de ética al grabar y hacer público lo producido en un ambiente privado y dirigido a una audiencia limitada aparecen mucho más frecuentemente cuando se trabaja con textos orales que con escritos. Por lo tanto, una transcripción de la lengua hablada debe incluir un mínimo de información contextual y no siempre es sencillo determinar qué es relevante dentro de una interacción comunicativa.

El habla también conlleva varios problemas estructurales. A diferencia de un texto escrito, el habla debe analizarse sobre un eje temporal. Su inicio y final pueden ser difíciles de determinar y su composición interna poco sencilla de definir. La mayoría de los investigadores concuerda en que las contribuciones individuales o turnos de palabra de cada hablante forman un componente estructural de gran importancia en la mayoría de tipos de

habla, pero estas unidades raras veces se comportan estructuralmente de manera tan regular como los párrafos u otras unidades análogas en los textos escritos: los hablantes se interrumpen frecuentemente entre sí, utilizan gestos como si fueran palabras, dejan producciones lingüísticas sin terminar, etc. Por otro lado, el habla, aunque se pueda representar con palabras, a menudo contiene elementos tales como pausas vocalizadas que, a pesar de ser semiléxicas, tienen una inmensa importancia al analizar el texto oral. Incluso los elementos no vocálicos, como los gestos, pueden ser considerados como componentes del texto oral para ciertos propósitos analíticos. Por debajo del nivel del turno de palabra y el nivel del enunciado individual, el habla puede ser segmentada en unidades definidas según fenómenos fonológicos, prosódicos o sintácticos. Sin embargo, no hay un acuerdo general sobre estas unidades, ni siquiera en los nombres apropiados para designarlas.

Por otra parte, la definición de los límites de un texto oral es a menudo una cuestión de convenciones arbitrarias o de conveniencia. En contextos públicos o semipúblicos, un texto puede considerarse sinónimo de, por ejemplo, un discurso, una emisión en algún medio de comunicación, una reunión, etc., mientras que en contextos informales o privados, un texto puede ser simplemente una conversación que incluya a un grupo específico de participantes. Alternativamente, los investigadores pueden también optar por delimitar los textos orales en términos de su tiempo de duración o el número de palabras. (Sperberg-McQueen y Burnard 2004: 10.1)

2.1.1 Sistemas de transcripción

Una transcripción lejos de ser un reflejo objetivo y exhaustivo de los eventos de una interacción, es fundamentalmente selectiva e interpretativa. El transcriptor se encuentra constantemente tomando decisiones sobre cuál información incluir, qué categorías descriptivas utilizar, cuáles aspectos de la interacción se relacionan entre sí y cuáles de ellos son los más importantes. Pero, sobre todo, tiene que tomar decisiones sobre cómo expresar todo lo anterior dentro de los límites del medio grafémico-espacial de la transcripción. Si una transcripción se adapta bien a los objetivos del investigador y a la orientación teórica, puede ser de gran ayuda para encontrar regularidades de interés sin la distracción de detalles irrelevantes. Si no, puede impedir la detección de patrones interesantes y conducir a impresiones erróneas. (Edwards 1995: 19-20)

Muchas veces, las convenciones utilizadas en una investigación o publicación simplemente se presentan como información complementaria, sin una discusión detallada en la que se explique las razones de su selección, a pesar de que las convenciones de una transcripción

influyen directamente en la metodología para el tratamiento de los datos. Lo ideal es presentar dichas convenciones de manera explícita y enumerar sistemáticamente sus principios e implicaciones para la investigación.

Los objetivos deseables en cada sistema son, por una parte, que la transcripción conserve la información necesaria para el investigador, de manera que corresponda con la naturaleza de la interacción, y, por otra parte, que las convenciones sean prácticas con respecto a la manera de administrar y analizar los datos, por ejemplo, fáciles de leer, aplicables a otros conjuntos de datos y, si es necesario, útiles para otros fines adicionales.

La principal aportación de Edwards (1993: 5-10, 1995: 21-25) es la definición de un marco de principios a los que se sujetan los distintos sistemas de transcripción. Este marco permite hacer una comparación entre los distintos sistemas como paso previo a la selección de un determinado diseño que satisfaga las necesidades de la investigación.

Estos principios se clasifican en tres grandes grupos: el diseño de categorías, la legibilidad y el procesamiento computacional.

Para que una transcripción cumpla con el principio de diseño correcto de categorías, hay tres propiedades que son particularmente importantes. Primero, las categorías deben ser *sistemáticamente discriminables*, en el sentido de que, por cada caso, quede claro si una categoría se aplica o no. Segundo, las categorías necesitan ser *exhaustivas*, es decir, para cada caso particular en los datos debe haber una categoría que lo incluya. Tercero, muy importante, las categorías deben ser *sistemáticamente contrastantes*. Para esto, al elegir un conjunto de categorías descriptivas, el investigador tiene que explotar su contrastividad, de manera que revele las propiedades de interés para la investigación.

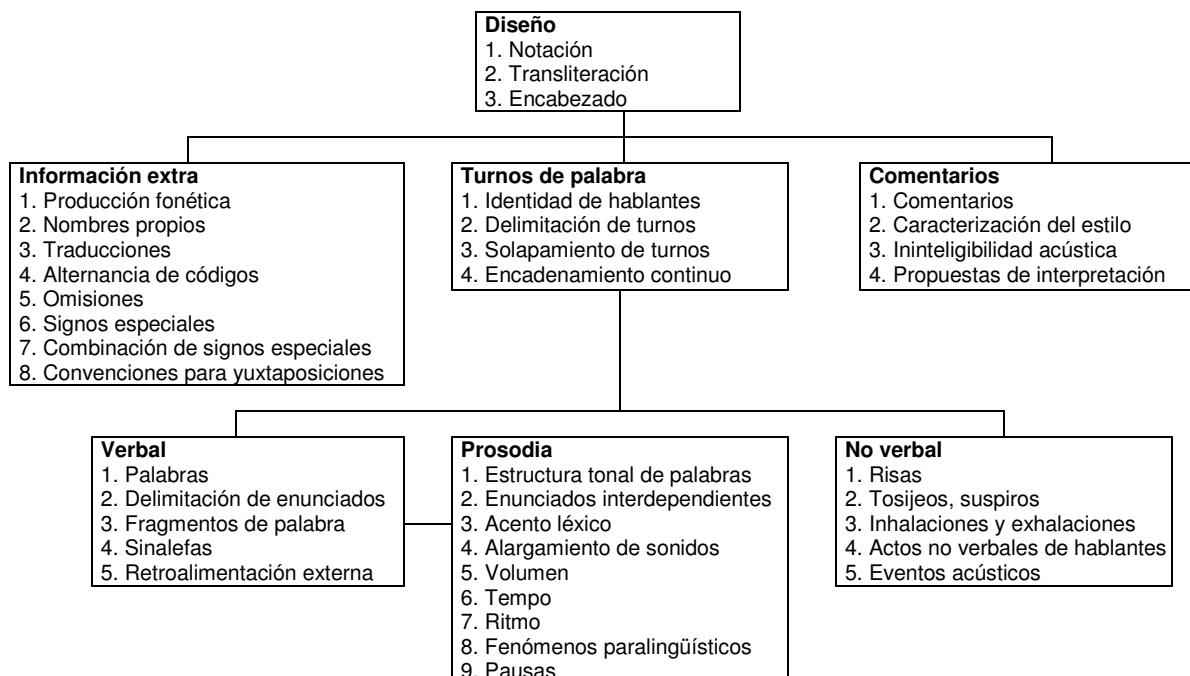
En cuanto a los principios de legibilidad, Edwards (1993: 6-9) hace referencia a dos técnicas utilizadas ampliamente en los medios gráficos para canalizar la atención del lector y dar forma a las precepciones sobre la relativa importancia de los diferentes tipos de información y su grado de interdependencia: la *prominencia visual* y el *acomodo espacial*. Del mismo modo, tras la comparación de los distintos sistemas de transcripción, se puede identificar seis técnicas utilizadas para resaltar la información y relaciones de interés: i) la *proximidad de eventos relacionados*, como el señalamiento del acento colocado justo a un lado o sobre la sílaba relevante; ii) la *separabilidad visual de eventos disímiles*, como las sílabas enfatizadas por medio de mayúsculas o negrita para señalar la prominencia; iii) la *iconicidad espacio-temporal*, por ejemplo si la secuencia temporal se representa de manera relativamente directa como una secuencia espacial de caracteres; iv) la *prioridad lógica*, por ejemplo cuando la contextualización precede al enunciado afectado; v) el *señalamiento mnemónico*, cuando las

categorías se codifican con abreviaturas interpretables o caracteres simbólicos, como el señalamiento de tonos ascendentes con ‘/’ y descendentes con ‘\’; y vi) la *eficiencia y brevedad*, marcando las distinciones con la menor cantidad posible de signos.

Por último, para que los sistemas de transcripción sean operables en computadoras, la codificación debe ser *sistemática y predecible*. Dado que las computadoras interpretan las instrucciones literalmente, se deben evitar dos errores de recuperación de información que pueden afectar los resultados: la *subselección*, es decir, pasar por alto aspectos relevantes en los datos, y la *sobreselección*, que es la recuperación de datos irrelevantes junto con los que sí son relevantes, como las marcas que contienen las mismas secuencias de caracteres. (Edwards 1993: 9-10)

Los principios abstractos que deben seguir los sistemas de transcripción también necesitan tener una serie correspondiente de categorías concretas con las que se puedan analizar y confrontar los distintos sistemas de notación de la lengua hablada. Para este fin, Dittmar (2002: 88-100) propone siete dimensiones para la comparación entre los distintos sistemas de transcripción que se resumen en la siguiente ilustración:

Ilustración 2-1: Dimensiones para la comparación entre sistemas de transcripción



Fuente: Diseño propio, basado en Dittmar 2002: 89

El *diseño* se refiere al formato general de la transcripción, según el cual se rige la elección de la manera en la que se representan las múltiples formas y funciones de los *turnos de palabra* registrados por medio de símbolos apropiados. Los turnos se subdividen en niveles *verbales*

(segmentales), *prosódicos* (suprasegmentales) y *no verbales* de los eventos comunicativos. A cada uno de estos tres niveles comunicativos corresponde un inventario de categorías, aunque las divisiones son muy relativas. Los *comentarios* son la información metacomunicativa que utiliza el transcriptor para describir con mayor detalle las propiedades de los turnos de palabra, y la *información extra* agrupa las aplicaciones específicas para intereses especiales. (Dittmar 2002: 88)

Más adelante se tratará de manera más detallada el diseño de sistemas de transcripción, cuando se apliquen estos criterios a la propuesta propia, cuyo propósito es el de hacer uso eficiente de la tecnología computacional. Como paso previo, veremos a continuación cómo ha influido el uso de las computadoras en la investigación lingüística, específicamente en la transcripción y estudio de la lengua hablada.

2.1.2 La computadora como medio de apoyo

En los últimos años, las computadoras han cambiado dramáticamente la vida profesional del lingüista, alterando el tipo de cosas que se hacen, la manera en que se hacen e incluso la manera en la que pensamos sobre ellas. El cambio hasta ahora ha sido gradual, creciente y mayoritariamente experimental, pero se encuentra en un continuo aceleramiento y no parece que vaya a terminar pronto. (Lawler y Dry 1998: 1)

El contacto con computadoras está dando forma a la manera en la que se conceptualiza la lingüística y la lengua, dado que la lengua es en realidad una forma de *software*. Así como el cerebro humano fue el modelo para el *hardware* de las computadoras, el lenguaje humano fue el modelo para el *software*. Por eso, las relaciones entre las computadoras y los lingüistas son mucho más profundas e interesantes de lo que podría sugerir el simple avance tecnológico.

No es de sorprender que los lingüistas hayan sido unos de los primeros científicos, aparte de los pertenecientes a áreas estrictamente técnicas, en aventurarse en los terrenos computacionales. Las tecnologías computacionales ofrecen a los lingüistas significantes beneficios, tanto a nivel individual como disciplinar, facilitándoles las tareas de investigación y docencia, de recolección rápida de información, y el análisis eficiente de *corpora* de datos. Al mismo tiempo, estas tecnologías le están dando una nueva forma a la disciplina, sacando a la luz nuevas áreas de investigación, nuevos tipos de datos y nuevas herramientas de análisis. (Lawler y Dry 1998: 1-2)

Desde mediados de los 80, el mundo de las computadoras ha pasado por una serie de notables revoluciones, cada una de ellas introduciendo nuevas posibilidades y retos. El poder de procesamiento de las computadoras de escritorio actuales hace parecer diminutas las

capacidades de las gigantescas computadoras centrales de hace unas décadas; las nuevas máquinas vienen provistas de tecnologías audiovisuales inimaginables hace tan sólo unos años; y los dispositivos tales como el CD-ROM y el DVD ofrecen una enorme capacidad de almacenamiento a precios razonables.

En el caso específico de la transcripción de la lengua oral, la introducción de las computadoras por una parte sirvió para cumplir el objetivo original de las transcripciones hechas a mano: fijar por escrito las fugaces impresiones auditivas y visuales atrapadas mediante la grabación de una conversación, sobre las cuales el investigador pudiera hacer análisis detallados. La tecnología actual ha abierto la posibilidad del acceso a audio y video digitalizado desde enlaces dentro de los textos de las transcripciones. De hecho, la transcripción es actualmente el punto de partida a una nueva realidad exploratoria por medio de la cual se puede acceder a la totalidad de multimedios interactivos, incluso a través de Internet. (Schmidt 2002: 3, MacWhinney 2000: 1-4)

Por otra parte, tras la introducción de las computadoras, se desarrollaron también nuevas posibilidades de análisis de datos. Las evaluaciones estadísticas y la búsqueda sistemática de datos en la transcripción se automatizaron, lo que a su vez permitió que se analizaran grandes cantidades de datos para, entre otras cosas, justificar cuantitativamente los resultados cualitativos. También ha habido numerosos proyectos, como el Child Language Data Exchange System, CHILDES (MacWhinney 2000), una base de datos compartida que se ha desarrollado continuamente desde 1984. Sin embargo, los investigadores tuvieron que irse deshaciendo de la idea de que la base de datos era únicamente un archivo o registro histórico, para darse cuenta de que la posibilidad de utilizar formatos de transcripción, códigos y programas de análisis compartidos podía llevarlos a avances teóricos y metodológicos.

En la siguiente sección se dedicará especial atención al *Extensible Markup Language* (XML), un lenguaje de etiquetado de información textual jerárquica que puede ser adaptado para que sea lo suficientemente flexible como para representar las complejas estructuras prosódicas de la lengua oral.

2.2 XML (*Extensible Markup Language*)

La razón por la cual en esta investigación se tiene la seguridad de emplear el XML es porque las ideas centrales de esta nueva tecnología en realidad son muy antiguas y han demostrado ser válidas a lo largo de varias décadas y miles de proyectos. La mejor manera de entender estas ideas es regresando a sus orígenes, el *Standard Generalized Markup Language* (SGML).

XML es en realidad un subconjunto racionalizado del SGML, por lo tanto, el avance alcanzado por el SGML, también lo es del XML. (Bray *et al.* 2000, Goldfarb y Prescod 2000: 6, Sperberg-McQueen y Burnard 2004: 2)

El XML proviene de una rica historia de los sistemas de procesamiento de texto. El procesamiento de texto es la subdisciplina de la ciencia computacional dedicada a crear sistemas computacionales que puedan automatizar partes del proceso de creación y publicación de documentos. El *software* para procesar texto incluye desde simples procesadores de palabras, hasta avanzadas bases de datos de noticieros, sistemas de presentación de documentos de hipertexto y otras herramientas de publicación.

La primera ola de procesamiento de texto automatizado fue la tipografía por computadora. Los autores escribían un documento y describían cómo les gustaría que fuera el formato. La computadora imprimía un documento con el texto y el formato descrito de manera adjunta.

Al tipo de archivo que contenía la mezcla de datos concretos del documento más la descripción del formato deseado se le llamó representación (*rendition*). El sistema convertía la representación en algo físicamente perceptible a un ser humano – una presentación. El medio de presentación ha sido históricamente el papel, pero eventualmente también la pantalla electrónica ha pasado a serlo.

Los sistemas tipográficos aceleraron el proceso de publicación de documentos y se desarrolló en lo que ahora conocemos como publicación de escritorio. Los nuevos programas como Microsoft Word y Adobe PageMaker aún trabajan con representaciones, pero le ofrecen a los autores una interface más agradable para manipularlas. La interface de representación para el usuario (el archivo incluyendo los códigos de formato) está diseñada para verse como la presentación (el producto final sobre papel). A esto es a lo que se le llama publicación ‘*What You See Is What You Get*’ (WYSIWYG). Dado que la representación describe únicamente una presentación, es importante que la interface de usuario refleje el producto final. Esta nueva forma de notación tipográfica se llama etiquetado de formato (*formatting markup*).

El gran problema en todo esto es que no hay manera de que la computadora ‘entienda’ el texto. La solución sería ‘enseñarle’ a la computadora tanto sobre el documento como fuera posible. Por supuesto, la computadora no entenderá el texto en el sentido real, pero puede pretender hacerlo, del mismo modo en que pretende entender simples datos o números decimales (dado que todo lo entiende en series de ceros y unos). Esta tarea se puede realizar reduciendo la complejidad del documento a unos pocos elementos estructurales elegidos a partir de un vocabulario común.

Una vez que se haya enseñado a las computadoras sobre los documentos, entonces podrán ser programadas para que hagan cosas que de otro modo no hubieran podido realizar. Utilizando su nueva ‘inteligencia’ pueden ayudar a los usuarios a navegar a través de grandes documentos, organizarlos y formatearlos automáticamente para su publicación en distintos medios, como el hipertexto, la imprenta o los discos ópticos. (Goldfarb y Prescod 2000: 9-10; Berners-Lee *et al.* 2001)

Una forma de etiquetado estructural, llamado lenguaje de etiquetado generalizado (*Generalized Markup Language*, GML), es la alternativa al etiquetado de formato o WYSIWYG (que a manera de broma se ha parodiado como ‘*what you see is all you get*”).

Aunque el XML, que es un tipo de GML, permitiría capturar las características de presentación del documento, no es necesario ni recomendado hacerlo, debido a la ambigüedad que implica la aplicación de formato. Esta información únicamente sobrecargaría la representación abstracta del documento.

Por supuesto, siempre hay la posibilidad de generar reproducciones impresas y digitales de alta calidad del documento. Los lectores no tienen por qué leer directamente el texto en XML. En lugar de insertar directamente los comandos de formato en el documento de XML, se le puede decir a la computadora cómo generar reproducciones formateadas *a partir* de la abstracción de XML.

Las reglas se especifican en un archivo llamado hoja de estilo (*stylesheet*) que es donde los diseñadores humanos pueden expresar su creatividad y conocimiento de las convenciones de formato. La hoja de estilo le permite a la computadora convertir automáticamente el documento de la abstracción a una reproducción con formato. (Goldfarb y Prescod 2000: 11-12)

2.2.1 Conceptos fundamentales

Hay tres características que distinguen al XML de otros lenguajes de etiquetado: a) su énfasis en el **etiquetado descriptivo** en lugar del etiquetado de procesos, lo que abre la posibilidad de que el mismo documento pueda ser procesado de muy distintas maneras por medio de hojas de estilo, utilizando únicamente las partes de él que se consideren relevantes o procesando la misma parte del archivo en distintos procesos; b) su concepto de **tipo de documento**, pues se considera que los documentos pueden ser de distintos tipos y cada tipo de documento debe definirse formalmente según sus partes constituyentes y su estructura; y c) su **independencia** de cualquier sistema de *hardware* o *software*, ya que todo documento de XML, sin importar el idioma o sistema de escritura que se emplee, utiliza el mismo método de codificación

subyacente de caracteres en cifras binarias, implementado por el estándar internacional Unicode, que provee un conjunto de caracteres que cubre la mayoría de sistemas de escritura actuales e históricos de todo el mundo. (Sperberg-McQueen y Burnard 2004: 2.1-2.1.3)

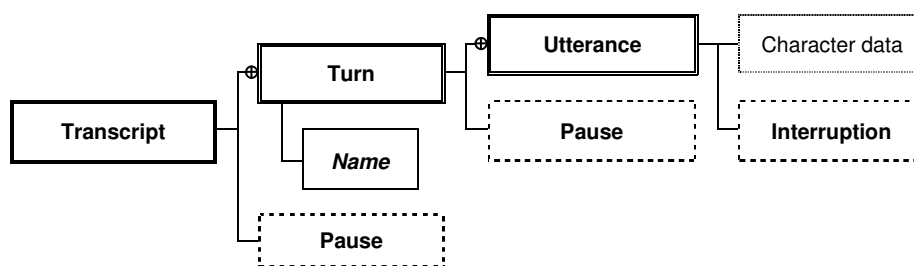
El XML se utiliza para la representación digital de documentos. Esto significa que el objetivo que se persigue al utilizarlo es producir documentos en un tipo de notación legible para computadoras, gracias al cual se podrá guardar, procesar, buscar, transmitir, desplegar e imprimir dichos documentos. Para que una computadora haga cosas útiles con un documento solamente es necesario informarle sobre la estructura del documento.

Antes de poder describir exactamente cómo representar documentos, se debe tener preconcebida la estructura del documento. La mayoría de los documentos (por ejemplo, las transcripciones de la lengua oral) pueden dividirse en componentes (turnos y enunciados), y estas a su vez pueden dividirse en subcomponentes (frases entonativas, pies, sílabas y segmentos).

En XML estos componentes se llaman elementos. Cada elemento representa un componente lógico de un documento. En lugar de que cada elemento describa información sobre el formato, describe el papel lógico de sus elementos, la abstracción que representa. Dicha abstracción debe ser lo suficientemente descriptiva y debe ser elegida de manera adecuada para que los usos particulares del documento (tales como impresión, selección y búsqueda) puedan ser automatizados completamente en el momento en que la computadora actúe sobre los elementos. (Goldfarb y Prescod 2000: 13)

Los elementos pueden contener otros elementos y también pueden contener las palabras, frases y otros signos que conforman lo que comúnmente se llama texto del documento. En XML a este texto se le llama datos de caracteres (*character data*). La estructura jerárquica más sencilla de un documento de transcripción se ejemplifica en la Ilustración 2-2.

Ilustración 2-2: Estructura arbórea de los elementos básicos de una transcripción



Fuente: Diseño propio

En la estructura arbórea de la Ilustración 2-2, al elemento que contiene a todos los demás, <Transcript>, se le denomina elemento raíz. Los elementos subordinados a la raíz son sus

subelementos. Estos a su vez también pueden contener subelementos. Si los tienen, se llaman ramas, y si no, se llaman hojas. Por lo tanto, los elementos `<Turn>` y `<Utterance>` son ramas, mientras que, `<Interruption>` y el elemento `<Pause>`, aunque este último puede aparecer a distintos niveles, son hojas. Las diferencias entre estos tipos de elementos son dos: `<Utterance>` contiene datos de caracteres y su aparición es obligatoria, mientras que `<Interruption>` y `<Pause>` son elementos vacíos y, además, opcionales. Por último, se puede especificar si los elementos tienen atributos que complementen a los elementos con información descriptiva (como el atributo necesario *Name* del elemento `<Turn>`) y la frecuencia con la que los elementos puedan aparecer (como en el caso de `<Turn>` y `<Utterance>` que deben aparecer una o más veces). El elemento raíz también puede llamarse elemento del documento porque engloba todo el documento lógico (Goldfarb y Prescod 2000: 24-25).

Los textos están formados a base de caracteres. Sin embargo, se debe decidir cómo se representará los caracteres a nivel de bits y bytes. A esto se le denomina codificación de caracteres (*character encoding*). También debe decidirse qué caracteres se aceptará en los documentos. Este es el conjunto de caracteres (*character set*). El conjunto de caracteres más restrictivo permitiría únicamente letras mayúsculas, mientras que el conjunto de caracteres más amplio permitiría el uso de signos de cualquier alfabeto y sistema de escritura, incluidos los caracteres del alfabeto fonético internacional, IPA.

El conjunto básico contiene únicamente 128 caracteres (cincuenta y dos caracteres en mayúsculas y minúsculas, además de algunos símbolos, signos de puntuación y caracteres acentuados) y es conocido como ASCII de 7 bits (7 cifras binarias). El ASCII, *American Standard Code for Information Interchange*, es a la vez un conjunto de caracteres y una codificación de caracteres, pues define qué conjunto de caracteres está disponible y cómo deben codificarse en términos de bits y bytes.

Como se dijo anteriormente, el conjunto de caracteres más común para el XML es el Unicode, que incluye miles de caracteres utilizados en la mayoría de las lenguas del mundo. Sin embargo, los primeros 128 caracteres del UTF-8 (*Universal Character Set Transformation Format 8*) del Unicode son compatibles con el ASCII. Esta característica permite que los autores en XML utilicen los editores estándar de texto básico, como, por ejemplo, el Notepad de Windows. (Goldfarb y Prescod 2000: 27)

Un documento de XML se define, por lo tanto, como una serie de caracteres. Además, el XML provee un mecanismo que permite que el texto se organice de manera no lineal y potencialmente en varias piezas. El procesador lo reorganizará siempre todo en una estructura

lineal. La ‘pieza de texto’ se llama entidad (*entity*), debe tener un nombre que la identifique, y su tamaño puede ser desde un solo carácter hasta miles de caracteres. Para utilizar una entidad, se inserta una referencia a la entidad en alguna parte del documento. Y el procesador reemplazará la referencia con la entidad misma. A esto se le llama texto de reemplazo.

Hay además una característica de XML, la entidad externa, que permite una de las ideas originales del hipertexto: que el texto se pueda reutilizar automáticamente en varios contextos distintos. De modo que una actualización en un lugar se pueda propagar a lo largo de todas las aplicaciones del texto.

Gracias a esta característica, un documento de XML puede dividirse en distintos archivos en el disco duro y, según la terminología de XML, cada uno de ellos será una entidad. Esas entidades podrían incluso estar dispersas en Internet. Mientras que los elementos de XML describen la estructura lógica del documento, las entidades siguen la huella de la ubicación de las piezas de bytes que conforman un documento de XML. Esta es la estructura física del documento. (Goldfarb y Prescod 2000: 28)

2.2.2 Etiquetado

Hasta ahora se ha discutido sobre el modelo conceptual del XML (la estructura jerárquica de elementos), su estrategia para codificar caracteres (el Unicode) y su mecanismo para administrar el tamaño y complejidad de los documentos (las entidades). Sobre lo que no se ha hablado aún es sobre la manera en la que se representa la estructura lógica del documento y en la que se vinculan todas las entidades físicas: el etiquetado (*markup*).

Un lenguaje de etiquetas o marcas es una serie de convenciones utilizadas en conjunto para codificar textos. Un lenguaje de etiquetas debe especificar qué etiquetas son permitidas, qué etiquetas son obligatorias, cómo se distingue el etiquetado del texto y qué significan las etiquetas. El XML provee los medios para realizar las primeras tres condiciones y para la tercera es necesario que las etiquetas, en la medida de lo posible, sean autodescriptivas y que se creen manuales o vocabularios que las describan. De los creadores de vocabularios de XML dependerá la elección de nombres inteligibles para los elementos que identifican y la definición de su uso correcto en el texto del etiquetado. De esta necesidad de elegir nombres de elementos que indiquen su función proviene el término técnico de ‘identificador genérico’ (*generic identifier*, GI), con el que se denomina específicamente al nombre del tipo de elemento. (Sperberg-McQueen y Burnard 2004: 2, 2.3.1)

Un documento de XML se puede crear en editores de texto estándar y está hecho exclusivamente a base de etiquetado y datos de caracteres, ambos en Unicode. Los dos juntos

se llaman ‘texto de XML’. La computadora únicamente procesa el etiquetado, mientras que los datos de caracteres sólo pueden ser leídos por los humanos. (Goldfarb y Prescod 2000: 30-31)

Las características generales de la sintaxis del etiquetado en XML son las siguientes:

- El etiquetado se diferencia de los datos de caracteres por el uso de caracteres especiales llamados delimitadores. Se puede decir que una etiqueta (*tag*) es todo lo que se encuentre entre un signo de menor que (<) y uno de mayor que (>), o entre un signo de *et* (&) y un punto y coma (;). (Goldfarb y Prescod 2000: 30)
- El nombre de las etiquetas es sensible a la diferencia entre mayúsculas y minúsculas. Por lo tanto, <TAG>, <Tag> y <TaG> se interpretarían como tres etiquetas distintas.
- Los espacios en blanco fuera del etiquetado siempre se conservan, mientras que los espacios entre los delimitadores se pueden preservar o ignorar. (Goldfarb y Prescod 2000: 724-726, Sperberg-McQueen y Burnard 2004: 2.3.1)

Otros lenguajes de etiquetado que utilizan una sintaxis similar son, por ejemplo, el HTML y otros lenguajes basados en el SGML. El HTML es el lenguaje con el cual se compara más frecuentemente al XML, pues siempre se habían escrito las páginas web en HTML hasta que el XML comenzó a reemplazarlo. En resumen, comparado con el HTML, el XML tiene tres importantes características: a) el XML es extensible: no contiene un conjunto fijo de etiquetas, b) los documentos de XML deben estar bien estructurados de acuerdo a una sintaxis predefinida y que puede ser validada formalmente, y c) el XML hace hincapié en el significado de los datos, no en su presentación. (Sperberg-McQueen y Burnard 2004: 2.1)

Para ejemplificar un documento de XML tomaremos un fragmento de grabación realizada en un negocio de artículos fotográficos en la Ciudad de México. Los participantes son dos sujetos de sexo masculino:

(2.1)

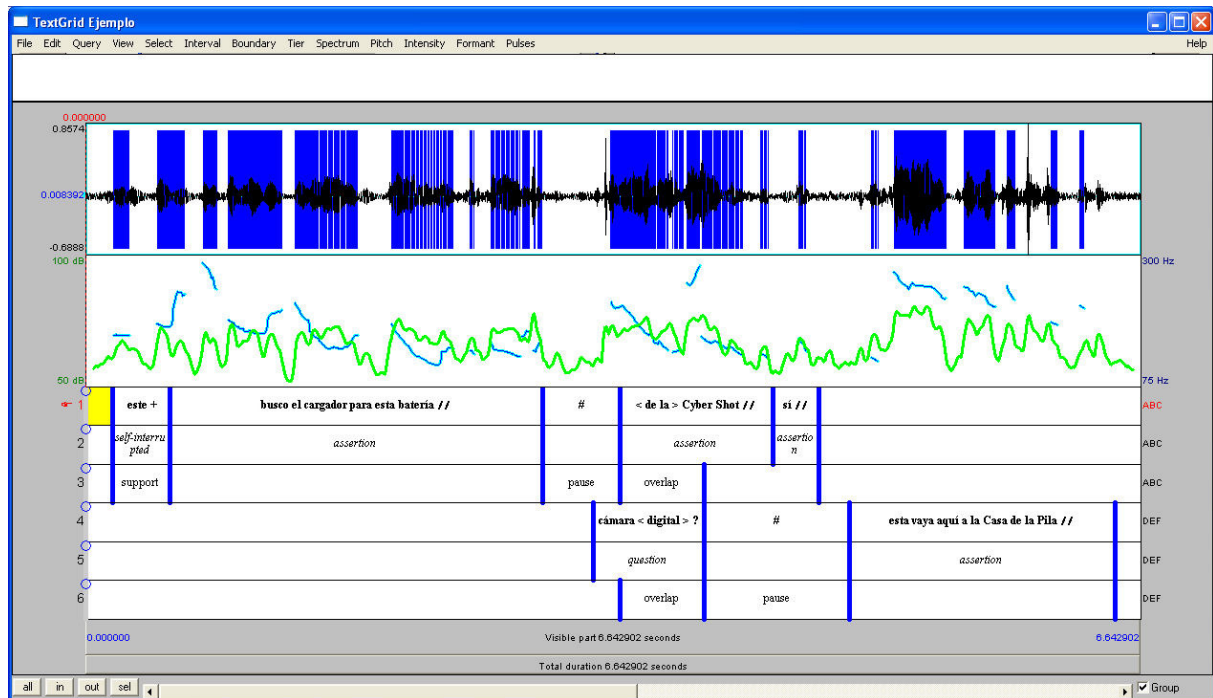
<pre>*ABC: este + busco el cargador para esta batería // # *DEF: cámara < digital > ? # *ABC: [<] < de la > Cyber Shot // sí // *DEF: esta vaya aquí a la Casa de la Pila //</pre>

En esta transcripción basada en el sistema CHAT (véase 2.3.3.1), se puede ver que el participante ABC en el primer turno se autointerrumpe (+), produce un tono descendente al final del enunciado, típico del contorno declarativo (//), y realiza una breve pausa (#). En el segundo turno, el participante DEF produce un enunciado interrogativo (?) seguido de una pausa (#). Sin embargo, mientras él produce la palabra *digital*, ABC produce simultáneamente las sílabas *de la*. Este enunciado y los dos siguientes son declarativos.

Fonéticamente se puede analizar este fragmento de grabación, al que denominaremos Ejemplo.wav, en el programa Praat, mediante el uso de distintas hileras (*tiers*) en las que, además del texto de la transcripción, se pueda recopilar la metainformación representada por medio de los signos gráficos.

En la Ilustración 2-3 se puede apreciar en la primera franja horizontal la representación gráfica de la onda sonora, dividida por pulsos o ciclos por segundo, y en la segunda franja la frecuencia fundamental en hercios (75 a 300 Hz) y la intensidad en decibeles (50 a 100 Db). Las siguientes seis hileras conforman la rejilla de texto (*text grid*): las hileras 1 y 4 contienen el texto de la transcripción, las hileras 2 y 5 incluyen la información sobre el contorno final de cada enunciado, y las hileras 3 y 6 describen otros fenómenos lingüísticos.

Ilustración 2-3: Análisis fonético del archivo de audio Ejemplo.wav en Praat



Fuente: Diseño propio

Utilizando la información del ejemplo (2.1) y la Ilustración 2-3, junto con la estructura presentada en la Ilustración 2-2, obtenemos el siguiente documento de XML:

```
(2.2) <?xml version="1.0">
<!DOCTYPE Transcript SYSTEM "Transcript.DTD">
<Transcript>
  <Turn Name="ABC">
    <Utterance>este +<Interruption/> </Utterance>
    <Utterance>busco el cargador para esta batería //</Utterance>
    <Pause/>
  </Turn>
  <Turn Name="DEF">
    <Utterance>cámara &lt; digital &gt; ?</Utterance>
    <Pause/>
  </Turn>
```

```

<Turn Name="ABC">
  <Utterance>[&lt;] &lt; de la &gt; Cyber Shot //</Utterance>
  <Utterance>sí //</Utterance>
</Turn>
<Turn Name="DEF">
  <Utterance>esta vaya aquí a la Casa de la Pila //</Utterance>
</Turn>
</Transcript>

```

Un documento de XML se divide en dos partes principales: un prólogo y un ejemplo de documento (*document instance*). El prólogo, las primeras dos líneas del ejemplo (2.2), provee la información necesaria para la interpretación del ejemplo de documento. La primera línea es una declaración sencilla de XML que indica la versión de XML a la cual se apega el documento (actualmente 1.0); una declaración más compleja podría verse así: `<?xml version="1.0" encoding="UTF-8" standalone="no"?>`, en la cual también se declara el tipo de codificación de caracteres utilizado (en este caso, el Unicode de 8 bits, UTF-8) y la poco usual declaración sobre la dependencia de un documento externo en el que se definan las normas de etiquetado (`standalone="yes"` se utiliza únicamente cuando el documento funcione de manera independiente, como cuando se experimenta con etiquetas dentro de un único documento). En la segunda línea se especifica el tipo de documento según el cual se rige el ejemplo de documento (en nuestro caso, un tipo de documento llamado `Transcript`, cuya definición se encuentra en el mismo directorio de la computadora, `SYSTEM`, y se llama `Transcript.dtd`). (Sperberg-McQueen y Burnard 2004: 2.10.1-2.10.3)

A continuación, se encuentra el ejemplo de documento que contiene los datos del documento propiamente dicho, organizados como una jerarquía de elementos. Para identificar esta estructura, lo que la computadora hace es mirar únicamente las etiquetas iniciales y finales de los elementos y los organiza en una estructura arbórea similar a la representada en la Ilustración 2-2.

En teoría, los elementos pueden contener datos de caracteres, elementos o ambos, y tienen una etiqueta inicial y una final, por ejemplo, `<Turn>` y `</Turn>`. Pero hay un tipo especial de elementos que no contiene ni caracteres ni elementos, a los que se les llama elementos vacíos (cfr. 2.2.1). Una manera de representarlos es simplemente eliminando el contenido entre las etiquetas inicial y final, mientras que la otra consiste en insertar una única etiqueta, útil en el caso de eventos puntuales en la transcripción, como una interrupción, que se vería así: `<Interruption/>` (Véase 2.2.3). (Goldfarb y Prescod 2000: 734)

Además, a los elementos puede agregárseles información extra a manera de atributos. Los atributos tienen nombres y valores con los que describen propiedades o características de los elementos. (Goldfarb y Prescod 2000: 738)

Por ejemplo, en nuestra transcripción no basta con incluir elementos `<Turn>`, sino que por medio del valor del atributo `Name` se puede identificar al participante que produjo los enunciados en dicho turno, como el hablante ABC en: `<Turn Name="ABC">`.

Finalmente, las entidades predefinidas `<` (*lesser than*) y `>` (*greater than*) hacen referencia a los signos `<` y `>` utilizados en el ejemplo (2.1) al señalar los textos simultáneos, para evitar confusiones con las marcas de los elementos dentro del documento de XML.

2.2.3 DTD (*Document Type Definition*)

En una definición del tipo de documento (DTD) se organiza una serie de especificaciones sobre los tipos de elementos, atributos y entidades. En la DTD se declara cuáles de ellos son válidos dentro del documento y en qué posiciones son válidos. Un documento puede declarar estar conformado según una DTD particular en su declaración del tipo de documento (*document type declaration*), que es abreviada normalmente como DOCTYPE para evitar confusiones.

Las DTD son poderosas herramientas de estandarización organizacional, al igual que las plantillas y guías de estilo. Las DTD, además, son importantes para permitir al *software* un procesamiento eficiente del documento. Por eso, al diseñar la estructura de los documentos, se debe tener cuidado de no especificarlos en términos de su formato en lugar de hacerlo en términos de su estructura abstracta, debido a la costumbre de diseñar los documentos en términos de su presentación. La razón de esta costumbre es que antes del SGML no existía una manera de crear un documento sin crear una presentación. El proceso de crear una DTD ofrece la oportunidad de repensar los documentos en términos de su estructura, como abstracciones. (Goldfarb y Prescod 2000: 29-34)

Al crear una DTD, el diseñador de documentos debe ser tan laxo o tan restrictivo como sea necesario. Se debe encontrar un equilibrio entre la conveniencia de utilizar reglas sencillas y la complejidad de trabajar con textos reales. Toda DTD es resultado de una interpretación del texto. No hay un esquema único que represente la verdad absoluta de un texto, aunque puede ser conveniente darle privilegio a algunos esquemas para tipos particulares de análisis.

El uso de reglas sencillas puede simplificar en gran medida la tarea del etiquetado cuidadoso de elementos de textos regulados de manera menos rígida. Al hacer estas reglas explícitas, el investigador reduce su carga al etiquetar y verificar el texto electrónico, pero al mismo tiempo se ve forzado a hacer explícita una interpretación de la estructura y particularidades significativas del texto que se está codificando. (Sperberg-McQueen y Burnard 2004: 2.4)

Veamos ahora la declaración del tipo de documento `Transcript` (cfr. Ejemplo 2.3):


```
(2.3) 01 <!ELEMENT Transcript      ( Turn+ | Pause )* >
02 <!ELEMENT Turn           ( Utterance+ | Pause )* >
03 <!ATTLIST Turn Name      CDATA      #REQUIRED>
04 <!ELEMENT Utterance     ( #PCDATA | Interruption )>
05 <!ELEMENT Interruption  EMPTY>
06 <!ATTLIST Interruption Type ( slf | ext ) #IMPLIED>
07 <!ELEMENT Pause         EMPTY>
08 <!ATTLIST Pause Duration CDATA      #IMPLIED
09      Type      ( short | long | xlong ) "short">
```

La DTD en el ejemplo (2.3) es una entidad externa (cfr. 2.2.1) a la que le daremos el nombre de `Transcript.dtd`. Esta DTD externa tiene que ser invocada en el prólogo de cada tipo de documento que se apegue a estas normas, como en (2.2). Sin embargo, la DTD también puede insertarse directamente en el prólogo de cada documento:

```
(2.4) <?xml version="1.0"?>
<!DOCTYPE Transcript [
  <!ELEMENT Transcript      ( Turn+ | Pause )* >
  <!ELEMENT Turn           ( Utterance+ | Pause )* >
  <!ATTLIST Turn Name      CDATA      #REQUIRED>
  <!ELEMENT Utterance     ( #PCDATA | Interruption )>
  <!ELEMENT Interruption  EMPTY>
  <!ATTLIST Interruption Type ( slf | ext ) #IMPLIED>
  <!ELEMENT Pause         EMPTY>
  <!ATTLIST Pause Duration CDATA      #IMPLIED
      Type      ( short | long | xlong ) "short">
]>

<Transcript>
  <Turn Name="ABC">
    <Utterance>este +<Interruption/> </Utterance>
    <Utterance>busco el cargador para esta batería //</Utterance>
    <Pause/>
  </Turn>
  <Turn Name="DEF">
    <Utterance>cámara &lt; digital &gt; ?</Utterance>
    <Pause/>
  </Turn>
  <Turn Name="ABC">
    <Utterance>[&lt;] &lt; de la &gt; Cyber Shot //</Utterance>
    <Utterance>sí //</Utterance>
  </Turn>
  <Turn Name="DEF">
    <Utterance>esta vaya aquí a la Casa de la Pila //</Utterance>
  </Turn>
</Transcript>
```

Cuando se trabaja en un solo documento, la DTD interna evita tener que alternar constantemente entre dos distintas ventanas, una con el documento de XML y otra con la DTD. Sin embargo, cuando la misma DTD sea aplicada a varios documentos, es preferible hacer únicamente referencia a la ubicación de la DTD en la misma computadora o en Internet. Del ejemplo (2.3) se pueden comentar varios aspectos sobre la sintaxis de las DTD. A saber, la declaración de cada elemento está delimitada por paréntesis angulares (signos de ‘menor

que', <, y 'mayor que', >). El primer carácter después del paréntesis de apertura debe ser un signo de exclamación e inmediatamente después una palabra clave predefinida con la que se especifica el tipo de objeto que se está declarando. Las declaraciones de las líneas 01, 02, 04, 05 y 07 en el ejemplo (2.3) comienzan con la palabra clave `ELEMENT`, que indica que se trata de la declaración de un elemento. Cada una de ellas consiste de dos partes: un nombre (o identificador genérico) y un modelo de contenido. Los componentes de la declaración se separan con caracteres de espacio en blanco (*whitespace*), tabuladores o saltos de línea. (Sperberg-McQueen y Burnard 2004: 2.4.2)

El identificador genérico puede contener caracteres alfabéticos en mayúsculas o minúsculas (se consideran caracteres distintos), dígitos, guiones y puntos, pero debe comenzar con un carácter alfabético. La segunda parte de la declaración, encerrada entre paréntesis, se llama modelo de contenido (*content model*) y especifica lo que legítimamente puede contener el elemento definido. El contenido puede especificarse en términos de otros elementos o utilizando palabras reservadas.

Además, el modelo de contenido puede incluir más de un componente, pero necesita especificar adicionalmente el orden en el que aparecerán los elementos. Este orden se determina por medio de un conector que se coloca entre los componentes de la lista. Existen dos tipos de conectores: la coma (,), que representa una secuencia, y la barra vertical (|), que representa una alternativa. Por ejemplo, en (2.3), se especifica que `<Transcript>` exige uno o más elementos `<Turn>`, pero también puede haber uno o varios `<Pause>`. (Sperberg-McQueen y Burnard 2004: 2.4.2.1-.5)

La más común de las palabras reservadas es `#PCDATA`, como en la declaración de `<Utterance>` en este ejemplo. Esta abreviatura significa que el elemento definido puede contener cualquier tipo de caracteres, pero nada de elementos, a menos que se especifiquen otros elementos como alternativas en el modelo de contenido, como en el caso de `<Interruption>`. Si recordamos la Ilustración 2-2 y partimos de la raíz `<Transcript>`, siguiendo las ramas principales `<Turn>` y las secundarias `<Utterance>`, al llegar al final, las hojas serán generalmente `#PCDATA`. Las ramas formarán la estructura del documento, pero son las hojas formadas por `#PCDATA` las que conforman el texto del documento.

Otra palabra reservada muy frecuente es `EMPTY`, que no aparece entre paréntesis e indica que el elemento carecerá de contenido. En el ejemplo (2.3), `<Interruption>` y `<Pause>` son elementos vacíos.

En el modelo de contenido de la declaración de `<Transcript>`, se especifica por medio de un indicador de frecuencia (*occurrence indicator*), el signo de adición (+), que `<Turn>` puede

aparecer una o más veces en el documento. Hay otros dos indicadores de frecuencia, el signo de interrogación (?), que significa una o ninguna vez, y el asterisco (*), indica que el elemento puede no aparecer o aparecer una o más veces.

Anteriormente se dijo que los atributos permiten al autor agregar información descriptiva adicional a los elementos de un documento. Los atributos representan datos de segunda importancia y a menudo es información sobre la información, es decir, metainformación.

En la DTD también se declaran los atributos de cada tipo específico de elemento, utilizando una especificación de lista de atributos que comienza con la palabra clave `ATTLIST`, como en las líneas 03, 06 y 08 del ejemplo (2.3). Inmediatamente después del carácter de espacio en blanco se coloca el nombre del elemento en cuestión, después el nombre del atributo, su tipo de valor y un valor predeterminado. Por ejemplo, el elemento `<Pause>` tiene dos atributos:

(2.5)

```
<!ELEMENT Pause EMPTY>
<!ATTLIST Pause Duration CDATA #IMPLIED
                Type (short|long|xlong) "short" >
```

El primer atributo se llama *Duration* y es válido sólo en los elementos `<Pause>`, su valor debe ser a base de datos de caracteres y está implícito, es decir, su aparición no es obligatoria. Para evitar tener que predecir si el valor de un atributo será únicamente a base de caracteres alfabéticos, alfanuméricos, numéricos, etc., generalmente se utiliza `CDATA` pues permite casi cualquier tipo de carácter. Con la secuencia `#IMPLIED` se permite en el ejemplo de documento que el elemento `<Pause>` aparezca con o sin atributo: `<Pause/>` o `<Pause Duration="0.49"/>`. El segundo atributo es *Type* y tiene tres valores opcionales (`short`, `long` o `xlong`), por ejemplo: `<Pause Duration="2.10" Type="xlong"/>`. En caso de que este atributo no aparezca se asumirá el valor predeterminado `short`.

En el caso del atributo *Name* del elemento `<Turn>`, la introducción de un valor es obligatorio. Por lo tanto, la declaración del atributo debe verse de la siguiente manera:

(2.6)

```
<!ATTLIST Turn Name CDATA #REQUIRED>
```

Estas últimas piezas de información en la declaración de los atributos, `#REQUIRED`, `#IMPLIED` o `"short"` indicarán a un programa de análisis sintáctico (*parser*) cómo debe interpretar la ausencia del atributo y su respectivo valor.

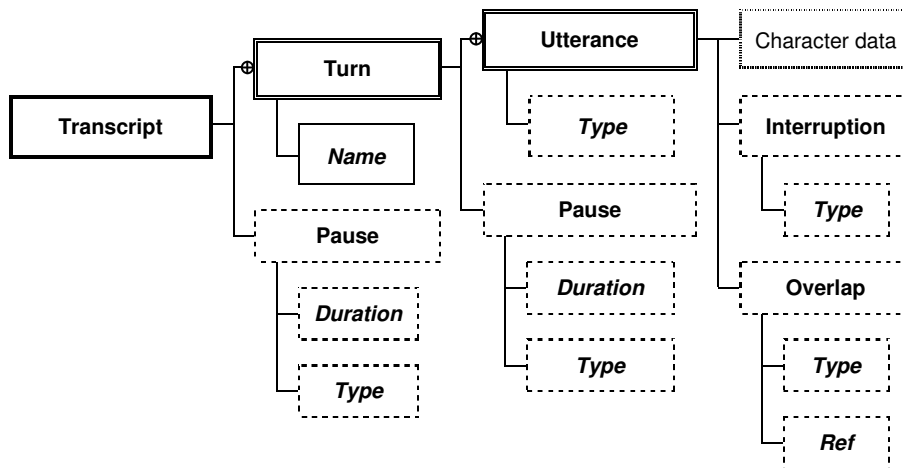
Por último, nótese que el paréntesis angular se cierra sólo después de haber declarado completamente la lista de atributos, como en el ejemplo (2.5).

Aunque es posible dar el mismo nombre a atributos de distintos elementos (por ejemplo, en (2.3), los atributos *Type* de `<Interruption>` y `<Pause>`, en las líneas 06 y 09), siempre son considerados distintos y pueden tener asignados diferentes valores según el elemento al que

pertenezcan. Si se ha declarado que un elemento tiene atributos, los valores se especificarán en el ejemplo de documento dentro de la etiqueta de inicio de dicho elemento cada que aparezca. Una etiqueta final no puede contener una especificación de valor de atributo porque sería redundante. El orden en que los atributos y sus respectivos valores se colocan dentro de la etiqueta no tiene importancia, pero deben estar separados por al menos un carácter de espacio en blanco. El valor del atributo debe colocarse entre comillas sencillas o dobles: `<Turn Name="DEF"> </Turn>`. (Sperberg-McQueen y Burnard 2004: 2.6)

Los atributos son una poderosa herramienta del XML, pues, a diferencia de los elementos que solamente proveen información sobre la estructura del documento, son los valores de los atributos los que realmente nos dan la posibilidad de describir ilimitadamente a los elementos. Por esa razón, se podría completar la estructura básica presentada en la Ilustración 2-2 con los atributos necesarios para poder incluir todos los fenómenos representados en el ejemplo (2.1).

Ilustración 2-4: Estructura jerárquica enriquecida con atributos



Fuente: Diseño propio

Antes de explicar la adición del elemento `<Overlap>`, veamos la DTD correspondiente a esta nueva estructura:

```
(2.7) 01 <!ELEMENT Transcript ( Turn+ | Pause )*>
02 <!ELEMENT Turn ( Utterance+ | Pause )*>
03 <!ATTLIST Turn Name CDATA #REQUIRED>
04 <!ELEMENT Utterance ( #PCDATA | Interruption | Overlap )*>
05 <!ATTLIST Utterance Type ( nt | qst | ssp | emp | none )
06 #IMPLIED>
07 <!ELEMENT Overlap ( #PCDATA )>
08 <!ATTLIST Overlap Type ( act | pas ) #REQUIRED
09 Ref CDATA #IMPLIED>
10 <!ELEMENT Interruption EMPTY>
11 <!ATTLIST Interruption Type ( slf | ext ) #IMPLIED>
12 <!ELEMENT Pause EMPTY>
13 <!ATTLIST Pause Duration CDATA #IMPLIED
14 Type ( short | long | xlong ) "short">
```

En el ejemplo (2.7) se han resaltado las añadiduras con respecto a la DTD presentada en el ejemplo (2.3). Primeramente, en la línea 04 se agregó un elemento más al modelo de contenido del elemento `<Utterance>`, lo que le permite tener un contenido mixto, es decir, a base de `#PCDATA` y elementos. Esto permite que los elementos `<Utterance>` acepten intercalaciones de subelementos `<Interruption>`, `<Overlap>` y datos de caracteres. También se especifica por medio del asterisco que el contenido puede no aparecer o aparecer una o más veces. La única condición de los grupos de modelos mixtos es que `#PCDATA` se declare como primera opción de una serie de alternativas, señalada por medio de barras verticales.

En las líneas 05-06 se agregó un atributo `Type` al elemento `<Utterance>` que nos permitirá especificar el tipo de enunciado por medio de un conjunto cerrado de valores: `nt`, para enunciados declarativos, correspondientes a los signos `//` del ejemplo (2.1), `qst`, para enunciados interrogativos, `ssp`, para enunciados suspendidos, `emp`, para enunciados enfáticos, y `none`, para enunciados de difícil clasificación.

El elemento `<Overlap>` (líneas 07-09), dentro del cual se colocarán los datos de caracteres solapados, se podrá especificar por medio del atributo `Type` si el texto encerrado fue solapado, `pas`, o si se trata del texto que se solapa a la producción de otro participante, `act`. Además, en el caso de que hubiera varios solapamientos, por medio del atributo `Ref` se les podría asignar un número o clave de referencia.

Por último, aunque el atributo `Type` de `<Interruption>` (línea 11) ya aparecía en el ejemplo (2.3), con él se puede especificar si el hablante se interrumpe a sí mismo, `slf`, o si es interrumpido por otro participante, `ext`.

A pesar de que el documento del ejemplo (2.2) sigue siendo compatible con esta nueva DTD, podemos enriquecerlo con los valores de los atributos recién incorporados:

```
(2.8) <?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE Transcript SYSTEM "Transcript.DTD">
<Transcript>
  <Turn Name="ABC">
    <Utterance Type="ssp">este +<Interruption Type="slf"/> </Utterance>
    <Utterance Type="nt">busco el cargador para esta bater&#237;a //</Utterance>
    <Pause Type="short" Duration="0.49"/>
  </Turn>
  <Turn Name="DEF">
    <Utterance Type="qst">c&#225;mara <Overlap Type="pas" Ref="1">&lt; digital
      &gt;</Overlap> ?</Utterance>
    <Pause Type="long" Duration="0.92"/>
  </Turn>
  <Turn Name="ABC">
    <Utterance Type="nt"><Overlap Type="act" Ref="1">[&lt;] &lt; de la
      &gt;</Overlap> Cyber Shot // </Utterance>
    <Utterance Type="nt">s&#237; //</Utterance>
  </Turn>
  <Turn Name="DEF">
    <Utterance Type="nt">esta vaya aqu&#237; a la Casa de la Pila //</Utterance>
  </Turn>
</Transcript>
```

En el prólogo de este ejemplo se optó por la declaración compleja, en la que se especifica que el sistema de codificación de caracteres es el Unicode de 8 bits (UTF-8), para garantizar su correcta visualización en todos los navegadores y sistemas operativos. Esto nos obliga a sustituir todos los caracteres no incluidos entre los 128 caracteres del sistema ASCII (véase 2.2.1) por su número de referencia. Por ejemplo: `í` equivale a `í` y `á` equivale a `á`.

2.2.4 Corrección y validez

En todas las lenguas, según sus normas gramaticales, se puede discernir entre lo que es válido dentro del sistema y lo que no. En las lenguas humanas hay dos nociones distintas de lo que es 'correcto'. Por una parte, la secuencia de segmentos formando sílabas y palabras tienen una pronunciación (o espectro de pronunciaciones) específica que es correcta, y por otra, esas palabras pueden combinarse de manera que se produzcan enunciados válidos, es decir, gramaticales. Del mismo modo, en XML el etiquetado puede ser simplemente inteligible, dando lugar a un documento bien formado (*well-formed*), y sin embargo, además tiene que ser sintácticamente correcto. A esto es a lo que se le llama validez (*validity*). Un documento es válido si declara apegarse a una DTD en una declaración del tipo de documento (DOCTYPE, cfr. 2.2.3) y si en verdad respeta las reglas de dicha DTD. (Goldfarb y Prescod 2000: 34-39)

Algunos documentos no tienen DOCTYPE. Esto no significa que no estén conformados según cierto tipo de documento, simplemente significa que no declaran conformarse según un tipo de documento formalmente definido. Estos documentos que no tienen una declaración del tipo de documento en realidad no son inválidos, dado que no violan las reglas de ninguna DTD, pero tampoco son válidos, porque no pueden ser cotejados según una DTD.

El XML es ideal para representar datos, pero eventualmente los datos tienen que ser procesados y para esto se requiere del uso de uno o más programas. Una de las grandes ventajas del XML es que hay una multitud de aplicaciones que son fáciles de conseguir y muchas veces a un muy bajo costo, por ejemplo, el XMLwriter, que es el editor de XML y analizador sintáctico o *parser* utilizado en esta investigación.

Un *parser* se encarga de separar la representación textual de un documento y convertirla en un conjunto de objetos conceptuales. Más específicamente, puesto que los documentos pueden clasificarse en distintos tipos, una vez que se haya provisto a un *parser* con una definición consistente de un tipo de documento, éste puede verificar que cualquier documento que declare ser de ese tipo realmente se apegue a la especificación. Un *parser* puede verificar que todos los elementos especificados para un tipo particular de documento, y ni uno más, estén presentes, que se combinen de manera apropiada, que se ordenen correctamente, etc. Pero

sobre todo, se encarga de que los diferentes documentos del mismo tipo sean procesados de manera uniforme. Como consecuencia, se puede escribir programas que aprovechen el conocimiento encapsulado en la información estructural del documento para que se comporten de manera ‘inteligente’. (Goldfarb y Prescod 2000: 38-39, Sperberg-McQueen y Burnard 2004: 2.1.2)

Por ejemplo, un *parser* que procesara el documento del ejemplo (2.8) reconocería la secuencia de caracteres ‘<Transcript>’ como una etiqueta de inicio, y sabría que ésta señala el inicio de un elemento <Transcript>. La etiqueta es parte de la representación, mientras que el elemento es el objeto conceptual. Si el *parser* validara este documento de acuerdo a la DTD (2.7), se aseguraría de que, por ejemplo, los elementos <Turn> estuvieran colocados después del elemento <Transcript>, y que <Turn> al menos incluyera un elemento <Utterance>.

Es conveniente mencionar que para la declaración de los tipos de documentos no solamente existen las DTD. A pesar de su amplia aceptación, a la sintaxis de las DTD se les critica estar escritas en un lenguaje distinto al XML y, sobre todo, el hecho de que no provee el soporte necesario para trabajar con direcciones electrónicas, tipos de datos específicos y estructuras no necesariamente jerárquicas. Como las principales alternativas han aparecido varios lenguajes esquemáticos, como el *XML Schema Language* (Biron y Malhotra 2004, Brown *et al.* 2001, Fallside y Walmsley 2004), recomendado por el *w3c Consortium*, y Relax NG (Clark y Murata 2001, Clark 2002), propuesto por la *Organization for the Advancement of Structured Information Standards*, OASIS. Sin embargo, estos lenguajes son compatibles en gran medida con las DTD y las ventajas que ofrecen sólo son necesarias para sitios web escritos totalmente en XML con necesidades específicas, como formularios de pedidos, manejo de multimedia, etc. Para los fines de esta investigación es suficiente utilizar la DTD.

2.2.5 XSL / XSLT

El *Extensible Stylesheet Language* (XSL) es una especificación aún en desarrollo dentro del World Wide Web Consortium (el organismo encargado de recomendar y estandarizar las tecnologías informáticas relacionadas con Internet) para aplicar formato a los documentos de XML de manera estandarizada. Una primera parte del estándar de XSL, llamada *XSL Transformations* (XSLT) se ha desarrollado en un útil lenguaje independiente para transformar un documento XML en otro documento de XML con información sobre el formato o directamente en HTML, es decir, una página *web*. La segunda parte define la semántica del formato, a manera de un catálogo de objetos de formato (*formatting objects*). Sin embargo, en este trabajo se hará uso exclusivo del XSLT. (Clark 1999)

Como se dijo, el XSL se diseñó para aplicar estilo a los documentos de XML. Se supone que estos documentos estarán etiquetados de acuerdo a su estructura abstracta y carecerán de etiquetas concebidas específicamente para la aplicación de estilo o para cualquier otro tipo de proceso. Por lo tanto, el XSL es el ‘eslabón perdido’ entre los datos abstractos, concebidos para que la computadora los procese, y la representación con formato para una cómoda lectura. En términos técnicos, un procesador de XSL transforma un documento de una abstracción a una representación. (Goldfarb y Prescod 2000: 873-874)

Muchas de las implementaciones de XSLT permiten transformaciones de XML a HTML, es decir, el XSLT procesa el etiquetado de XML y lo complementa con etiquetado de HTML. Esta es una gran ventaja, pues se puede aprovechar los navegadores de Internet para visualizar las representaciones de los documentos de XML como si hubieran sido escritos originalmente en HTML. Los navegadores más actuales (a partir de las versiones Internet Explorer 6.0, Netscape 5.0 y Mozilla 1.0) hacen la conversión directamente, sin necesidad de programas auxiliares.

Las hojas de estilo en XSL son simplemente una forma especializada de etiquetado de XML, diseñado para especificar el formato de otros documentos de XML, pero para fines prácticos pueden ser consideradas como un tipo de documento diferente. El XSL define tipos de elementos y atributos, los obliga a aparecer en sitios particulares y describe la apariencia que deben de tener.

El elemento `<xsl:stylesheet>` es normalmente el elemento raíz de las hojas de estilo en XSL. La razón de utilizar el prefijo `xsl:` es simplemente convencional, para facilitar la identificación del tipo de documento. La declaración de la hoja de estilo es la siguiente:

(2.9)

```

<xsl:stylesheet version="1.0"
                xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
    <!-- Aquí se debe introducir todas las reglas de plantilla -->
</xsl:stylesheet>
```

El elemento `<xsl:stylesheet>` contiene las reglas de plantilla (*template rules*) que describen cómo dar formato a los elementos en el documento de origen. Durante el proceso de transformación, cada elemento, carácter, comentario e instrucción de proceso en un documento de XML se procesa por medio de una regla de plantilla. (Clark 1999)

Antes de continuar, se debe tomar en cuenta que los documentos de XML deben especificar la existencia de una hoja de estilo a la cual se apegan. El prólogo del documento en el ejemplo (2.8), al cual llamaremos a partir de ahora `Transcript.xml`, tendría que especificar el nombre de la entidad de XSL que se encargará de darle formato, `Transcript.xsl`:


```
(2.10) <?xml version="1.0"?>
<?xml:stylesheet type="text/xsl" href="Transcript.xsl"?>
```

Por otra parte, las reglas de plantilla constan de dos partes, el nivel de aplicación (*pattern*) y la plantilla (*template*). Una plantilla no es lo mismo que una regla de plantilla. El nivel de aplicación describe cuáles nodos (*nodes*) en el documento de origen (elementos, secuencias de datos textuales, comentarios o instrucciones de proceso) serán procesados por la regla. La plantilla describe la estructura de XML o HTML que se generará cuando se encuentren los nodos que correspondan al nivel de aplicación.

En la hoja de estilo de XSL, la regla de plantilla aparece como un elemento `<xsl:template>`. El nivel de aplicación es el valor del atributo `match` del elemento `<xsl:template>`, y la plantilla propiamente dicha es el contenido del elemento (Goldfarb y Prescod 2000: 876-879, Clark 1999: 5).

El primer paso es diseñar en HTML la estructura básica de formato del documento:

```
(2.11) <xsl:template match="Transcript">
  <html>
    <body>
      <font size="4"><b><!-- Título de la página --></b></font>
      <hr/>
      <!-- Incluir el texto de los turnos -->
    </body>
  </html>
</xsl:template>
```

La plantilla en esta regla de plantilla genera un `<body>` de HTML que servirá de soporte al contenido de `Transcript.xml`. Los comentarios, señalados por medio de `<!--` y `-->`, en el ejemplo (2.11), indican los lugares en los que se deberá introducir los datos. Con el elemento `<html>` se indica al procesador de XSLT que todos los subelementos son resultados literales que se copiarán al producto final. Estos elementos describen la forma del documento resultante ordenando el contenido y generando elementos de resultado literal. Con el elemento `` se especifican las características del tipo de letra, el elemento `` hace que su contenido se represente en negrita y `<hr>` introduce una barra horizontal.

El título de la página puede ser fijo para todas las transcripciones, por eso lo escribiremos directamente:

```
(2.12) <font size="4"><b>Transcript:</b></font>
```

El texto formado por los datos de caracteres de los elementos `<Utterance>`, subordinados a los elementos `<Turn>`, será extraído del documento de XML por medio de la instrucción `<xsl:apply-templates>`. Este texto abstraído se puede especificar para que seleccione cada `<Turn>` y le agregue siempre al final un salto de línea `
`. Para que esta acción se repita las

veces necesarias, recurriremos a la instrucción `<xsl:for-each>`, a cuyo atributo `select` daremos como valor el nivel de aplicación, es decir, el elemento `<Turn>`:

```
(2.13) <xsl:for-each select="Turn">
<xsl:apply-templates/><br/>
</xsl:for-each>
```

Mientras que `<xsl:apply-templates>` aplicaría las características especificadas en la plantilla al nodo que concordara con el nivel de aplicación, `<xsl:for-each>` se encargaría de repetir la acción por cada uno de los elementos, en este caso `<Turn>`, a lo largo de todo el documento.

Si ahora extraemos el valor del atributo `Name` de cada `<Turn>` por medio de la instrucción `<xsl:value-of>` y su atributo `select`, que indica el nivel de aplicación, tendremos que especificar el valor '@Name' (la arroba indica que se trata de un atributo y no de un subelemento de `<Turn>`). Además, si queremos que después del nombre haya dos puntos y que todo se despliegue en letra negrita, la regla de plantilla completa se verá así:

```
(2.14) <xsl:for-each select="Turn">
<b><xsl:value-of select="@Name"/>: </b>
<xsl:apply-templates/><br/>
</xsl:for-each>
```

Ahora, si se aplicaran los cambios presentados en los ejemplos (2.12) y (2.14) al documento del ejemplo (2.11) en los espacios reservados, se obtendría la siguiente estructura:

```
(2.15) <xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

<xsl:template match="Transcript">
<html>
<body>
<font size="4"><b>Transcript:</b></font>
<hr/>
<xsl:for-each select="Turn">
<b><xsl:value-of select="@Name"/>: </b>
<xsl:apply-templates/><br/>
</xsl:for-each>
</body>
</html>
</xsl:template>
</xsl:stylesheet>
```

La transformación de esta entidad daría como resultado el siguiente documento de HTML:

```
(2.16) <html>
<body>
<font size="4"><b>Transcript:</b></font>
<hr>
<b>ABC: </b>este + busco el cargador para esta batería //<br>
<b>DEF: </b>cámara &lt; digital &gt; ?<br>
<b>ABC: </b>[&lt;] &lt; de la &gt; Cyber Shot // sí //<br>
<b>DEF: </b>esta vaya aquí a la Casa de la Pila //<br>
</body>
</html>
```

Y este documento de HTML se desplegaría en cualquier navegador de la siguiente manera:

(2.17) **Transcript:**

ABC: este + busco el cargador para esta batería //
DEF: cámara < digital > ?
ABC: [<] < de la > Cyber Shot // sí //
DEF: esta vaya aquí a la Casa de la Pila //

Sin embargo, el formato de texto no es lo único que se puede realizar por medio de XSLT. También se puede numerar cada turno y hacer, por ejemplo, un sumario en el que se enliste el número de veces que aparece cada fenómeno prosódico. El documento `Transcript.xsl` modificado se verá así:

```
(2.18) <xsl:stylesheet version="1.0"
          xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

  <xsl:template match="Transcript">
    <html>
      <body>
        <font size="4"><b>Transcript:</b></font>
        <hr/>
        <ol type="1">
          <xsl:for-each select="Turn">
            <li/>
            <b><xsl:value-of select="@Name"/>: </b>
            <xsl:apply-templates/>
            <br/>
          </xsl:for-each>
        </ol>
        <hr/>
        <font size="4"><b>Summary:</b></font>
        <br/><font size="-1"><b>Turns: </font>
        <b><xsl:value-of select="count(..//Turn)" /></b>
        <br/><font size="-1"><b>Utterances: </font>
        <b><xsl:value-of select="count(..//Utterance)" /></b>
        <br/><font size="-1"><b>Overlaps: </font>
        <b><xsl:value-of select="count(..//Overlap[@Type='act'])" /></b>
        <br/><font size="-1"><b>Pauses: </font>
        <b><xsl:value-of select="count(..//Pause)" /></b>
        <br/><font size="-1"><b>Assertive utterances: </font>
        <b><xsl:value-of select="count(..//Utterance[@Type='nt'])" /></b>
        <br/><font size="-1"><b>Interrogative utterances: </font>
        <b><xsl:value-of select="count(..//Utterance[@Type='qst'])" /></b>
        <br/><font size="-1"><b>Suspended utterances: </font>
        <b><xsl:value-of select="count(..//Utterance[@Type='ssp'])" /></b>
        <br/><font size="-1"><b>Self-interrupted utterances: </font>
        <b><xsl:value-of select="count(..//Interruption[@Type='slf'])" /></b>
        <br/><font size="-1"><b>Ext. interrupted utterances: </font>
        <b><xsl:value-of select="count(..//Interruption[@Type='ext'])" /></b>
      </body>
    </html>
  </xsl:template>
</xsl:stylesheet>
```

Para la numeración de los turnos se debe abrir una lista ordenada (*ordered list*, ``) que abarque completamente el proceso `<xsl:for-each select="Turn">` y simplemente agregar un elemento `` (*list item*) al inicio de cada línea.

Tanto el título `Summary` como el nombre de cada uno de los fenómenos se escriben directamente con datos de caracteres. El valor negativo `"-1"` en `` es solamente por cuestiones de estilo, para que el tamaño de la fuente sea ligeramente más

pequeña que la del texto de la transcripción. El valor correspondiente al número de veces que aparece un fenómeno en el documento se obtiene por medio de la instrucción general:

```
<xsl:value-of select="count(..//Elemento[@Atributo='Valor'])"/>.
```

El resultado de esta transformación se desplegaría en un navegador de Internet de este modo:

(2.19) **Transcript:**

-
1. **ABC:** este + busco el cargador para esta batería //
 2. **DEF:** cámara < digital > ?
 3. **ABC:** [<] < de la > Cyber Shot // sí //
 4. **DEF:** esta vaya aquí a la Casa de la Pila //
-

Summary:

Turns: **4**

Utterances: **6**

Overlaps: **1**

Pauses: **2**

Assertive utterances: **4**

Interrogative utterances: **1**

Suspended utterances: **1**

Self-interrupted utterances: **1**

Ext. interrupted utterances: **0**

En conclusión, en los ejemplos anteriores se han visto sólo las funciones básicas del XSLT. Estos son los primeros pasos para la formación de un formato visual en el que se representen textos sencillos de manera legible y se apliquen procesos simples. Más adelante se verá cómo se puede utilizar algunas instrucciones del HTML dinámico (DHTML) para hacer que los documentos de las transcripciones sean más llamativos y así facilitar el enlace interactivo a multimedia y a la información sobre los distintos fenómenos lingüísticos.

2.2.6 Resumen

El XML es un lenguaje sencillo y lo suficientemente flexible para poder aplicarse a la lingüística en general y a la transcripción de la lengua hablada en particular. Para poder gozar de todas las ventajas que ofrece el XML deberán definirse los tipos de elementos que aparecerán a lo largo de la transcripción, así como sus posibles atributos, y declararlos en una DTD. Consecuentemente, se deberá crear las hojas de estilo en XSL que le den a las transcripciones el formato ideal para una investigación detallada.

En la siguiente sección presentaré unas propuestas de DTD y XSL que podrán ser aplicadas a las transcripciones de un *corpus* de conversaciones grabadas en la Ciudad de México y en

Madrid. Este *corpus* de documentos de XML con la información prosódica etiquetada formará la base de mi análisis de ambas variedades del español.

2.3 Transcripción prosódica de la lengua oral en XML

De manera intuitiva, los hablantes de una lengua son capaces de identificar con mayor o menor precisión las características que hacen que su propia variedad de la lengua y la de otro hablante sean distintas. Por este motivo, en esta investigación se propone un sistema de transcripción de la lengua oral que permita identificar de manera clara las características prosódicas que dan a cada dialecto del español su identidad única y que, por medio de la comparación de estas características con las de otro dialecto, se pueda sistematizar las diferencias dialectales al nivel de la prosodia.

En las secciones anteriores se fue introduciendo paso a paso el lenguaje XML y se demostraron algunas de sus posibilidades en cuanto al manejo de la estructura de la lengua oral en tanto que se le trate como un sistema jerárquico. En esta sección se profundizará en la adaptación y aplicación del XML a textos de mayor extensión y riqueza en cuanto a la variedad de fenómenos de interés para el estudio de la prosodia comparada.

En la estructura resultante se permitirá la introducción de los fenómenos cuya importancia ya se había señalado en la sección 1, por ejemplo: 1) duración, tempo relativo, frecuencia fundamental, intensidad, volumen relativo, prominencia y foco de cada sílaba [1.1.1, 1.3.3, 1.1.2.1], 2) distribución de grupos rítmicos [1.3.4], 3) tonos de inicio, prenucleares, nucleares y finales de cada unidad entonativa [1.2.4, 1.2.4.3], 4) tipo de enunciado [1.1.2.2] y 5) otros fenómenos tales como pausas, solapamientos, interrupciones, fragmentos ininteligibles, etc.

2.3.1 Antecedentes

La idea de la representación de la lengua humana en un lenguaje de etiquetas no es nueva. Distintos equipos de trabajo con una mayoría de especialistas en informática se dedican a hacer posible la navegación por la web semántica (por ejemplo, Berners-Lee *et al.* 2001), basada en una clasificación conceptual de la información, mientras otros se ocupan de la síntesis del habla (*Speech Synthesis Markup Language*, SSML, Burnett *et al.* 2004) y el reconocimiento de voz (*Voice Extensible Markup Language*, VoiceXML, McGlashan *et al.* 2004). De mayor interés para los científicos del lenguaje son las aplicaciones EXMARaLDA y TEI, cuyos principios se relacionan en cierta medida con mi propuesta. En las siguientes dos secciones se presentará ambos sistemas a grandes rasgos.

2.3.1.1 EXMARaLDA

EXMARaLDA, acrónimo de ‘*Extensible Markup Language for Discourse Annotation*’, es un sistema computacional de conceptos, formatos de datos y herramientas para la transcripción y etiquetado de la lengua oral, desarrollado en la Universität Hamburg como componente arquitectónico central de la base de datos ‘*Mehrsprachigkeit*’.

Las características principales de EXMARaLDA son: a) el uso del XML para las transcripciones, lo que asegura que los datos se puedan utilizar flexiblemente y archivar a largo plazo; b) las herramientas de software para la creación y procesamiento de documentos (Partitur-Editor y Corpus-Manager) a base de Java, un lenguaje computacional compatible con todos los sistemas operativos actuales (Windows, Macintosh, Linux, Unix); y c) su interoperabilidad a base del Annotation Graph Framework (Bird y Liberman 2001), que asegura el intercambio y reutilización de los datos de las transcripciones, así como su creación por medio de otras herramientas basadas en el mismo principio.

Además, los documentos de EXMARaLDA se pueden visualizar como presentaciones impresas o electrónicas en los formatos RTF, HTML y PDF. Por último, este sistema es compatible con varios de los sistemas de transcripción más comunes (HIAT, DIDA, GAT, CHAT). (Schmidt 2005)

Sin embargo, EXMARaLDA no es un sistema de transcripción ni un programa, sino un lenguaje para describir transcripciones discursivas. El programa Partitur-Editor, al igual que el sistema de transcripción HIAT (*HalbInterpretative ArbeitsTranskriptionen*, Ehlich 1993), utiliza la llamada notación de partitura para visualizar la secuencia temporal y la simultaneidad entre las producciones lingüísticas de distintos hablantes, entre distintas modalidades (comportamiento verbal y no verbal) y entre los fenómenos segmentales y no segmentales (prosódicos).

Las relaciones estructurales en una partitura se visualizan por medio de una subdivisión vertical en filas o hileras (*tiers*) que representan una asignación de identidades a los distintos hablantes y a las diferentes categorías de análisis. La hilera principal, divisible en unidades más pequeñas ordenadas según la secuencia temporal, es la del comportamiento verbal y se utiliza como punto de referencia para todas las demás entidades. La hilera de las descripciones del comportamiento no verbal incluye los eventos en unidades indivisibles, independientes de la información en las otras hileras. Finalmente, las hileras de anotaciones describen rasgos adicionales del comportamiento verbal no capturados en las hileras de la transcripción (prosodia, traducciones, etc.) y no son entidades independientes. Las entidades en las diferentes hileras de la partitura se relacionan con una línea común de tiempo: los eventos

simultáneos o pares de entidad y rasgo se colocan en la misma posición horizontal y la dirección de izquierda a derecha corresponde a la secuencia temporal. (Schmidt 2004a)

En una versión simplificada de la DTD del sistema EXMARaLDA para las transcripciones básicas se puede observar la manera en la que se representa la estructura temporal:

```
(2.20) <!ELEMENT basic-transcription (head, basic-body)>
<!ELEMENT head (speakertable)>
<!ELEMENT speakertable (speaker*)>
<!ELEMENT speaker EMPTY>
<!ATTLIST speaker id ID #REQUIRED>
<!ELEMENT basic-body (common-timeline, tier*)>
<!ELEMENT common-timeline (tli*)>
<!ELEMENT tli EMPTY>
<!ATTLIST tli id ID #REQUIRED
time CDATA #IMPLIED>
<!ELEMENT tier (event*)>
<!ATTLIST tier speaker IDREF #IMPLIED
category CDATA #REQUIRED
type (t|d|a) #REQUIRED>
<!ELEMENT event (#PCDATA)>
<!ATTLIST event start IDREF #REQUIRED
end IDREF #REQUIRED>
```

Fuente: Schmidt 2004a

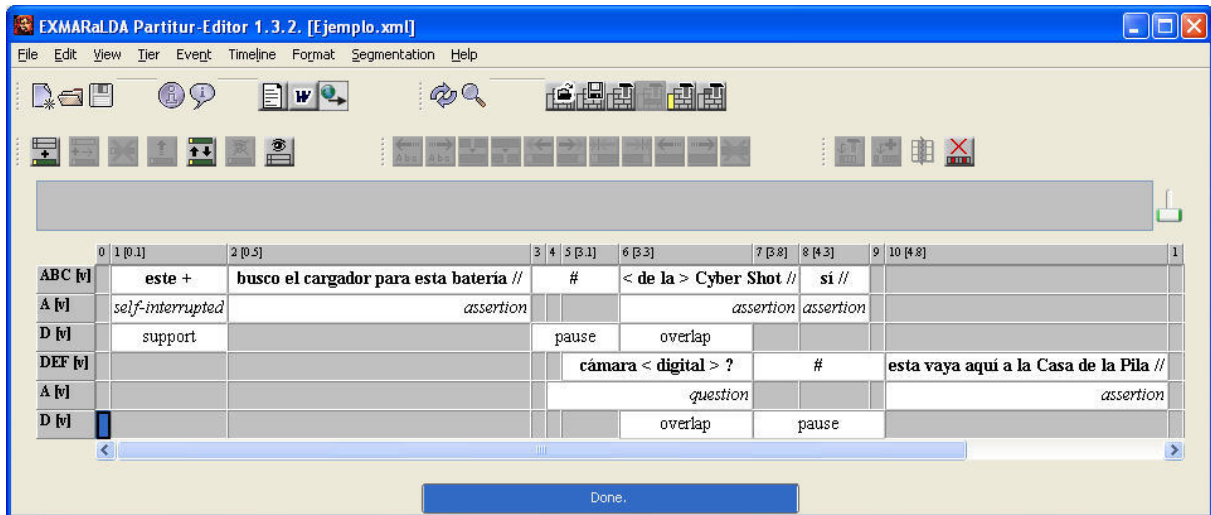
Según esta DTD, las unidades más pequeñas de la transcripción de partitura se representan como eventos (elementos `<event>`). Los eventos se refieren por medio de los atributos `start` y `end` a las claves de identidad de los elementos individuales, `<tli>`, de una línea común de tiempo, `<common-timeline>`, y se agrupan en hileras, `<tier>`. A cada `<tier>` se le asignan los atributos `speaker` y `category`, así como uno de los tres tipos de hilera descritos anteriormente: `t` (*transcription*), `d` (*description*) o `a` (*annotation*). Adicionalmente, a cada elemento de la línea de tiempo se le puede asignar un valor absoluto de tiempo por medio del atributo opcional `time` y apuntar así a una posición en una grabación de audio o video.

Esta transcripción básica en la que se describe la estructura temporal de los eventos concretos del discurso se hace normalmente siguiendo alguno de los principios tradicionales de diseño. Todos ellos requieren únicamente una línea de tiempo común para todos los eventos. Sin embargo, cuando se necesita agregar información a estos datos principales (por ejemplo, traducciones de los enunciados o la clasificación sintáctica de sus componentes), por lo general, la estructura temporal ya no es tan importante como la estructura lingüística. Esto puede resultar en nuevos puntos en la línea de tiempo que no se pueden acomodar de manera unequívoca. El caso típico es el que surge cuando los fragmentos solapados de dos o más hablantes contienen más de una unidad lingüística. En este caso, un formato más complejo de EXMARaLDA, el de transcripción segmentada, permite líneas de tiempo múltiples que se intersectan de manera parcial para hacer posible la anotación de unidades arbitrarias sin tener

que acomodarlas en un orden estricto (que, de cualquier modo, en la mayoría de los casos sería imposible para el transcriptor). (Schmidt 2001)

Para darnos una idea clara de la estructura íntera de EXMARaLDA, podemos retomar el ejemplo (2.1), cuyo análisis en Praat se había presentado en la Ilustración 2-3. Podemos aprovechar la estructura temporal de Praat importando los datos desde el Partitur-Editor. El resultado es el siguiente:

Ilustración 2-5: Exportación de la información de Praat a EXMARaLDA



Fuente: Diseño propio

Si guardamos esta información procesada por el Partitur-Editor como un nuevo archivo de XML, obtendremos la siguiente estructura:

```
(2.21) <?xml version="1.0" encoding="UTF-8"?>
<!-- (c) http://www.rrz.uni-hamburg.de/exmaralda -->
<basic-transcription>
<head>
  <speaker-table>
    <speaker id="ABC"/>
    <speaker id="DEF"/>
  </speaker-table>
</head>
<basic-body>
  <common-timeline>
    <tli id="T0" time="0.0"/>
    <tli id="T1" time="0.17162455452589195"/>
    <tli id="T2" time="0.5310996307672728"/>
    <tli id="T3" time="2.8765271773913645"/>
    <tli id="T9" time="3.1951453800829968"/>
    <tli id="T4" time="3.3657268085142613"/>
    <tli id="T8" time="3.8901290317113237"/>
    <tli id="T5" time="4.326213550430376"/>
    <tli id="T6" time="4.61497221823083"/>
    <tli id="T10" time="4.80944234144338"/>
    <tli id="T11" time="6.477170973841917"/>
    <tli id="T7" time="6.642902494331066"/>
  </common-timeline>
  <tier id="TIE0" speaker="ABC" category="v" type="t" display-name="ABC [v]">
    <event start="T1" end="T2">este +</event>
    <event start="T2" end="T3">busco el cargador para esta batería //</event>
    <event start="T3" end="T4">#</event>
    <event start="T4" end="T5">&lt; de la &gt; Cyber Shot //</event>
    <event start="T5" end="T6">sí //</event>
  </tier>
```



```

<tier id="TIE1" speaker="ABC" category="v" type="a" display-name="A [v]">
  <event start="T1" end="T2">self-interrupted</event>
  <event start="T2" end="T3">assertion</event>
  <event start="T4" end="T5">assertion</event>
  <event start="T5" end="T6">assertion</event>
</tier>
<tier id="TIE2" speaker="ABC" category="v" type="d" display-name="D [v]">
  <event start="T1" end="T2">support</event>
  <event start="T3" end="T4">pause</event>
  <event start="T4" end="T8">overlap</event>
</tier>
<tier id="TIE3" speaker="DEF" category="v" type="t" display-name="DEF [v]">
  <event start="T9" end="T8">cámara &lt; digital &gt; ?</event>
  <event start="T8" end="T10">#</event>
  <event start="T10" end="T11">esta vaya aquí a la Casa de la Pila //</event>
</tier>
<tier id="TIE4" speaker="DEF" category="v" type="a" display-name="A [v]">
  <event start="T9" end="T8">question</event>
  <event start="T10" end="T11">assertion</event>
</tier>
<tier id="TIE5" speaker="DEF" category="v" type="d" display-name="D [v]">
  <event start="T4" end="T8">overlap</event>
  <event start="T8" end="T10">pause</event>
</tier>
</basic-body>
</basic-transcription>

```

En este ejemplo de documento podemos observar como `<basic-transcription>` contiene un elemento `<head>`, donde se recopilan los datos de los hablantes, y un elemento `<basic-body>`, que contiene una sección denominada `<common-timeline>` y que enumera los elementos individuales de la línea común de tiempo. Después hay una serie de elementos `<tier>` que consta de los distintos `<event>` donde se deposita el texto de la hilera, señalando su inicio y final por medio de atributos que hacen referencia a los elementos `<tli>`.

Según Schmidt (2002), una subdivisión en eventos y una organización de estos eventos en un eje temporal ordenado siempre debe ser un componente del modelo. A diferencia de la clasificación lingüística de las unidades conversacionales, el ordenamiento temporal tiene la ventaja de hacer posible un análisis con unidades relativamente objetivas, sin la necesidad de hacer suposiciones lingüístico-teóricas, aunque esto precisamente también podría representar una desventaja. Por estos motivos, el modelo debe restringirse a ofrecer una posibilidad de describir, además de la estructura temporal, una estructura lingüística mínimamente definida de una conversación, sin comprometerse con unas categorías determinadas.

2.3.1.2 TEI

La *Text Encoding Initiative* (TEI) es un proyecto internacional de investigación, nacido en 1987, cuya meta es definir un conjunto de normas generales (las llamadas *TEI Guidelines*) para la representación de materiales textuales de manera electrónica. El proyecto fue patrocinado y organizado por la *Association for Computational Linguistics* (ACL), la *Association for Literary and Linguistic Computing* (ALLC) y la *Association for Computing and the Humanities* (ACH). Pero el interés de la TEI no sólo se limita al uso del texto

electrónico como paso previo para la producción de documentos impresos, y la palabra ‘text’ no debe ser entendida de manera tan literal. La TEI se ocupa tanto de recursos electrónicos textuales como de los no textuales, ya sea como constituyentes de una base de datos o como componentes de publicaciones no impresas.

En la investigación lingüística, los textos suelen ser muy heterogéneos. Sin embargo, los materiales textuales de los orígenes y características más diversas suelen ser integrados en un solo conjunto de materiales. Las necesidades de la investigación imponen que esta integración se realice con una mínima pérdida de información y al mismo tiempo con una complejidad mínima: en cualquiera de los casos, el ‘texto’ resultante estará muy alejado de la noción convencional de la obra impresa.

Los objetivos del proyecto TEI inicialmente tuvieron un enfoque dual: establecer **cuáles** rasgos textuales deberían de codificarse (hacerse explícitos) en un texto electrónico, y **cómo** se debería representar esa codificación para facilitar el intercambio sin pérdida de información y de manera independiente de la plataforma. Para lograrlo, se eligió el SGML como el vehículo apropiado para las normas, inicialmente por la razón puramente pragmática de que la creación de un lenguaje formal con una expresividad y versatilidad comparables sería otro proyecto de investigación por sí solo.

La estandarización de la manera en la que se almacena y representa la información es la solución a varios problemas estrechamente relacionados que son de interés para los usuarios de las tecnologías de la información tanto académicos como comerciales. En particular, para los creadores de recursos relacionados con el lenguaje, la estandarización ya no plantea la dificultad de asegurar que la información sea reutilizable, la dificultad de asegurar que la información representada de distintas maneras pueda ser integrada fácilmente, y la dificultad de facilitar el intercambio sin pérdida de información entre la más amplia gama de distintas plataformas, diferentes sistemas de aplicación y distintos idiomas.

Al estandarizar a partir del nivel de la representación textual, sin afinidad por ninguna corriente teórica, con la TEI se espera mantener la flexibilidad necesaria para desarrollar nuevas aplicaciones, asegurándose al mismo tiempo de que las antiguas sigan funcionando.

Por este motivo, las TEI Guidelines no intentan proveer un conjunto de reglas normativas para aplicaciones particulares, sino más bien un marco modular y extensible dentro del cual se pueda definir las normas específicas para una aplicación particular. (Burnard 1995)

La primera versión oficial de las normas (también conocido como TEI P3) se publicó en 1994 en un manual de referencia con el nombre de *Guidelines for Text Encoding for Interchange* (Sperberg-McQueen y Burnard 1994 y 1999). En este documento se incluía alrededor de 400

definiciones de rasgos textuales, expresados como elementos y atributos de SGML, con documentación asociada y ejemplos. Estos elementos se agruparon en conjuntos de etiquetas (*tag sets*) y juntos constituían un esquema modular que podía configurarse para proveer un soporte, independiente de *hardware*, *software* y otras aplicaciones, para la codificación de todo tipo de textos en lenguas modernas e históricas.

Los distintos conjuntos de etiquetas pueden ser utilizados en una gran variedad de combinaciones, de acuerdo a las necesidades del usuario y la naturaleza del material que tiene que ser codificado. Hay un conjunto de etiquetas nuclear y otro para el encabezado. El primero define los elementos que pueden ser considerados como universales, pues no son específicos de tipos particulares de textos. El segundo define los elementos que permiten la introducción de información documental o bibliográfica sobre el texto electrónico mismo (metadatos). Además de estos, hay los conjuntos de etiquetas que se aplican a los distintos tipos específicos de texto: prosa, drama, poesía, transcripción de la lengua oral, diccionarios, etc. También hay varios conjuntos adicionales de etiquetas diseñados para ser utilizados en tipos particulares de procesos o investigaciones, como la transcripción de fuentes literarias primarias o crítica textual. De este modo, el usuario puede adaptar el sistema a sus necesidades individuales, seleccionando de la inmensidad de elementos disponibles sólo aquellos que sean más relevantes para el material que será codificado. Además, estos conjuntos de etiquetas se basan en los sistemas de codificación existentes, pero no se limitan a ellos, pues están diseñados para ser comprensivos y extensibles.

En 2002 apareció la segunda versión de las *Guidelines* (TEI P4), que, aparte de la corrección de varios errores y omisiones, se diferenciaba de la TEI P3 en que estaba expresada en XML y ya no en SGML, aunque las DTD de las dos versiones seguían siendo compatibles. La versión más actual, la TEI P5, aparecida en 2005, ya no es compatible con la DTD de las versiones anteriores. Varios capítulos de las *Guidelines* fueron eliminados, otros fueron agregados y el documento completo fue sometido a revisión. El principal cambio estructural es que para la declaración del tipo de documento se utilizan paralelamente una DTD y un esquema en el lenguaje Relax NG cuya sintaxis se conforma a las normas de XML (Driscoll, 2004). Los fragmentos aquí presentados fueron simplificados y reajustados a las normas sintácticas de una DTD por cuestiones de continuidad sistemática.

El conjunto de elementos y atributos necesarios para la codificación de la lengua oral deben incluirse en la declaración de cada documento, haciendo referencia a las entidades paramétricas y secciones que los contienen. En (2.22) podemos ver cómo se incluyen

TEI.XML y TEI.spoken al esquema global de la TEI (tei2.dtd) que ya especifica las características básicas para cualquier tipo de texto:

```
(2.22) <?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE TEI PUBLIC "-//TEI P5//DTD Main Document Type//EN"
"tei2.dtd" [
  <!ENTITY % TEI.XML      'INCLUDE' >
  <!ENTITY % TEI.spoken  'INCLUDE' >
]>
```

De esta misma manera se pueden incluir también conjuntos adicionales de elementos y atributos útiles para la sincronización o segmentación de los elementos básicos.

Según el esquema de la TEI, presentado por Sperberg-McQueen y Burnard (2004: 10), un texto oral puede contener cualquiera de los siguientes componentes:

- Enunciados: <u>
- Pausas: <pause>
- Fenómenos vocalizados pero no léxicos, tales como las risas: <vocal>
- Fenómenos kinésicos (ni verbales, ni léxicos), tales como los gestos: <kinesic>
- Eventos completamente alingüísticos que ocurren durante el transcurso de la conversación y posiblemente influyen en ella: <events>
- Presentación oral de textos escritos, considerada como una clase especial de evento en la que se puede transcribir: <writing>
- Alteraciones o cambios en la calidad de la voz: <shift>

Veamos a continuación de manera más detallada la especificación de cada uno de los elementos. Un enunciado, <u>, se define como un fragmento de conversación precedido y seguido por un silencio o por un cambio de hablante. El enunciado puede contener producciones léxicas entrelazadas con pausas y sonidos vocálicos no léxicos. Además, durante un enunciado pueden suceder eventos no lingüísticos y se pueden presentar materiales escritos. Por lo tanto, el elemento <u> puede contener cualquiera de los otros elementos de la lista anterior entrelazados con unidades léxicas. Los otros elementos pueden aparecer entre los enunciados o uno al lado del otro, pero a excepción de <writing>, ninguno de los elementos contiene otros elementos o datos:

```
(2.23) <!ELEMENT u ( #PCDATA | pause | vocal | kinesic | event | writing
| shift )*>
<!ATTLIST u trans (smooth|latching|overlap|pause) "smooth"
who CDATA #IMPLIED
start CDATA #IMPLIED
end CDATA #IMPLIED
dur CDATA #IMPLIED >
```

El atributo específico de `<u>` es *trans* (transición), que indica la naturaleza de la transición entre un enunciado y el previo. El atributo *who*, que proporciona una identificación del hablante o grupo de hablantes previamente especificado en el encabezado, aparece en la lista de atributos de todos los elementos del conjunto `TEI.spoken`.

Los atributos *start*, *end* y *dur* también forman parte de la lista de atributos de los elementos `<pause>`, `<vocal>`, `<kinesic>` y `<event>`. Por medio de estos atributos se puede registrar los valores abstractos o relativos de la posición del elemento en la línea del tiempo.

Los elementos `<pause>` pueden aparecer entre enunciados o dentro de ellos:

```
(2.24) <!ELEMENT pause EMPTY>
<!ATTLIST pause type CDATA #IMPLIED
           who CDATA #IMPLIED
           start CDATA #IMPLIED
           end CDATA #IMPLIED
           dur CDATA #IMPLIED >
```

En (2.24) podemos ver que en la lista de atributos de `<pause>` también se encuentran *start*, *end* y *dur*. Su atributo específico es *type* que sirve para categorizar de algún modo los distintos tipos de pausas.

Con `<vocal>` se codifican todos los fenómenos vocalizados pero no necesariamente léxicos, por ejemplo, pausas vocalizadas, reacciones aléxicas, etc.

```
(2.25) <!ELEMENT vocal EMPTY>
<!ATTLIST vocal who CDATA #IMPLIED
           iterated (true|false|unknown) "false"
           desc CDATA #IMPLIED
           start CDATA #IMPLIED
           end CDATA #IMPLIED
           dur CDATA #IMPLIED >
```

Además de los atributos explicados anteriormente, `<vocal>` (al igual que `<kinesic>` y `<event>`) tiene dos atributos especiales: *iterated*, que indica si el fenómeno se repite o no, y *desc*, por medio del cual se puede proporcionar una descripción convencional para el fenómeno.

El elemento `<kinesic>` se utiliza para representar todo fenómeno comunicativo que no necesariamente sea vocalizado, por ejemplo, un gesto, un fruncimiento de ceño, etc.

```
(2.26) <!ELEMENT kinesic EMPTY>
<!ATTLIST kinesic who CDATA #IMPLIED
           iterated (true|false|unknown) "false"
           desc CDATA #IMPLIED
           start CDATA #IMPLIED
           end CDATA #IMPLIED
           dur CDATA #IMPLIED >
```

Por medio de `<event>` se codifica cualquier fenómeno o suceso no necesariamente vocalizado o comunicativo, por ejemplo, ruidos incidentales u otros eventos que afecten la comunicación.

(2.27)

```
<!ELEMENT event EMPTY>
<!ATTLIST event who CDATA #IMPLIED
                 iterated (true|false|unknown) "false"
                 desc CDATA #IMPLIED
                 start CDATA #IMPLIED
                 end CDATA #IMPLIED
                 dur CDATA #IMPLIED >
```

El elemento `<writing>` se utiliza cuando se revela a los participantes un pasaje de un texto escrito durante el transcurso de un texto oral.

(2.28)

```
<!ELEMENT writing (#PCDATA)>
<!ATTLIST writing who CDATA #IMPLIED
                 type CDATA #IMPLIED
                 script CDATA #IMPLIED
                 gradual (true|false|unknown) #IMPLIED >
```

Los dos atributos que no habían formado parte de la lista de atributos de ninguno de los elementos anteriores son *script*, por medio del cual se apunta a una cita bibliográfica en el encabezado donde se hace una descripción completa de la fuente del escrito, y *gradual*, que indica si el escrito se revela en su totalidad de una sola vez o de manera gradual.

Por último, `<shift>` señala el punto en el cual alguno de los hablantes cambia un rasgo paralingüístico de una serie de enunciados.

(2.29)

```
<!ELEMENT shift EMPTY>
<!ATTLIST shift who CDATA #IMPLIED
                 feature (tempo|loud|pitch|tension|rhythm|voice)
                       #IMPLIED
                 new CDATA "normal" >
```

Por medio del atributo *feature* se puede especificar el rasgo paralingüístico que se altera: *tempo* (velocidad), *loud* (volumen), *pitch* (altura tonal), *tension* (tensión), *rhythm* (ritmo) o *voice* (calidad de la voz). En el valor del atributo *new* se indica el nuevo estado del rasgo paralingüístico especificado por medio de un conjunto semicerrado de categorías.

Utilizando todos estos elementos juntos, obtendríamos la siguiente estructura jerárquica de nuestro ejemplo introducido en (2.1):

(2.30)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI PUBLIC "-//TEI P5//DTD Main Document Type//EN" "tei2.dtd"
[
  <!ENTITY % TEI.XML 'INCLUDE' >
  <!ENTITY % TEI.spoken 'INCLUDE' >
]>
<TEI.2>
<teiHeader>
  <fileDesc/>
  <profileDesc>
    <particDesc>
      <person id="ABC"/>
      <person id="DEF"/>
    </particDesc>
  </profileDesc>
</teiHeader>
<text>
  <u who="ABC"><vocal desc="vocalized pause"/>este +</u>
```

```

<u who="ABC">busco el cargador para esta batería //</u>
<pause who="ABC" type="short"/>#
<u who="DEF">cámara &lt; digital &gt; ?</u>
<pause who="DEF" type="short"/>#
<u who="ABC" trans="overlap">&lt; de la &gt; Cyber Shot //</u>
<u who="ABC">sí //</u>
<u who="DEF">esta vaya aquí a la Casa de la Pila //</u>
</text>
</TEI.2>

```

Un texto codificado en XML está concebido generalmente como una jerarquía, es decir, cada elemento del documento contiene otros componentes de una manera estrictamente jerárquica sin solapamientos. A pesar de ser útil y poderoso, este modelo no representa todas las estructuras que contienen los documentos de interés para los usuarios. Algunas de estas estructuras son los textos paralelos (tales como un texto y su traducción a otro idioma) con estructuras similares que necesitaran alinearse y la correspondencia temporal de dos textos orales producidos al mismo tiempo.

La representación de múltiples niveles paralelos se puede hacer de distintas maneras. Una de ellas es por medio del uso de mojones (*milestones*) o marcas. Estas marcas son señales coordinadas que pueden colocarse en el texto y pueden ser referidas por medio de apuntadores contenidas en las estructuras separadas. Dado que el número de apuntadores a una misma marca es ilimitado, se puede organizar varios niveles paralelos de este modo. Por ejemplo, es posible representar análisis paralelos de los rasgos de un texto utilizando varias clases distintas de apuntadores que hagan referencia a un conjunto de coordenadas en el texto. Las estructuras analíticas no necesitan estar insertadas en el texto principal, sino que pueden ser colocadas fuera de él. (Barnard y Ide 1995)

Para dar cuenta de los textos paralelos, debemos agregar un conjunto más de elementos a la declaración presentada en (2.22), `<!ENTITY % TEI.linking 'INCLUDE' >`, que incluye lo siguiente:

```

(2.31) <!ELEMENT timeline (when+)>
<!ATTLIST timeline origin (IDREF|CDATA) #IMPLIED
          unit CDATA #IMPLIED
          interval CDATA #IMPLIED>
<!ELEMENT when EMPTY>
<!ATTLIST when id ID #REQUIRED
          synch IDREF #IMPLIED
          absolute CDATA #IMPLIED
          unit CDATA #IMPLIED
          interval CDATA #IMPLIED
          since IDREF #IMPLIED>
<!ELEMENT anchor EMPTY>
<!ATTLIST anchor id ID #IMPLIED
          synch IDREF #IMPLIED>

```

El elemento `<timeline>` agrupa y especifica una serie de elementos `<when>` a los que se les asigna un identificador único, *id*, y, entre otras cosas, el valor absoluto de tiempo, *absolute*,

en alguna unidad específica, por ejemplo, *unit="seconds"*. Finalmente, en la posición indicada dentro del texto se irán colocando las marcas, *<anchor>*, que bien pueden recibir un identificador o que especificarán una sincronización, *synch*, haciendo referencia directa a los identificadores. Finalmente, nuestro ejemplo (cfr. ejemplos B8 y B21) se vería entonces así:

(2.32)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI PUBLIC "-//TEI P5//DTD Main Document Type//EN" "tei2.dtd" [
  <!ENTITY % TEI.XML 'INCLUDE' >
  <!ENTITY % TEI.spoken 'INCLUDE' >
  <!ENTITY % TEI.linking 'INCLUDE' >
]>
<TEI.2>
<teiHeader>
  <profileDesc>
    <particDesc>
      <person id="ABC"/>
      <person id="DEF"/>
    </particDesc>
  </profileDesc>
</teiHeader>
<text>
  <timeline>
    <when id="T0" absolute="0.0" unit="seconds"/>
    <when id="T1" absolute="0.17162"/>
    <when id="T2" absolute="0.53109"/>
    <when id="T3" absolute="2.87652"/>
    <when id="T4" absolute="3.19514"/>
    <when id="T5" absolute="3.36572"/>
    <when id="T6" absolute="3.89012"/>
    <when id="T7" absolute="4.32621"/>
    <when id="T8" absolute="4.61497"/>
    <when id="T9" absolute="4.80944"/>
    <when id="T10" absolute="6.47717"/>
    <when id="T11" absolute="6.64290"/>
  </timeline>
  <u who="#ABC" start="#T1" end="#T2"><vocal desc="vocalized pause"/>este +</u>
  <u who="#ABC" start="#T2" end="#T3">busco el cargador para esta batería //
  </u>
  <pause who="#ABC" type="short" start="#T3" end="#T4"/>#
  <u who="#DEF" start="#T4" end="#T6">cámara <anchor synch="#T5"/>
  &lt; digital &gt;<anchor synch="#T6"/> ?</u>
  <pause who="#DEF" type="short" start="#T6" end="#T7"/>#
  <u who="#ABC" trans="overlap" start="#T5" end="#T7"><anchor synch="#T5"/>
  &lt; de la &gt;<anchor synch="#T6"/> Cyber Shot //</u>
  <u who="#ABC" start="#T7" end="#T8">sí //</u>
  <u who="#DEF" start="#T9" end="#T10">esta vaya aquí a la Casa de la Pila //</u>
</text>
</TEI.2>
```

Sin embargo, a pesar de las posibilidades de captura detallada de datos sobre fenómenos vocálicos, kinésicos, paralingüísticos y lingüísticos, dentro de una estructura que permite un orden jerárquico y temporal, no deja de parecer curiosa la falta de un conjunto estricto de etiquetas para la codificación de información detallada no sólo entonativa (cfr. Sperberg-McQueen y Burnard 2004: 10.3.4), sino también de otros rasgos fonéticos abstractos al nivel de la sílaba (cfr. 1.1.1). Además, un sistema que permite a cada usuario hacer su propia ampliación del inventario, podría provocar fácilmente la creación de ‘dialectos’ ininteligibles entre sí, que iría en contra del objetivo original de la estandarización.

2.3.1.3 Resumen

En los últimos años, el gran desarrollo de la tecnología de la información parece abrir cada vez más posibilidades a los investigadores del lenguaje. Las aplicaciones de los lenguajes de etiquetas, en especial el XML, permiten la integración de textos y multimedia de maneras antes inimaginables. Sin embargo, la creciente oferta hace difícil mantener el paso, sobre todo cuando distintos enfoques teóricos se enfrentan entre sí y se desarrollan intencionalmente en direcciones opuestas.

A pesar de que los sistemas de transcripción de la lengua con estructura temporal parecen a primera vista incompatibles con los sistemas jerárquicos, Schmidt (2005) demuestra que el acoplamiento de las dos estructuras es posible y que ofrece nuevos retos. Estos retos suelen someterse a las necesidades específicas de cada investigación, y de cada equipo de trabajo dependerá que se avance en dirección del aislamiento o en dirección de la estandarización.

En las siguientes páginas se presentará mi propuesta personal de transcripción y análisis de la prosodia de español oral, que se sustenta en los modelos, unidades estructurales y áreas de estudio de la prosodia descritos en la Sección 1. Este sistema tiene una estructura jerárquica y ha comprobado ser lo suficientemente complejo y flexible como para satisfacer completamente las necesidades de la investigación que se desarrollará en la Sección 3. Cada uno de los distintos pasos en el desarrollo del sistema tuvieron distintas influencias, así como ideas originales que lo hacen parcialmente incompatible con sistemas individuales como, por ejemplo, TEI o EXMARaLDA. De cualquier modo, los dos sistemas anteriores son relativamente ampliables y, siguiendo el ejemplo de Schmidt (2005), tras haber comprobado dentro del marco de esta investigación la utilidad de mis propuestas, se podría intentar hacer modificaciones a las estructuras e integrarlas armónicamente.

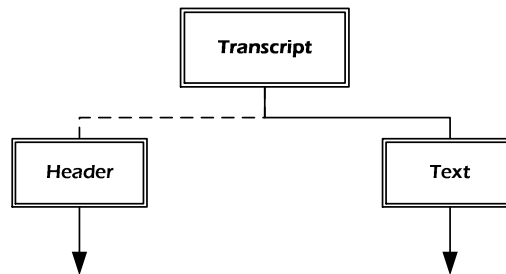
2.3.2 Propuesta de estructura jerárquica prosódica para documentos de XML

En esta sección se hará una presentación de una estructura jerárquica prosódica que nos permitirá tanto el señalamiento de los fenómenos prosódicos dentro del texto como su especificación por medio de valores acústicos reales. De manera paralela, se irá presentando la definición del tipo de documento (DTD) correspondiente a la estructura abstracta.

A esta nueva DTD le daremos el nombre de `PROSODY_XML.DTD`. Basándonos en el documento `Transcript.dtd`, que se presentó íntegramente en la Ilustración 2-4 y en el ejemplo (2.7), en esta nueva estructura el elemento raíz también se llamará `<Transcript>`. La primera diferencia con respecto a nuestro ejemplo simplificado, será que el elemento `<Transcript>`

constará de dos subelementos, encabezado, <Header>, y texto, <Text>, como se muestra en la siguiente Ilustración:

Ilustración 2-6: Elemento <Transcript>



Fuente: Diseño propio

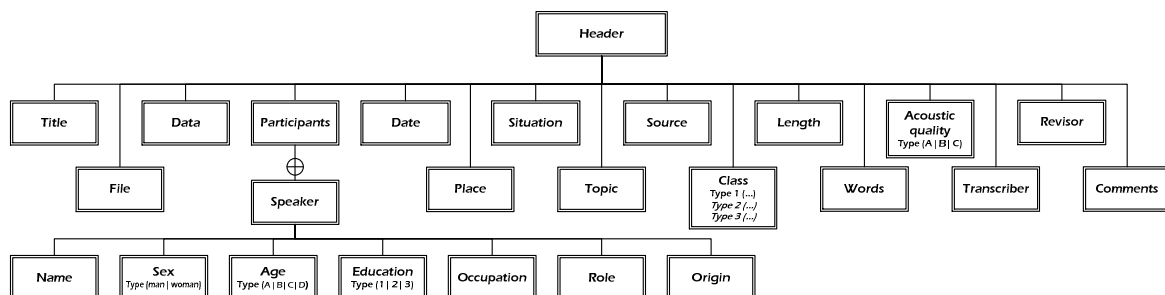
Sin embargo, la línea punteada hacia el elemento <Header> indica que la ruta es opcional, para permitir la existencia de textos con o sin encabezado. Dentro de <Header> se podrá almacenar la información necesaria para organizar y archivar los documentos de XML, aunque habrá casos en los que seguramente sea posible prescindir de esta información, por ejemplo, cuando se trabaje con fragmentos de distintas grabaciones. Esta estructura principal se representa en la DTD de la siguiente manera:

```
(2.33) <!-- Root element -->
<!ELEMENT Transcript (Header?,Text)>
```

Por medio del indicador de frecuencia ‘?’ se especifica que el elemento <Header> puede aparecer o no, y en caso de hacerlo, irá seguido del elemento <Text>.

A su vez, <Header> tendrá la siguiente estructura:

Ilustración 2-7: Elemento <Header>



Fuente: Diseño propio

Todos los subelementos de <Header> son hojas, excepto <Participants>, dado que no tienen más elementos subordinados y su contenido consta de datos de caracteres. Los elementos <Class> y <Acoustic_quality> tienen atributos. En letra de estilo regular se indicará de aquí en adelante los atributos necesarios y en letra cursiva los atributos opcionales. El atributo *Type* de <Acoustic_quality> tiene tres valores posibles predefinidos

(A, B o C), mientras que el atributo `Type1` de `<Class>` tendrá un valor mixto: a base de datos de caracteres o los dos valores predefinidos `informal` y `formal`, que, por cuestiones de espacio, sólo se representan en (2.34). `Type2` y `Type3` también tienen valores predefinidos.

El elemento `<Participants>` puede tener uno o más elementos `<Speaker>`, los cuales tienen a su vez una serie de subelementos, y algunos de ellos tienen sus propios atributos. Con estos elementos se podrá proveer un mínimo de información por cada participante de la conversación. Las características de dicha información se detallarán más adelante.

La estructura de `<Header>` se representa en la DTD de este modo:

```
(2.34) <!-- Declaration of Header-relevant tags -->
<!ELEMENT Header      (Title|File|Data|Participants|Date|Place|Situation|
                        Topic|Source|Class+|Length|Words|Acoustic_quality|
                        Transcriber|Revisor|Comments)*>

<!ELEMENT Title       (#PCDATA)>
<!ELEMENT File        (#PCDATA)>
<!ELEMENT Data        (#PCDATA)>
<!ELEMENT Participants (Speaker+)>
  <!ELEMENT Speaker   (#PCDATA|Name|Sex|Age|Education|Occupation|
                        Role|Origin)*>
    <!ELEMENT Name     (#PCDATA)>
    <!ELEMENT Sex      (#PCDATA)>
    <!ATTLIST Sex       Type   (woman|man|x)          #REQUIRED >
    <!ELEMENT Age      (#PCDATA)>
    <!ATTLIST Age       Type   (A|B|C|D|x)          #REQUIRED >
    <!ELEMENT Education (#PCDATA)>
    <!ATTLIST Education Type   (1|2|3|x)          #REQUIRED >
    <!ELEMENT Occupation (#PCDATA)>
    <!ELEMENT Role      (#PCDATA)>
    <!ELEMENT Origin    (#PCDATA)>
  <!ELEMENT Date      (#PCDATA)>
  <!ELEMENT Place     (#PCDATA)>
  <!ELEMENT Situation (#PCDATA)>
  <!ELEMENT Topic     (#PCDATA)>
  <!ELEMENT Source    (#PCDATA)>
  <!ELEMENT Class     (#PCDATA)>
  <!ATTLIST Class Type1  (CDATA|informal|formal) #REQUIRED >
  <!ATTLIST Class Type2  (familiar-private|public|
                        formal_in_natural_context|media|telephone)
                        #IMPLIED >
  <!ATTLIST Class Type3  (monologue|dialogue|conversation) #IMPLIED >
  <!ELEMENT Length    (#PCDATA)>
  <!ELEMENT Words     (#PCDATA)>
  <!ELEMENT Acoustic_quality (#PCDATA)>
  <!ATTLIST Acoustic_quality Type (A|B|C)          #REQUIRED >
  <!ELEMENT Transcriber (#PCDATA)>
  <!ELEMENT Revisor   (#PCDATA)>
  <!ELEMENT Comments  (#PCDATA)>
```

En este ejemplo se puede ver que los subelementos de `<Speaker>` con atributos, como `<Sex>`, `<Age>` y `<Education>`, no son elementos vacíos. Esto podría parecer redundante, puesto que una vez que se haya provisto un valor de atributo no hay necesidad de agregar contenido a base de datos de caracteres (`#PCDATA`). No obstante, esta medida es útil en el caso de que, por alguna razón, se decida proveer esta información como valor de atributo y además como

texto, o simplemente por si hubiera un error tipográfico que impidiera la localización de la información.

Por la misma razón, en el caso de los atributos *Type1*, *Type2* y *Type3* de `<Class>`, podemos ver que habría tres opciones para proveer la información: como el valor de cada atributo; como datos de caracteres sólo en el valor del atributo *Type1* (gracias a `CDATA`), y como datos de caracteres complementarios en el contenido dentro de las etiquetas `<Class></Class>`. Para que estas opciones de introducción de los datos queden claras, véase los valores de los atributos, escritos en cursiva, y los contenidos a base de datos de caracteres, escritos en negrita, según se muestra en el siguiente ejemplo:

```
(2.35) <Header>
<Title>Las repcionistas</Title>
<File>df11-2004</File>
<Data>hotel_1, df1-04-1</Data>
<Participants>
  <Speaker>
    EDU, <Name>EDU</Name>
    (<Sex Type="man">man</Sex>,
    <Age Type="B">B</Age>,
    <Education Type="3">3</Education>,
    <Occupation>linguist</Occupation>,
    <Role>participant</Role>,
    <Origin>Guadalajara, MX</Origin>)
  </Speaker>
  <Speaker>
    ADR, <Name>ADR</Name>
    (<Sex Type="woman">woman</Sex>,
    <Age Type="A">A</Age>,
    <Education Type="3">3</Education>,
    <Occupation>hotel receptionist</Occupation>,
    <Role>participant</Role>,
    <Origin>Mexico City</Origin>)
  </Speaker>
  <Speaker>
    MIR, <Name>MIR</Name>
    (<Sex Type="woman">woman</Sex>,
    <Age Type="A">A</Age>,
    <Education Type="3">3</Education>,
    <Occupation>hotel receptionist</Occupation>,
    <Role>participant</Role>,
    <Origin>Mexico City</Origin>)
  </Speaker>
</Participants>
<Date>05/08/2004</Date>
<Place>Mexico City</Place>
<Situation>Free-style interview at hotel reception, not hidden,
researcher interviewer</Situation>
<Topic>Mexico City's sights, everyday life and work</Topic>
<Source>DF-2004</Source>
<Class Type1="informal" Type2="public" Type3="conversation">
informal, public, conversation</Class>
<Length>1585.830022675737</Length>
<Words>2867</Words>
<Acoustic_quality Type="B">B</Acoustic_quality>
<Transcriber>Eduardo</Transcriber>
<Revisor>Eduardo; Eduardo (prosody)</Revisor>
<Comments/>
</Header>
```

Ahora, basándonos en este mismo ejemplo, vayamos viendo cuáles son las características que tendrá la información de cada uno de los elementos del encabezado. <Title> estará redactado en español y tendrá pocas palabras que ayudarán a reconocer la grabación. <File> indicará el nombre único o más característico con el que se ha clasificado los archivos, mientras que <Data> contendrá otros nombres utilizados para los archivos en formatos .wav, .TextGrid o .mp3 correspondientes a la misma grabación.

<Participants> contendrá los subelementos <Speaker>, correspondientes al número de participantes que se desee analizar en la grabación. En caso de que se desconozca algún dato correspondiente a los subelementos de <Speaker>, este será sustituido por x. En el caso de <Name>, dentro del cual normalmente se colocará un código formado por tres letras mayúsculas extraídas del nombre, seguido del nombre propio o por el mismo código, cuando el dato se desconozca, dicho código será sustituido por XYZ, ABC o alguna otra combinación que se prefiera.

El atributo `Type` de <Sex> tendrá `woman` y `man` como valores predeterminados. Por otra parte, los valores del atributo `Type` de <Age> corresponden a 4 distintos rangos de edad: `A` para los hablantes entre 18 y 25 años, `B` para aquellos entre 26 y 40, `C` para quienes tengan entre 41 y 60, y `D` para los mayores de 60. Por último, el valor `1` de `Type` dentro del elemento <Education> corresponderá a la educación básica o nula, `2` para aquellos con educación media y `3` para los estudiantes o graduados universitarios.

Para terminar con los subelementos de <Speaker>, dentro de <Occupation> se hará referencia a la actividad profesional, empleo u ocupación del participante; en <Role> se dará a conocer el rol comunicativo que jugó cierta persona en la conversación (`participant`, `interviewer`, etc.) y, para finalizar, dentro de <Origin> se introducirá la ciudad o región de procedencia del hablante.

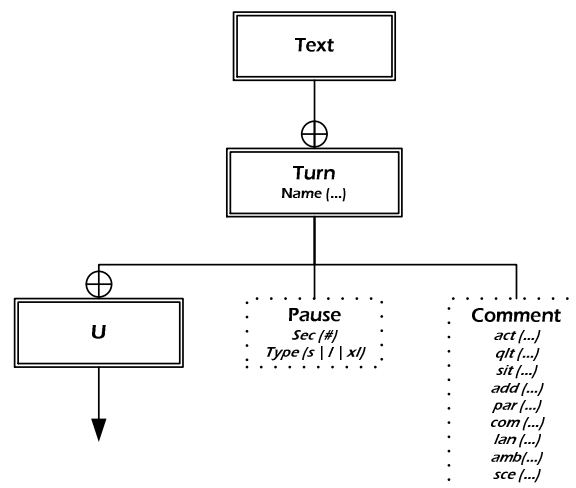
La ubicación espacio-temporal se manifestará por medio de <Date>, donde se colocará la fecha de grabación, y de <Place>, donde se indicará el lugar geográfico en el que se realizó la conversación, mientras que con <Situation> se podrá introducir información sobre la ubicación, el ambiente y la situación comunicativa, y con <Topic> se podrán incluir los temas y subtemas tratados en el texto.

El origen o pertenencia de la grabación a un *corpus* determinado se indicará con <Source> y se podrá clasificar con los tres atributos de <Class>. Con <Length> se podrá especificar la duración de la grabación en segundos, y con <Words>, el número de palabras sin contar fragmentos o fenómenos no lingüísticos. La calidad de la grabación puede ser evaluada como `A`: muy buena, `B`: buena o `C`: suficientemente buena.

Finalmente, con `<Transcriber>` se dará a conocer el nombre del transcriptor responsable del texto, con `<Revisor>` se puede introducir los nombres de revisores y transcriptores de la segunda fase, y se podrán agregar comentarios generales con el elemento `<Comments>`.

Por otra parte, en la Ilustración 2-6, tenemos también al elemento `<Text>` como descendiente directo de `<Transcript>` pero que, comparado con `<Header>`, es absolutamente necesario. En la Ilustración 2-8 se señala que el único descendiente directo de `<Text>` es `<Turn>`, que tiene un indicador de frecuencia que especifica que debe aparecer una o más veces y que tiene el atributo `Name`, que es un atributo obligatorio que admite cualquier tipo de datos de caracteres, lo cual también nos permitirá introducir los códigos a base de tres mayúsculas asignados a los nombres de cada participante de la conversación.

Ilustración 2-8: Elementos `<Text>` y `<Turn>`



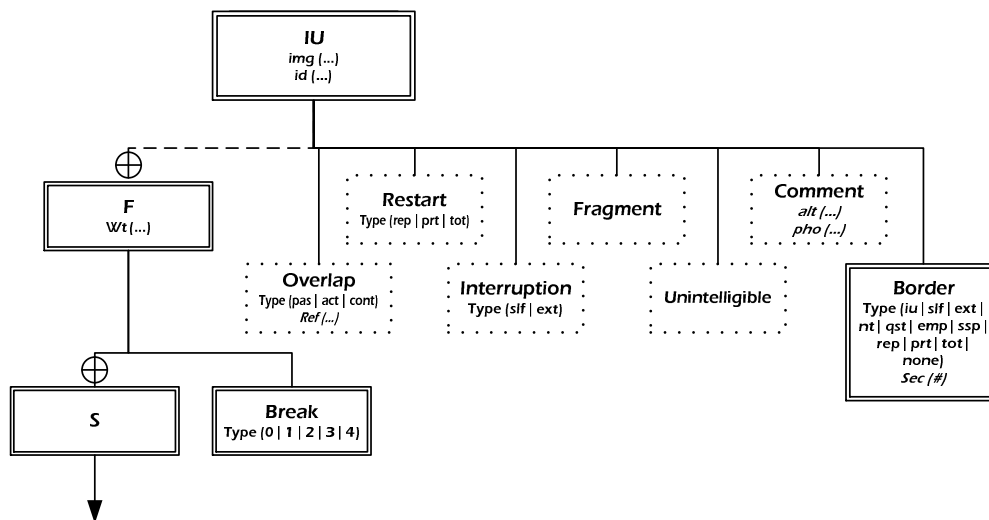
Fuente: Diseño propio

A su vez, cada `<Turn>` tendrá necesariamente un subelemento `<U>` (cfr. `<Utterance>` en la Ilustración 2-4) que debe aparecer una o más veces y, además, puede contener elementos `<Pause>` y `<Comment>`. El elemento `<Pause>` concuerda casi completamente con el elemento descrito en `Transcript.dtd` (Ilustración 2-4), excepto por la abreviación de los valores a `s`, `l` y `xl` (cfr. `short`, `long` y `xlong`). Por otra parte, los atributos opcionales de `<Comment>` servirán para clasificar distintos fenómenos o eventos ocurridos durante la conversación y que se consideren relevantes, por ejemplo: acciones de los interlocutores que se consideren independientes del texto (`act`), alteraciones importantes de la calidad de la voz (`qlt`), situaciones o sucesos relevantes en la comunicación (`sit`), cambio de interlocutor a quien se dirige el texto del turno de palabra (`add`), gestos o elementos paralingüísticos del hablante (`par`), comentarios del transcriptor (`com`), lengua alternativa (`lan`), y, por último, para el caso

Cada elemento <U> tendrá dos atributos opcionales y tres necesarios. *Type* puede ser utilizado para definir la categoría sintáctica del enunciado: declarativo o neutro (*nt*), interrogativo (*qst*), exclamativo o enfático (*emp*), suspendido (*ssp*), interrumpido por el interlocutor (*ext*) o autointerrumpido (*slf*). En *Subtype* se puede introducir directamente algún otro tipo de categorías. Los atributos necesarios *Quest*, *Emph* y *Susp* hacen referencia a los rasgos fonológicos utilizados por Cantero (2002: 135-143; cfr. 1.2.5.6) para caracterizar los distintos tonemas del castellano. En este fragmento de la DTD ya no es necesario repetir la declaración de <Pause>.

A continuación veremos en la Ilustración 2-10 la estructura de la unidad entonativa, <IU>, que es el elemento más complejo pues es en este nivel de la jerarquía en el que se incluirá, a manera de subelementos, una gran variedad de fenómenos prosódicos.

Ilustración 2-10: Elementos <IU> y <F>



Fuente: Diseño propio

Los atributos *img* (*image*) e *id* (*identity number*) serán utilizados para asignar la clave única de cada unidad entonativa para que se pueda enlazar el texto con sus respectivos archivos de audio e imágenes, y que faciliten la selección y análisis de cada frase.

En realidad, todos los subelementos de <IU>, excepto <Border>, son opcionales, sobre todo con la finalidad de permitir documentos de XML con distintos niveles de detalle.

Con <Overlap> se deberá indicar el inicio y el final de los fragmentos producidos simultáneamente, de manera pasiva (*pas*), activa (*act*) o si alguno de los textos se extiende a lo largo de más de un turno (*cont*). También se podrá asignar un número o nombre de referencia en caso de que haya varios solapamientos en el texto, por medio de *Ref*.

Los valores de atributo del elemento `<Restart>` corresponden a la repetición (*rep*), reformulación parcial (*prt*) o total (*tot*); el elemento `<Unintelligible>` señalará los fragmentos ininteligibles y codificados con *xxx* en el texto de la transcripción.

Aquí es donde se ubica el elemento `<Interruption>` que ya se había presentado en la Ilustración 2-4 y el ejemplo (2.7). Con `<Fragment>` se marcarán las producciones fragmentadas y con `<Comment>` se indicarán las propuestas para errores de producción y variantes no estándar de difícil identificación (*alt*), así como fenómenos fonéticos o fonológicos de interés (*pho*). Estos dos nuevos atributos aparecerán únicamente en este nivel.

`<Border>` es un elemento con el cual se podrá aprovechar los signos gráficos permitidos al final de cada enunciado, según las normas de transcripción que se presentarán más adelante. Los valores del atributo *Type*, corresponden a los dos distintos tipos de interrupción, al límite de unidad entonativa y a los diferentes tipos de enunciados (cfr. `<U>`). En caso de ser necesario para el análisis, se podrá suministrar la duración absoluta del intervalo entre dos producciones lingüísticas de un mismo hablante, por medio del atributo *Sec*.

Las líneas punteadas que nos llevan a `<F>` (*foot*) significan que no será absolutamente necesario introducir la información sobre la estructura rítmica o silábica, por ejemplo, cuando se trate de una transcripción básica sin separación de sílabas.

El fragmento de DTD correspondiente a esta estructura tendrá el siguiente formato:

```
(2.38) <!-- IU = Intonation Unit -->
<!ELEMENT IU (#PCDATA|F|S|Overlap|Restart|Break|Fragment|
             Unintelligible|Interruption|Comment|Border)*>
<!ATTLIST IU img      CDATA          #IMPLIED
             id       CDATA          #IMPLIED>

<!ELEMENT Overlap (#PCDATA|F|S|Fragment|Mora|Comment|Unintelligible)*>
<!ATTLIST Overlap  Type (pas|act|cont) #IMPLIED
             Ref     CDATA          #IMPLIED>

<!ELEMENT Restart (#PCDATA)>
<!ATTLIST Restart  Type (rep|prt|tot) #REQUIRED>
<!ELEMENT Fragment (#PCDATA)>
<!ELEMENT Unintelligible (#PCDATA)>
<!ELEMENT Interruption EMPTY>
<!ATTLIST Interruption Type (slf|ext) #REQUIRED>
<!-- Tone-Unit-level comments -->
<!ATTLIST Comment  alt    CDATA          #IMPLIED
             pho    CDATA          #IMPLIED>

<!ELEMENT Border (#PCDATA)>
<!ATTLIST Border   Type (slf|ext|iu|nt|qst|emp|ssp|rep|prt|tot|none)
             #REQUIRED
             Sec    CDATA          #IMPLIED>
```

En caso de que se incluya la estructura rítmica de los grupos fónicos o de los pies métricos, se podrá echar mano del atributo *wt* (*weight*) de `<F>`, con el cual podrá especificarse el peso de las sílabas con la notación que se prefiera, dado que admite todo tipo de datos de caracteres. Al nivel de la sílaba se podrá echar mano del elemento `<Break>`, cuyos valores corresponden a las características descritas en el inventario del Sp-ToBI:

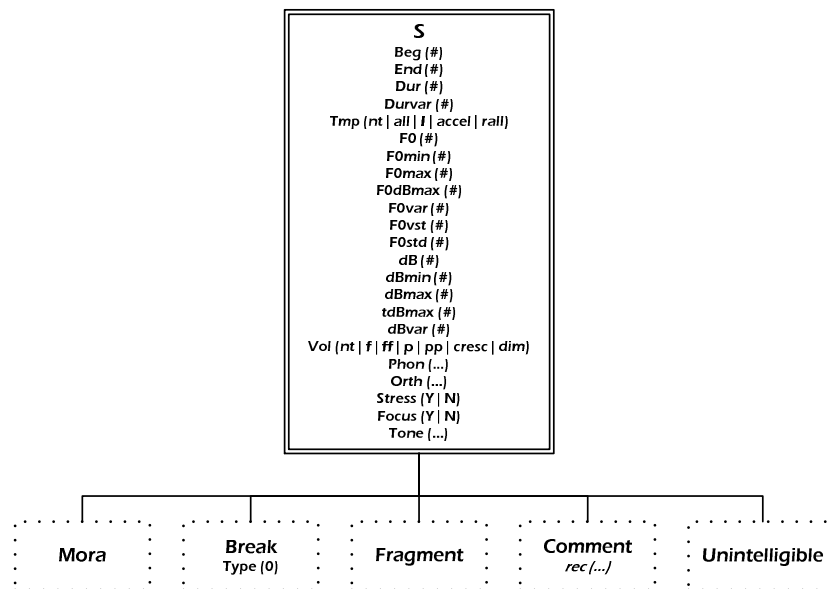
(2.39)

```
<!-- F = Foot -->
<!ELEMENT F (#PCDATA|S|Break|Overlap)*>
<!ATTLIST F Wt CDATA #REQUIRED>

<!ELEMENT Break EMPTY>
<!ATTLIST Break Type (0|1|2|3|4) #REQUIRED>
```

Finalmente, en la Ilustración 2-11 se podrá apreciar la serie de atributos del elemento <S> (*syllable*) y sus elementos subordinados. Cada <S> tiene una serie muy importante de atributos que darán cuenta del papel fundamental de la sílaba en el análisis acústico, por ejemplo: tiempo de inicio y final, duración, variación relativa y tempo (*Beg*, *End*, *Dur*, *Durvar*, *Tmp*); su frecuencia fundamental, valor mínimo y máximo, valor de la frecuencia fundamental en el máximo de intensidad, variación relativa de frecuencia tanto en porcentaje como en semitonos, y frecuencia estandarizada (cfr. Cantero 2002: 147-153) (*F0*, *F0min*, *F0max*, *F0dBmax*, *F0var*, *F0vst*, *F0std*); la intensidad promedio, mínima y máxima, tiempo en el que la intensidad alcanza su máximo, variación relativa y volumen (*dB*, *dBmin*, *dBmax*, *tdBmax*, *dBvar*, *Vol*), así como representación fonética y ortográfica, la presencia de prominencia acentual, focalización y tono, preferentemente según el inventario del Sp-ToBI (*Phon*, *Orth*, *Stress*, *Focus*, *Tone*).

Ilustración 2-11: Elemento <S>



Fuente: Diseño propio

El subelemento <Mora> carecerá de contenido y con el se señalarán la frontera entre una mora y otra en los casos en los que una sílaba pueda o deba analizarse como compuesta por dos o incluso tres morae. Cada mora será analizada como una sílaba común y corriente, por eso es necesario al menos destacar las fronteras. Los elementos <Break> y <Comment> ya habían sido declarados previamente, pero con el valor 0 del atributo *Type* de <Break> se podrá

señalar los límites morfológicos de dos sílabas dentro de un hiato, y con el nuevo atributo *rec* de `<Comment>` se podrá sugerir la reconstrucción normativa de una producción lingüística dentro de los límites de una sílaba. Los elementos `<Fragment>` y `<Unintelligible>`, que también son subelementos de `<IU>`, no necesitan ser declarados nuevamente en el siguiente y último fragmento de la DTD:

```
(2.40) <!-- S = Syllable -->
<!ELEMENT S (#PCDATA|Mora|Break|Comment|Fragment|Unintelligible|Overlap) *>
<!ATTLIST S
  Beg          CDATA          #REQUIRED
  End          CDATA          #REQUIRED
  Dur          CDATA          #REQUIRED
  Durvar       CDATA          #REQUIRED
  Tmp (nt|all|l|accel|rall) #REQUIRED
  F0min        CDATA          #REQUIRED
  F0max        CDATA          #REQUIRED
  F0dBmax      CDATA          #REQUIRED
  F0           CDATA          #REQUIRED
  F0var        CDATA          #REQUIRED
  F0vst        CDATA          #REQUIRED
  F0std        CDATA          #REQUIRED
  dBmin        CDATA          #REQUIRED
  dBmax        CDATA          #REQUIRED
  tdBmax       CDATA          #REQUIRED
  dB           CDATA          #REQUIRED
  dBvar        CDATA          #REQUIRED
  Vol (nt|f|ff|p|pp|cresc|dim) #REQUIRED
  Phon         CDATA          #REQUIRED
  Orth         CDATA          #REQUIRED
  Stress       (Y|N)          #REQUIRED
  Focus        (Y|N)          #REQUIRED
  Tone         CDATA          #REQUIRED>
<!-- Spanish-ToBI: L*+H, L+H*, H+L*, H*, *, L*+!H, L+!H*, !L*+H, !L+H*,
L%, M%, H% - upstep must be marked with upside-down exclamation mark!-->

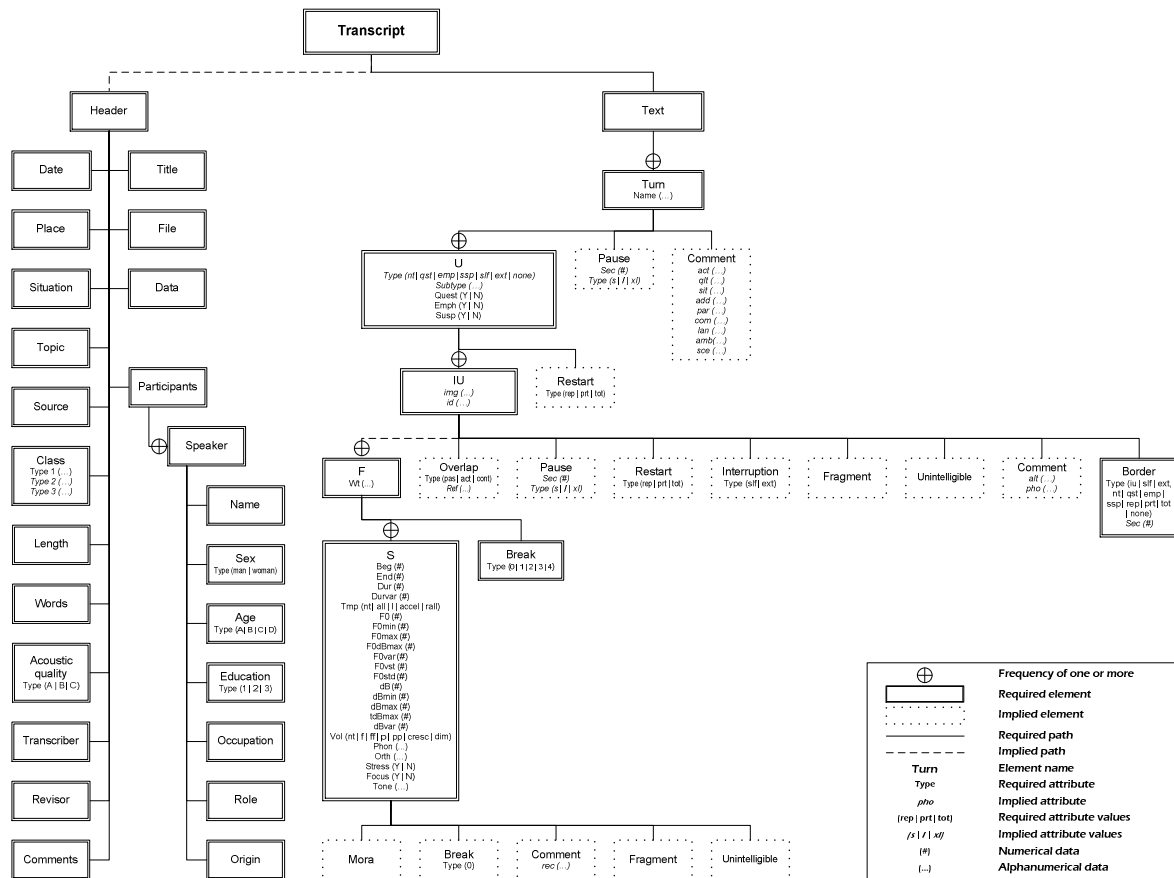
<!ELEMENT Mora      EMPTY>
<!ATTLIST Break     Type      (0)          #IMPLIED>
<!-- Syllable-level comments -->
<!ATTLIST Comment   rec       CDATA          #IMPLIED>
```

Los valores de los atributos de *Vol* y *Tmp*, provenientes de las convenciones de notación musical, también son utilizadas en la TEI (Sperberg-McQueen y Burnard 2004: 10.2.6) y en el sistema de transcripción GAT (Selting 1995: X-XI; Selting *et al.* 1998: 91-122).

Si, ahora que se han presentado todos los elementos que componen la estructura jerárquica subyacente de la DTD a la que dimos el nombre de `PROSODY_XML.DTD` (véase el documento completo en el Anexo 2), procediéramos a juntar las piezas en una sola estructura jerárquica global, obtendríamos algo comparable a la Ilustración 2-12.

Ahora bien, esta estructura, aunque puede aplicarse sin problemas a conjuntos de fragmentos pertenecientes a una misma o a distintas grabaciones, puede simplificarse según se muestra en la Ilustración 2-13 (Anexo 3).

Ilustración 2-12: Estructura jerárquica de PROSODY_XML.DTD



Fuente: Diseño propio

Las únicas tres diferencias entre la PROSODY_XML.DTD y la nueva estructura, a la que denominaremos PROSODY_XML_SET.DTD (Ilustración 2-13), son:

a) la transcripción ya no será un texto a base de turnos de palabra, sino un conjunto, <Set>, de fragmentos o muestras de grabación <Sample>:

```
(2.41) <!-- Root element -->
<!ELEMENT Transcript (Header?,Set)>
```

b) el encabezado, al no pertenecer a una sola grabación, puede ser simplificado sustancialmente, eliminando, sobre todo, los elementos referentes a los participantes (2.42):

```
(2.42) <!-- Declaration of Header-relevant tags -->
<!ELEMENT Header (Title|File|Data|Date|Place|Source|Transcriber|
Revisor|Comments)*>

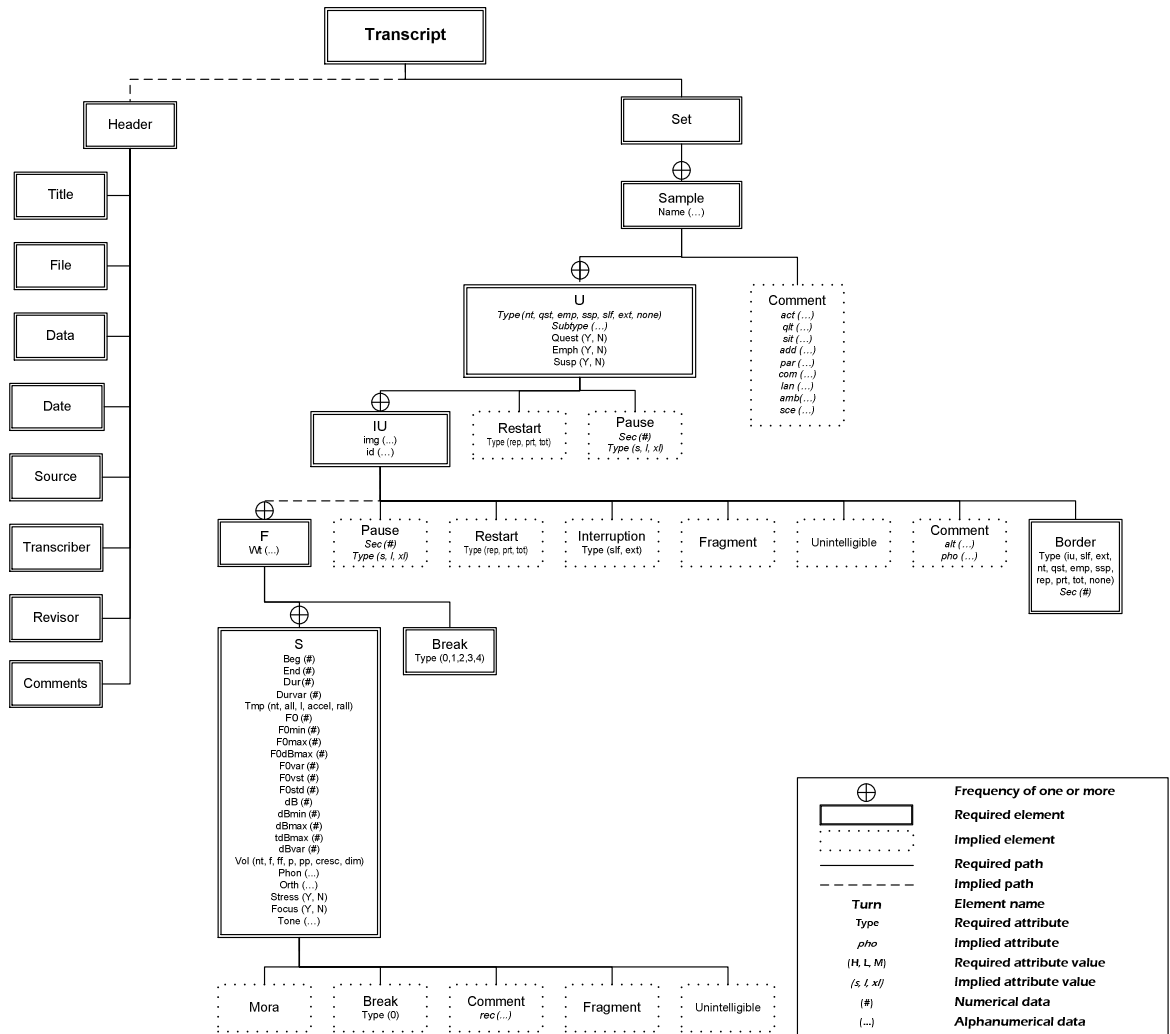
<!ELEMENT Title (#PCDATA)>
<!ELEMENT File (#PCDATA)>
<!ELEMENT Data (#PCDATA)>
<!ELEMENT Date (#PCDATA)>
<!ELEMENT Source (#PCDATA)>
<!ELEMENT Transcriber (#PCDATA)>
<!ELEMENT Revisor (#PCDATA)>
<!ELEMENT Comments (#PCDATA)>
```

c) puesto que para los análisis acústicos detallados los solapamientos son un problema, se elimina el elemento <Overlap> de la lista de elementos de IU (2.43).

```
(2.43) <!ELEMENT IU (#PCDATA|F|Pause|Restart|Fragment|Unintelligible|
Interruption|Comment|Border)*>
```

El resto de los elementos permanecerán inalterados.

Ilustración 2-13: Estructura jerárquica de PROSODY_XML_SET.DTD



Fuente: Diseño propio

En la siguiente sección se explicará la manera en la que, respetando tanto la PROSODY_XML.DTD como la PROSODY_XML_SET.DTD, se podrá crear automáticamente documentos de XML a partir de transcripciones básicas en rejillas de texto (*text grids*) del programa Praat.

2.3.3 Recolección de datos en Praat y transformación en XML

La razón por la cual se prefiere utilizar el programa Praat en este proceso es que con Praat (Boersma 2001, Boersma y Weenink 2008) se puede realizar una gran cantidad de tipos de análisis acústicos (espectral, frecuencial, de formantes, de intensidad, etc.) y, además, se puede sincronizar grabaciones con rejillas de texto etiquetables y segmentables en multiples

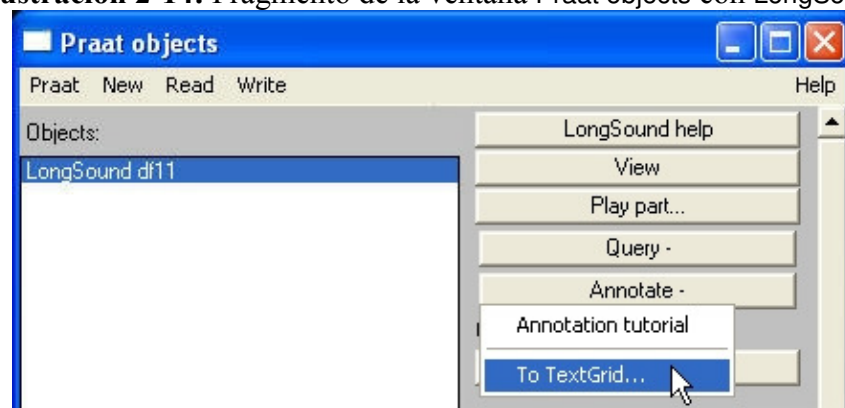
hileras (*tiers*), con la posibilidad de utilizar directamente el alfabeto fonético. El programa también tiene la capacidad de aplicar logaritmos, producir gráficas de alta calidad y estadísticas. Praat puede ser utilizado, por ejemplo, para proyectos de síntesis del habla, experimentos perceptivos, manipulación del habla, entre otros. Sin embargo, las mejores características de Praat son la portabilidad, debido a su uso de archivos binarios independientes de la plataforma informática y su capacidad para leer distintos formatos de audio, así como la posibilidad de permitir que el usuario configure la plataforma y programe las funciones originales por medio de un lenguaje de programación literal.

Por medio del uso de subprogramas (*scripts*) se puede extraer los fragmentos de grabación correspondientes a cada unidad entonativa o enunciado y crear las imágenes correspondientes del análisis acústico. Estas imágenes, en las que se puede plasmar la frecuencia fundamental, el espectrograma y el contenido de las filas de texto, serán desplegadas más adelante en la página *web* resultante. Los subprogramas también permiten la asignación de variables a los parámetros acústicos de cada sílaba y la repetición de este proceso por cada sílaba conformando una unidad entonativa. Estos valores podrán también ir anexándose continuamente a un archivo de texto, utilizando una sintaxis similar a la del XML.

Puesto que los documentos de XML constan únicamente de texto, no hay necesidad de conversiones o adaptaciones de formatos.

A manera de introducción en el manejo de Praat, se puede ejemplificar sus múltiples funciones con los siguientes ejemplos. En la Ilustración 2-14, podemos ver la ventana principal del programa Praat, denominada Praat objects. Para comenzar la transcripción de una grabación, bajo el menú Read seleccionaremos la opción Open LongSound file... y para crear su respectiva rejilla de texto, haremos clic en el botón Annotate y seleccionaremos la opción To TextGrid...:

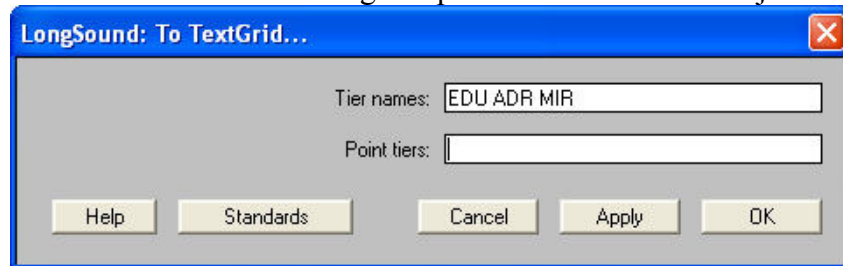
Ilustración 2-14: Fragmento de la ventana Praat objects con LongSound



Fuente: Diseño propio

En la ventana emergente (Ilustración 2-15), separados por espacios en blanco, escribiremos en el espacio de Tier names el nombre, el orden y el número de hileras o Tiers con las que queremos que cuente el objeto TextGrid. Por el momento no necesitamos definir ninguna de estas hileras como hileras de puntos o Point tiers, es decir, serán automáticamente creadas como hileras de intervalos, Interval tiers.

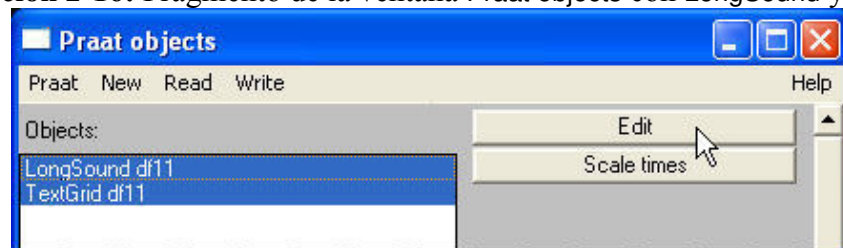
Ilustración 2-15: Ventana emergente para la creación de un objeto TextGrid



Fuente: Diseño propio

Ahora, en la ventana Praat objects tendremos un objeto TextGrid con el mismo nombre que el objeto LongSound. Si seleccionamos ambos objetos, veremos que en el menú dinámico, formado por los botones a la derecha, se reduce a dos opciones. Para este experimento seleccionaremos Edit como se muestra en la siguiente Ilustración:

Ilustración 2-16: Fragmento de la ventana Praat objects con LongSound y TextGrid



Fuente: Diseño propio

Se abrirá una nueva ventana denominada TextGrid Editor. En esa ventana haremos un breve experimento para mostrar el lenguaje literal de programación de Praat, así como la configuración de la plataforma. En la ventana del editor, tendremos que ir al menú File y seleccionar la opción New editor script. En la nueva ventana, untitled script, copiaremos el siguiente texto:

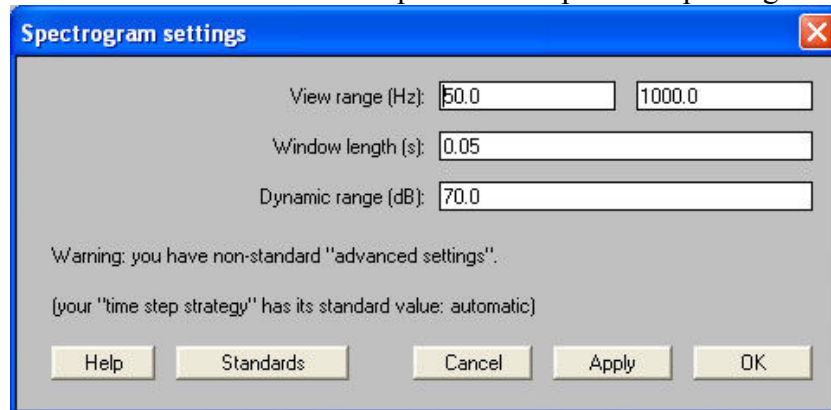
(2.44)

```
Spectrogram settings... 50 1000 0.05 70
Advanced spectrogram settings... 1000 500 Fourier Gaussian yes 100 6 0
Pitch settings... 50 500 Hertz "Intonation (AC method)" automatic
Advanced pitch settings... 50 500 no 15 0.03 0.45 0.01 0.35 0.14
Intensity settings... 50 100 "mean dB" yes
Formant settings... 1000 3 0.05 30 0.5
```

Cada una de las instrucciones del subprograma de (2.44) se escribe con el mismo nombre con el que aparecen las opciones bajo cada uno de los menús del editor: Spectrum, Pitch, Intensity y Formant. La serie de números y palabras separadas por espacios en blanco y colocadas delante de cada instrucción corresponden a las celdas en blanco, botones y opciones dentro la ventana

emergente que se abriría si seleccionáramos la opción desde el menú y exactamente en la misma secuencia. Por ejemplo, compárense los valores delante de Spectrogram settings... y las celdas en la Ilustración 2-17.

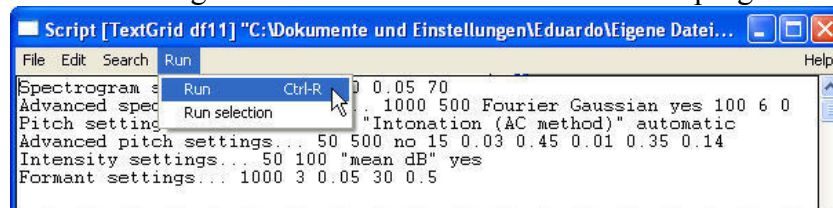
Ilustración 2-17: Ventana de preferencias para el espectrograma



Fuente: Diseño propio

A continuación correremos este subprograma yendo al menú Run y seleccionando la opción con el mismo nombre, como se muestra en la siguiente Ilustración:

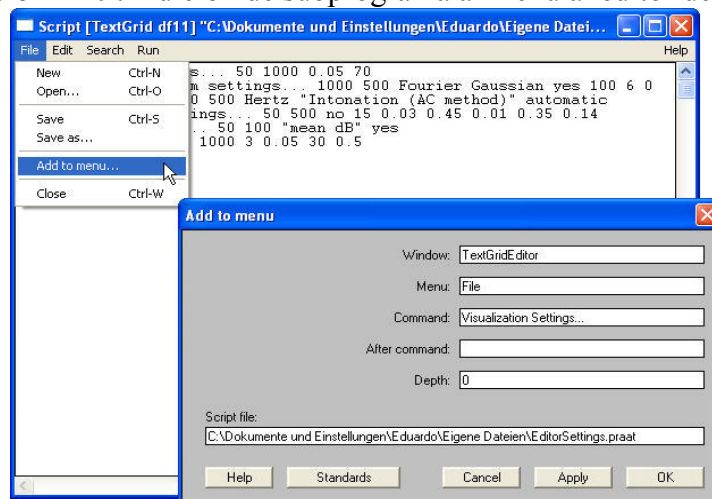
Ilustración 2-18: Fragmento de la ventana de edición de subprogramas (*scripts*)



Fuente: Diseño propio

Al hacer esto, habremos automatizado el proceso de establecer manualmente las preferencias de visualización de una por una. Si, por razones prácticas, se decide agregar este u otro *script* a alguno de los menús, se deberá seleccionar la opción Add to menu... bajo el menu File, y especificar en la ventana emergente el nombre y las condiciones bajo las que aparecerá:

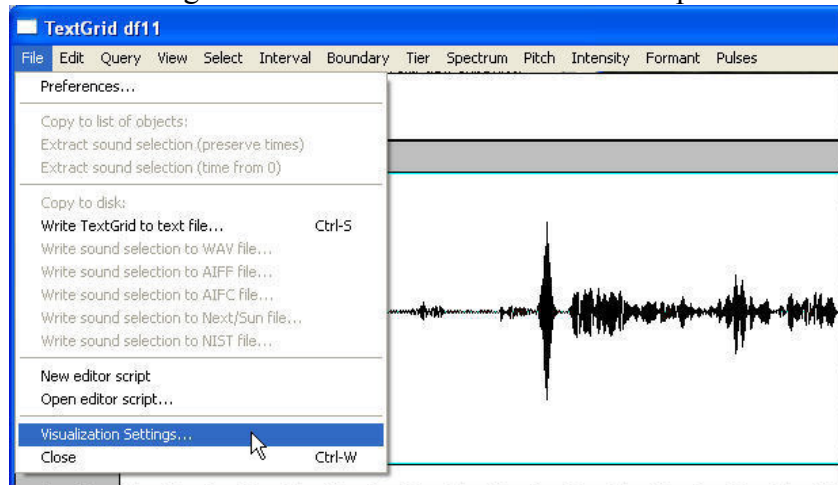
Ilustración 2-19: Adición de subprograma al menú al editor de TextGrids



Fuente: Diseño propio

Así, si se especificó que este subprograma aparezca en la ventana del TextGridEditor, bajo el menú File, con el nombre de Visualization Settings..., el menú modificado se desplegará así:

Ilustración 2-20: Fragmento del TextGrid Editor con nueva opción en el menú File



Fuente: Diseño propio

Una vez que hemos visto de manera detallada y sencilla los pasos a seguir para escribir un subprograma y su eventual incorporación a uno de los menús, a continuación se irá describiendo paso a paso el proceso de transcripción y recolección de datos en Praat por medio de subprogramas, así como su inserción y transformación en documentos de XML.

2.3.3.1 Transcripción

En las Ilustraciones 2-3 y 2-21 podemos apreciar a primera vista que en una rejilla de texto la transcripción y la grabación se rigen por la línea de tiempo. A diferencia de una transcripción sobre papel, en la cual los turnos de los hablantes son colocados en una nueva línea, en una rejilla de texto de Praat puede identificarse las producciones de los hablantes limitadas por pausas. Además, el uso de una hilera por cada participante de la conversación, permite representar fenómenos tales como los solapamientos de una manera sencilla y precisa. Por otro lado, las posibilidades de acercamiento a detalle sobre fragmentos de la grabación, hacen innecesaria la tediosa transcripción tradicional ‘a ciegas’ desde dispositivos y *software* de reproducción de audio con una línea de tiempo muy limitada, que, además, consume mucho tiempo de trabajo.

Sin embargo, lo ideal sería que cuando se transcribe una conversación ortográficamente, se aproveche para codificar de manera breve y sencilla fenómenos lingüísticos, paralingüísticos y situacionales que puedan ser relevantes para la investigación.

Por esta razón adapté algunos de los signos gráficos y convenciones utilizadas por Cresti y Moneglia (2005, 31-47), que a su vez hicieron una adaptación del sistema CHAT (McWhinney 1999, 2000), y agregué otros específicos para los fines de esta investigación.

En general, cada turno o producción lingüística entre pausas será considerada como la unidad máxima dentro de la conversación, aunque esta suele ser una medida muy abstracta. En la mayoría de los casos, un turno puede medir décimas de segundo o incluso minutos enteros, sobre todo mientras no haya ninguna interrupción por parte de los interlocutores. Para fines prácticos, la unidad máxima dentro de una TextGrid será simplemente una producción lingüística entre pausas. Estos intervalos deben ser divididos en enunciados (con o sin pausas de por medio) e, internamente, estos enunciados pueden contener una o más unidades entonativas. Por lo tanto, es muy común que una serie de unidades entonativas con entonación suspendida estén separadas por pausas y sean consideradas como producciones lingüísticas separadas pero interdependientes.

Los solapamientos entre las producciones lingüísticas de los distintos interlocutores deberán ser señaladas, así como las interrupciones, reinicios y otros fenómenos. Para la óptima identificación electrónica de los caracteres utilizados en combinación con otros signos gráficos para indicar distintos fenómenos, es indispensable respetar los espacios en blanco, representados aquí por medio de □ (véase, por ejemplo, el uso de la barra diagonal en □/□, /□, □//, □+/, etc.).

A continuación se hará una presentación de las convenciones gráficas necesarias según el tipo de fenómeno o evento relevante dentro de la interacción comunicativa.

Enunciados

- ..□// Indica que un enunciado es declarativo o neutro.
- ..□? Señala que un enunciado es interrogativo.
- ..□... Representa una suspensión del flujo de la comunicación interpretada como el final de un enunciado.
- ..□! Se utiliza cuando un enunciado sea considerado como exclamativo.

Unidades tonales

- ..□/□... Sirve para indicar los límites de cada unidad tonal, lo cual también da lugar a una nueva frase entonativa. Al final de una unidad tonal se puede originar un alargamiento vocálico. Las marcas de final de enunciado implican también un final de unidad tonal.

Interrupciones

Desde el punto de vista discursivo, una interrupción es de gran importancia. Si se utilizara la no marca, sería imposible identificarlas automáticamente. Las interrupciones aparecen únicamente al final de la producción lingüística y se distingue entre los siguientes dos tipos:

- ..□+ Autointerrupción
- ..□+/ Interrupción por parte del interlocutor

Reinicio

Aquí se agrupa distintos tipos de reinicio, como son las repeticiones y las reformulaciones.

- ..□[/]□.. Repetición, p.ej. voy para [/] para la casa
- ..□[//]□.. Reformulación parcial, p.ej. voy para [//] a la casa
- ..□[///]□.. Reformulación total, p.ej. voy para [///] no / olvídale //

Pausa

- # Es un silencio durante un espacio temporalmente muy perceptible (aprox. >1 sec) y que no se puede considerar ni interrupción ni suspensión ni reinicio. En la primera fase de transcripción es preferible dejar en blanco los intervalos de pausa.

Apoyo silábico y muletillas

Las muletillas y otros elementos de apoyo silábico deben aparecer normalmente como unidades entonativas independientes (aislados entre barras diagonales sencillas).

Comentarios

En caso de ser relevante, dentro de un intervalo se incluirán comentarios de distintos tipos:

- [%act:□..] Acciones de los interlocutores independientes del texto, p.ej. [%act: laugh]
- [%sit:□..] Situación o suceso relevante en la comunicación: [%sit: telephone rings]
- [%add:□..] Interlocutor a quien se dirige el texto del turno, p.ej. [%add: ABC]
- [%par:□..] Gestos o elementos paralingüísticos del hablante: [%par: pointing to a map]
- [%com:□..] Comentarios del transcriptor, p.ej. [%com: noise]
- [%lan:□..] Lengua alternativa, p.ej. la gente [%lan: English] nice
- [%rec:□-.-.-] Reconstrucción de palabras, sin espacio en blanco, □, ya sea antes o después de los paréntesis, p.ej. (%rec: en-)tonces, verdá(%rec: -d)
- [%alt:□..] Propuestas para errores de producción y variantes no estándar de difícil identificación, p.ej. edá (%alt: verdad)
- [%pho:□..] Fenómeno fonético muy marcado, p.ej. lol [%pho: assimilation_of_s] lugares
- [%qlt:□..]□(□..□) Calidad de voz del interlocutor; el texto afectado se señala entre paréntesis, p.ej. [%qlt: laughing] (eso no te lo creo) !
- [%amb:] Ambientación
- [%sce:] Escena

Sonidos no reconocidos como signos lingüísticos

- xxx** Se utiliza en casos de ininteligibilidad.
- ..- Se coloca tras segmentos fónicos considerados como fragmentos o sonidos carentes de significado y función lógica. No forman parte del recuento de palabras. P.ej. no / pero m- [/] me voy a dar una vuelta + (= 8 palabras)

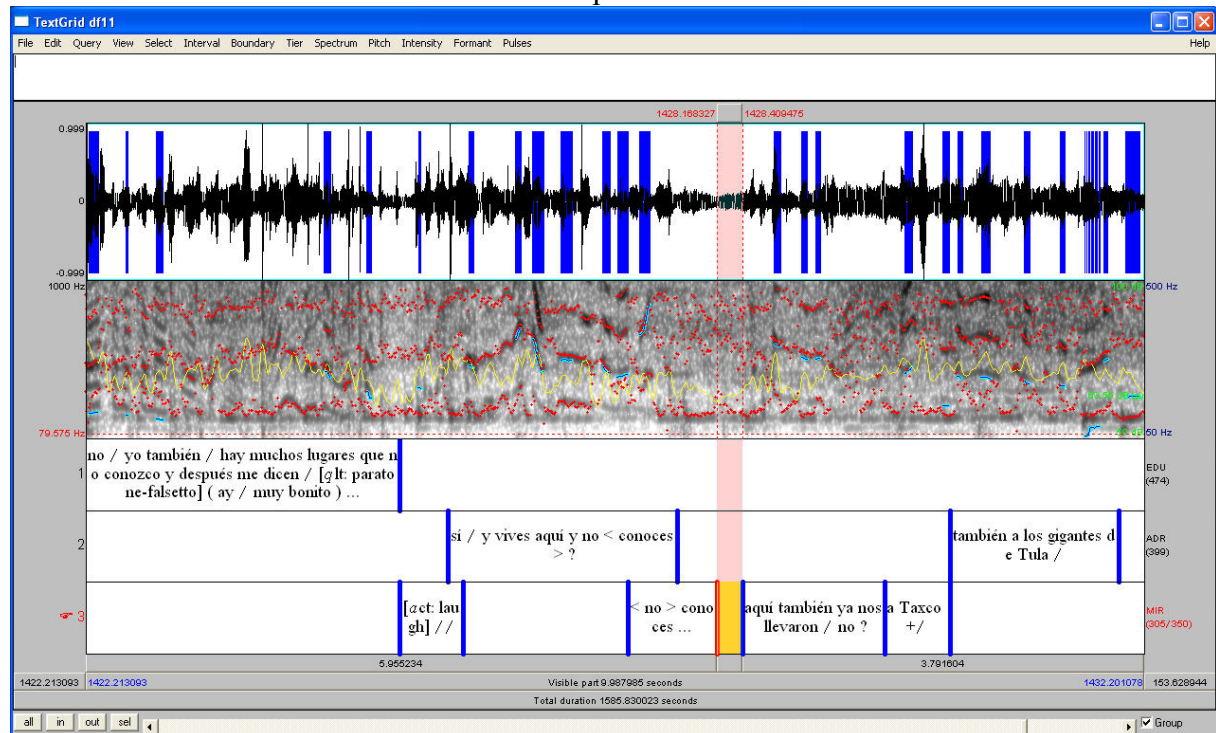
Solapamiento

<□. .□> Indica que el texto encerrado ahí dentro ha sido producido simultáneamente con respecto al texto del interlocutor. > se coloca siempre antes de la marca del final de turno o de unidad tonal, p.ej. . .□>□?

[<]□<□. .□> Señala que el próximo texto se ha solapado al texto de otro interlocutor quien tenía la palabra en ese momento. Es preferible evitar los cortes entre palabras.

En el fragmento de la grabación *df11.wav* presentado en la Ilustración 2-21 se puede identificar varios de los signos gráficos explicados anteriormente.

Ilustración 2-21: Transcripción en el editor de TextGrids



Fuente: Diseño propio

En el Anexo 1 se hace un breve recuento de los signos gráficos necesarios para las transcripciones básicas en TextGrids de Praat, según se presentan en esta sección.

2.3.3.2 Documentos básicos de XML

Una vez que se haya transcrito completamente una grabación o la parte de ella que se considere relevante, será necesario extraer los fragmentos en los que se encuentren cada una de las producciones lingüísticas. Además, sería práctico hacer una imagen por cada fragmento y vaciar la información a un documento de XML. Hacerlo a mano sería casi imposible.

ExtractTextGrid+Sound+XML.praat (véase el Anexo 5) es un programa que automatiza todas estas funciones. Primero desplegará un formulario donde el usuario deberá: 1) proveer una identificación única que será utilizada como prefijo para cada uno de los fragmentos, 2) indicar el directorio en el que los nuevos archivos serán guardados, 3) especificar qué tipo de

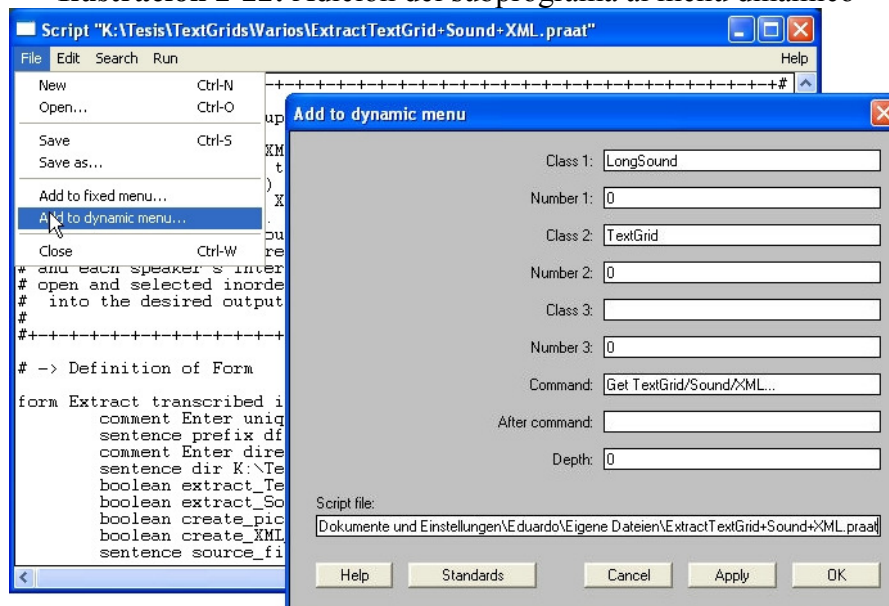
2.3 Transcripción prosódica de la lengua oral en XML

archivos deben ser producidos (TextGrids, archivos de audio, imágenes y/o un documento de XML), 4) agregar un mínimo de metainformación sobre la conversación para el encabezado.

Antes de apretar el botón ‘OK’ deberá de haberse abierto y seleccionado un archivo largo de audio (LongSound) y su correspondiente rejilla de texto (TextGrid), la cual deberá contener la transcripción básica de la conversación, indicando por medio de intervalos los enunciados y las pausas de cada participante de la conversación y, por medio de hileras, las aportaciones de cada hablante. El programa extraerá la información necesaria para crear los distintos tipos de archivos seleccionados.

Si, como en el caso de esta investigación, se desea trabajar con varias grabaciones, podría ser práctico añadir el Script a un menú dinámico que apareciera cuando un objeto LongSound y un objeto TextGrid estén seleccionados, como se muestra en la siguiente Ilustración:

Ilustración 2-22: Adición del subprograma al menú dinámico



Fuente: Diseño propio

A continuación se irá describiendo cada una de las secciones de las que se compone este subprograma. En el Anexo 4 vuelve a aparecer esta enumeración de los procesos con referencia a los números de línea en los que se encuentran. Por cuestiones de espacio, omito descripciones detalladas de la sintaxis de los *scripts*. Para mayor información, la mejor referencia es el Manual de Praat (Boersma y Weenink 2008).

- Definición del formulario

Como ya se había comentado al inicio de esta sección, el formulario (Ilustración 2-23) permite la interacción con el usuario, quien, en este caso, puede seleccionar las partes del subprograma que se activarán, por ejemplo, para la selección de los archivos a producir. Además, el usuario puede introducir la metainformación que no está contenida en ninguna parte dentro de la rejilla de texto.

Ilustración 2-23: Formulario del subprograma `ExtractTextGrid+Sound+XML.praat`

Fuente: Diseño propio

- Selección y asociación de TextGrid y LongSound

En este proceso se guardarán los nombres de estos dos objetos en variables que serán utilizadas a lo largo del subprograma, además, se contará el número de hileras y se extraerá la duración total de la grabación.

- Impresión del encabezado del documento de XML

Con este proceso inicia la creación de un documento de texto conforme a la `PROSODY.DTD`. La información del encabezado será colocada en la posición necesaria por medio de variables.

- Inicio del ciclo de extracción de información de cada una de las hileras

Se contará el número de intervalo en la hilera actual y se extraerá la clave de tres letras del hablante al que corresponden las producciones lingüísticas de dicha hilera.

- Inicio del ciclo de inspección y asociación de cada uno de los intervalos en la hilera

En este proceso se detectará si el intervalo actual está precedido y seguido de pausas o de intervalos con contenido.

- Asignación de los puntos inicial y final del intervalo actual [`st` y `et`]

Con estas dos variables se podrá extraer más adelante la información específica del intervalo actual, independientemente de su asociación o no con otros intervalos.

- Asignación de variable para el tipo de frontera del último intervalo que no esté vacío [`border0$`]

Con la finalidad de detectar el tipo de enunciado en el intervalo previo, se interpretará los signos utilizados al final del texto.

- Asignación de variable para el tipo de frontera del penúltimo intervalo que no esté vacío [`border00$`]

En caso de que este enunciado previo también esté precedido de otro enunciado, se determinará el tipo al que pertenece.

- Asignación de variable para el tipo de frontera del intervalo actual que no esté vacío [border1\$]

Aquí se detectará el tipo de enunciado contenido en el intervalo actual.

- Inspección de hasta ocho intervalos consecutivos para localizar intervalos vacíos y transcritos

Con este proceso se detectarán los intervalos correspondientes a las pausas al final de la serie de intervalos transcritos.

- Asociación de intervalos consecutivos sin vacíos intermedios

Los intervalos consecutivos y sin pausas intermedias serán asociados para tratarse de aquí en adelante como una unidad. Esta es la razón por la cual, en este nivel de análisis, las pausas no deben etiquetarse dentro de la rejilla de texto.

- Asignación de variable para los puntos inicial y final del grupo asociado de intervalos [t_1 y t_2]

Con este proceso se establecerán las fronteras del grupo, aunque dicho ‘grupo’ puede también estar formado por sólo un intervalo.

- Asignación de variable para los puntos inicial y final del intervalo vacío después del grupo [t_4 y t_5]

Por medio de estas variables se podrá medir la duración de la pausa posterior al grupo.

- Adición de 0.05 segundos antes y después de los límites del grupo [t_0 y t_3]

Para dejar espacio al inicio y al final del grupo, se agregará 0.05 segundos.

- Asignación de variable para el tipo de frontera del grupo [border\$]

Se detectará el tipo de enunciado al que pertenece el último intervalo, para ser considerado como la frontera del grupo.

- Estandarización de formato de las variables de tiempo con ceros a la izquierda

Con la finalidad de que la computadora pueda leer de manera ordenada los archivos que se crearán más adelante utilizando las variables de tiempo, se normalizarán todas las cifras para que sean números de cuatro dígitos y dos decimales (####.##).

- Extracción del fragmento de TextGrid entre t_0 y t_3 , registro con nombre especial (prefijo_PuntoInicialNormalizado-PuntoFinalNormalizado_NúmeroDeHileras_frontera)

En este punto del proceso se extraerá el fragmento de TextGrid que incluye el grupo de intervalos más los márgenes. El fragmento contendrá el total de hileras.

- Extracción la hilera actual del fragmento de TextGrid y registro con nombre especial (prefijo_PuntoInicialNormalizado-PuntoFinalNormalizado_NombreDeHilera_frontera)

Se guardará, además, una versión en la que se eliminen todas las hileras excepto la actual.

- Extracción del fragmento correspondiente del archivo de audio y registro (prefijo_PuntoInicialNormalizado-PuntoFinalNormalizado_NombreDeHilera_frontera)

Del archivo de audio se extraerá el fragmento correspondiente al punto de inicio y al final.

- Extracción de la frecuencia (Pitch) del archivo de audio actual y cálculo de la frecuencia mínima y máxima

Aunque en este nivel de análisis Praat no analizará detalladamente la información sobre la frecuencia fundamental, se extraerán los valores máximo y mínimo de la frecuencia para determinar el rango tonal.

- Redondeo de la frecuencia mínima para establecer una base más holgada para las imágenes

El valor mínimo de la frecuencia fundamental se redondeará para que Praat calcule la base con espacio para cada una de las imágenes dependiendo de su rango tonal.

- Redondeo de la frecuencia máxima para establecer una cima más holgada para las imágenes

Este es el proceso inverso al anterior para agregar espacio entre el margen superior de la imagen y la cima más alta de la frecuencia fundamental.

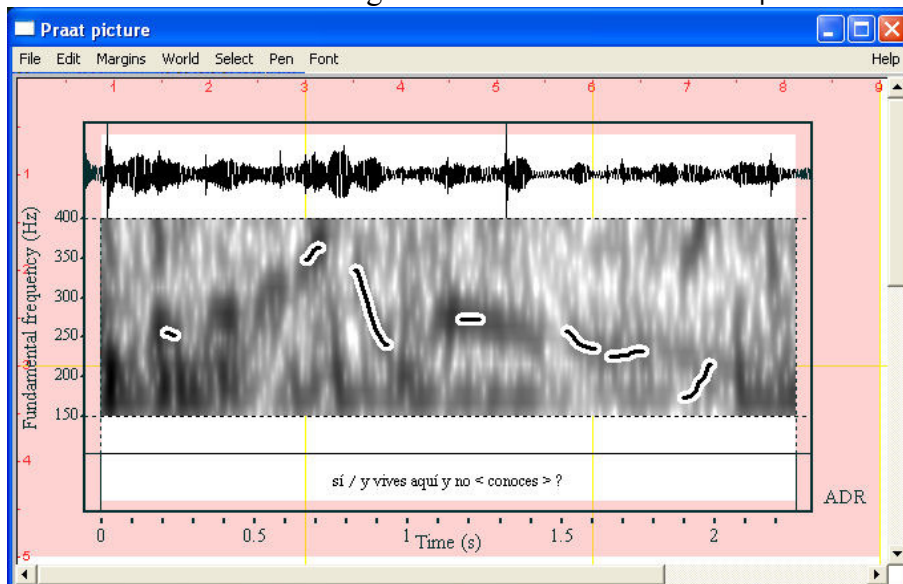
- Extracción del espectrograma (Spectrogram) del archivo de audio actual

Se extraerá la porción visible del espectrograma para el segmento de la grabación. Dado que en la mayoría de las grabaciones utilizadas para esta investigación el segundo formante era, por lo general, más visible que el primero, este proceso toma los valores del rango tonal y los multiplica por dos.

- Creación de imágenes

En este paso se creará una imagen por cada uno de los grupos de intervalos con las mismas características. Un ejemplo de estas imágenes es la Ilustración 2-24.

Ilustración 2-24: Imagen creada en la ventana Praat picture



Fuente: Diseño propio

- Escritura de la imagen a archivos (Metaarchivo de Windows / Archivo de Imagen de Praat)

Las imágenes serán guardadas en dos formatos distintos, primero como metaarchivo de Windows (.EMF/.WMF), que puede ser leído y procesado en cualquier editor de imágenes, y después como archivo de imagen de Praat (.PRAAPIC) que solamente puede ser leído por Praat, y, a su vez, el único formato de imagen que lee Praat.

- Limpieza de la ventana de imágenes antes de pasar al siguiente intervalo

Para evitar imágenes encimadas, deberá borrarse siempre el contenido de Praat picture.

- Extracción de información del TextGrid para crear un archivo compatible con XML

Una vez que se ha concluido el proceso de creación de fragmentos de TextGrids, archivos de audio e imágenes, siempre y cuando se hayan seleccionado en el formulario de inicio, se procederá a extraer la información de la rejilla de texto relativa al grupo de intervalos.

- Impresión del inicio de la unidad entonativa, enunciado o turno de palabra, tomando en cuenta la frontera de los intervalos previos

Se continuará con la impresión en la Info Window de las etiquetas necesarias según el tipo de unidad actual, con respecto a la frontera de las unidades en los intervalos previos.

- Reemplazo de signos al inicio y dentro del intervalo actual por código compatible con XML (solapamientos, comentarios, producciones ininteligibles, fragmentos, reinicios y fronteras de unidades entonativas intermedias)

Debido a que algunos signos que se utilizan para transcribir los fenómenos prosódicos arriba mencionados también forman parte del inventario cerrado de signos sensibles en XML (por ejemplo, los signos < > con los que se señalan los solapamientos, un *parser* interpretaría su contenido como una etiqueta de apertura y haría que el documento dejara de funcionar. Todos los signos sensibles serán sustituidos por código compatible.

- Reemplazo de signos al final del intervalo actual por código compatible con XML (tipo de frontera, final de unidad entonativa, final de enunciado, final de turno de palabra - seguido inmediatamente de pausa)

Se interpreta y sustituye los signos colocados a manera de frontera al final del intervalo.

- Etiquetado explícito de pausas únicamente cuando sean mayores a 1 segundo

Dado que en la transcripción básica cada uno de los intervalos está separado por una pausa de mayor o menor longitud, se intenta filtrar la mayoría de ellas etiquetando sólo las pausas mayores a un segundo para evitar sobrecargar innecesariamente el código resultante.

- Reemplazo de signos al final del intervalo actual por código compatible con XML (para intervalos que no estén seguidos inmediatamente de pausa)

Se señalarán las fronteras de las unidades contenidas en intervalos intermedios para que tengan su estatus correspondiente en la estructura de XML a pesar de estar unidas a otro intervalo, sobre todo los enunciados sin pausa de por medio.

- Eliminación de archivos generados por los análisis automáticos según las tareas seleccionadas (Extract TextGrids, Extract Sounds, Create Pictures, Create XML)

En cada uno de los procesos del subprograma se fueron creando distintos archivos temporales que, sobre todo en el caso de los procesos cíclicos, pueden llegar a interrumpir el subprograma y bloquear Praat completamente. Por otra parte, como en los procesos cíclicos hay otros archivos que se utilizan constantemente, no se puede eliminar todos los archivos de la ventana de objeto de una sola vez y por eso se eliminará un conjunto determinado de archivos por cada proceso seleccionado.

- Selección de los archivos TextGrid y LongSound para analizar el siguiente intervalo

Se selecciona nuevamente la rejilla de texto y el archivo de sonido para que el proceso cíclico de análisis pueda continuar.

- Fin del ciclo de intervalos

Aquí es donde termina el proceso de análisis de cada uno de los intervalos en una hilera y se pasa a la siguiente hilera.

- Fin del ciclo de hileras

En esta sección se selecciona la siguiente hilera, hasta que el número de hileras analizadas sea igual al total.

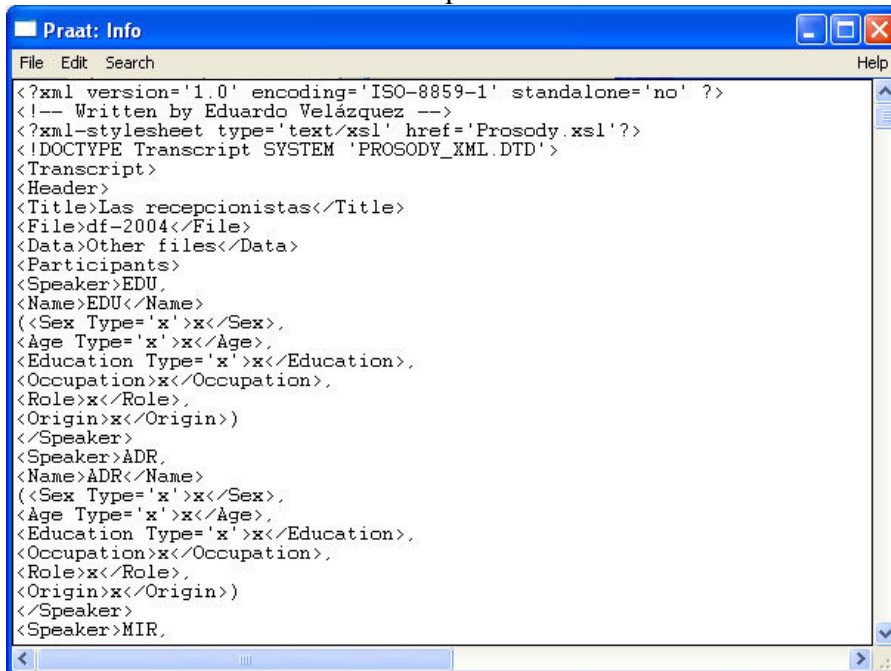
- Impresión del final del documento de XML

Una vez que se han ido analizando los intervalos transcritos de cada una de las hileras, cuya información se ha ido escribiendo en la Info Window, se escriben las etiquetas finales de la estructura jerárquica de XML.

- Escritura del contenido de la Info Window a un archivo de texto con terminación .XML

El documento resultante (véase la Ilustración 2-25) se guardará como un archivo de texto común pero con la terminación .XML. Puesto que el documento se había ido escribiendo según la sintaxis requerida por la PROSODY.DTD, el archivo de texto será leído de ahora en adelante como un documento bien formado de XML.

Ilustración 2-25: Contenido compatible con XML en la ventana Info



```

<?xml version='1.0' encoding='ISO-8859-1' standalone='no' ?>
<!-- Written by Eduardo Velázquez -->
<?xml-stylesheet type='text/xsl' href='Prosody.xsl'?>
<!DOCTYPE Transcript SYSTEM 'PROSODY_XML.DTD'>
<Transcript>
<Header>
<Title>Las recepcionistas</Title>
<File>df-2004</File>
<Data>Other files</Data>
<Participants>
<Speaker>EDU,
<Name>EDU</Name>
(<Sex Type='x'>x</Sex>,
<Age Type='x'>x</Age>,
<Education Type='x'>x</Education>,
<Occupation>x</Occupation>,
<Role>x</Role>,
<Origin>x</Origin>)
</Speaker>
<Speaker>ADR,
<Name>ADR</Name>
(<Sex Type='x'>x</Sex>,
<Age Type='x'>x</Age>,
<Education Type='x'>x</Education>,
<Occupation>x</Occupation>,
<Role>x</Role>,
<Origin>x</Origin>)
</Speaker>
<Speaker>MIR,

```

Fuente: Diseño propio

- Fin del subprograma

Aquí termina el proceso de creación del documento de XML, pero también de la extracción de los fragmentos de rejilla de texto y archivo de sonido, además de la creación de las imágenes que, como veremos en el siguiente apartado, podrán ser utilizadas en el documento de HTML resultante de la transformación del documento de XML con una hoja de estilo.

2.3.3.3 Hoja de estilo para documentos básicos de XML

Del proceso total de creación de un documento de XML, la aplicación de estilo es seguramente la que requiere del conocimiento de una mayor cantidad de lenguajes, herramientas y aplicaciones computacionales. Para profundizar en la combinación de XSLT y el lenguaje de localización de nodos y contenido, Xpath, se recomienda revisar Clark y DeRose (1999) o Niedermair y Niedermair (2001: 155-222). Por el lado del formato, la combinación de HTML y su versión sintácticamente más estricta XHTML ampliados con subprogramas en JavaScript y especificaciones de estilo en CSS se recomienda Bos *et al.* (1998), Hanke (2000: 185-233, 285-307) y Steyer (2003: 62-123, 332-400, 434-490). En esta sección se presentará a grandes rasgos las partes que constituyen la hoja de estilo que denominaremos *Prosody.xsl* y que se reproduce completamente en el Anexo 10. Los elementos básicos de la hoja de estilo son los siguientes:

- Plantilla principal que se aplicará a todo el elemento `Transcript`

Con esta función se selecciona el elemento `Transcript`, a partir del cual se establecerá el formato general de todo el documento, así como las propiedades del texto.

- Encabezado con título, enlaces al resumen y a las convenciones, así como una tabla con los datos de los participantes

En esta sección se define el formato y contenido de los elementos que compondrán el encabezado. En la Ilustración 2-26 se muestra un ejemplo de encabezado en formato HTML.

Ilustración 2-26: Encabezado de la grabación en HTML

The screenshot shows a web browser window displaying the HTML header of a transcript document. The page title is "Transcript of File: df11-2004". Below the title is a summary and conventions section, followed by a table of participants.

Transcript of File: df11-2004
(Summary and conventions at the bottom)

Title: Las recepcionistas
Data: hotel_1_df1-04-1
Place: Mexico City
Date: 05/08/2004
Topic: Mexico City's sights, everyday life and work
Situation: Free-style interview at hotel reception, not hidden, researcher interviewer
Source: DF-2004
Class: informal, public, conversation
Length (in seconds): 1585.83
Words: 2867
Acoustic quality: B
Transcriber: Eduardo
Revisor: Eduardo; Eduardo (prosody)
Comments:

Participants:

Name	Sex	Age	Education	Occupation	Role	Origin
EDU	man	B	3	linguist	participant	Guadalajara, MX
ADR	woman	A	3	hotel receptionist	participant	Mexico City
MIR	woman	A	3	hotel receptionist	participant	Mexico City

x = unknown
Age: A (18-25), B (26-40), C (40-60), D (>60)
Education: 1 (primary school or less), 2 (high school), 3 (graduated or university students), x (unknown)

Contract All | Expand All

Fuente: Diseño propio

La hoja de estilo extraerá automáticamente la información de distintos elementos y la adaptará al formato especificado.

- Botones para contraer o expandir todas las tablas de imagen y sonido

Haciendo uso del subprograma `dynamic.js` (Anexo 13), escrito en el lenguaje JavaScript, se añadirá unos botones con enlaces a instrucciones para contraer o expandir todas las áreas especificadas y que, idealmente, contendrán cada una de las tablas de imagen y sonido correspondientes a cada enunciado de la grabación.

- Estructura básica de los turnos de palabra y los enunciados, ordenados según la clave de identificación de cada unidad entonativa

Por medio de tablas, se ordena la información de cada turno en distintas columnas que mantendrán las mismas proporciones a pesar de las diferencias de tamaño de las ventanas de navegador (*browser*).

- Alternativa para evitar los turnos de palabra producidos por EDU, así como los enunciados solapados, comentados con `act` (risas, etc.) o ininteligibles

Este proceso debe sacarse de las marcas de comentarios para que, en lugar de que se seleccione y ordene todos los fragmentos de la grabación, se seleccione únicamente los fragmentos sin solapamientos y sin irregularidades tales como risas o segmentos ininteligibles, y se excluya los fragmentos producidos por el investigador.

- Proceso para permitir que cada una de las tablas de imagen y sonido pueda abrirse de manera independiente

Además de los botones para contraer o expandir en bloque las áreas donde se ubican las tablas de imagen y sonido, también se incluye este proceso que permite que las tablas se contraigan o expandan de manera individual, por ejemplo, cuando se desee hacer una segunda selección.

- Impresión del texto de cada enunciado en el turno de palabra

Con esta instrucción se coloca en el lugar indicado el texto de cada enunciado del documento.

- Asignación de variables para los atributos `img` e `id` de cada unidad entonativa

Puesto que las claves de identidad de cada enunciado y unidad entonativa serán utilizados en el documento resultante para establecer enlaces a las imágenes y archivos de sonido correspondientes, se asignan variables que permitan que estos valores cambien con cada ciclo del proceso.

- Tablas alterables de imagen y sonido

En esta sección se establece el formato que tendrán las tablas de imagen y sonido que podrán contraerse o expandirse de manera individual o en bloque.

- Impresión de la clave de identidad de la unidad entonativa actual

Para facilitar la búsqueda y selección de un fragmento determinado, se proveerá su clave de identidad delante de cada tabla de imagen y sonido.

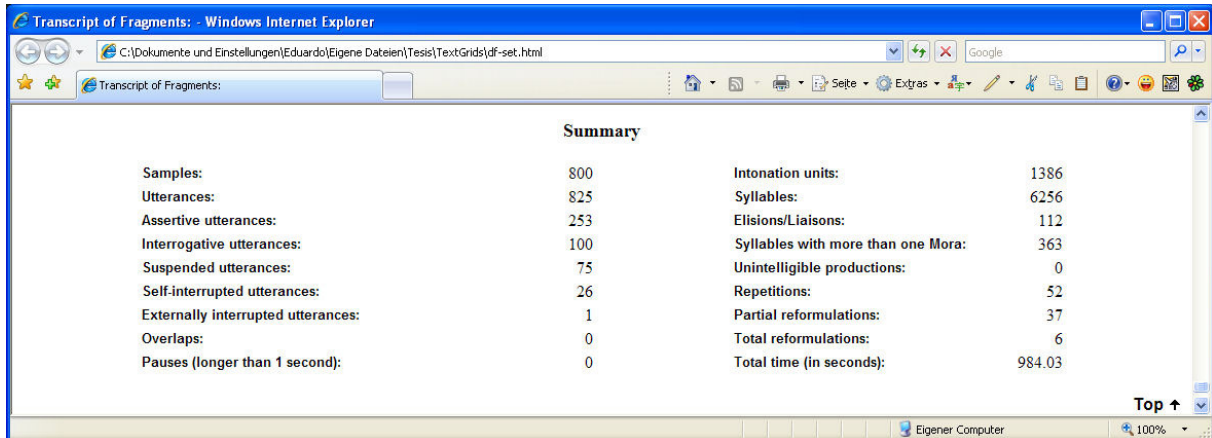
- Fragmento que deberá sacarse de los comentarios cuando el filtro de turnos y enunciados esté activo

Con estas instrucciones se cierra el ciclo de filtrado, en caso de que haya sido activado.

- Resumen de fenómenos lingüísticamente relevantes

En esta área se especifica el formato de una tabla que contendrá un listado de fenómenos lingüísticamente relevantes que aparecen a lo largo de todo el documento. En las columnas correspondientes se desplegará el resultado de la contabilización de cada uno de los fenómenos. En la Ilustración 2-27 se puede ver un ejemplo de esta tabla.

Ilustración 2-27: Resumen de fenómenos relevantes en el documento



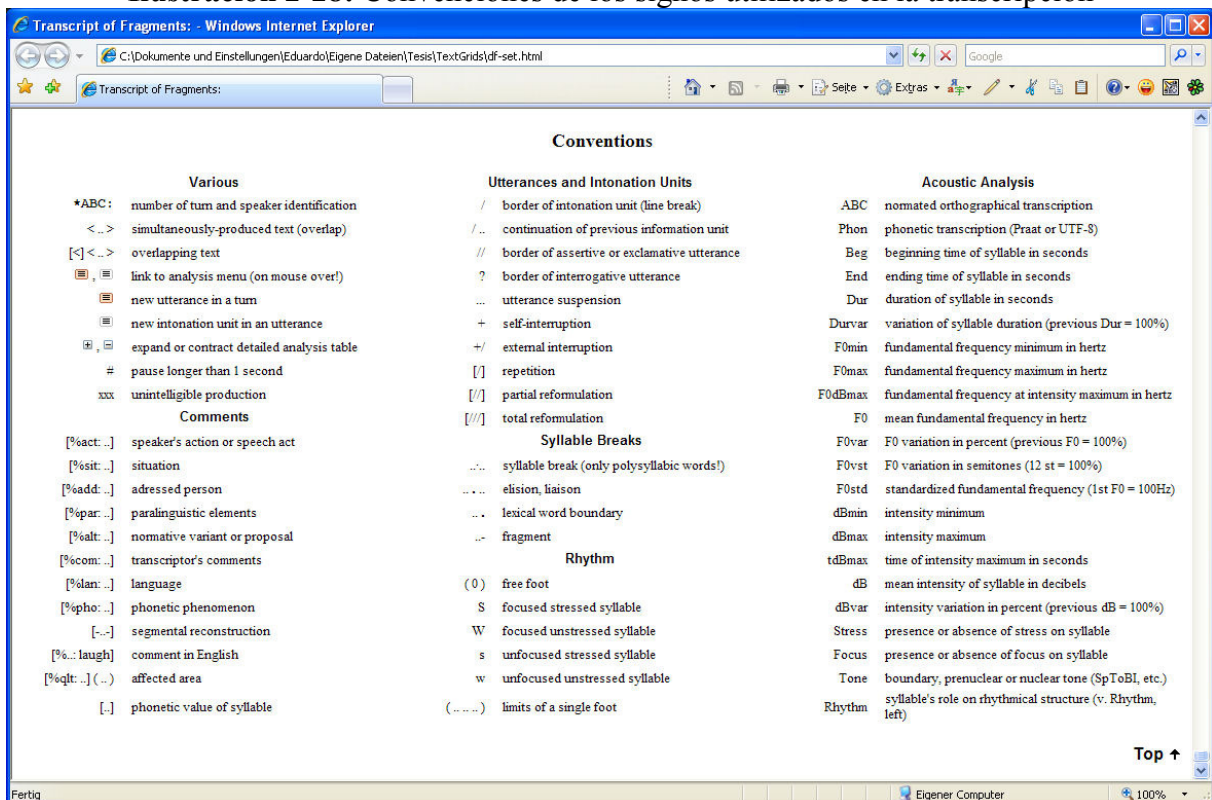
Summary			
Samples:	800	Intonation units:	1386
Utterances:	825	Syllables:	6256
Assertive utterances:	253	Elisions/Liaisons:	112
Interrogative utterances:	100	Syllables with more than one Mora:	363
Suspended utterances:	75	Unintelligible productions:	0
Self-interrupted utterances:	26	Repetitions:	52
Externally interrupted utterances:	1	Partial reformulations:	37
Overlaps:	0	Total reformulations:	6
Pauses (longer than 1 second):	0	Total time (in seconds):	984.03

Fuente: Diseño propio

- Convenciones de los signos utilizados en la transcripción para representar los fenómenos lingüísticamente relevantes

Al igual que en la sección anterior, aquí se incluye una tabla que aparecerá al final de cada documento, pero, a diferencia de aquella sección, aquí el contenido de la tabla no variará de un documento a otro. En esta tabla (Ilustración 2-28) se incluye una lista de todos los signos gráficos convencionales utilizados en la transcripción y despliegue del contenido del texto.

Ilustración 2-28: Convenciones de los signos utilizados en la transcripción



Conventions		
Various	Utterances and Intonation Units	Acoustic Analysis
*ABC: number of turn and speaker identification	/ border of intonation unit (line break)	ABC normated orthographical transcription
< . > simultaneously-produced text (overlap)	/.. continuation of previous information unit	Phon phonetic transcription (Praat or UTF-8)
[<] < . > overlapping text	// border of assertive or exclamative utterance	Beg beginning time of syllable in seconds
[] , [] link to analysis menu (on mouse over!)	? border of interrogative utterance	End ending time of syllable in seconds
[] new utterance in a turn	? utterance suspension	Dur duration of syllable in seconds
[] new intonation unit in an utterance	+ self-interruption	Durvar variation of syllable duration (previous Dur = 100%)
[] expand or contract detailed analysis table	+/- external interruption	F0min fundamental frequency minimum in hertz
# pause longer than 1 second	[] repetition	F0max fundamental frequency maximum in hertz
xxx unintelligible production	[/] partial reformulation	F0dBmax fundamental frequency at intensity maximum in hertz
	[/ /] total reformulation	F0 mean fundamental frequency in hertz
Comments	Syllable Breaks	F0var F0 variation in percent (previous F0 = 100%)
[%act: ..] speaker's action or speech act	... syllable break (only polysyllabic words!)	F0vst F0 variation in semitones (12 st = 100%)
[%sit: ..] situation	... elision, liaison	F0std standardized fundamental frequency (1st F0 = 100Hz)
[%add: ..] adressed person	... lexical word boundary	dBmin intensity minimum
[%par: ..] paralinguistic elements	.. fragment	dBmax intensity maximum
[%alt: ..] normative variant or proposal	Rhythm	tdBmax time of intensity maximum in seconds
[%com: ..] transcripator's comments	(0) free foot	dB mean intensity of syllable in decibels
[%lan: ..] language	S focused stressed syllable	dBvar intensity variation in percent (previous dB = 100%)
[%pho: ..] phonetic phenomenon	W unfocused unstressed syllable	Stress presence or absence of stress on syllable
[...] segmental reconstruction	s unfocused stressed syllable	Focus presence or absence of focus on syllable
[%: laugh] comment in English	w unfocused unstressed syllable	Tone boundary, prenuclear or nuclear tone (SpToBI, etc.)
[%: laugh] comment in English	(... ..) limits of a single foot	Rhythm syllable's role on rhythmical structure (v. Rhythm, left)
[%qt: ..] (..) affected area		
[] phonetic value of syllable		

Fuente: Diseño propio

- Fin de la platilla principal

Con esta instrucción se da fin al ciclo que da formato a todo el documento.

- Despliegue de la duración de la pausa, anclada al signo #

El signo #, con el cual se designa la presencia de una pausa en el texto, puede ser utilizado para que, cuando el usuario pase el puntero del ratón por encima de él, despliegue en una cinta amarilla la duración de la pausa.

- Asignación de una imagen diferente delante de las unidades entonativas en posición distinta a la primera

Para destacar gráficamente la subordinación de las unidades entonativas pertenecientes a un mismo enunciado, puesto que serán desplegadas cada una en una nueva línea, se colocará la imagen ☰ delante de ellas.

- Asignación de una imagen diferente delante de los enunciados en posición distinta a la primera

Del mismo modo, para destacar gráficamente la existencia de más de un enunciado dentro de un fragmento de grabación, se colocará la imagen ☰ delante de la primera unidad entonativa del enunciado.

- Fin de la hoja de estilo

Con esta instrucción finaliza el proceso de transformación del documento de XML en un nuevo documento de HTML según las reglas de plantilla recopiladas en `Prosody.xml`. Aunque ya se ha visto en las Ilustraciones 2-26 a 2-28 ejemplos del producto final en HTML, en la Ilustración 2-29 se ejemplifica el formato que tendrá el texto de la transcripción en el cual se podrá expandir y contraer las tablas de imagen y sonido.

Ilustración 2-29: Documento básico de XML transformado en HTML

The screenshot shows a web browser window titled "Transcript of File: df11-2004 - Windows Internet Explorer". The address bar shows "K:\Tesis\TextGrids\df11.html". The main content area displays a transcription of an audio file with the following structure:

- *EDU: ☰ no /
 - ☐ yo también /
 - ☐ hay muchos lugares que no conozco y después me dicen /
 - ☐ [%qlt: paratone-falsetto] (ay /
 - ☐ muy bonito) ...
 - #
- *MIR: ☰ [%act: laugh] //
- #
- *ADR: ☰ si /
 - ☐ y vives aquí y no < conoces > ?
 - #

Below the transcription is an audio player interface. It features a spectrogram with the text "si / y vives aquí y no < conoces > ?" and a "Play audio fragment" button. The x-axis is labeled "Time (s)" and ranges from 0 to 2. The y-axis is labeled "Fundamental frequency (Hz)" and ranges from 150 to 400. The spectrogram shows the frequency components of the speech signal over time.

At the bottom of the page, there is a list of transcription units with their corresponding time intervals and markers:

- *MIR: ☰ < no > conoces ...
- *MIR: ☰ aquí también ya nos llevaron /
 - ☐ no ?
 - ☐ a Taxco +/
 - #
- *ADR: ☰ también a los gigantes de Tula /
 - #
 - ☐ a Puebla < también > ...
 - #

Fuente: Diseño propio

2.3.3.4 Análisis detallado en Praat

El proceso que hasta ahora se ha descrito, exceptuando el proceso de delimitación y transcripción de los intervalos de la rejilla de texto, es completamente automático y con sólo hacer un par de clics el usuario obtendrá una serie de archivos de imagen, fragmentos de audio y rejillas de texto, aparte de los documentos que contendrán toda la información y los enlaces a cada uno de los archivos. Sin embargo, tras echar un vistazo a las imágenes con los resultados del análisis frecuencial, el usuario se dará cuenta de que, especialmente cuando se trata de grabaciones de habla espontánea sin las condiciones ideales de grabación, tal y como sucedió en esta investigación, los análisis automáticos no siempre tienen la calidad suficiente como para obviar la labor del lingüista.

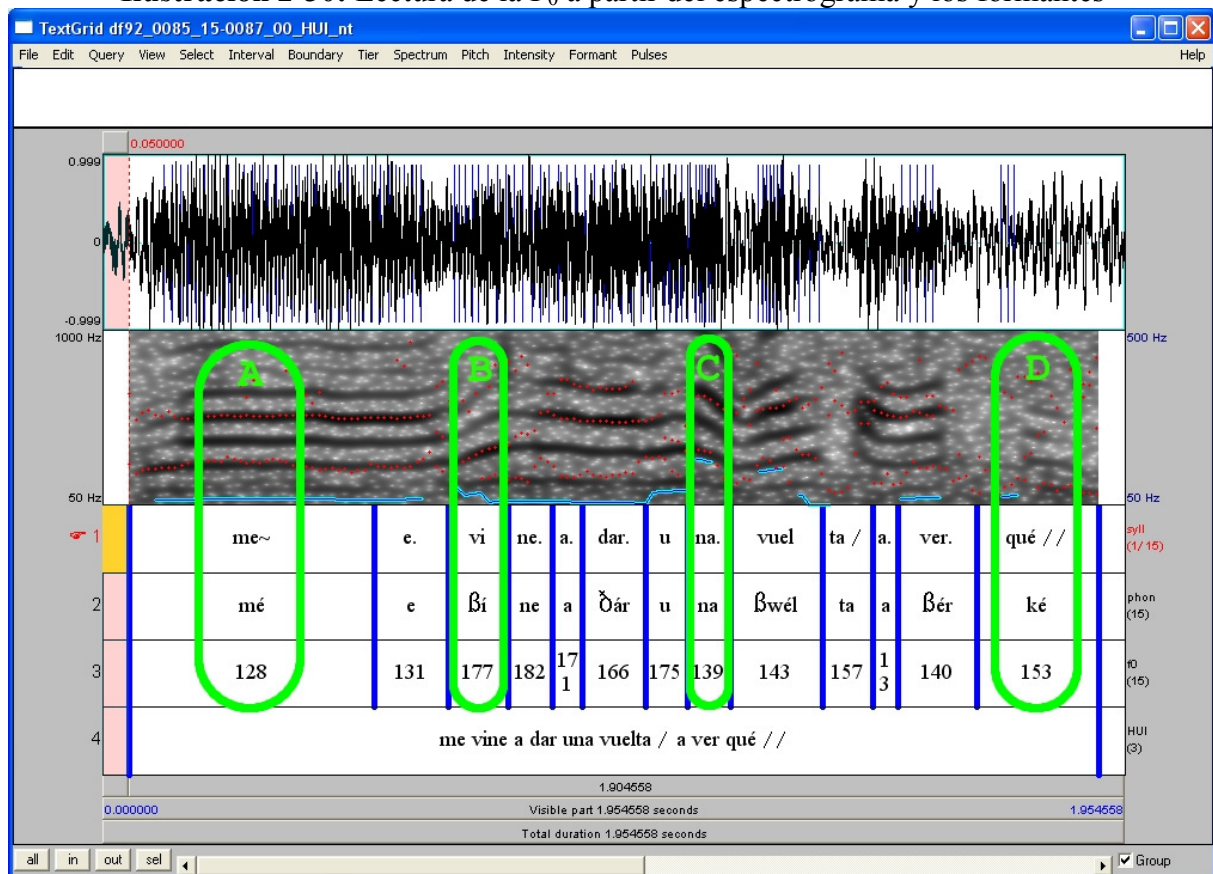
Aunque, por ejemplo, el contorno de la frecuencia fundamental presente irregularidades, huecos, saltos de octavas o trazos aparentemente imposibles de interpretar (véase la Ilustración 2-30), hay distintas técnicas con las que el usuario puede obtener el valor real de F_0 . Una de ellas es la que propone Font Rotchés (2005: 155-158), que consiste en verificar que el número de pulsos seleccionados por Praat como candidatos idóneos concuerde con el número real de pulsos visibles en la onda sonora y calcular el valor real de la F_0 por medio de una regla de tres. Sin embargo, con algunas de las grabaciones del *corpus* de esta investigación, esta técnica es difícil de aplicar porque Praat no siempre puede extraer automáticamente el número de pulsos (véase, por ejemplo, el área D en la Ilustración 2-30).

Con las preferencias de visualización que fueron previamente establecidas (ejemplo 2.44 e Ilustraciones 2-17 a 2-20), en el espectrograma de la Ilustración 2-30 se puede apreciar las franjas oscuras formadas por las frecuencias formánticas (en realidad, frecuencias armónicas) y los puntos rojos que representan los mejores candidatos de los valores para cada uno de los ‘formantes’. Lo que llama la atención en esta Ilustración es que, aunque la frecuencia fundamental sea tan irregular, las franjas negras en el espectrograma tienen, por lo general, un contorno completamente regular. Este patrón de frecuencias se repite con la misma regularidad en otras frecuencias armónicas.

La frecuencia fundamental (Pitch) que se puede ver según estas preferencias es la que se encuentra entre el rango de 50 a 500 Hz, mientras que el espectrograma y los armónicos se despliegan en un rango de 50 a 1000 Hz, lo que hace que, por ejemplo, debajo de la F_0 del área C se vea en realidad el segundo armónico y el primer armónico (la F_0) en el espectrograma es la franja que pasa a la mitad de la altura por debajo del contorno de la F_0 , es decir, la línea azul. El segundo armónico es el que normalmente se ve con más claridad y el que servirá de base para calcular todos los valores de F_0 faltantes. En el caso del área A,

solamente se pueden percibir claramente dos armónicos, aunque el segundo sea más regular que el primero. Si apretáramos la tecla F5, obtendríamos la frecuencia fundamental promedio de [e]: 64.21 Hz. Utilizando la técnica de Font Rotchés (2005), veríamos que Praat únicamente detectó la mitad de los pulsos y que este valor tendría que ser multiplicado por dos, lo que daría un resultado de 128.42 Hz. Ahora, si apretáramos la tecla F1, obtendríamos el primer formante (segundo armónico): 255.90 Hz que, en este caso, tendría que ser dividido entre dos y daría un resultado de 127.95 Hz. Con la tecla F2 obtendríamos el segundo formante (tercer armónico), en este caso, correspondiente a la cuarta frecuencia del espectrograma: 516.88 Hz que dividiríamos entre cuatro y nos daría como resultado 129.22 Hz. Entonces, para este intervalo tenemos tres evidencias de que la frecuencia fundamental es de aproximadamente 128 Hz. Para el área marcada con B lo más conveniente sería seleccionar el final del intervalo, si apretáramos la tecla F1, obtendríamos la frecuencia fundamental o el resultado de apretar la tecla F2 tendríamos que dividirlo entre dos. En resumen, no importa cuál técnica se utilice, siempre y cuando se identifique de cuál frecuencia se trata, se obtendrá un valor muy similar.

Ilustración 2-30: Lectura de la F_0 a partir del espectrograma y los formantes



Fuente: Diseño propio

Por lo tanto, los documentos básicos de XML y su representación en HTML pueden ser útiles para ordenar todos los intervalos transcritos de una grabación y para facilitar la selección de los fragmentos de interés para una investigación. Sin embargo, lo ideal es que estos fragmentos también reflejen la estructura silábica y permitan, por ejemplo, análisis de la entonación y el ritmo. Para este fin, se creó en el marco de esta investigación el subprograma `Create_XML(U+).praat` (Anexo 6) que leerá todos los archivos de sonido y de rejilla de texto en un directorio especificado y, en caso de que estos últimos aún no hayan sido segmentados silábicamente, el subprograma guiará al usuario para que realice dicho análisis. Una vez terminado, el subprograma extraerá toda una gama de valores acústicos que serán impresos con una sintaxis similar a la del XML en la ventana de información (Info Window).

A continuación se presentará a grandes rasgos los procesos que este subprograma realiza:

- Definición del formulario

En esta ventana se solicitará información que puede dividirse en cuatro áreas: 1) ubicación del directorio, 2) selección de opciones que permitan o eviten la creación de imágenes, el análisis de saltos de octavas y la segmentación manual de sílabas, y 3) ubicación donde se guardarán los archivos de rejilla de texto, imágenes y XML:

Ilustración 2-31: Formulario de `Create_XML(U+).praat`

Fuente: Diseño propio

- Limpieza de la Info Window

Para evitar errores de sobrescritura en el nuevo documento de XML, primeramente se limpiará todo el contenido de la Info Window.

- Impresión del encabezado de documento de XML

Con este proceso se escribe en la Info Window las primeras líneas obligatorias de un documento de XML conforme a la definición `PROSODY_XML_SET.DTD` (véase Anexo 3)

- Creación de lista de archivos de sonido en el directorio especificado

Para iniciar con el ciclo de procesamiento de los fragmentos preseleccionados de una grabación, se creará una lista de los archivos de sonido en el directorio especificado en el formulario.

- Inicio del ciclo para procesar todos los archivos especificados y lectura de los archivos TextGrid y de sonido con el mismo nombre de base

Con este proceso se seleccionará y abrirá el primer par de archivos de sonido y de rejilla de texto. Los siguientes procesos se repetirán por cada par de archivos.

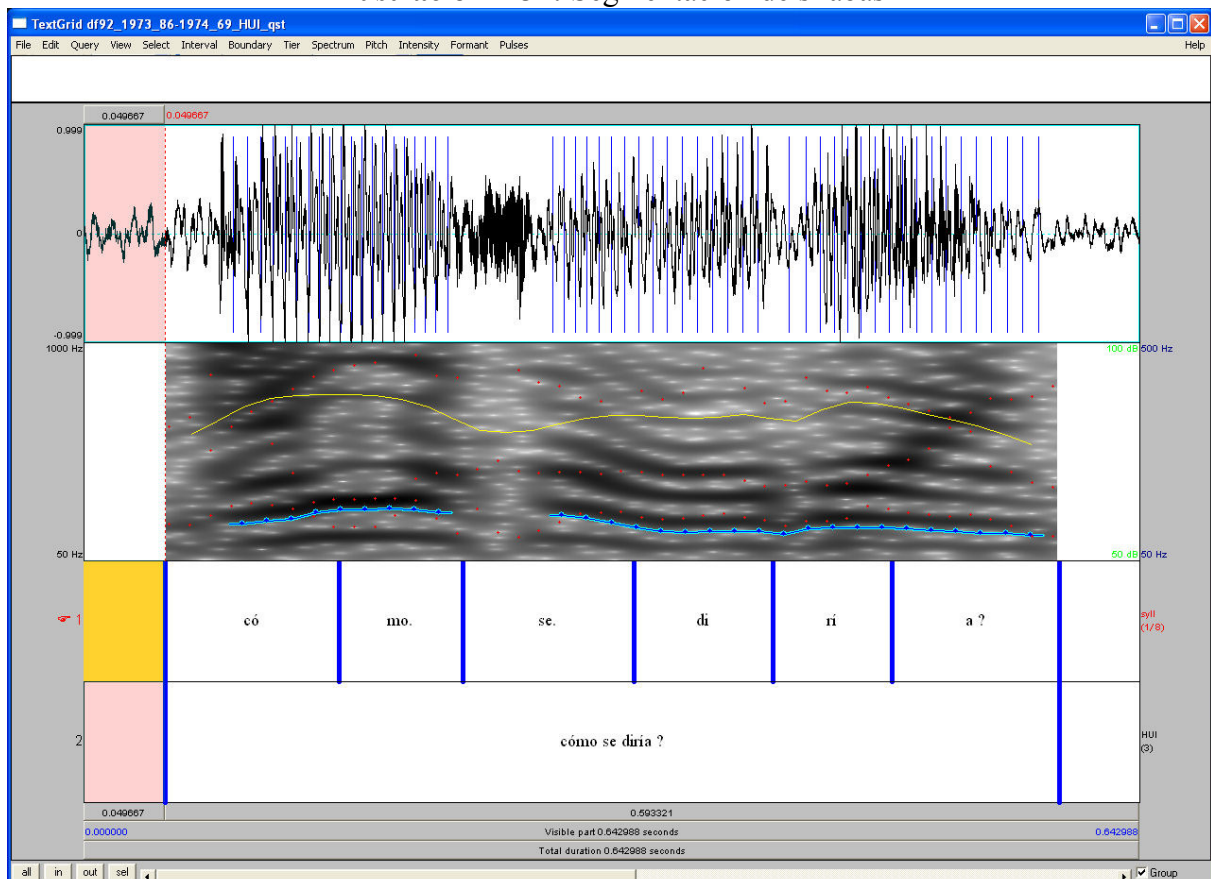
- Ajuste de la opción de segmentación manual a las características de la TextGrid

A pesar de la opción del formulario donde se puede elegir si se desea segmentar los fragmentos por sílabas o no, el subprograma revisará si el archivo de TextGrid actual está completamente segmentado y analizado, si no lo está, procederá a prepararlo para el proceso de segmentación.

- Segmentación de sílabas y asignación de valores paso a paso

Aquí es donde inicia el proceso paso a paso con el que se podrá segmentar las sílabas que componen un fragmento (marcando las fronteras de palabra con ‘.’), transcribirlas fonéticamente; anotar su frecuencia fundamental, dejar que la computadora calcule la variación de la frecuencia fundamental de cada sílaba con respecto a la anterior, y que le asigne el valor estándar de la F_0 comenzando cada enunciado con 100 Hz (véase 1.1.1.2 y 3.2.4), así como anotar la estructura rítmica y tonal. En la Ilustración 2-32 se muestra una ventana del editor de TextGrids en el que ya se han separado las sílabas del enunciado.

Ilustración 2-32: Segmentación de sílabas



Fuente: Diseño propio

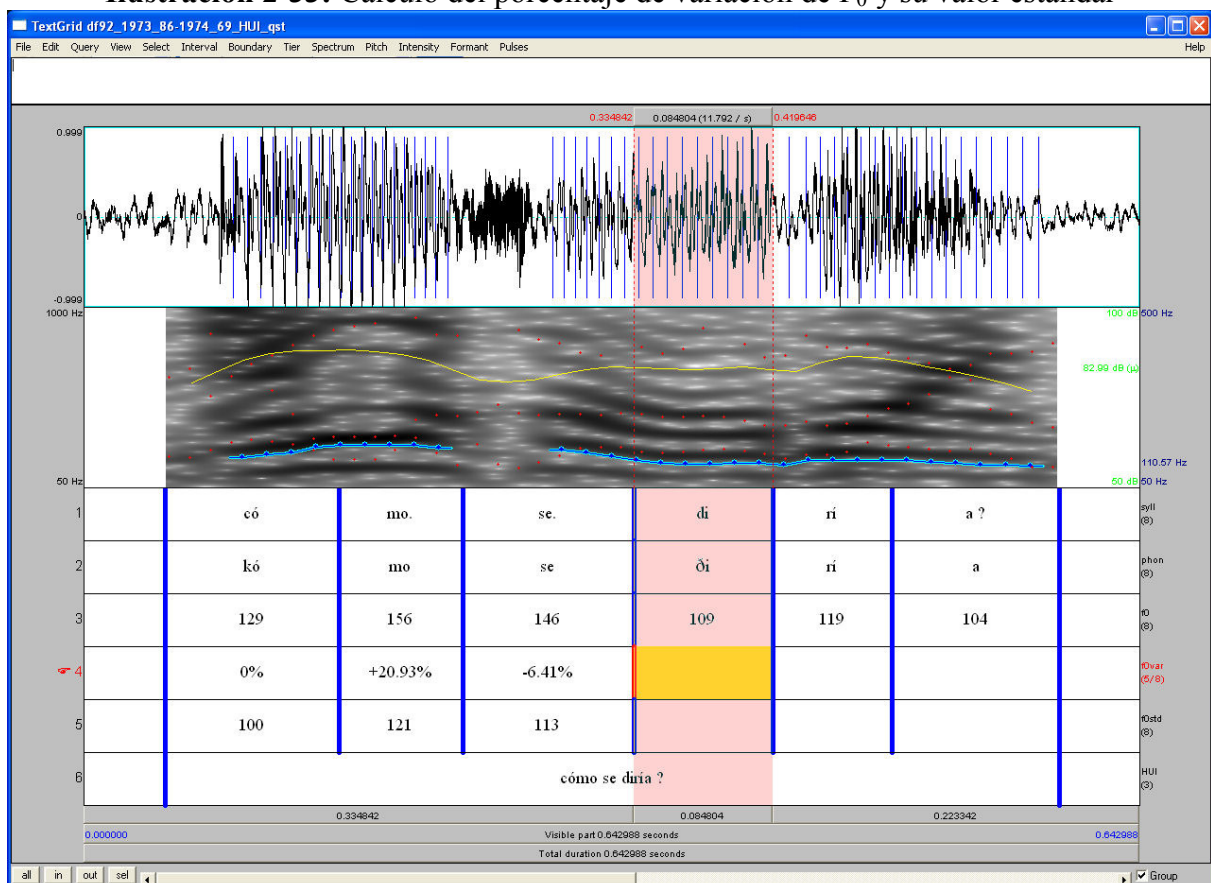
- Localización y extracción de la F_0 del intervalo actual, el siguiente y los dos anteriores

La anotación de los valores de F_0 deberá hacerse manualmente para evitar errores debidos al análisis automático. El subprograma localizará y asignará variables a los valores de F_0 de la sílaba actual y las dos anteriores. El primero y el último intervalo siempre estarán vacíos.

- Cálculo del porcentaje de variación y valores estándar de los valores introducidos en la hilera f_0 y extracción de valores para las variables de frecuencia

Con estas instrucciones se calculará la variación porcentual de F_0 de una sílaba con respecto a la sílaba anterior y se les asignará un valor estandarizado. El valor de F_0 de los intervalos etiquetados sin frecuencia perceptible deberá marcarse con 0, así, la F_0 estandarizada, f_0std , será el promedio de los intervalos anterior y posterior y se escribirá entre paréntesis. En las TextGrids que contengan más de un enunciado, los valores de f_0var y f_0std se reestablecerán en 0% y 100. La frecuencia estandarizada de cada uno de los fragmentos que contenga múltiples enunciados será analizada de manera individual. En la Ilustración 2-33 se puede apreciar el editor de TextGrids a la mitad del proceso de cálculo del porcentaje de variación de F_0 y la escritura de valores estandarizados (véase 1.1.1.2 y 3.2.4).

Ilustración 2-33: Cálculo del porcentaje de variación de F_0 y su valor estándar



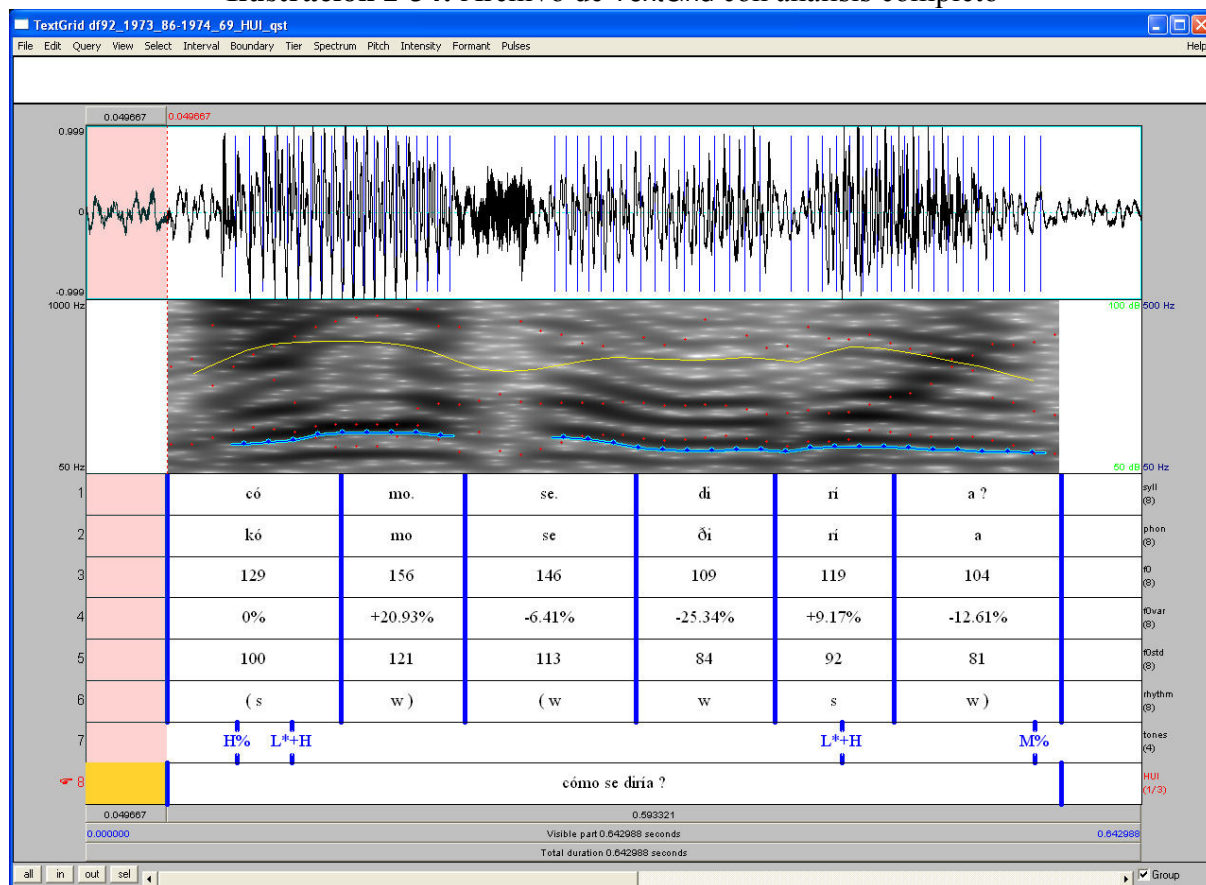
Fuente: Diseño propio

- Escritura del archivo de TextGrid completado

Después de que se ha completado manualmente cada una de las hileras se guardará la nueva versión del archivo TextGrid. Las ocho hileras de las que se compondrá un archivo TextGrid son: syll – transcripción de las sílabas según las normas descritas en 2.3.3.1 y Anexo 1, phon – transcripción fonética, f0 – frecuencia fundamental, f0var – porcentaje de variación de F_0 , f0std – valor estandarizado de F_0 , rhythm – estructura rítmica según la notación descrita en 1.3.4,

tones – estructura tonal según el Sp-ToBI, descrito en 1.2.4.1 y 1.2.4.3, así como la transcripción original del fragmento de grabación, como se ejemplifica en la Ilustración 2-34.

Ilustración 2-34: Archivo de TextGrid con análisis completo



Fuente: Diseño propio

- Extracción de las hileras phon, f0, rhythm, tones y la ortográfica. Conversión de las hileras en una nueva TextGrid para utilizarse en las imágenes

En este proceso se eligen cinco de las ocho hileras y se asocian en una nueva TextGrid que después será utilizada en la creación de las imágenes de análisis acústico.

- Elaboración de los objetos Pitch, Intensity y Spectrogram

Para la creación de las imágenes también se extraerán los objetos de altura tonal, intensidad y el espectro.

- Redondeo de la frecuencia mínima para establecer una base más holgada y redondeo de la frecuencia máxima para establecer una cima más holgada para las imágenes

Para que en las imágenes la frecuencia fundamental y el espectrograma se desplieguen dentro del marco de la ventana lo más ampliamente posible pero con un ligero margen superior e inferior, se redondearán los valores de la frecuencia mínima y máxima. En el caso del espectrograma, este valor será duplicado para que la imagen de fondo corresponda a la F_1 , que en las grabaciones del *corpus* de esta investigación suele ser más clara.

- Creación de imágenes

En esta sección se combinan los distintos objetos con un formato preestablecido para crear una imagen por cada fragmento de grabación.

- Escritura de la imagen a archivos

La imagen creada en la ventana Picture se guardará como metaarchivo de Windows – EMF o WMF, que después será transformada al formato GIF y comprimida para que la imagen pueda utilizarse en un navegador de HTML. Además, se guardará una copia de seguridad en el formato nativo de Praat – .PRAAPIC.

- Limpieza de la ventana de imágenes antes de pasar al siguiente intervalo

Este proceso se realiza para evitar que se generen imágenes defectuosas.

- Extracción de información de la TextGrid

A partir de aquí se extraerá la información sobre las variables acústicas de cada intervalo, correspondientes a sílabas o pausas intermedias en el archivo de TextGrid.

- Predefinición de variables en posición 0

Al igual que los valores de las variables de F_0 , se establecen valores predefinidos para las variables de intensidad y duración.

- Inicio del ciclo que procesa todos los intervalos de la hilera ortográfica silábica, *syll* (1)

Con este ciclo se representa en el documento de XML cada intervalo como una sílaba con sus respectivas variables acústicas.

- Interpretación de los rasgos de frontera y de enunciado en el intervalo actual

En este proceso se identifica e interpreta los signos convencionales colocados al final de cada sílaba y crea el código correspondiente para representarlo dentro del documento de XML.

- Obtención de los valores de la TextGrid para el intervalo actual, el previo y los dos posteriores

Para poder identificar las pausas delimitadas apenas durante el proceso de segmentación, se extraen valores del intervalo actual, el anterior y los dos siguientes.

- Extracción e interpretación de la hilera rítmica, *rhythm* (6)

Los valores de la hilera *rhythm* son clasificados según su prominencia silábica, *s* vs. *w*, y según la presencia o ausencia de foco, *s* vs. *S*, para obtener el peso silábico e indicar la focalización (véase la Tabla 1-7) que serán recogidos en variables.

- Extracción de los puntos en la hilera tonal, *tones* (7), durante el intervalo actual

En cada intervalo se contabiliza el número de tonos transcritos durante la producción de la sílaba actual y, en caso de haberlos, se extraerán y recogerán en variables.

- Extracción e interpretación de los valores de las variables de duración

Para interpretar los valores de la variable *durvar*, se asigna de manera intuitiva y provisoria el valor relativo a una variable denominada *tmp\$*.

- Extracción e interpretación de los valores de las variables de intensidad

De manera similar, se interpreta el valor de la variable *dBvar* y, según la intensidad relativa con respecto a la sílaba anterior, se asignará el valor a la nueva variable *vol\$*.

- Extracción y simplificación de los valores de las variables de frecuencia

Todos los valores de las variables de frecuencia que por algún motivo pudieran haber obtenido un valor *undefined* serán cambiados por 0.

- Inicio de la impresión de las muestras de grabación

Se procederá a escribir el código necesario para marcar el inicio de cada fragmento de grabación.

- Inicio de la impresión de código compatible con XML por cada sílaba

Haciendo uso de todas las variables especificadas con anterioridad, se imprimirá en la Info Window el texto necesario para que los atributos de `s` y sus valores correspondientes sean interpretados como código bien formado y válido de XML.

- Interpretación de la puntuación en la hilera ortográfica por sílabas para obtener los valores de sus disyunciones (*breaks*)

Con estas instrucciones se establece el tipo de disyunción entre la sílaba actual y la siguiente según las marcas interléxicas (.) agregadas durante la segmentación.

- Interpretación e impresión de la frontera de unidad entonativa o enunciado

Este proceso lee los signos gráficos utilizados al final de la sílaba para indicar el final de una unidad entonativa o enunciado y se crea el código necesario.

- Selección del archivo TextGrid para analizar el siguiente intervalo

Con el fin de que el proceso cíclico continúe, es necesario volver a seleccionar el archivo TextGrid, puesto que este objeto estaba seleccionado antes de que iniciara el ciclo.

- Fin del ciclo de intervalos (*syll*)

Con esta instrucción se termina el análisis automático de sílaba por sílaba.

- Impresión del final de una muestra de grabación individual

En la Info Window se imprimirá las etiquetas necesarias para que el texto sea interpretado como el final de un fragmento individual de grabación.

- Selección y eliminación de todos los objetos generados por Praat para la presente muestra de grabación

Para evitar que el subprograma se colapse debido al número de objetos abiertos en la ventana Praat objects, se seleccionará y eliminará los objetos generados automáticamente por Praat.

- Fin del ciclo de archivos (*i_file*)

Aquí termina el análisis e impresión de un fragmento individual de grabación y se pasará a procesar el siguiente.

- Impresión del final del documento de XML

Con esta instrucción se imprime las etiquetas necesarias para indicar el final de un documento de XML.

- Escritura del contenido de la Info Window a un archivo de texto con terminación .XML y fin del subprograma

El documento resultante se guardará en el directorio especificado con el nombre predeterminado y el subprograma se cerrará.

Además del subprograma `Create_XML(U+).praat`, durante el transcurso de esta investigación se desarrolló y aplicó una serie de herramientas para facilitar la revisión, corrección y división de los fragmentos de grabación y rejillas de texto.

Primeramente, `ReviewTextGrids.praat` desplegará un formulario en el que se solicita al usuario que especifique el directorio donde se encuentran los archivos de audio y de rejilla de texto. Todos los archivos de audio en el directorio especificado se enlistarán y ordenarán. Un ciclo procesará los archivos encontrados, los leerá y los abrirá. Un mensaje de alerta aparecerá: `'When ready, press 'Continue''`. Al hacer clic sobre dicho botón, el archivo `TextGrid` actual será guardado y se procederá a abrir el siguiente hasta terminar con la lista.

Por otra parte, `RecalculateF0var&F0std.praat` es una herramienta creada para corregir los archivos de rejilla textual que ya han sido segmentados silábicamente y que contienen 8 hileras, donde los valores de `f0` están en la hilera 3, los de `f0var` en la hilera 4 y los de `f0std` en la 5. Los valores de estas dos últimas hileras serán calculados nuevamente. La ventana del editor de `TextGrids` deberá cerrarse antes de apretar el botón o correr el programa y volverá a abrirse automáticamente cuando el proceso esté listo.

Finalmente, `ExtractSound&TextGrid.praat` extraerá los fragmentos de sonido y de rejilla de texto seleccionados dentro de la ventana del editor de `TextGrids`. Debido a que no existe una función inherente para realizar este proceso directamente en la ventana del editor de `TextGrids`, únicamente se extraerá la selección de sonido y aparecerá un mensaje en el que se pedirá al usuario que seleccione la `TextGrid` correspondiente en la ventana de objetos. Los dos nuevos archivos también deberán ser guardados manualmente por el usuario.

2.3.3.5 Documentos de XML resultantes

El producto de la transformación de cada uno de los pares de archivos de sonido y de `TextGrid` por medio del subprograma `Create_XML(U+).praat` tendrá una estructura completamente compatible con la definición de tipo de documento `PROSODY_XML_SET.DTD` (Anexo 3, Ilustración 2-13, Ejemplo 2.45): después de las tres líneas de la declaración del tipo de documento (`DOCTYPE`) se abrirá el elemento raíz, en este caso `Transcript`, compuesto por un elemento `Set`, que a su vez se compone de varios elementos `Sample` (en el ejemplo 2.45 sólo se reproduce uno). Cada elemento `Sample` tendrá un atributo `Name` y uno o varios subelementos `U`, donde se habrán escrito los valores de los atributos `Quest`, `Emph` y `Susp` (cfr. 1.2.5.6 y 1.2.5.7). Los elementos `U` se compondrán de uno o más elementos `IU` con atributos `img` e `id`, que se complementarán con el elemento `Border`, ubicado al final de cada unidad entonativa y enunciado. Por su parte, los elementos `IU` tendrán como descendientes directos a los elementos `F`, que incluirán cada uno solamente una sílaba o elemento `S`. En `F` se especificará el peso silábico, mientras que el subelemento `S` tendrá una lista de atributos acústicos (véase 2.3.2) con sus respectivos valores. El contenido de `S` serán datos de

caracteres y será lo que componga el texto de la transcripción. Finalmente, habrá un elemento `Break` con el que se especificará el tipo de disyunción de cada sílaba.

(2.45)

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<!-- Written by Eduardo Velázquez -->
<?xml-stylesheet type='text/xsl' href='Prosody.xsl'?>
<!DOCTYPE Transcript SYSTEM "PROSODY_XML_SET.DTD">
<Transcript>
  <Set>
    <Sample Name="HUI">
      <U Quest="Y" Emph="N" Susp="N">
        <IU img="df92_1973.86-1974.69_HUI_qst" id="df92_1973.86-1974.69_HUI_qst">
          <F Wt="( s">
            <S Beg="0.168" End="0.274" Dur="0.106" Durvar="0" Tmp="nt" F0min="129"
              F0max="154" F0dBmax="153" F0="129" F0var="0" F0vst="0" F0std="100"
              dBmin="77" dBmax="88" tdBmax="0.268" dB="84" dBvar="0" Vol="nt"
              Stress="Y" Focus="N" Tone="H% L*+H" Phon="kó" Orth="có">có</S>
            <Break Type="0"/>
          </F>
          <F Wt="w )">
            <S Beg="0.274" End="0.349" Dur="0.076" Durvar="-28.35" Tmp="all"
              F0min="140" F0max="157" F0dBmax="155" F0="156" F0var="20.93"
              F0vst="2.51" F0std="121" dBmin="82" dBmax="88" tdBmax="0.278" dB="86"
              dBvar="2.31" Vol="nt" Stress="N" Focus="N" Tone="" Phon="mo"
              Orth="mo.">mo.</S>
            <Break Type="1"/>
          </F>
          <F Wt="( w">
            <S Beg="0.349" End="0.453" Dur="0.104" Durvar="37.36" Tmp="rall"
              F0min="121" F0max="146" F0dBmax="126" F0="146" F0var="-6.41" F0vst="-
              0.77" F0std="113" dBmin="79" dBmax="84" tdBmax="0.442" dB="81"
              dBvar="-5.89" Vol="pp" Stress="N" Focus="N" Tone="" Phon="se"
              Orth="se.">se.</S>
            <Break Type="1"/>
          </F>
          <F Wt="w">
            <S Beg="0.453" End="0.538" Dur="0.085" Durvar="-18.40" Tmp="nt"
              F0min="109" F0max="113" F0dBmax="110" F0="109" F0var="-25.34"
              F0vst="-3.04" F0std="84" dBmin="83" dBmax="84" tdBmax="0.524" dB="83"
              dBvar="2.44" Vol="nt" Stress="N" Focus="N" Tone="" Phon="dhi"
              Orth="di">di</S>
            <Break Type="0"/>
          </F>
          <F Wt="s">
            <S Beg="0.538" End="0.610" Dur="0.072" Durvar="-14.71" Tmp="nt"
              F0min="105" F0max="119" F0dBmax="119" F0="119" F0var="9.17"
              F0vst="1.10" F0std="92" dBmin="81" dBmax="87" tdBmax="0.587" dB="84"
              dBvar="1.58" Vol="nt" Stress="Y" Focus="N" Tone="L*+H" Phon="rí"
              Orth="rí">rí</S>
            <Break Type="0"/>
          </F>
          <F Wt="w )">
            <S Beg="0.610" End="0.713" Dur="0.102" Durvar="41.38" Tmp="rall"
              F0min="104" F0max="117" F0dBmax="117" F0="104" F0var="-12.61"
              F0vst="-1.51" F0std="81" dBmin="75" dBmax="85" tdBmax="0.620" dB="80"
              dBvar="-4.72" Vol="pp" Stress="N" Focus="N" Tone="M%" Phon="a"
              Orth="a ?">a ?</S>
            <Break Type="4"/>
          </F>
        <Border Type="qst"/>
      </IU>
    </U>
  </Sample>
</Set>
</Transcript>
```

A diferencia de los documentos básicos del apartado 2.3.3.2, estos documentos tienen la suficiente cantidad de datos como para permitir análisis descriptivos y estadísticos más detallados. En la siguiente sección se presentará una hoja de estilo básica con la que se podrá

dar un formato general a los documentos detallados de manera que utilice la mayor cantidad posible de información de la base de datos y la presente en un formato práctico y leíble.

2.3.3.6 Hojas de estilo para cada necesidad

Como se dijo en el apartado 2.2.5, los documentos de XML no están hechos para que los usuarios tengan que leerlos e identificar la información escondida entre y dentro de las etiquetas. Por esta razón, necesitaremos una hoja de estilo que transforme el documento de XML en una presentación de HTML que saque el mejor provecho de la información que más probablemente sea utilizada por el usuario. A continuación se describirá cada uno de los componentes de la hoja de estilo `Prosody-set.xml` (Anexo 11). Esta es una propuesta con la que se intenta presentar la mayor cantidad posible de información en el documento de HTML.

- Plantilla principal que se aplicará a todo el elemento `Transcript`

Con esta instrucción se selecciona el elemento raíz, `Transcript`, que es el nivel al que se aplicarán las reglas generales de formato.

- Predefinición del estilo y color de fuente para los enlaces y el menú emergente

Para hacer la página aún más dinámica y no tener que abrir y cerrar ventanas continuamente, se puede hacer con CSS (*Cascading Style Sheets*, Bos *et al.* 1998) que, al pasar el apuntador del ratón por una sílaba o marca de frontera, el color de fuente cambie a rojo y despliegue automáticamente una cinta amarilla con la misma información que en las ventanas (véase la Ilustración 2-36).

- Código externo de JavaScript `dynamic.js` para el menú emergente y las tablas expandibles de imagen y sonido

Aquí se especifica que el subprograma externo, `dynamic.js`, escrito en el lenguaje JavaScript se utilizará en el documento resultante de HTML para que permita que los valores de todas las sílabas del fragmento se ordenen por categorías y se desplieguen por medio de un menú emergente (véase la Ilustración 2-35). Además, se especifica en este subprograma las instrucciones necesarias para que las tablas de imagen y sonido puedan contraerse y expandirse.

- Título y enlaces al resumen y a las convenciones

Con esta serie de instrucciones se escribirá el título y se colocará un enlace al resumen de fenómenos lingüísticos relevantes y otro a la lista de convenciones que serán colocadas al final del documento.

- Botones para contraer o expandir todas las tablas de imagen y sonido

Aquí se especificará cuáles imágenes deberán utilizarse para representar gráficamente que las tablas de imagen y sonido han sido contraídas o expandidas.

- Estructura básica de los turnos de palabra y los enunciados, ordenados según la clave de identificación de cada unidad entonativa

En esta sección se define el formato y el orden que tendrán los fragmentos de grabación y los enunciados.

- Proceso para permitir que cada una de las tablas de imagen y sonido pueda abrirse de manera independiente

Es aquí donde se realiza el proceso de selección del área que podrá expandirse o contraerse de manera individual por cada fragmento de grabación.

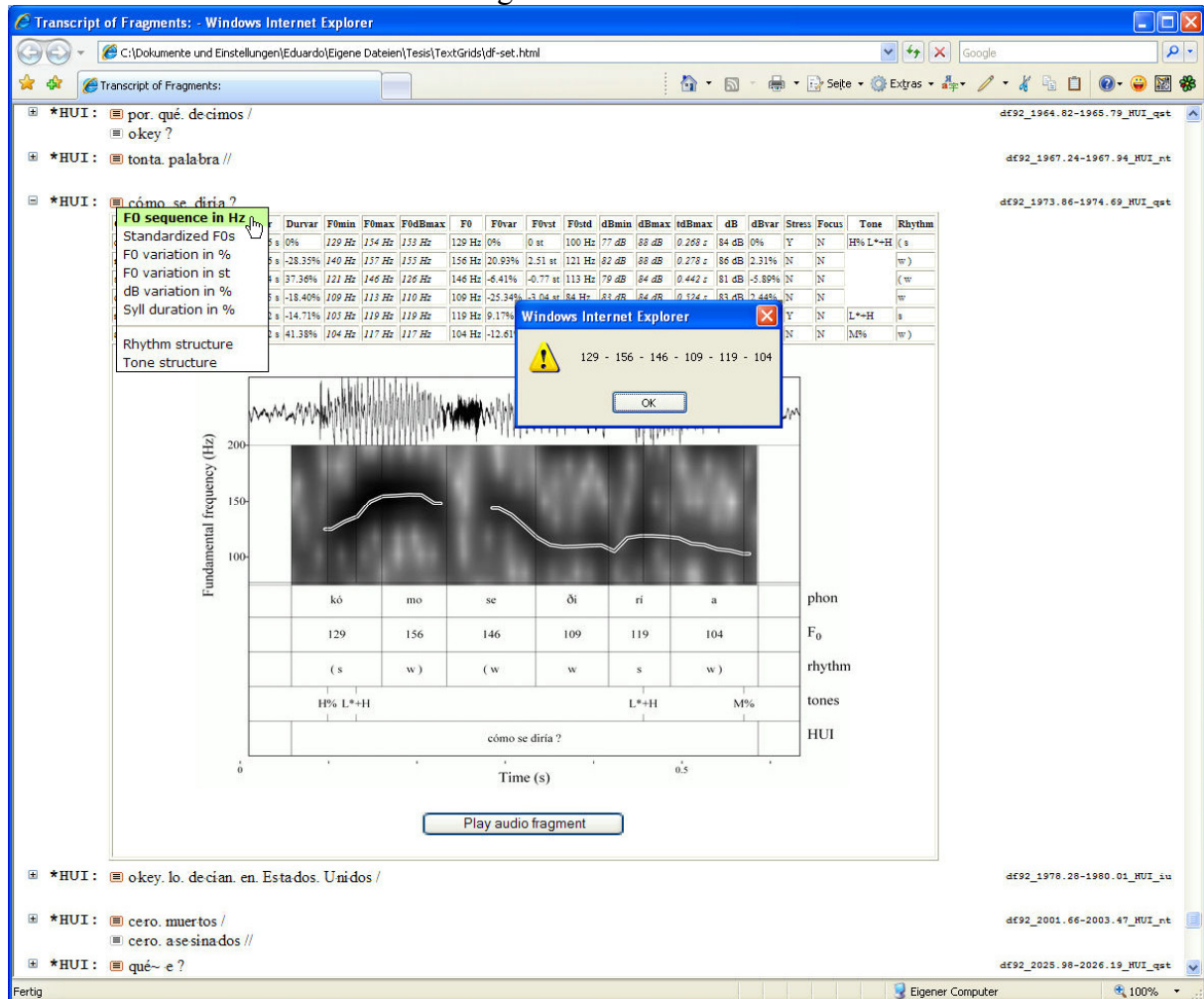
- Asignación de variables a los valores de F_0 , F_0std , F_0vst , F_0var , $dBvar$, $Durvar$, $rhythm$ y $tone$ de cada sílaba

Con estas instrucciones se extrae la información contenida en los atributos del elemento `s` y se guardarán temporalmente en variables que se adaptarán a las características de cada sílaba.

- Valor de cada variable según la manera de ordenarse y desplegarse en el menú emergente

Los valores de estas variables se ordenarán y se formatearán para que puedan ser desplegados en un menú emergente junto con los valores de todas las demás sílabas en el fragmento de grabación. En la Ilustración 2-35 se puede ver cómo se despliega el menú emergente en el documento resultante de la transformación en HTML.

Ilustración 2-35: Menú emergente en el documento transformado en HTML



Fuente: Diseño propio

- Asignación de variables para los atributos `img` e `id` de cada unidad entonativa

Con estas instrucciones se extraerá los nombres de identidad de cada unidad entonativa y serán recogidos en variables temporales.

- Tabla de visión de conjunto con todos los atributos fonéticos de la sílaba

Dentro del área expandible de las tablas de imagen y sonido de cada fragmento de grabación se creará una tabla donde se vaciará los atributos fonéticos de las sílabas que lo componen.

- Tabla de imagen y sonido

También se colocará un enlace a la imagen del análisis acústico y otro al archivo de audio.

- Impresión de la clave de identidad de la unidad entonativa actual

Para facilitar la búsqueda, navegación y administración de los fragmentos de grabación, se imprimirá el nombre de identidad a la misma altura que la tabla de datos acústicos.

- Resumen de fenómenos lingüísticamente relevantes

Al final del cuerpo del documento, conformado por las tablas expandibles de cada fragmento de grabación, se colocará una tabla en la que se contabilice el número de veces que aparezcan determinados fenómenos lingüísticos de interés.

- Convenciones de los signos utilizados en la transcripción para representar los fenómenos lingüísticamente relevantes

Esta sección se omite del código transcrito en el Anexo 11 por ser idéntica a la de `Prosody.xsl` (Anexo 10, líneas 571 a 1392).

- Fin de la platilla principal

Aquí termina el primer paso de la aplicación de formato.

- Despliegue de los datos fonéticos de cada sílaba cuando el apuntador pase sobre ellas y generación de ventanas de alerta cuando sean apretadas

Con estas instrucciones se realiza un proceso por medio del cual se desplegará los valores acústicos de cada sílaba al pasar el puntero del ratón por encima de su representación ortográfica (véase la Ilustración 2-36).


- Despliegue de la duración de la pausa, anclada al signo #

Del mismo modo, cuando haya algún signo que indique la existencia de una pausa, al pasar el puntero del ratón sobre él, se desplegará su duración.


- Interpretación y formato de las disyunciones entre sílabas (sinalefas -0- y disyunciones interléxicas -1-)

Con esta plantilla secundaria se establece el formato que deben presentar las fronteras intersilábicas.

- Asignación de una imagen diferente delante de las unidades entonativas en posición distinta a la primera

Para destacar gráficamente la subordinación de las unidades entonativas pertenecientes a un mismo enunciado, cada una desplegada en una nueva línea, se colocará la imagen  delante de ellas.

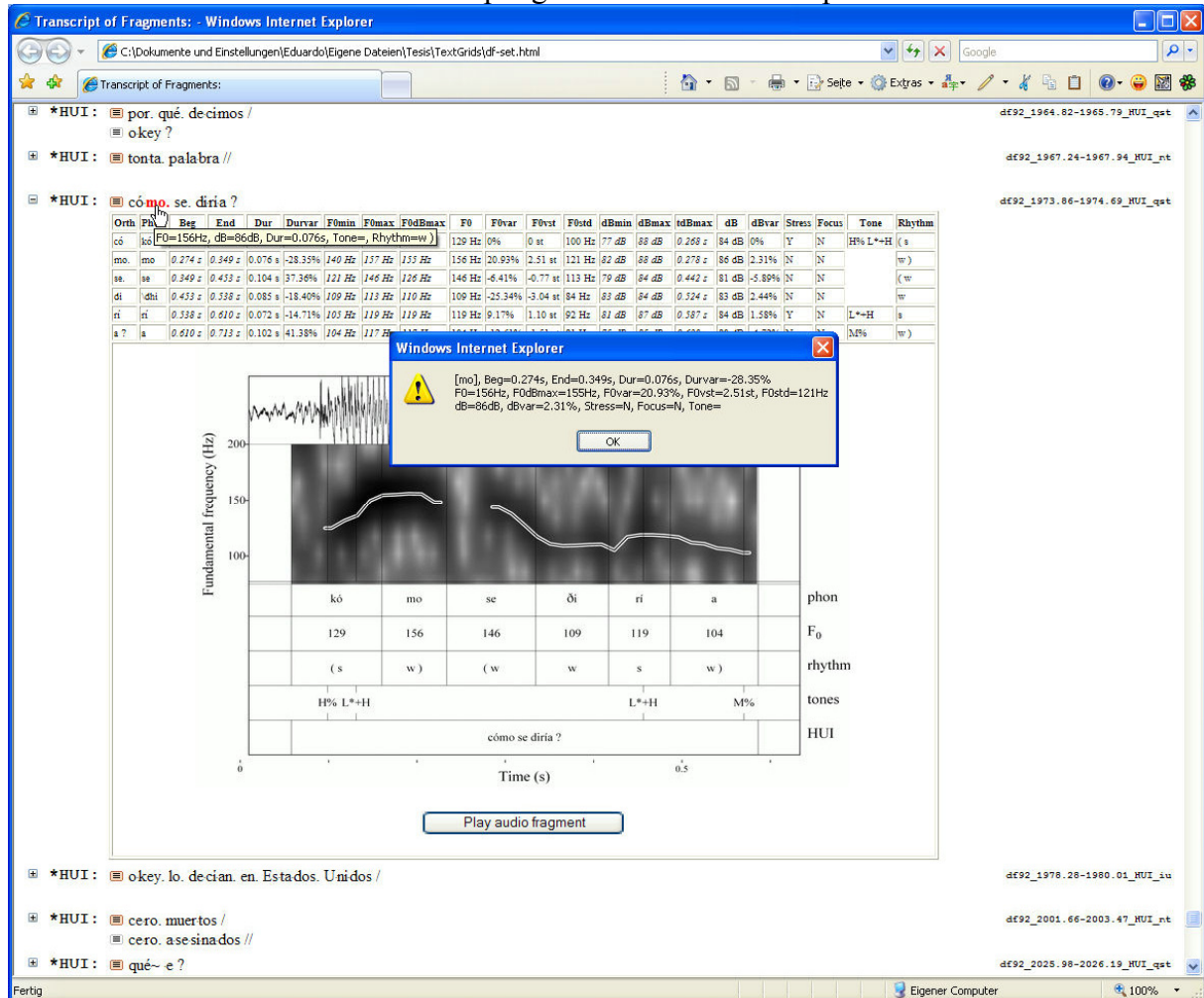
- Asignación de una imagen diferente delante de los enunciados en posición distinta a la primera

Del mismo modo, para destacar gráficamente la existencia de más de un enunciado dentro de un fragmento de grabación, se colocará la imagen  delante de la primera unidad entonativa del enunciado.

- Fin de la hoja de estilo

Con esta etiqueta se finaliza el proceso de aplicación de formato a documentos detallados de análisis prosódico en XML.

Ilustración 2-36: Despliegue de datos fonéticos por cada sílaba



Fuente: Diseño propio

Sin embargo, este es sólo un ejemplo de formato. Para obtener otros resultados con características específicas, se deberá modificar la hoja de estilo recién presentada o hacer una completamente nueva. Por ejemplo, la hoja de estilo `Prosody-set-table.xsl` (Anexo 12) producirá, a partir de la misma fuente, un documento como el de la Ilustración 2-37.

Ilustración 2-37: Resultado en HTML de la hoja de estilo `Prosody-set-table.xsl`

ID	Name	Orth	F0(Hz)	F0var	F0std	Tones	Rhythm	Quest	Emph	Susp
d92_1967.24-1967.94_HUI_nt	HUI	ton	161	0%	100	H% L+H*	(s	N	N	N
		ta.	143	-11.18%	89		w)			
		pa	114	-20.28%	71		(w			
		la	155	35.96%	96	L+H*	s			
		bra //	145	-6.45%	90	M%	w)			
d92_1973.86-1974.69_HUI_gst	HUI	có	129	0%	100	H% L*+H	(s	Y	N	N
		mo.	156	20.93%	121		w)			
		se.	146	-6.41%	113		(w			
		di	109	-25.34%	84		w			
		ri	119	9.17%	92	L*+H	s			
a ?	104	-12.61%	81	M%	w)					
d92_1978.28-1980.01_HUI_iu	HUI	o	163	0%	100	H% L+H*	(0)	N	N	Y
		key.	205	25.77%	126	¡L+H*	(0)			
		lo.	180	-12.20%	110		(w			
		de	132	-26.67%	81		w			

Fuente: Diseño propio

A diferencia del documento creado con la hoja de estilo básica, cuyo formato estaba subordinado a la división por áreas de contenido para permitir el uso de tablas expansibles, en estos nuevos documentos se hace un uso más eficiente de las tablas, en las que se ordena una selección del total de la información proveniente del documento de origen. Este nuevo formato puede ser simplemente copiado y pegado a una hoja en blanco de Excel, tal y como se puede observar en la Ilustración 2-38.

Ilustración 2-38: Formato automático en Excel

ID	Name	Orth	F0(Hz)	F0var	F0std	Tones	Rhythm	Quest	Emph	Susp
d92_1967.24-1967.94_HUI_nt	HUI	ton	161	0%	100	H% L+H*	(s	N	N	N
		ta.	143	-11.18%	89		w)			
		pa	114	-20.28%	71		(w			
		la	155	35.96%	96	L+H*	s			
		bra //	145	-6.45%	90	M%	w)			
d92_1973.86-1974.69_HUI_gst	HUI	có	129	0%	100	H% L*+H	(s	Y	N	N
		mo.	156	20.93%	121		w)			
		se.	146	-6.41%	113		(w			
		di	109	-25.34%	84		w			
		ri	119	9.17%	92	L*+H	s			
		a ?	104	-12.61%	81	M%	w)			
d92_1978.28-1980.01_HUI_iu	HUI	o	163	0%	100	H% L+H*	(0)	N	N	Y
		key.	205	25.77%	126	¡L+H*	(0)			
		lo.	180	-12.20%	110		(w			
		de	132	-26.67%	81		w			
		ci	194	46.97%	119	L+H*	s			
		an.	183	-5.67%	112		w)			
		en.	200	9.29%	123		(w			
		fc	141	-29.50%	87		w			

Fuente: Diseño propio

Estas hojas de cálculo en Excel serán la base del análisis estadístico que se describirá y desarrollará en la sección 3.

2.3.4 Resumen

A lo largo de esta sección se ha demostrado que, a pesar de la existencia de sistemas de análisis lingüístico basados en el lenguaje XML, por cuestiones de arbitrariedad y posibilidades de aplicación más amplias, los creadores de EXMARaLDA y TEI evitan comprometerse teóricamente, y como posible consecuencia, tampoco proponen una estructura lo suficientemente detallada para hacer análisis prosódicos a nivel silábico.

Por esta razón, antes de dedicarme a completar algún sistema existente, con sus posibles ventajas y desventajas, decidí crear un sistema autónomo que abarcara el proceso completo de estudio de la prosodia. Dicho proceso inicia con la transcripción básica de los fenómenos prosódicos relevantes en las grabaciones de interacciones comunicativas. Esta transcripción puede hacerse primeramente en un procesador de texto o directamente en rejillas de texto del programa Praat, donde toda esta información será procesada por medio de subprogramas.

La definición del tipo de documento con información prosódica fue especialmente diseñada para dar cabida a los modelos teóricos en los que sustentaré el estudio comparativo de las variedades de español de la Ciudad de México y Madrid. Sin embargo, el proceso no se detiene ahí, en la concepción del fondo, sino que también adquiere una forma que aleja el texto original de la transcripción del concepto común de documento de texto o página *web*, pues en el producto final no solamente el aspecto visual es importante, sino también el manejo de los datos de manera que faciliten la tarea de análisis. En este documento final se habrán conjuntado y combinado texto, bases de datos, archivos de audio e imágenes, para ofrecer una plataforma fácil de usar, pero muy potente.

Sin embargo, una vez que se haya probado la eficacia y funcionalidad de esta estructura, sería un buen reto intentar acoplar los tres sistemas de etiquetado lingüístico, para aprovechar las fortalezas de cada sistema y superar sus debilidades.

3. Análisis

En esta tercera y última sección se describirá primeramente la muestra de hablantes y el número y duración de sus producciones lingüísticas que fueron tomadas en cuenta para conformar la base de datos Prosodia.xml. La información contenida en esta base de datos se divide en dos *corpora*, uno con todos los datos concernientes a las grabaciones de Madrid, y otro con la información proveniente de las grabaciones de la Ciudad de México. Desde Prosodia.xml se extraerá los valores de las variables acústicas, para después poder ser recopiladas, filtradas, procesadas y promediadas para cada uno de los hablantes o grupos de hablantes.

En el apartado 3.2 se hará una descripción detallada, por medio de tablas de datos, de distintas modalidades de análisis de las variables acústicas provenientes de las grabaciones de hablantes de Madrid. Primero se hará un conteo detallado del número total de sílabas y enunciados que cada hablante aporta al *corpus*. Más adelante se hará la presentación de los resultados del procesamiento de los datos sobre la duración silábica, de donde se obtendrá también un índice sobre la velocidad del habla. Después, se mostrará los datos sobre la intensidad absoluta y se demostrará la utilidad de la estandarización de los valores de la intensidad para permitir las comparaciones entre los resultados de los distintos hablantes, independientemente de las diferencias debidas al uso de distintos equipos o técnicas de grabación. Del mismo modo se abordará el tema de la frecuencia y de las dificultades que la utilización de distintos hablantes implica para un estudio comparativo, pero también, al igual que en el estudio de la intensidad, la manera de solucionarlo por medio de la utilización de valores estandarizados. Además, se hará un recuento de los distintos tonos utilizados en cada uno de los enunciados y, a partir de los resultados totales, los valores se filtrarán por sexos y tipos de enunciados para permitir una mejor identificación de los inventarios de tonos utilizados en distintos contextos sintácticos. Por último, se hará un recuento de los grupos rítmicos más frecuentes en la producción de enunciados, así como del número de grupos rítmicos idénticos contiguos, pues eso es lo que da lugar al principio de isocronía y, por lo tanto, lo que es percibido como el ritmo propiamente dicho.

En el apartado 3.3 se hará el análisis paralelo de las variables acústicas provenientes de las grabaciones de hablantes de la Ciudad de México, aunque no será sino hasta el apartado 3.4 que se hará un análisis comparativo de manera sistemática entre los resultados de los dos *corpora* lingüísticos y se presentará las conclusiones.

3.1 Descripción de la muestra

A lo largo de las secciones anteriores se ha ido anticipando algunas de las características de la muestra de grabaciones de conversaciones espontáneas con o entre hablantes de la Ciudad de México y de Madrid. A continuación se verá con más detenimiento las características de cada uno de los conjuntos.

Las grabaciones de Madrid me fueron cedidas en verano de 2001 por el Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid, dirigido por el Dr. Antonio Moreno Sandoval. Las grabaciones fueron realizadas directamente en formato digital MiniDisc, después fueron copiadas como archivos .WAV y posiblemente fragmentadas.

De un total de 17 grabaciones, con una duración de 2:12:32 horas, se seleccionaron únicamente 6 de ellas, con una duración total de 47:35. Esta selección fue necesaria porque en el resto de las grabaciones participaban personas originarias de otras regiones de España o porque había un número excesivo de solapamientos entre las producciones lingüísticas de los hablantes, lo que hubiera impedido un óptimo análisis acústico de las frases aisladas.

Después de haber sido transcritas en Praat, se obtuvo un total de 14 hablantes, 7 mujeres y 7 hombres con rangos de edad entre los 18 y los 60 años. La distribución de hablantes por grabación y la duración de cada una de las grabaciones pueden ser apreciadas en la Tabla 3-1. Además, en la misma Tabla se puede observar el número de fragmentos seleccionados de cada grabación para hacer el análisis detallado y su duración total.

Tabla 3-1: Hablantes de Madrid

Grabación (Duración)	Participantes (Sexo)	Edad*	Educ.**	Ocupación	Fragmentos (Duración)
md1-LLI-UAM (9:14.71)	GUI (♂)	A	3	lingüista	17 (18.58s)
	MGL (♂)	B	2	peluquero	20 (22.43s)
	PAT (♀)	B	2	peluquera	29 (38.76s)
	ROS (♀)	B	3	docente	23 (29.81s)
md2-LLI-UAM (6:49.59)	FRA (♂)	A	2	inst. de aire acond.	18 (24.78s)
	PTR (♀)	C	3	telefonista	29 (41.83s)
md3-LLI-UAM (7:08.96)	MIG (♂)	A	2	taxista	32 (41.54s)
	PAC (♂)	B	2	inst. de aire acond.	19 (23.09s)
md4-LLI-UAM (9:10.78)	DAN (♂)	B	3	dis. de multimedios	35 (46.06s)
	LET (♀)	B	3	estudiante	41 (53.33s)
md5-LLI-UAM (7:29.16)	GEM (♀)	A	3	estudiante	35 (46.71s)
	SON (♀)	B	2	telefonista	58 (58.81s)
md6-LLI-UAM (7:42.62)	ABE (♂)	A	3	estudiante	9 (10.13s)
	HEL (♀)	A	3	estudiante	35 (54.78s)
Total:					400 (8:30.64)

* A = 18-25, B = 26-40, C = 41-60

** 1 = esc. primaria o secundaria, 2 = esc. preparatoria, 3 = estudiantes o graduados universitarios

Fuente: Diseño propio

Muy probablemente, FRA y PAC son la misma persona, pues coinciden en la edad aproximada, nivel de estudios y sobre todo en la ocupación (técnico en la instalación de

3.1 Descripción de la muestra

equipos de aire acondicionado), pero como en la clasificación original del LLI así es como aparecen, se decidió tratarlos aquí también como dos personas distintas.

Del total de duración de las grabaciones, 2855 segundos (47:35), se aprovecharon únicamente 510.64 (8:30.64), lo cual corresponde a un 17.89%.

Las grabaciones de la Ciudad de México, en cambio, con todas las ventajas y desventajas que esto implica, fueron hechas por mí en el verano de 2004. La situación comunicativa fue, en general, diferente de la que se dio en las grabaciones de Madrid, pues las grabaciones comenzaron como entrevistas libres, aunque la mayoría de las veces terminaron como conversaciones en las que los entrevistados también hacían preguntas y el grado de confianza fue aumentando. Sin embargo, como la muestra se suponía que fuera lo más heterogénea posible, se disponía de pocos días para realizar las grabaciones y sólo se contaba con un número reducido de contactos en la Ciudad de México, se consideró que las mejores condiciones de grabación se darían en la recepción del hotel, dentro de locales comerciales o en las partes del Parque de la Alameda más alejadas de las avenidas y puntos con mayor afluencia de personas. Pocas veces los resultados fueron completamente satisfactorios, pues el ruido de fondo creado por automóviles, fuentes cercanas, personas ajenas a la grabación, objetos con los que los hablantes intentaban disimular su nerviosismo, vendedores de discos con música a todo volumen, músicos callejeros, además de un defecto del micrófono Philips SBC ME570 utilizado en varias de las grabaciones, hizo necesario que las grabaciones en MiniDisc (Sharp IM-DR420H) fueran lo más largas posible, a fin de poder extraer una cantidad mínima de fragmentos con la calidad máxima.

Fue por estos y otros motivos, como la mayor o menor recurrencia de solapamientos o para mantener un número casi equitativo de hombres y mujeres, que de las 21 grabaciones originales, con una duración total de 9:07:02, se decidió elegir únicamente 10 grabaciones con una duración total de 4:19:36.

Estas grabaciones fueron transcritas en Praat y se obtuvo un total de 11 hablantes, 5 mujeres y 6 hombres, con edades entre los 18 y los 60 años y provenientes de distintas clases sociales. De este modo, si bien la muestra fue en un principio aleatoria, al igual que la de Madrid, las grabaciones también fueron elegidas siguiendo un criterio de representatividad. En la Tabla 3-2 se puede observar la distribución de hablantes por grabación y su duración original, su edad aproximada, nivel de estudios, ocupación y el número de fragmentos seleccionados de cada grabación para hacer el análisis detallado, así como la duración total de estos fragmentos.

Tabla 3-2: Hablantes de la Ciudad de México

Grabación (Duración)	Participantes (Sexo)	Edad*	Educ.**	Ocupación	Fragmentos (Duración)
df11-2004 (26:25.83)	ADR (♀) MIR (♀) ^{***}	A A	3 3	repcionista de hotel repcionista de hotel	40 (87.18s) 23 (32.31s)
df21-2004 (45:35.96)	MOI (♂) ROD (♂)	A A	3 3	recepc./ing. en sistemas repcionista de hotel	187 (213.45s) 11 (10.38s)
df22-2004 (21:21.09)	FRA (♂)	B	2	comerciante de artículos fotográficos	78 (122.71s)
df31-2004 (50:40.89)	MIR (♀) ^{***}	A	3	repcionista de hotel	70 (90.98s)
df62-2004 (21:08.92)	ALB (♂) CEC (♀)	B B	3 2	vendedor/ing. químico vendedora/educadora	52 (57.63s) 28 (26.92s)
df71-2004 (9:57.96)	FAB (♀)	A	3	estudiante	13 (14.27s)
df81-2004 (17:32.14)	MAR (♀)	C	1	servienta	72 (108.45s)
df82-2004 (15:44.45)	IVA (♂) ^{***}	C	2	plomero	19 (27.56s)
df83-2004 (14:20.84)	IVA (♂) ^{***}	C	2	plomero	63 (97.12s)
df92-2004 (36:48.47)	HUI (♂)	A	2	técnico en electrónica	144 (190.14s)
Total:					800 (17:59.10)

* A = 18-25, B = 26-40, C = 41-60

** 1 = esc. primaria o secundaria, 2 = esc. preparatoria, 3 = estudiantes o graduados universitarios

*** Misma persona en distintas grabaciones

Fuente: Diseño propio

De todas las grabaciones de la Ciudad de México seleccionadas, con una duración total de 15576.55 segundos (4:19:36.55), únicamente se seleccionaron 800 fragmentos con una duración total de 1079.1 segundos (17:59.10), lo que corresponde a un 6.93%. En comparación con el porcentaje de material de Madrid utilizado para el análisis detallado (17.89%), la proporción de material relevante de la Ciudad de México parece mucho más bajo. Sin embargo, mucho de la diferencia en el porcentaje se debe a mi participación como entrevistador, pues ninguno de estos fragmentos fue tomado en cuenta para el análisis.

Una vez que las grabaciones correspondientes a los hablantes de Madrid pasaron la segunda fase de transcripción y análisis, los fragmentos seleccionados fueron procesados según los pasos descritos en los apartados 2.3.3.4. a 2.3.3.6. y los datos resultantes fueron recolectados en una hoja de cálculo de Excel (véase la Ilustración 2-38), denominada *Prosodia.xls* (distinta de la hoja de estilo *Prosody.xsl*). A cada sílaba de un determinado enunciado le corresponde una fila y las distintas categorías de datos fueron organizados en grupos de columnas subordinadas a dichas filas. Para facilitar el análisis estadístico, se separó las producciones de los hablantes de Madrid y de la Ciudad de México primeramente en dos hojas diferentes, *md* y *df*. En cada una de estas hojas se incluyó la clave de identificación, el nombre del hablante, la transcripción ortográfica de la sílaba, su frecuencia fundamental, la variación porcentual con respecto a la sílaba anterior, la frecuencia estandarizada con valores inicial de 100 Hz, el tono

alineado a la sílaba y su peso rítmico. En el siguiente grupo de columnas se colocó los valores de los rasgos binarios de interrogación, exclamación y suspensión. En otra serie se agrupó los valores del peso rítmico, asignando cada grupo rítmico a una columna. También se separó los tonos para después ser colocados en una sola fila y luego cada tono en una columna diferente. Estos dos grupos de columnas fueron creadas con el fin de analizar las secuencias de grupos rítmicos y las secuencias tonales sin tomar en cuenta las fronteras silábicas. En el siguiente grupo de columnas se colocó la información correspondiente al inicio, final, duración, variación de la duración con respecto a la sílaba anterior, el tempo relativo, la variación de la frecuencia fundamental en semitonos, la intensidad en decibeles, la variación porcentual de la intensidad, el volumen relativo y la presencia de prominencia acentual y focalización por cada una de las sílabas. Finalmente, en una última serie de columnas, se agrupó la transcripción ortográfica de cada sílaba en unidades entonativas y fueron colocadas en una sola fila.

En un paso posterior se agregó tres pares de hojas complementarias: *df+dur*, *df+db* y *df+hz* para la Ciudad de México, así como *md+dur*, *md+db* y *md+hz* para Madrid. En la hojas terminadas en *+dur* fueron agrupados todos los valores de la duración y su variación porcentual. Por otra parte, en las hojas terminadas en *+db* se colocó, en una sola línea, los valores de la intensidad, su variación porcentual y la intensidad estandarizada, para facilitar los análisis secuenciales de los datos. Del mismo modo, en las hojas terminadas en *+hz* fueron colocados los datos de la frecuencia fundamental, la variación porcentual y la frecuencia estandarizada, en filas por cada enunciado de los *corpora* de Madrid y la Ciudad de México.

Por cada tipo de análisis hubo la necesidad de crear nuevas páginas separadas de fórmulas dependientes de las ocho hojas principales. En algunos casos, cada una de esas hojas de cálculo llegó a alcanzar unas dimensiones que eran difíciles de procesar en una computadora de escritorio con un procesador de 1.80 GHz y 1.00 GB de RAM. Este es un aspecto que deberá ser tomado en cuenta en investigaciones futuras con *corpora* de datos mucho mayores. En los siguientes apartados se presentará un resumen estadístico del número de sílabas, unidades entonativas y enunciados, así como de las variables de duración, intensidad, frecuencia fundamental y estandarizada, además de la interacción de las variables para dar cuenta de la entonación y el ritmo. En cada tipo de análisis se dará a conocer por separado los resultados de mujeres y hombres, para después dividir los resultados totales según el tipo de enunciado. Primeramente, se presentará los resultados de los fragmentos producidos por cada uno de los hablantes de Madrid y después se presentará los resultados de los hablantes de la Ciudad de México, para concluir con un análisis comparativo de los valores obtenidos para cada una de las dos variedades del español.

3.2 Corpus de Madrid

En este apartado se dará a conocer los resultados estadísticos de los fenómenos prosódicos de las producciones lingüísticas de los hablantes de Madrid. Primeramente, se hará un recuento de las sílabas y enunciados que conforman el *corpus* de Madrid. Más adelante se presentará los valores promediados de fenómenos tales como la duración silábica, la intensidad, la frecuencia fundamental y estandarizada, para concluir con un análisis de los distintos factores que interactúan en la percepción de la entonación y el ritmo.

3.2.1 Sílabas y enunciados

El número total de sílabas producidas por las mujeres de Madrid es de 2128 (véase Tabla 3-3), repartidas entre 300 enunciados, es decir, un promedio de 7.09 sílabas por enunciado. ROS fue la persona que menos sílabas produjo, 225 en sólo 25 enunciados, y SON fue la que más sílabas pronunció, 366 en 65 enunciados, lo que la coloca muy por debajo del promedio de sílabas por enunciado, con 5.63 sílabas. PTR es la mujer con los enunciados más largos con respecto al promedio de sílabas por enunciado, con un total de 9.39 sílabas.

Tabla 3-3: Promedio de sílabas por enunciado (mujeres)

	Sílabas	Enunciados	Promedio
GEM	324	42	7.71
HEL	325	49	6.63
LET	315	52	6.06
PAT	263	34	7.74
PTR	310	33	9.39
ROS	225	25	9.00
SON	366	65	5.63
Mujeres	2128	300	7.09

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres, ABE es quien está menos representado en el *corpus*, con tan sólo 59 sílabas repartidas en 9 enunciados, mientras que MIG produjo 306 sílabas del total de 1230 (véase Tabla 3-4).

Tabla 3-4: Promedio de sílabas por enunciado (hombres)

	Sílabas	Enunciados	Promedio
ABE	59	9	6.56
DAN	238	42	5.67
FRA	190	19	10.00
GUI	125	20	6.25
MGL	152	22	6.91
MIG	306	37	8.27
PAC	160	26	6.15
Hombres	1230	175	7.03

Fuente: Prosodia.xls (Base de datos propia)

DAN es quien más enunciados produjo, 42, del total de enunciados producidos por hombres de Madrid, que es de 175, dando como promedio general 7.03 sílabas por enunciado. DAN y FRA son quienes más distan del promedio con 5.67 y 10.00 sílabas por enunciado respectivamente.

Por otra parte, del total de sílabas en el *corpus* de Madrid, 3358, repartidas entre 475 enunciados, dan un promedio general de 7.07 sílabas por enunciado. Además, como podrá apreciarse en la Tabla 3-5, los enunciados exclamativos (/I +E -S/) son los menos representados, con únicamente 59 sílabas y 13 enunciados, dando como promedio 4.54 sílabas por enunciado, mientras que los enunciados declarativos o neutros (/I -E -S/) son los que tienen una mayor presencia: 1740 sílabas repartidas entre 239 enunciados. Intuitivamente, se puede afirmar que los enunciados suspendidos (/I -E +S/), con un promedio de 7.41 sílabas por enunciado, son más largos que los enunciados exclamativos, entre los que se cuentan las interjecciones cuya brevedad es muy característica.

Tabla 3-5: Promedio de sílabas por tipo de enunciado

	Sílabas	Enunciados	Promedio
/+I -E -S/	647	100	6.47
/-I +E -S/	59	13	4.54
/-I -E +S/	912	123	7.41
/-I -E -S/	1740	239	7.28
Total	3358	475	7.07

Fuente: Prosodia.xls (Base de datos propia)

En cuanto al número de enunciados de cada tipo que produjeron las mujeres, en la Tabla 3-6 se puede observar que, por ejemplo, del total de 57 enunciados interrogativos (/+I -E -S/), HEL y SON produjeron la mayor cantidad, 13, mientras que PTR sólo produjo uno. Cinco de las hablantes produjeron cada una 2 enunciados exclamativos y LET no produjo ninguno, aunque ella misma es quien más enunciados suspendidos produjo, 20 del total de 65, mientras que PAT no produjo ninguno. Finalmente, SON es quien más enunciados neutros produjo, 44, en contraste con ROS, que sólo produjo 12. Por lo tanto, en orden descendente, las mujeres produjeron una mayor cantidad de enunciados neutros, seguidos de enunciados suspendidos, después los enunciados interrogativos y una menor cantidad de enunciados exclamativos.

Tabla 3-6: Número de enunciados de cada tipo por persona (mujeres)

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres
/+I -E -S/	9	13	9	5	1	7	13	57
/-I +E -S/	2	2	0	2	1	2	2	11
/-I -E +S/	16	12	20	0	7	4	6	65
/-I -E -S/	15	22	23	27	24	12	44	167
Total	42	49	52	34	33	25	65	300

Fuente: Prosodia.xls (Base de datos propia)

Entre los hombres, de un total de 43 enunciados interrogativos (véase Tabla 3-7), PAC produjo 10 de ellos, mientras que MIG sólo produjo 2. También es PAC quien produjo los únicos 2 enunciados exclamativos del total de las producciones masculinas, lo que contrasta con sus 3 enunciados neutros, de un total de 72, mientras que MIG produce 21 de ellos. En cuanto a los enunciados suspendidos, es DAN quien produce la mayor cantidad de este tipo, 22, mientras que ABE y MGL sólo producen uno cada uno.

Tabla 3-7: Número de enunciados de cada tipo por persona (hombres)

	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres
/+I -E -S/	3	9	8	5	6	2	10	43
/-I +E -S/	0	0	0	0	0	0	2	2
/-I -E +S/	1	22	4	5	1	14	11	58
/-I -E -S/	5	11	7	10	15	21	3	72
Total	9	42	19	20	22	37	26	175

Fuente: Prosodia.xls (Base de datos propia)

Al igual que en el caso de las mujeres, los enunciados neutros son los que conforman el mayor grupo de enunciados, seguidos por los enunciados suspendidos e interrogativos, mientras que los enunciados exclamativos ocupan el último lugar.

3.2.2 Duración

La selección de los fragmentos idóneos producidos por cada uno de los hablantes dependió más de la calidad acústica que de estrictos criterios estadísticos de representatividad. Por tal motivo, la duración total de las sílabas producidas por una persona puede distar mucho de la de otra persona (véase Tabla 3-8). En el caso de las mujeres, SON es quien tiene una mayor presencia en el *corpus* con 52.743 segundos de un total de 296.367, mientras que de la producción lingüística de ROS sólo se seleccionaron 27.415 seg. Por otra parte, de entre los hombres, ABE sólo está representado con 9.232 seg, mientras que de DAN se analizaron 41.239 seg de un total de 168.788. Esta cifra coloca las producciones masculinas por debajo de las femeninas, dando un total de 465.155 segundos, 7:45.2 minutos.

Tabla 3-8: Duración total de sílabas por persona (en segundos y minutos)

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Duración total (s)	43.219	50.897	47.915	35.860	38.318	27.415	52.743	296.367	
Duración total (min)	00:43.2	00:50.9	00:47.9	00:38.9	00:38.3	00:27.4	00:52.7	04:56.4	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Duración total (s)	9.232	41.239	22.178	16.773	20.420	37.953	20.993	168.788	465.155
Duración total (min)	00:09.2	00:42.2	00:22.2	00:16.8	00:20.4	00:38.0	00:21.0	02:48.8	07:45.2

Fuente: Prosodia.xls (Base de datos propia)

Si se compara esta cifra con el total presentado en la Tabla 3-1, 8:30.64 min, se verá que hubo una diferencia de aproximadamente 45.52 seg debida a los silencios en los extremos y dentro de los fragmentos de las grabaciones.

Ahora, si nos enfocamos en los valores promediados de la duración máxima y mínima de todas las sílabas producidas por cada hablante, el promedio por persona y el promedio general (véase Tabla 3-9), veremos que, en el caso de las mujeres, el promedio de duración silábica máxima más alto es el de LET, con un promedio de 0.264 seg, mientras que el promedio es de 0.244 seg. La duración mínima promedio de las mujeres es de 0.086 segundos y el caso más extremo es el de PTR, con 0.066 seg. Así, en cuanto a la duración promedio de todas las sílabas producidas por cada mujer, LET, con 0.163 seg, y ROS, con 0.128 seg, son quienes más se alejan del promedio general de 0.149 segundos.

En el caso de los hombres, la duración máxima promedio es de 0.259 seg, la mínima promedio es de 0.089 seg y el promedio general es de 0.152. Los casos más extremos son DAN con una duración máxima promedio de 0.320 seg y un promedio general de 0.190, mientras que FRA tiene una duración mínima promedio de 0.069 y un promedio general de 0.127 segundos.

Tabla 3-9: Duración silábica máxima, mínima y promedio (segundos)

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Dur. máx. promedio	0.235	0.262	0.264	0.233	0.257	0.220	0.230	0.244	
Dur. mín. promedio	0.073	0.099	0.096	0.084	0.066	0.076	0.095	0.087	
Duración promedio	0.138	0.161	0.163	0.142	0.133	0.128	0.155	0.149	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Dur. máx. promedio	0.278	0.320	0.238	0.250	0.224	0.241	0.231	0.259	0.250
Dur. mín. promedio	0.089	0.111	0.069	0.084	0.083	0.079	0.093	0.089	0.088
Duración promedio	0.155	0.190	0.127	0.140	0.145	0.136	0.146	0.152	0.150

Fuente: Prosodia.xls (Base de datos propia)

Únicamente la duración máxima promedio de los hombres es significativamente mayor que la de las mujeres. En total, para Madrid hay una duración máxima promedio de 0.250 seg, una duración mínima promedio de 0.088 seg y un promedio general de 0.150 segundos.

Si tomamos en cuenta los distintos tipos de enunciados (véase Tabla 3-10), veremos que los enunciados suspendidos tienen claramente una duración máxima promedio, 0.301 seg, y un promedio general mayor a los demás, 0.175 seg, lo que nos indica que las sílabas en este tipo de enunciados son más largas. Por el contrario, los enunciados interrogativos, 0.083 seg, y los neutros, 0.084 seg, son los tipos de enunciados con la duración mínima promedio más baja. Sin embargo, en el promedio general son los enunciados exclamativos los que tienen una duración más corta, 0.130 seg, seguidos de los enunciados interrogativos y neutros.

Tabla 3-10: Duración silábica máxima, mínima y promedio por tipo de enunciado (segundos)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Dur. máx. promedio	0.227	0.207	0.301	0.235
Dur. mín. promedio	0.083	0.085	0.099	0.084
Duración promedio	0.138	0.130	0.175	0.143

Fuente: Prosodia.xls (Base de datos propia)

No obstante, si lo que se desea es calcular la velocidad aproximada a la que habla cada persona, es necesario invertir la fórmula (dividir 1 entre el promedio de la duración: $1/\overline{dur}$) para que la hoja de cálculo obtenga el número máximo, mínimo y promedio de sílabas por segundo. El resultado es que los madrileños en general producen un promedio de 6.656 sílabas por segundo, máximo 11.406 síl/seg y mínimo 4.003 síl/seg, aunque según las cifras presentadas en la Tabla 3-11, las mujeres hablan ligeramente más rápido que los hombres, es decir, todos los valores promediados de sílabas por segundo son ligeramente mayores. PTR, con 15.221 síl/seg, y FRA, con 14.460 síl/seg, son quienes tienen los valores más altos con respecto al número máximo de sílabas por segundo, mientras que LET, con 3.785 síl/seg, y DAN, con 3.124 síl/seg, tienen los valores más bajos con respecto al mínimo de sílabas. En promedio, ROS, con 7.785 síl/seg, y FRA, con 7.865 síl/seg, tienen los valores más altos, mientras que LET, con 6.121 síl/seg, y DAN, con 5.270 síl/seg, quienes de hecho participan en la misma conversación, se mantienen en la posición más baja.

Tabla 3-11: Número máximo, mínimo y promedio de sílabas por segundo

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Máx. sílabas x seg.	13.748	10.062	10.423	11.867	15.221	13.193	10.528	11.531	
Mín. sílabas x seg.	4.250	3.816	3.785	4.286	3.898	4.539	4.341	4.091	
Prom. sílabas x seg.	7.230	6.198	6.121	7.018	7.511	7.785	6.433	6.701	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Máx. sílabas x seg.	11.264	9.003	14.460	11.869	12.061	12.615	10.797	11.198	11.406
Mín. sílabas x seg.	3.593	3.124	4.202	4.006	4.469	4.148	4.330	3.862	4.003
Prom. sílabas x seg.	6.442	5.270	7.865	7.136	6.918	7.356	6.837	6.579	6.656

Fuente: Prosodia.xls (Base de datos propia)

Del mismo modo se obtiene la velocidad promedio de los distintos tipos de enunciados. Por ejemplo, en la Tabla 3-12 se observa que los enunciados interrogativos, con un máximo de 12.031 sílabas por segundo, es el que los madrileños producen con mayor rapidez, aunque en promedio son los enunciados exclamativos, con 7.713 síl/seg los más rápidos. Por otra parte, los enunciados suspendidos son claramente los que se producen más lentamente, pues el número mínimo de sílabas por segundo es de 3.319 y el promedio es de 5.702 síl/seg.

Tabla 3-12: Número máximo, mínimo y promedio de sílabas por segundo y por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Máx. sílabas x seg.	12.031	11.743	10.078	11.938
Mín. sílabas x seg.	4.403	4.840	3.319	4.253
Prom. sílabas x seg.	7.221	7.713	5.702	6.976

Fuente: Prosodia.xls (Base de datos propia)

Además, con la información de la base de datos se puede también obtener fácilmente el porcentaje de variación máxima y mínima promedio de la duración de una sílaba con respecto a la anterior (véase Tabla 3-13). En este caso se trata de una localización de los porcentajes

más altos o más bajos y se obtienen promedios separados. En caso de que se promediaran todos los valores juntos, los porcentajes positivos y los negativos se neutralizarían mutuamente. Lo que sí se puede hacer es restar el valor negativo al positivo y de este modo se obtendrá el rango de variación promedio de cada hablante. En el caso específico de la duración, este rango nos indicará si la producción lingüística de un determinado hablante es más regular o más irregular. De entre las mujeres, cuyo rango promedio es de 146.42%, PTR es quien produce las sílabas con una mayor irregularidad, mientras que SON utiliza un rango de variación de la duración promedio más estrecho.

El rango promedio de los hombres es más alto, 155.38%. Sin embargo, incluso el rango de variación promedio de FRA, 188.19%, es más estrecho que el de PTR, 216.06%. Por otro lado, ABE es quien produce las sílabas con mayor regularidad. El rango general de variación promedio de la duración silábica de los madrileños es de 149.72%.

Tabla 3-13: Variación máxima y mínima promedio de la duración silábica

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Var. máxima prom.	138.29%	102.71%	124.20%	79.50%	171.20%	103.97%	81.91%	111.92%	
Var. mínima prom.	-35.86%	-24.89%	-26.92%	-43.11%	-44.86%	-36.51%	-36.42%	-34.50%	
Rango var. prom.	174.15%	127.60%	151.12%	122.61%	216.06%	140.48%	118.33%	146.42%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Var. máxima prom.	118.80%	117.25%	153.66%	127.82%	101.02%	125.99%	120.95%	122.85%	115.95%
Var. mínima prom.	-16.92%	-23.05%	-34.53%	-38.15%	-45.34%	-33.73%	-34.92%	-32.53%	-33.78%
Rango var. prom.	135.72%	140.30%	188.19%	165.97%	146.36%	159.72%	155.86%	155.38%	149.72%

Fuente: Prosodia.xls (Base de datos propia)

Los tipos de enunciados también pueden ser analizados con esta fórmula: los enunciados exclamativos son los que hacen un uso más regular de la duración silábica, mientras que los enunciados suspendidos utilizan un rango de variación más amplio (véase Tabla 3-14).

Tabla 3-14: Variación máxima y mínima promedio de la duración silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Var. máxima promedio	120.57%	62.93%	136.98%	106.07%
Var. mínima promedio	-30.19%	-15.86%	-29.27%	-38.57%
Rango de var. prom.	150.76%	78.79%	166.25%	144.63%

Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, la variación de la duración silábica puede analizarse por medio de la desviación estándar, es decir, el porcentaje de variación que alcanza una sílaba con respecto a la duración promedio de todas las sílabas dentro de un enunciado (véase Tabla 3-15). La ventaja de este indicador es que al sumar los valores elevados al cuadrado de las desviaciones en la duración de cada sílaba, partiendo del promedio de todas las sílabas, se evita que las reducciones en la duración se compensen con aumentos de la misma magnitud.

Según este análisis, entre las mujeres es HEL quien presenta la desviación estándar máxima, 21.70%, pero también es ella quien logró la desviación estándar mínima de 0.00%, lo que significa que al menos dos sílabas contiguas de algún enunciado tuvieron exactamente la misma duración. La desviación estándar promediada para todas las mujeres es de 5.43% y fueron LET y ROS quienes más se apartaron de tal promedio con 6.07% y 4.69%, respectivamente.

En el caso de los hombres es DAN quien produjo la mayor desviación de la duración silábica, 18.31%, mientras que GUI también produjo al menos dos sílabas en un enunciado con exactamente la misma duración. También es DAN quien más se aleja del promedio general de 5.82% entre los hombres con un 7.74%, mientras que MGL alcanzó una desviación estándar promedio de sólo 4.88%.

Tabla 3-15: Desviación estándar máxima, mínima y promedio de la duración silábica

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Desv. est. máxima	10.57%	21.70%	19.55%	13.16%	13.40%	8.61%	13.10%	21.70%	
Desv. est. mínima	0.10%	0.00%	0.05%	0.70%	1.46%	1.00%	0.20%	0.00%	
Desv. est. promedio	5.20%	5.79%	6.07%	4.95%	5.89%	4.69%	5.10%	5.43%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Desv. est. máxima	15.88%	18.31%	8.71%	14.11%	11.88%	12.38%	8.18%	18.31%	21.70%
Desv. est. mínima	1.10%	1.00%	2.50%	0.00%	0.40%	1.11%	1.22%	0.00%	0.00%
Desv. est. promedio	6.01%	7.74%	5.29%	5.30%	4.88%	5.32%	4.91%	5.82%	5.57%

Fuente: Prosodia.xls (Base de datos propia)

En los enunciados suspendidos se registró una desviación estándar máxima de 21.70%, mientras que en los enunciados neutros fue donde se dio el caso de la producción de al menos dos sílabas consecutivas con la misma duración (véase Tabla 3-16). En promedio son nuevamente los enunciados suspendidos los que tienen una mayor desviación, 7.10%, mientras que los enunciados exclamativos sólo presentaron una desviación de 4.31%. Los resultados de la presente Tabla coinciden con los resultados de la Tabla 3-14.

Tabla 3-16: Desviación estándar máxima, mínima y promedio de la duración silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	17.44%	8.44%	21.70%	17.13%
Desv. est. mínima	0.10%	0.70%	0.33%	0.00%
Desv. est. promedio	5.00%	4.31%	7.10%	5.09%

Fuente: Prosodia.xls (Base de datos propia)

Como se ha podido observar, por medio de distintas fórmulas básicas se puede extraer de la información sobre la duración datos importantes sobre la longitud silábica, la velocidad del habla, el rango de variación y la desviación estándar. Algunas de estas fórmulas serán aplicadas a continuación a los conjuntos de datos referentes a otros parámetros acústicos tales como la intensidad y la frecuencia fundamental.

3.2.3 Intensidad

La intensidad es un índice difícil de medir y comparar. En la sección 1.4 se explicaba que para obtener una medición precisa es necesario que los hablantes se encuentren en un estudio de grabación o laboratorio y hablen constantemente a la misma distancia del micrófono. En las grabaciones que conforman el *corpus* éstas no fueron las condiciones y, posiblemente, las principales diferencias entre los hablantes simplemente indiquen las diferencias con respecto al micrófono en cada grabación. A continuación se presentarán los resultados de la medición automática de la intensidad en decibeles, obtenidos por medio del programa Praat. Más adelante se presentarán estos valores, pero estandarizados, para permitir una mejor comparación entre personas, sexos y variedades del español.

La intensidad máxima promedio se basa únicamente en la suma del valor máximo de cada enunciado y después se divide entre el número total de enunciados (véase Tabla 3-18). El promedio de intensidad máxima entre las mujeres de Madrid es de 64.62 decibeles y es HEL, con 68.88 dB, la mujer que alcanza el máximo nivel. La intensidad mínima promedio es de 54.08 dB y es ROS, con 45.84 dB, la persona que produce los enunciados con la menor intensidad. En comparación con el promedio de 59.69 dB son también HEL y ROS quienes hablan con mayor y menor intensidad, 64.74 dB y 52.96 dB respectivamente.

En el caso de los hombres, FRA es quien alcanza el valor más alto de la intensidad máxima promedio, 72.05 dB, mientras que MGL es quien alcanza el valor más bajo de la intensidad mínima promedio. Del promedio general masculino de 60.74 dB, son ABE, con 65.96 dB, y MGL, con 53.29 dB, quienes más se alejan, aunque las diferencias con respecto a FRA y GUI son realmente mínimas. El promedio general de la intensidad en Madrid es de 60.05 dB.

Tabla 3-17: Intensidad silábica máxima, mínima y promedio (decibeles)

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Int. máxima prom.	66.48	68.88	63.96	60.62	63.88	58.28	65.65	64.62	
Int. mínima prom.	56.14	60.10	53.67	47.91	50.52	45.84	56.75	54.08	
Int. promedio	61.52	64.74	59.04	54.86	57.75	52.96	61.32	59.69	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Int. máxima prom.	68.89	64.19	72.05	58.30	58.23	68.65	70.54	65.75	65.02
Int. mínima prom.	62.22	53.26	57.47	49.05	47.41	57.38	59.46	54.75	54.29
Int. promedio	65.96	59.07	65.85	53.58	53.29	63.93	65.20	60.74	60.05

Fuente: Prosodia.xls (Base de datos propia)

En la Tabla 3-18, por otra parte, tenemos los valores de la intensidad según los distintos tipos de enunciados. En ella se puede observar que los enunciados exclamativos son los que tienen el valor más alto de la intensidad máxima promedio, 67.15 dB. Además, los enunciados neutros son los que alcanzan el nivel más bajo dentro de la intensidad mínima promedio,

52.87 dB. En promedio son, igualmente, los enunciados exclamativos los que tienen una intensidad más alta y los enunciados neutros los que la tienen más baja.

Tabla 3-18: Intensidad silábica máxima, mínima y promedio por tipo de enunciado (dB)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Int. máxima prom.	65.98	67.15	66.73	63.65
Int. mínima prom.	55.41	57.31	55.98	52.87
Int. promedio	60.87	62.38	61.69	58.79

Fuente: Prosodia.xls (Base de datos propia)

Los promedios de variación de la intensidad entre una sílaba y la consecutiva nos permiten obtener los porcentajes máximos y mínimos. La distancia entre estos dos porcentajes será entonces el rango de variación promedio (véase Tabla 3-19). De este modo, entre las mujeres, PTR utiliza el rango más amplio, 20.91%, y HEL utiliza el más estrecho, 13.67%, mientras que el promedio de las mujeres es de 17.61% y el general es de 18.22%.

Entre los hombres las diferencias son más marcadas, pues FRA tiene un rango de variación promedio de 27.54%, mientras que el de ABE es más de cuatro veces más estrecho: 6.28%.

Tabla 3-19: Variación máxima y mínima de la intensidad silábica

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Var. máxima prom.	10.16%	8.28%	8.40%	8.83%	8.91%	8.27%	7.23%	8.47%	
Var. mínima prom.	-8.62%	-5.39%	-9.89%	-11.03%	-12.00%	-11.61%	-8.32%	-9.14%	
Rango var. prom.	18.78%	13.67%	18.29%	19.86%	20.91%	19.88%	15.55%	17.61%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Var. máxima prom.	4.05%	10.63%	13.96%	4.57%	7.35%	10.37%	9.85%	9.38%	8.91%
Var. mínima prom.	-2.23%	-10.56%	-13.58%	-8.58%	-10.11%	-10.13%	-8.52%	-9.78%	-9.31%
Rango var. prom.	6.28%	21.19%	27.54%	13.15%	17.46%	20.72%	18.37%	19.16%	18.22%

Fuente: Prosodia.xls (Base de datos propia)

Además, los enunciados suspendidos son los que utilizan un rango de variación de la intensidad más amplio, 19.00%, mientras que los enunciados exclamativos sólo utilizan un rango de 12.35%, como se muestra en la Tabla 3-20.

Tabla 3-20: Variación máxima y mínima de la intensidad silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Var. máxima promedio	9.04%	8.29%	9.25%	8.50%
Var. mínima promedio	-8.69%	-4.06%	-9.75%	-9.76%
Rango de var. prom.	17.73%	12.35%	19.00%	18.26%

Fuente: Prosodia.xls (Base de datos propia)

Como se vio en el apartado 3.2.2 sobre la duración, de manera adicional a la variación promedio, se puede analizar la variación intersilábica de la intensidad haciendo uso de la desviación estándar (véase Tabla 3-21). Así, de las mujeres, PAT es quien produce la desviación estándar máxima, 850%, y es también ella quien en promedio tiene la mayor desviación estándar promedio, 431%. En el caso de la desviación estándar mínima son cinco las madrileñas que obtienen un valor de 0%, es decir, que al menos produjeron dos sílabas con

exactamente la misma intensidad. En promedio, HEL es quien produjo la menor desviación estándar, 307%, mientras que el promedio general de las mujeres es de 357%.

Tabla 3-21: Desviación estándar máxima, mínima y promedio de la intensidad silábica

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Desv. est. máxima	641%	600%	718%	850%	708%	611%	718%	850%	
Desv. est. mínima	0%	0%	0%	50%	82%	0%	0%	0%	
Desv. est. promedio	341%	307%	355%	431%	419%	395%	323%	357%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Desv. est. máxima	416%	896%	748%	603%	671%	685%	765%	896%	896%
Desv. est. mínima	47%	0%	200%	0%	0%	0%	82%	0%	0%
Desv. est. promedio	227%	397%	460%	311%	379%	381%	379%	377%	364%

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres, DAN es quien produjo la desviación máxima, 896%, mientras que cuatro de los madrileños se disputan haber producido una desviación estándar mínima de 0%. Si tomamos en cuenta el promedio, al igual que en la Tabla 3-19, FRA, con 460%, y ABE, con sólo 227%, son quienes se encuentran a ambos extremos del promedio de 377%. El promedio general de desviación estándar para Madrid es de 367%.

Además, tomando en cuenta los tipos de enunciados (véase Tabla 3-22), es en los enunciados interrogativos en los que aparece la desviación máxima de 896%, mientras que en todos los tipos, excepto los exclamativos, hubo casos de desviación de 0%. En promedio son los enunciados exclamativos los que presentan mayor desviación estándar, 389%, mientras que los enunciados neutros son los que menos desviación presentan, 359%.

Tabla 3-22: Desviación estándar máxima, mínima y promedio de la intensidad silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	896%	850%	733%	795%
Desv. est. mínima	0%	150%	0%	0%
Desv. est. promedio	368%	389%	369%	359%

Fuente: Prosodia.xls (Base de datos propia)

Ahora, si tomamos en cuenta los valores estandarizados de la intensidad, fijando la intensidad inicial de cada enunciado en 60 dB y sumando la variación porcentual multiplicada por 0.60, podremos hacer una comparación más objetiva entre los distintos hablantes, sin importar su sexo, cantidad de enunciados analizados o distancia con respecto al micrófono.

En este caso (véase Tabla 3-23), entre las mujeres, ya no es HEL sino PAT quien produce la intensidad máxima promedio, lo que se podría explicar como una diferencia de distancia entre las hablantes y el micrófono, mientras que lo que más interesa es la diferencia relativa de intensidad producida dentro del enunciado. Del mismo modo, si la persona, como en el caso de ROS, en realidad produjo un cierto descenso o ascenso de la intensidad, se seguirá notando en los resultados. En promedio, por ejemplo, son PAT, con 62.99 dB, y ROS, con 60.30 dB, los extremos con respecto al promedio femenino de 61.74 dB.

Tabla 3-23: Intensidad silábica estandarizada máxima, mínima y promedio (dB inicial = 60)

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Int. máxima prom.	66.85	65.44	67.36	69.68	66.19	66.35	65.29	66.91	
Int. mínima prom.	56.43	57.10	56.39	54.87	54.56	52.22	56.52	55.82	
Int. promedio	61.86	61.51	62.12	62.99	62.52	60.30	61.03	61.74	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Int. máxima prom.	63.78	69.02	72.32	66.25	70.18	68.56	68.83	68.82	67.61
Int. mínima prom.	57.58	57.15	57.48	55.63	56.96	57.28	57.62	57.11	56.29
Int. promedio	61.06	63.47	66.03	60.84	64.15	63.85	63.41	63.48	62.38

Fuente: Prosodia.xls (Base de datos propia)

Entre los hombres, FRA sigue siendo, con 72.32 dB, quien más elevó la intensidad de la voz durante sus producciones lingüísticas, mientras GUI es quien más disminuyó la intensidad de la voz, con 55.63 dB. En promedio, GUI, con 60.84 dB, por un lado, así como FRA, con 66.03 dB, por el otro lado, son quienes más se alejan del promedio general de 62.38 dB.

En la Tabla 3-24 se puede apreciar, si se le compara con la Tabla 3-18, que la intensidad máxima promedio sigue siendo la producida en los enunciados suspendidos y la mínima en los enunciados neutros. Sin embargo, los extremos de la intensidad promedio cambian: los enunciados interrogativos, con 62.13 dB, se producen con la menor intensidad y los suspendidos, con 62.73, se producen con la mayor intensidad, aunque las diferencias son demasiado leves entre los distintos tipos de enunciados.

Tabla 3-24: Intensidad silábica estandarizada máxima, mínima y promedio por tipo de enunciado (dB inicial = 60)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Int. máxima prom.	66.47	67.27	67.93	67.52
Int. mínima prom.	56.41	57.02	56.82	55.94
Int. promedio	62.13	62.27	62.73	62.31

Fuente: Prosodia.xls (Base de datos propia)

Con estos mismos datos de la intensidad estandarizada se puede calcular la desviación estándar máxima, mínima y promedio (véase Tabla 3-25). Por lo general, los casos extremos de la desviación estándar promedio son similares a los presentados en la Tabla 3-21.

Tabla 3-25: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Desv. est. máxima	675%	614%	958%	1186%	914%	733%	686%	1186%	
Desv. est. mínima	0%	0%	0%	55%	84%	0%	0%	0%	
Desv. est. promedio	345%	293%	378%	508%	460%	448%	318%	376%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Desv. est. máxima	410%	1033%	863%	696%	875%	709%	1119%	1119%	1186%
Desv. est. mínima	43%	0%	194%	0%	0%	0%	78%	0%	0%
Desv. est. promedio	213%	432%	467%	356%	463%	381%	382%	402%	386%

Fuente: Prosodia.xls (Base de datos propia)

En el caso de la mujeres, HEL, con 293%, y ROS, con 508%, son quienes más se alejan del promedio femenino de 376%, mientras que ABE y FRA, con 213% y 467% respectivamente, son quienes más se alejan del promedio masculino de 402%. La desviación estándar promedio de los hablantes de Madrid es de 386%.

Si tomamos en cuenta los tipos de enunciados, veremos que los resultados de la Tabla 3-26 difieren mucho de los de la Tabla 3-22. En este caso, en promedio, la desviación estándar de las sílabas de los enunciados exclamativos es la más baja, 381%, mientras que la de los enunciados suspendidos es la más alta, 411%.

Tabla 3-26: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	1119%	1186%	758%	973%
Desv. est. mínima	0%	134%	0%	0%
Desv. est. promedio	385%	411%	381%	387%

Fuente: Prosodia.xls (Base de datos propia)

Más adelante se comprobará con mayor detenimiento las ventajas que ofrece el uso de valores estandarizados de la intensidad.

3.2.4 Frecuencia

En cuanto a la información sobre la frecuencia fundamental, su variación intersilábica y la frecuencia fundamental estandarizada. Al igual que en el apartado anterior, la estandarización se hace necesaria en casos en los que, desde un inicio, los valores acústicos de los distintos hablantes son completamente dispares. La estandarización es una solución sencilla para poder llegar a conclusiones generales y hacer comparaciones objetivas (véase 1.1.1.2).

Veamos, por ejemplo, la frecuencia fundamental máxima, mínima y promedio, además del rango frecuencial promedio de cada uno de los hablantes de Madrid (Tabla 3-27). No sólo los valores promedio generales de mujeres y hombres son muy distintos (208.25 Hz vs. 121.86, por ejemplo), sino también entre los distintos hablantes las frecuencias mínimas y máximas varían en mayor o menor medida. Estos valores son muy útiles para caracterizar acústicamente a cada hablante, pero no para hacer un estudio comparativo.

De este modo, entre las mujeres, PAT es quien alcanza frecuencias más altas, 288.41 Hz si se promedia sólo los valores más altos, es decir, es quien alcanza notas más agudas, mientras que LET es quien alcanza notas más graves, 145.85 Hz en promedio de las frecuencias más bajas. Sin embargo, al promediar todos los valores, son PAT y ROS, con una ligera diferencia, 236.77 Hz y 235.83 Hz respectivamente, quienes hablan más agudo, mientras que PTR y LET, con 184.85 Hz y 186.96 Hz, son quienes hablan más grave. Sin embargo, es interesante

ver que, por ejemplo, mientras que GEM utiliza en promedio un rango frecuencial de 100.62 Hz, SON sólo utiliza 58.64 Hz. Como vemos, los valores promedio generales sirven poco.

Entre los hombres es mucho más claro, pues MIG, con una frecuencia máxima promedio de 194.57 Hz y mínima de 126.16 Hz, un promedio general de 164.97 Hz y un rango frecuencial de 68.41 Hz, es quien habla más agudo y hace un mayor uso de la escala de frecuencia, mientras que DAN, con una frecuencia máxima promedio de 113.40 Hz, una mínima de tan sólo 75.57 Hz, un promedio de 97.35 Hz y un rango frecuencial de únicamente 37.83 Hz, es quien habla más grave y, además, hace un uso muy limitado del rango de frecuencia.

Tabla 3-27: Frecuencia fundamental máxima, mínima y promedio (Hertz)

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Frec. máx. promedio	263.43	249.53	236.75	288.41	224.09	283.00	242.66	252.17	
Frec. mín. promedio	162.81	165.18	145.85	190.85	152.06	203.84	184.02	170.27	
Frecuencia promedio	213.66	205.44	186.96	236.77	184.85	235.83	210.25	208.25	
Rango frec. prom.	100.62	84.35	90.90	97.56	72.03	79.16	58.64	81.90	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Frec. máx. promedio	120.00	113.40	156.11	137.60	126.14	194.57	140.92	143.99	212.32
Frec. mín. promedio	100.11	75.57	103.63	90.70	69.91	126.16	101.35	95.42	142.69
Frecuencia promedio	108.74	97.35	127.65	116.78	103.87	164.97	119.57	121.86	176.42
Rango frec. prom.	19.89	37.83	52.48	46.90	56.23	68.41	39.57	48.57	69.63

Fuente: Prosodia.xls (Base de datos propia)

Los valores totales de hombres y mujeres sólo servirían para especificar aproximadamente los valores promedio de los hablantes de un determinado grupo o región, pero es preferible no utilizarlo. Del mismo modo, en la Tabla 3-28 se puede apreciar que, tomando esos valores promedio, son los enunciados exclamativos los que tienen las frecuencias máxima, mínima y promedio más altas, mientras que los enunciados suspendidos tienen las frecuencias máxima, mínima y promedio más bajas. También es en los enunciados exclamativos en los que se hace un uso más amplio del rango tonal, 82.61 Hz, mientras que en los enunciados interrogativos y los suspendidos es en los que se hace un uso más estrecho, 72.35 Hz y 73.05 Hz respectivamente.

Tabla 3-28: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (Hz)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. máx. promedio	217.73	261.92	205.28	210.97
Frec. mín. promedio	145.38	179.31	132.23	144.96
Frecuencia promedio	179.55	212.62	166.11	178.45
Rango frec. promedio	72.35	82.61	73.05	66.01

Fuente: Prosodia.xls (Base de datos propia)

Estos valores, que a simple vista parecen muy raros, se explican fácilmente como un truco estadístico si los datos de mujeres y hombres se analizan por separado (véase Tablas 3-29 y 3-30). Por ejemplo, en el caso de las mujeres, son, en efecto, los enunciados exclamativos los que alcanzan las frecuencias más altas, pero, aunque los enunciados neutros y los suspendidos

se disputan las frecuencias más bajas, los enunciados neutros y los suspendidos son los que hacen un uso más restringido del rango tonal.

Tabla 3-29: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (mujeres) (Hz)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. máx. promedio	271.67	289.73	257.86	240.83
Frec. mín. promedio	184.07	192.91	162.25	167.19
Frecuencia promedio	224.73	231.87	203.74	202.82
Rango frec. promedio	87.60	96.82	95.62	73.64

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres, los resultados son muy diferentes, pues los enunciados suspendidos son los que alcanzan las frecuencias más altas, aunque las diferencias con respecto a los enunciados interrogativos y neutros son, por lo general, mínimas. Por el contrario, los enunciados exclamativos son los que utilizan claramente las frecuencias más bajas. En cuanto al rango frecuencial, los enunciados exclamativos hacen tan sólo un uso de 4.50 Hz, mientras que los interrogativos hacen el uso más amplio, 52.14 Hz. Sin embargo, esto obedece nuevamente a un engaño visual, pues en el *corpus* de Madrid sólo se encontraron dos enunciados exclamativos producidos por hablantes del sexo masculino.

Tabla 3-30: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (hombres) (Hz)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. máx. promedio	146.23	109.00	146.36	141.72
Frec. mín. promedio	94.09	104.50	98.59	93.42
Frecuencia promedio	119.61	106.75	123.94	121.92
Rango frec. promedio	52.14	4.50	47.77	48.30

Fuente: Prosodia.xls (Base de datos propia)

En cuanto a la variación intersilábica de la frecuencia fundamental (véase Tabla 3-31), vemos que los valores promedio de hombres y mujeres son similares, aunque el rango de variación promedio, 33.12% en mujeres y 29.52% en hombres, muestra ligeras diferencias. En el caso de las mujeres, PAT es quien hace un mayor uso del rango de la variación promedio de la F_0 , 40.46%, mientras que SON utiliza solamente un 23.85% del rango.

Tabla 3-31: Variación máxima y mínima promedio de la frecuencia fundamental

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Var. máxima prom.	24.27%	18.67%	24.86%	23.12%	22.06%	19.12%	12.97%	20.21%	
Var. mínima prom.	-9.24%	-14.20%	-12.57%	-17.34%	-17.63%	-10.24%	-10.88%	-12.91%	
Rango var. prom.	33.51%	32.87%	37.43%	40.46%	39.69%	29.36%	23.85%	33.12%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Var. máxima prom.	6.49%	16.31%	23.40%	19.63%	20.81%	18.24%	16.21%	17.91%	19.36%
Var. mínima prom.	-5.21%	-14.04%	-16.61%	-7.01%	-12.22%	-15.70%	-4.46%	-11.61%	-12.43%
Rango var. prom.	11.70%	30.35%	40.01%	26.64%	33.03%	33.94%	20.67%	29.52%	31.79%

Fuente: Prosodia.xls (Base de datos propia)

De entre los hombres, FRA utiliza el mayor rango de variación, 40.01%, mientras que ABE sólo utiliza un rango de 11.70%. El promedio general para Madrid es de 31.79%.

Los enunciados interrogativos son los que hacen uso de un mayor rango de variación, 34.76%, es decir, la frecuencia promedio entre dos sílabas de estos enunciados varía alrededor de dicho porcentaje, mientras que los enunciados neutros utilizan sólo un rango de 30.40% (véase Tabla 3-32).

Tabla 3-32: Variación máxima y mínima promedio de la frecuencia fundamental por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Var. máxima promedio	27.61%	19.42%	19.13%	16.03%
Var. mínima promedio	-7.15%	-14.11%	-12.77%	-14.37%
Rango de var. prom.	34.76%	33.53%	31.90%	30.40%

Fuente: Prosodia.xls (Base de datos propia)

Ahora, si comparamos estos resultados con la desviación estándar de la frecuencia fundamental de las sílabas consecutivas (véase Tabla 3-33), veremos que, por ejemplo, en el caso de las mujeres, PAT es quien más se aleja del promedio, con 12512% de desviación máxima y un promedio de 3329%, mientras que HEL produjo al menos dos sílabas con una desviación de exactamente 0%. Sin embargo, en promedio es SON quien menos desviación presenta, 2113%, mientras que el promedio general de las mujeres es de 2807%.

Entre los hombres, FRA es quien presenta una mayor desviación, 6074%, y PAC presenta una desviación de 0%. Sin embargo, en promedio son MIG, con 2175%, y ABE, con 660%, quienes representan los extremos del promedio masculino de 1581%. Nuevamente, el promedio de general de hombres y mujeres es de poca utilidad.

Tabla 3-33: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Desv. est. máxima	8432%	11800%	8697%	12512%	6919%	5650%	8475%	12512%	
Desv. est. mínima	273%	0%	300%	450%	400%	460%	200%	0%	
Desv. est. promedio	3174%	3111%	3074%	3329%	2376%	2697%	2113%	2807%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Desv. est. máxima	1722%	4201%	6074%	4602%	5561%	5638%	4625%	6074%	12512%
Desv. est. mínima	150%	200%	617%	450%	232%	510%	0%	0%	0%
Desv. est. promedio	660%	1287%	1668%	1441%	1827%	2175%	1364%	1581%	2355%

Fuente: Prosodia.xls (Base de datos propia)

Si utilizamos estos valores promedio generales, obtenemos los resultados de la Tabla 3-34:

Tabla 3-34: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	11800%	5850%	9207%	12512%
Desv. est. mínima	438%	150%	0%	0%
Desv. est. promedio	2645%	3007%	2341%	2205%

Fuente: Prosodia.xls (Base de datos propia)

Según esto, los enunciados neutros presentan la mayor desviación con respecto del promedio, 12512%, pero también estos enunciados, junto con los suspendidos, presentan al menos un caso de desviación 0%. En promedio, sin embargo, son los enunciados exclamativos los que presentan una mayor desviación, 3007%, mientras que los neutros tienen la menor desviación. Estos resultados deben ser comprobados con un análisis por separado de la desviación estándar de la F_0 de mujeres y hombres. En el caso de las primeras (véase Tabla 3-35), son los enunciados neutros los que, a la vez, presentan la mayor y la menor desviación estándar, aunque en promedio presenten la desviación estándar más baja, mientras que los enunciados exclamativos son los que más se desvían.

Tabla 3-35: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (mujeres)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	11800%	5850%	9207%	12512%
Desv. est. mínima	438%	450%	200%	0%
Desv. est. promedio	3314%	3512%	3106%	2470%

Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, en el caso de los hombres (véase Tabla 3-36), son los enunciados interrogativos los que presentan la mayor desviación estándar, también en promedio, mientras que los enunciados exclamativos son los que tienen la menor desviación en general.

Tabla 3-36: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (hombres)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	6074%	300%	4602%	5638%
Desv. est. mínima	450%	150%	0%	150%
Desv. est. promedio	1758%	225%	1484%	1591%

Fuente: Prosodia.xls (Base de datos propia)

Para evitar las diferencias de frecuencias fundamentales absolutas de cada hablante, se recurre en esta investigación a la frecuencia estandarizada, asignando el valor de 100 Hz a la primera sílaba de cada enunciado. A este valor se le suma o resta la variación porcentual de las sílabas subsecuentes. Los valores estandarizados se presentan en la Tabla 3-37.

Tabla 3-37: Frecuencia estandarizada máxima, mínima y promedio (F_0 inicial = 100)

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Frec. est. máx. prom.	126.55	121.65	129.06	127.18	123.27	120.00	116.00	123.06	
Frec. est. mín. prom.	88.67	83.55	86.96	87.24	86.30	86.72	88.03	86.81	
Frec. estándar prom.	104.30	100.99	102.69	105.10	102.30	100.21	100.54	102.19	
Rango frec. prom.	37.88	38.10	42.10	39.94	36.97	33.28	27.97	36.25	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Frec. est. máx. prom.	105.89	114.69	125.32	117.75	123.82	127.51	125.19	121.16	122.36
Frec. est. mín. prom.	89.89	88.64	86.58	90.75	90.55	87.43	92.96	89.35	87.75
Frec. estándar prom.	96.66	100.57	102.45	101.59	105.48	108.42	106.84	103.90	102.82
Rango frec. prom.	16.00	26.05	38.74	27.00	33.27	40.08	32.23	31.81	34.61

Fuente: Prosodia.xls (Base de datos propia)

Con este método se puede apreciar que los valores promedio para mujeres y hombres son muy similares. Sin embargo, se puede seguir apreciando que algunos de ellos hacen un uso mayor o menor de la variación frecuencial. Por ejemplo, LET alcanza una frecuencia máxima promedio de 129.06 Hz, mientras que HEL alcanza una frecuencia mínima promedio de 83.55 Hz, valores más extremos que los de MIG y FRA, aunque son MIG, con 108.42 Hz, y ABE, con 96.66 Hz, quienes representan los extremos del promedio general de hombres y mujeres. Por último, se puede observar que, mientras LET utiliza un rango frecuencial estandarizado de 42.10 Hz, DAN sólo usa un rango frecuencial de 26.05 Hz.

También con estos valores promedio se puede observar en la Tabla 3-38 que es en los enunciados interrogativos en los que se alcanza la frecuencia máxima promedio de 128.90 Hz y un promedio general de 106.55 Hz, mientras que en los enunciados neutros encontramos la frecuencia mínima promedio de 86.96 Hz y un promedio general de 100.90 Hz, lo que, en realidad, refuerza la idea intuitiva general de la altura tonal en los distintos tipos de enunciados. Por si fuera poco, el rango frecuencial promedio de los enunciados interrogativos es de 39.79 Hz, mientras que el de los enunciados neutros es de únicamente 30.65 Hz.

Tabla 3-38: Frecuencia estandarizada máxima, mínima y promedio por tipo de enunciado (F_0 inicial = 100)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. est. máx. promedio	128.90	127.31	125.76	117.61
Frec. est. mín. promedio	89.11	90.00	87.93	86.96
Frec. estándar promedio	106.55	105.26	103.27	100.90
Rango frec. promedio	39.79	37.31	37.83	30.65

Fuente: Prosodia.xls (Base de datos propia)

Si recurrimos, además, al método de la desviación estándar, veremos confirmados algunos de los resultados anteriores. Por ejemplo, HEL es la persona que presenta la mayor desviación general y, al mismo tiempo, una desviación de 0%, al igual que PAC. Sin embargo, en promedio, son LET, con 1447%, y ABE, con 554%, los que representan los extremos del promedio general de 1193%.

Tabla 3-39: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada

	GEM	HEL	LET	PAT	PTR	ROS	SON	Mujeres	Total
Desv. est. máxima	3400%	7300%	5074%	4920%	3780%	2450%	3990%	7300%	
Desv. est. mínima	160%	0%	200%	200%	250%	212%	100%	0%	
Desv. est. promedio	1278%	1428%	1447%	1351%	1215%	1128%	1006%	1262%	
	ABE	DAN	FRA	GUI	MGL	MIG	PAC	Hombres	
Desv. est. máxima	1045%	3121%	4752%	2853%	2884%	4365%	3900%	4752%	7300%
Desv. est. mínima	150%	200%	454%	357%	232%	250%	0%	0%	0%
Desv. est. promedio	554%	922%	1280%	933%	1120%	1329%	1074%	1077%	1193%

Fuente: Prosodia.xls (Base de datos propia)

Por último, si tomamos en cuenta los distintos tipos de enunciados (véase Tabla 3-40), es en los enunciados interrogativos en los que se da la mayor desviación estándar, mientras que los suspendidos y los neutros presentan casos de desviación de 0%. Sin embargo, en promedio, los enunciados interrogativos son que hacen un mayor uso de la desviación, 1465%, mientras que los enunciados neutros sólo hacen uso de una desviación estándar de 1043%.

Tabla 3-40: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	7300%	2708%	5074%	4920%
Desv. est. mínima	212%	150%	0%	0%
Desv. est. promedio	1465%	1372%	1247%	1043%

Fuente: Prosodia.xls (Base de datos propia)

Con los resultados de la frecuencia estandarizada, al igual que en el caso de la intensidad estandarizada, queda claro que en un estudio comparativo como el presente, la mejor opción es utilizar valores estandarizados. Los valores absolutos siguen siendo útiles para caracterizar acústicamente las producciones lingüísticas de los hablantes, pero los valores estandarizados dejan entrever claramente la variación intersilábica y permiten estudios comparativos.

3.2.5 Entonación

Primeramente, se hará un recuento de los distintos tonos que forman parte del *corpus* de Madrid. La clasificación corresponderá mayoritariamente a la propuesta por el Sp-ToBI: tonos de frontera iniciales, tonos prenucleares, tonos nucleares y tonos de frontera finales. En otras obras los tonos de frontera normalmente no son subclasificados sistemáticamente en iniciales y finales, pero como para esta investigación era más práctico fragmentar las grabaciones en enunciados, éstas siempre tienen tonos de frontera que coinciden también con el inicio y el final del fragmento, además, mientras no se compruebe que no son relevantes para hacer estudios comparativos de la entonación, seguirán siendo analizados así.

En la Tabla 3-41 se puede observar que el porcentaje de tonos iniciales H% es de 84.42%. Si bien el porcentaje parece muy alto, esto se debe posiblemente a que, por cuestiones de sistematicidad, se tuvo que establecer una frecuencia de base para cada hablante antes de haber podido acceder a resultados estadísticos. Por lo general, después de haber analizado las frecuencias de varios de los fragmentos de una persona, para los hombres se establecían frecuencias de base de 100, 125 o 150 Hz y para las mujeres 150, 200 o incluso 250. Dado que el tono inicial depende directamente de este valor absoluto, puede ser que alguno de estos valores asignado *a priori* haya resultado demasiado bajo y, por consiguiente, todos los tonos iniciales entre la frecuencia de base real y la asignada hayan sido analizados como H%.

Tabla 3-41: Número y porcentaje de cada tipo de tono inicial

	H%	L%	Total
Tonos iniciales	401	74	475
Porcentaje	84.42%	15.58%	100%

Fuente: Prosodia.xls (Base de datos propia)

Como tonos prenucleares se clasificó a todos los tonos después del tono inicial y antes del tono nuclear. De manera aislada, los tonos prenucleares más comunes en el habla de Madrid son *, con 31.36%, L+H* y H+L*, con igual porcentaje, 14.54%, seguidos por L*+H, con 13.32% y ¡L*+H, con 13.19%. El tono * representa la ausencia o poca perceptibilidad de acento sobre la sílaba tónica. Dentro de este *corpus* se consideró como * todo aquel tono que no mostrará una variación promedio de al menos $\pm 5\%$. Además, se agrupó los distintos tonos según la clasificación presentada al final del apartado 1.2.4.1, en el que se denomina \acute{S} a todos aquellos tonos cuyo pico frecuencial coincide con la sílaba acentuada (las cuatro primeras columnas de la Tabla 3-42: L+H*, L+!H*, ¡L+H* y H*), S' a los tonos cuyo pico frecuencial se encuentra después de la sílaba tónica (L*+H, L*+!H y ¡L*+H), 'S a los tonos con pico frecuencial antes de la sílaba acentuada (H+L*), y S a los tonos sin pico frecuencial o muy poco perceptible (*). Por lo tanto, los tonos S representan el mayor número de tonos prenucleares, 31.36%, seguidos de \acute{S} , con 27.20%, después 'S, con 26.91%, y por último S', con 14.54% del total de tonos prenucleares.

Tabla 3-42: Número y porcentaje de cada tipo de tono pre nuclear

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
Tonos prenucleares	108	2	69	23	99	3	98	108	233	743
Porcentaje	14.54%	0.27%	9.29%	3.10%	13.32%	0.40%	13.19%	14.54%	31.36%	100%

Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, a primera vista los tonos nucleares más comunes son L+H*, con 25.89%, *, con 24%, H+L*, con 15.79%, y L*+H, con 13.47% (véase Tabla 3-43). Sin embargo, los tonos nucleares del tipo \acute{S} abarcan un 35.58%, los tonos subordinados al tipo 'S un 24.63%, los del tipo S un 24.00%, y los tonos del tipo S' están representados con un 15.79%.

Tabla 3-43: Número y porcentaje de cada tipo de tono nuclear

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
Tonos nucleares	123	2	34	10	64	0	53	75	114	475
Porcentaje	25.89%	0.42%	7.16%	2.11%	13.47%	0.00%	11.16%	15.79%	24.00%	100%

Fuente: Prosodia.xls (Base de datos propia)

Por último, con respecto a los tonos finales (véase Tabla 3-44), la mayoría de ellos son del tipo M%, 91.79%, seguidos de L%, con 5.89%, y al último H%, con 2.32%. Para hacer la clasificación, se tomó como auditivamente relevante un ascenso o descenso porcentual de la variación de la frecuencia fundamental de 25%. Entonces, el último tono de cada enunciado que tuviera una variación final entre -25% y +25% sería considerado como M%.

Posiblemente se compruebe más adelante que este valor arbitrario haya sido demasiado generoso y se tendrá que hacer las adecuaciones necesarias.

Tabla 3-44: Número y porcentaje de cada tipo de tono final

	H%	M%	L%	Total
Tonos finales	11	436	28	475
Porcentaje	2.32%	91.79%	5.89%	100%

Fuente: Prosodia.xls (Base de datos propia)

Ahora, si asociamos los tonos iniciales con los distintos tipos de enunciados (véase Tablas 3-45 y 3-46), veremos que, en orden descendente, los enunciados neutros son los más comunes en el *corpus* de Madrid (véase Tabla 3-5), seguidos por los enunciados suspendidos y los interrogativos. Los enunciados exclamativos son los menos representados.

Tabla 3-45: Número de tonos iniciales por tipo de enunciado

	H%	L%	Total
/+I -E -S/	89	11	100
/-I +E -S/	12	1	13
/-I -E +S/	104	19	123
/-I -E -S/	196	43	239
Total	401	74	475

Fuente: Prosodia.xls (Base de datos propia)

En los cuatro casos, con porcentajes entre 92.31% (enunciados exclamativos) y 82.01% (neutros), la mayoría de los enunciados fueron producidos con tono inicial tipo H%.

Tabla 3-46: Porcentaje de tonos iniciales por tipo de enunciado

	H%	L%	Total
/+I -E -S/	89.00%	11.00%	100%
/-I +E -S/	92.31%	7.69%	100%
/-I -E +S/	84.55%	15.45%	100%
/-I -E -S/	82.01%	17.99%	100%
Total	84.42%	15.58%	100%

Fuente: Prosodia.xls (Base de datos propia)

En las Tablas 3-47 y 3-48 se puede observar la asociación de tonos nucleares con los tipos de enunciados. Individualmente, los tonos nucleares más comunes en enunciados interrogativos son L+H*, L*+H, ¡L*+H y *; en los enunciados exclamativos son *, L*+H, ¡L*+H y L+H*; en los enunciados suspendidos son *, H+L*, ¡L*+H, ¡L+H* y L+H*, y en los enunciados neutros son L+H*, *, H+L* y L*+H.

Tabla 3-47: Número de tonos nucleares por tipo de enunciado

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
/+I -E -S/	32	0	8	3	22	0	15	8	12	100
/-I +E -S/	2	0	1	0	3	0	3	0	4	13
/-I -E +S/	16	0	17	2	10	0	21	28	29	123
/-I -E -S/	73	2	8	5	29	0	14	39	69	239
Total	123	2	34	10	64	0	53	75	114	475

Fuente: Prosodia.xls (Base de datos propia)

Mientras que si los agrupamos, resulta que, en el caso de los enunciados interrogativos, los tonos más comunes son del tipo \acute{S} y $\prime S$, con 43% y 37% respectivamente; en los enunciados exclamativos los únicos tipos que existen son $\prime S$, S y \acute{S} , con 46.16%, 30.77% y 23.07% respectivamente; en los enunciados suspendidos la distribución es bastante equilibrada \acute{S} , con 28.46%, $\prime S$, con 25.20%, S , con 23.58%, y S' , con 22.76%, mientras que en los enunciados neutros los tipos más sobresalientes son \acute{S} , con 35.58%, y S , con 24%.

Tabla 3-48: Porcentaje de tonos nucleares por tipo de enunciado

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
/+I -E -S/	32.00%	0%	8.00%	3.00%	22.00%	0%	15.00%	8.00%	12.00%	100%
/-I +E -S/	15.38%	0%	7.69%	0%	23.08%	0%	23.08%	0%	30.77%	100%
/-I -E +S/	13.01%	0%	13.82%	1.63%	8.13%	0%	17.07%	22.76%	23.58%	100%
/-I -E -S/	30.54%	0.84%	3.35%	2.09%	12.13%	0%	5.86%	16.32%	28.87%	100%
Total	25.89%	0.42%	7.16%	2.11%	13.47%	0%	11.16%	15.79%	24.00%	100%

Fuente: Prosodia.xls (Base de datos propia)

En las Tablas 3-49 y 3-50 se asocian los distintos tonos finales con los tipos de enunciados. Como era de esperarse, todos los tipos de enunciados tienen una mayor presencia de tonos finales del tipo M%, seguidos siempre de los tonos L%, mientras que los tonos H% son los más escasos.

Tabla 3-49: Número de tonos finales por tipo de enunciado

	H%	M%	L%	Total
/+I -E -S/	3	92	5	100
/-I +E -S/	0	8	5	13
/-I -E +S/	5	113	5	123
/-I -E -S/	3	223	13	239
Total	11	436	28	475

Fuente: Prosodia.xls (Base de datos propia)

El rango de promedios de presencia del tono M% varía entre 93.31% (neutros) y 61.54% (exclamativos), aunque de estos últimos hay solamente 13 en el *corpus*.

Tabla 3-50: Porcentaje de tonos finales por tipo de enunciado

	H%	M%	L%	Total
/+I -E -S/	3.00%	92.00%	5.00%	100%
/-I +E -S/	0%	61.54%	38.46%	100%
/-I -E +S/	4.07%	91.87%	4.07%	100%
/-I -E -S/	1.26%	93.31%	5.44%	100%
Total	2.32%	91.79%	5.89%	100%

Fuente: Prosodia.xls (Base de datos propia)

Pasando específicamente al tema de la combinación de los tonos nucleares con los distintos tipos de tonos finales, con esta serie de datos se puede identificar las secuencias más comunes y, en general, dar cuenta de las posibilidades combinatorias (véase Tablas 3-51 y 3-52). Por ejemplo, los tonos nucleares más comunes en compañía del tono final H% son L^*+H , $¡L^*+H$, $¡L+H^*$ y $L+H^*$; los tonos nucleares más combinados con M% son $L+H^*$ y $*$, con igual

número de casos, seguidos de H+L*, L*+H y ¡L*+H, mientras que los tonos nucleares asociados con L% más comunes son L+H*, ¡L+H*, ¡L*+H y H+L*.

Tabla 3-51: Número de tonos nucleares combinados con cada tipo de tono final

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	1	0	2	0	5	0	3	0	0	11
M%	112	2	27	9	57	0	46	71	112	436
L%	10	0	5	1	2	0	4	4	2	28
Total	123	2	34	10	64	0	53	75	114	475

Fuente: Prosodia.xls (Base de datos propia)

Si, además, agrupamos los distintos tonos en los 4 distintos ‘architonemas’, resulta que los tipos más comunes en enunciados interrogativos son ‘S, con 72.72% de presencia, y \acute{S} , con un total de 27.27%; los tres tipos de tonos más comunes en los enunciados terminados en M% son \acute{S} , con 34.40%, S, con 25.69%, y ‘S, con 23.62%, mientras que los más comunes en compañía de los tonos L% son \acute{S} y ‘S, con 57.14% y 21.43% respectivamente.

Tabla 3-52: Porcentaje de tonos nucleares combinados con cada tipo de tono final

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	9.09%	0%	18.18%	0%	45.45%	0%	27.27%	0%	0%	100%
M%	25.69%	0.46%	6.19%	2.06%	13.07%	0%	10.55%	16.28%	25.69%	100%
L%	35.71%	0%	17.86%	3.57%	7.14%	0%	14.29%	14.29%	7.14%	100%
Total	25.89%	0.42%	7.16%	2.11%	13.47%	0%	11.16%	15.79%	24.00%	100%

Fuente: Prosodia.xls (Base de datos propia)

Ahora bien, si, además, a los valores de las dos Tablas anteriores les aplicamos un filtro para que nos muestren únicamente los tonos nucleares en combinación con los distintos tonos finales en enunciados interrogativos, nos dará como resultado las series de valores que se plasma en las Tablas 3-53 y 3-54.

Tabla 3-53: Número de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0	0	0	0	2	0	1	0	0	3
M%	32	0	6	2	19	0	14	7	12	92
L%	0	0	2	1	1	0	0	1	0	5
Total	32	0	8	3	22	0	15	8	12	100

Fuente: Prosodia.xls (Base de datos propia)

Entonces, en los enunciados interrogativos terminados en tono de frontera H%, los únicos tonos dentro del *corpus* son L*+H y ¡L*+H, es decir, ambos del tipo ‘S. Por otra parte, en los enunciados terminados en M%, los tonos más comunes son L+H*, L*+H, ¡L*+H y *, lo que equivale a 43.47% de tonos del tipo \acute{S} , 35.87% del tipo ‘S, 13.04% del tipo S y 7.61% del tipo S’. Además, que en los enunciados terminados en L%, los únicos tonos representados son ¡L+H*, H*, L*+H y H+L*, que equivale a 60% de tonos del tipo \acute{S} , 20% del tipo ‘S y también 20% del tipo S’.

Tabla 3-54: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0%	0%	0%	0%	66.67%	0%	33.33%	0%	0%	100%
M%	34.78%	0%	6.52%	2.17%	20.65%	0%	15.22%	7.61%	13.04%	100%
L%	0%	0%	40.00%	20.00%	20.00%	0%	0%	20.00%	0%	100%
Total	32.00%	0%	8.00%	3.00%	22.00%	0%	15.00%	8.00%	12.00%	100%

Fuente: Prosodia.xls (Base de datos propia)

Del mismo modo se puede restringir el total de tonos nucleares y finales para que muestren las combinaciones posibles en enunciados exclamativos (véase Tablas 3-55 y 3-56).

Tabla 3-55: Número de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0	0	0	0	0	0	0	0	0	0
M%	1	0	0	0	2	0	2	0	3	8
L%	1	0	1	0	1	0	1	0	1	5
Total	2	0	1	0	3	0	3	0	4	13

Fuente: Prosodia.xls (Base de datos propia)

Así, vemos que entre los 475 enunciados analizados de los hablantes de Madrid, no hubo ningún caso de enunciado exclamativo con tono final H%. Los tonos nucleares combinados con M% son *, L*+H, ¡L*+H y L+H*, que agrupados corresponden a 50% del tipo 'S, 37.50% del tipo S y 12.50% del tipo Ś. Por último, de los tonos nucleares combinados con L%, hay un caso por cada uno de los siguientes tonos: L+H*, ¡L+H*, L*+H, ¡L*+H y *, lo que equivale a 40% tanto del tipo Ś como del 'S, además de 20% del tipo S.

Tabla 3-56: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0%	0%	0%	0%	0%	0%	0%	0%	0%	*100%
M%	12.50%	0%	0%	0%	25.00%	0%	25.00%	0%	37.50%	100%
L%	20.00%	0%	20.00%	0%	20.00%	0%	20.00%	0%	20.00%	100%
Total	15.38%	0%	7.69%	0%	23.08%	0%	23.08%	0%	30.77%	100%

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los tonos nucleares y finales en enunciados suspendidos, los segundos más extensos en número, obtenemos los resultados de las Tablas 3-57 y 3-58:

Tabla 3-57: Número de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0	0	2	0	1	0	2	0	0	5
M%	14	0	14	2	9	0	19	27	28	113
L%	2	0	1	0	0	0	0	1	1	5
Total	16	0	17	2	10	0	21	28	29	123

Fuente: Prosodia.xls (Base de datos propia)

En tales Tablas se observa que los únicos tonos nucleares en combinación con H% son ¡L+H*, ¡L*+H y L*+H, lo que equivale a 60% de tonos del tipo 'S y 40% del tipo Š. Por otro lado, los tonos nucleares combinados con M% más comunes son *, H+L*, ¡L*+H, L+H* y ¡L+H*. Del total de combinaciones de tonos se obtienen unos valores muy equitativos para cada grupo: 26.55% del tipo Š, 24.78% del tipo S, 24.77% del tipo 'S y 23.89% del tipo S'. Por último, los tonos nucleares combinados con L% son L+H*, ¡L+H*, H+L* y *, que viene a representar un 60% de tonos del tipo Š, 20% del tipo S' y 20% del tipo S.

Tabla 3-58: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0%	0%	40.00%	0%	20.00%	0%	40.00%	0%	0%	100%
M%	12.39%	0%	12.39%	1.77%	7.96%	0%	16.81%	23.89%	24.78%	100%
L%	40.00%	0%	20.00%	0%	0%	0%	0%	20.00%	20.00%	100%
Total	13.01%	0%	13.82%	1.63%	8.13%	0%	17.07%	22.76%	23.58%	100%

Fuente: Prosodia.xls (Base de datos propia)

Para terminar con los tipos de tonos, en las Tablas 3-59 y 3-60 se presenta el subconjunto de casos de tonos nucleares y sus combinaciones con cada uno de los distintos tonos finales en enunciados neutros:

Tabla 3-59: Número de tonos nucleares combinados con cada tipo de tono final en enunciados neutros

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	1	0	0	0	2	0	0	0	0	3
M%	65	2	7	5	27	0	11	37	69	223
L%	7	0	1	0	0	0	3	2	0	13
Total	73	2	8	5	29	0	14	39	69	239

Fuente: Prosodia.xls (Base de datos propia)

A pesar de ser el grupo más numeroso de enunciados, las combinaciones de los tonos nucleares son mayoritariamente con tonos finales del tipo M%, de las cuales *, L+H*, H+L* y L*+H son las más comunes, es decir, 35.43% de tonos del tipo Š, 30.94% del tipo S, 17.04% del tipo 'S y 16.59% del tipo S'. Las únicas combinaciones de tonos nucleares con tonos finales H% fueron con L*+H y L+H*, lo que equivale a un 66.67% de tonos del tipo 'S y 33.33% del tipo Š. Por último, L+H*, ¡L*+H, H+L* y ¡L+H* son los únicos tonos nucleares que se combinan con L%, que equivalen a 61.54% del tipo Š, 23.08% de 'S y 15.38% de S'.

Tabla 3-60: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados neutros

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	33.33%	0%	0%	0%	66.67%	0%	0%	0%	0%	100%
M%	29.15%	0.90%	3.14%	2.24%	12.11%	0%	4.93%	16.59%	30.94%	100%
L%	53.85%	0%	7.69%	0%	0%	0%	23.08%	15.38%	0%	100%
Total	30.54%	0.84%	3.35%	2.09%	12.13%	0%	5.86%	16.32%	28.87%	100%

Fuente: Prosodia.xls (Base de datos propia)

Para terminar con la entonación, los valores totales también fueron filtrados según si habían sido producidos por mujeres (véase Tablas 3-61 y 3-62) o por hombres (véase Tablas 3-63 y 3-64), para poder observar si había diferencias sustanciales en la elección de tonos.

Tabla 3-61: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	1	0	2	0	3	0	2	0	0	8
M%	67	1	15	5	41	0	34	50	61	274
L%	6	0	4	1	1	0	4	2	0	18
Total	74	1	21	6	45	0	40	52	61	300

Fuente: Prosodia.xls (Base de datos propia)

Entre las mujeres, las únicas combinaciones de tonos nucleares y el tono final H% fueron con L*+H, ¡L+H*, ¡L*+H y L+H*, lo que equivale a 62.50% de tonos del tipo 'S y 37.50% del tipo Ś. Al tono final M% se le combinó principalmente con tonos nucleares L+H*, *, H+L*, L*+H y ¡L*+H, pero todos los tonos agrupados tendrían el siguiente orden de importancia: Ś, con 32.10%, 'S, con 27.37%, S, con 22.26%, y S', con 18.25%. Por último, las principales combinaciones con L% se dan con tonos L+H*, ¡L+H*, ¡L*+H y H+L*, que, en proporción, pertenecen 61.11% al grupo Ś, 27.78% al 'S y 11.11% al S'.

Tabla 3-62: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	12.50%	0%	25.00%	0%	37.50%	0%	25.00%	0%	0%	100%
M%	24.45%	0.36%	5.47%	1.82%	14.96%	0%	12.41%	18.25%	22.26%	100%
L%	33.33%	0%	22.22%	5.56%	5.56%	0%	22.22%	11.11%	0%	100%
Total	24.67%	0.33%	7.00%	2.00%	15.00%	0%	13.33%	17.33%	20.33%	100%

Fuente: Prosodia.xls (Base de datos propia)

En comparación, los enunciados producidos por los hombres arrojan los siguientes resultados:

Tabla 3-63: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0	0	0	0	2	0	1	0	0	3
M%	45	1	12	4	16	0	12	21	51	162
L%	4	0	1	0	1	0	0	2	2	10
Total	49	1	13	4	19	0	13	23	53	175

Fuente: Prosodia.xls (Base de datos propia)

Las únicas combinaciones de tonos nucleares y tono final H% se dieron con L*+H y ¡L*+H, es decir, todos ellos pertenecientes al grupo 'S. Por otro lado, las combinaciones con M% se dieron principalmente con *, L+H*, H+L*, L*+H, ¡L+H* y ¡L*+H, lo que equivale a 38.28% de tonos del tipo Ś, 31.48% del tipo S, 17.29% del tipo 'S y 12.96% del S'. Finalmente, los principales tonos nucleares combinados con L% son L+H*, H+L* y *, que agrupados darían un resultado de 50% de Ś, 20% tanto de S' como de S, y 10% de 'S.

Tabla 3-64: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0%	0%	0%	0%	66.67%	0%	33.33%	0%	0%	100%
M%	27.78%	0.62%	7.41%	2.47%	9.88%	0%	7.41%	12.96%	31.48%	100%
L%	40.00%	0%	10.00%	0%	10.00%	0%	0%	20.00%	20.00%	100%
Total	28.00%	0.57%	7.43%	2.29%	10.86%	0%	7.43%	13.14%	30.29%	100%

Fuente: Prosodia.xls (Base de datos propia)

Con estas últimas Tablas se dan por concluida la serie de combinaciones entre los distintos tipos de tonos y variables tales como el tipo de enunciado y el sexo de los hablantes. Se descartaron otros tipos de combinaciones porque, a pesar de contar con 475 enunciados en el *corpus* de Madrid, las Tablas empezaban a mostrar en mayor medida la falta de una cantidad significativa de ejemplos para cada caso y esto, a su vez, arrojaba porcentajes poco fiables. Sin embargo, la posibilidad de continuar con las combinaciones deberá de repensarse cuando la base de datos contenga una cantidad mucho mayor de ejemplos.

3.2.6 Ritmo

La misma información sobre los acentos alrededor de los cuales se originan los movimientos tonales analizados en el apartado anterior nos sirve para estudiar las configuraciones compuestas por una sílaba acentuada y, opcionalmente, de una hasta cinco sílabas átonas dependientes de ella, según el esquema propuesto en 1.3.4, donde se trataba el tema de los ritmos abstractos. En aquel apartado, así como en la Tabla 1-7, se hizo un recuento de todos los grupos rítmicos teóricamente posibles en español.

Dentro de este texto, hacer referencia a los distintos tipos de configuración sería confuso, por eso a partir de ahora nos referiremos a los grupos rítmicos según el número de identificación del 0 al 20 que aparecerá en la segunda columna de las Tablas 3-65 a 3-71. Cabe aclarar que para la recopilación de los resultados se dejó de lado la oposición entre focalizado y no focalizado, es decir, (S W) se interpretó como idéntico a (S W) – grupo rítmico 1.

En las Tablas 3-65 a 3-69 se presenta los resultados del conteo de los distintos grupos rítmicos y su equivalencia porcentual, primero en los enunciados producidos por las mujeres (Tablas 3-65 y 3-66), después los producidos por los hombres (Tablas 3-67 y 3-68), y, por último, los valores totales por grupo y en general (Tabla 3-69).

En el caso de las mujeres, en los valores de la Tabla 3-65 se observa el inventario y la cantidad de grupos rítmicos empleados por cada hablante. En todos los casos, el grupo rítmico más común es el 1, mientras que de los grupos 10, 11, 15, 16 y 17 (todos ellos con una secuencia final de al menos 3 sílabas átonas) no se produjo ningún caso.

Tabla 3-65: Número de grupos rítmicos en enunciados producidos por mujeres

Grupo rítmico		GEM	HEL	LET	PAT	PTR	ROS	SON
(0)	0	17	23	19	9	15	10	22
(s w)	1	35	40	33	33	26	23	49
(w s)	2	9	14	16	12	11	11	21
(s w w)	3	4	5	5	3	1	1	6
(w s w)	4	16	19	12	11	24	13	25
(w w s)	5	6	6	6	2	5	1	7
(s w w w)	6	0	2	0	1	0	0	0
(w s w w)	7	1	3	3	2	4	4	0
(w w s w)	8	10	10	18	17	12	10	7
(w w w s)	9	6	0	2	0	2	0	4
(s w w w w)	10	0	0	0	0	0	0	0
(w s w w w)	11	0	0	0	0	0	0	0
(w w s w w)	12	0	2	2	2	1	0	3
(w w w s w)	13	11	5	3	3	6	8	4
(w w w w s)	14	0	0	1	1	0	0	1
(s w w w w w)	15	0	0	0	0	0	0	0
(w s w w w w)	16	0	0	0	0	0	0	0
(w w s w w w)	17	0	0	0	0	0	0	0
(w w w s w w)	18	1	0	1	0	1	0	1
(w w w w s w)	19	2	0	0	1	3	1	0
(w w w w w s)	20	0	1	0	0	0	0	0
Total		118	130	121	97	111	82	150

Fuente: Prosodia.xls (Base de datos propia)

Con los valores porcentuales se puede observar que GEM, en orden descendente, utilizó principalmente los grupos 1, 0, 4, 13, 8; HEL empleó 1, 0, 4, 2, 8; LET produjo mayoritariamente 1, 0, 8, 2, 4; PAT hizo mayor uso de 1, 8, 2, 4, 0; PTR utilizó sobre todo 1, 4, 0, 8, 2; mientras que ROS produjo especialmente los grupos 1, 4, 2, 0 y 8, 13; por último, SON hizo uso de 1, 4, 0, 2, 5 y 8. A pesar de que ninguna de las mujeres obtuvo siquiera la misma secuencia de grupos rítmicos, sí hay una preferencia por los grupos 1 (23.42% - 34.02%), 4 (9.92% - 21.62%), 0 (9.28% - 17.69%), 2 (7.63% - 14%) y 8 (4.67% - 17.53%), que se caracterizan por ser muy cortos. La presencia del grupo 0, monosilábico, es engañosa, ya que esta estructura difícilmente tiene la capacidad de agruparse en configuraciones rítmicas. Por otra parte, en todos los casos hay una preferencia clara de 1 sobre 2, es decir, del troqueo sobre el yambo, de la acentuación grave sobre la aguda. Esta preferencia se confirma en la alta incidencia de los grupos 4 y 8, que, a pesar de ser trisilábico y tetrasilábico respectivamente, ambos tienen una acentuación grave.

Algo que deberá verificarse con una base de datos aún mayor es si los grupos 10, 11, 15, 16 y 17 son solamente raros o si son absolutamente imposibles. En el caso del grupo 6 sí se dan tres casos (dos de HEL y uno de PAT) de sílaba tónica seguida de tres sílabas átonas. Según la clasificación tradicional de la acentuación léxica en aguda, grave, esdrújula y sobreesdrújula, ésta última tiene tres o más sílabas después de la tónica. Sin embargo, parece

poco posible que grupos como 10 y 16 (con cuatro sílabas átonas), o incluso 15 (con cinco) se produzcan en el habla espontánea.

Tabla 3-66: Porcentaje de grupos rítmicos en enunciados producidos por mujeres

Grupo rítmico		GEM	HEL	LET	PAT	PTR	ROS	SON
(0)	0	14.41%	17.69%	15.70%	9.28%	13.51%	12.20%	14.67%
(s w)	1	29.66%	30.77%	27.27%	34.02%	23.42%	28.05%	32.67%
(w s)	2	7.63%	10.77%	13.22%	12.37%	9.91%	13.41%	14.00%
(s w w)	3	3.39%	3.85%	4.13%	3.09%	0.90%	1.22%	4.00%
(w s w)	4	13.56%	14.62%	9.92%	11.34%	21.62%	15.85%	16.67%
(w w s)	5	5.08%	4.62%	4.96%	2.06%	4.50%	1.22%	4.67%
(s w w w)	6	0%	1.54%	0%	1.03%	0%	0%	0%
(w s w w)	7	0.85%	2.31%	2.48%	2.06%	3.60%	4.88%	0%
(w w s w)	8	8.47%	7.69%	14.88%	17.53%	10.81%	12.20%	4.67%
(w w w s)	9	5.08%	0%	1.65%	0%	1.80%	0%	2.67%
(s w w w w)	10	0%	0%	0%	0%	0%	0%	0%
(w s w w w)	11	0%	0%	0%	0%	0%	0%	0%
(w w s w w)	12	0%	1.54%	1.65%	2.06%	0.90%	0%	2.00%
(w w w s w)	13	9.32%	3.85%	2.48%	3.09%	5.41%	9.76%	2.67%
(w w w w s)	14	0%	0%	0.83%	1.03%	0%	0%	0.67%
(s w w w w w)	15	0%	0%	0%	0%	0%	0%	0%
(w s w w w w)	16	0%	0%	0%	0%	0%	0%	0%
(w w s w w w)	17	0%	0%	0%	0%	0%	0%	0%
(w w w s w w)	18	0.85%	0%	0.83%	0%	0.90%	0%	0.67%
(w w w w s w)	19	1.69%	0%	0%	1.03%	2.70%	1.22%	0%
(w w w w w s)	20	0%	0.77%	0%	0%	0%	0%	0%

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres no es tan clara la preferencia por el grupo rítmico 1, pues FRA y MIG tienen mayoría de 4, con 19 observaciones, aunque éste último tiene 18 observaciones de 0. GUI presenta el caso contrario: mayoría de 1, con 13 observaciones, pero 12 casos del grupo 4. La preferencia de grupos rítmicos por persona es la siguiente: ABE utiliza principalmente 1, 4 y 5, 2 y 8; DAN produce especialmente 1, 0, 4 y 8, 5; mientras que FRA hace mayor uso de 4, 8 y 13, así como 0, 1 y 2; mientras que GUI utiliza 1, 4, 0, 2, 5; MGL usa los grupos 4, 1, 8, 0 y 2, 13; y, por último, PAC produce mayoritariamente 1, 4, 0, 2, 8 y 13. Nuevamente, ninguna de las preferencias por orden descendente coincide con la de alguien más, ni siquiera con alguna de las mujeres, aunque las preferencias rítmicas de GUI y PAC son semejantes a las de SON.

Dentro de los enunciados de los hombres no apareció ningún caso del grupo 6, además de que se confirma la ausencia de los grupos 10, 15, 16 y 17. No obstante, hay dos ejemplos del grupo 11, uno de ellos producido por ABE y otro por DAN, mientras que se agrega a la lista de grupos raros al 20, para el cual sí había una observación (de HEL) entre las mujeres, un grupo de cinco sílabas átonas y una vocal tónica colocada justo al final del grupo rítmico. Seguramente es la acentuación aguda la que lo hace tan poco común. En general, los grupos

rítmicos más comunes entre los hombres son 1 (11.48% - 36.36%), 4 (11.32% - 23.53%), 0 (4.76% - 17.65%), 8 (5.88% - 18.87%) y 2 (3.41% - 11.76%).

Tabla 3-67: Número de grupos rítmicos en enunciados producidos por hombres

Grupo rítmico		ABE	DAN	FRA	GUI	MGL	MIG	PAC
(0)	0	1	13	7	9	7	10	9
(sw)	1	6	32	7	13	18	18	23
(ws)	2	2	3	7	6	2	10	6
(sww)	3	0	2	0	1	0	3	0
(wsw)	4	4	10	13	12	6	19	14
(wws)	5	4	4	3	4	4	2	3
(swww)	6	0	0	0	0	0	0	0
(wsww)	7	0	2	1	1	1	5	0
(wwsw)	8	2	10	11	3	10	15	4
(wwws)	9	0	3	1	1	0	1	0
(swwww)	10	0	0	0	0	0	0	0
(wswww)	11	1	1	0	0	0	0	0
(wwsww)	12	0	1	0	0	0	5	0
(wwwsw)	13	1	3	11	1	2	7	4
(wwwws)	14	0	2	0	0	1	0	0
(swwwww)	15	0	0	0	0	0	0	0
(wswwww)	16	0	0	0	0	0	0	0
(wwswww)	17	0	0	0	0	0	0	0
(wwwwsw)	18	0	0	0	0	0	1	0
(wwwwsw)	19	0	2	0	0	2	3	1
(wwwwws)	20	0	0	0	0	0	0	0
Total		21	88	61	51	53	99	64

Fuente: Prosodia.xls (Base de datos propia)

Tabla 3-68: Porcentaje de grupos rítmicos en enunciados producidos por hombres

Grupo rítmico		ABE	DAN	FRA	GUI	MGL	MIG	PAC
(0)	0	4.76%	14.77%	11.48%	17.65%	13.21%	10.10%	14.06%
(sw)	1	28.57%	36.36%	11.48%	25.49%	33.96%	18.18%	35.94%
(ws)	2	9.52%	3.41%	11.48%	11.76%	3.77%	10.10%	9.38%
(sww)	3	0%	2.27%	0%	1.96%	0%	3.03%	0%
(wsw)	4	19.05%	11.36%	21.31%	23.53%	11.32%	19.19%	21.88%
(wws)	5	19.05%	4.55%	4.92%	7.84%	7.55%	2.02%	4.69%
(swww)	6	0%	0%	0%	0%	0%	0%	0%
(wsww)	7	0%	2.27%	1.64%	1.96%	1.89%	5.05%	0%
(wwsw)	8	9.52%	11.36%	18.03%	5.88%	18.87%	15.15%	6.25%
(wwws)	9	0%	3.41%	1.64%	1.96%	0%	1.01%	0%
(swwww)	10	0%	0%	0%	0%	0%	0%	0%
(wswww)	11	4.76%	1.14%	0%	0%	0%	0%	0%
(wwsww)	12	0%	1.14%	0%	0%	0%	5.05%	0%
(wwwsw)	13	4.76%	3.41%	18.03%	1.96%	3.77%	7.07%	6.25%
(wwwws)	14	0%	2.27%	0%	0%	1.89%	0%	0%
(swwwww)	15	0%	0%	0%	0%	0%	0%	0%
(wswwww)	16	0%	0%	0%	0%	0%	0%	0%
(wwswww)	17	0%	0%	0%	0%	0%	0%	0%
(wwwwsw)	18	0%	0%	0%	0%	0%	1.01%	0%
(wwwwsw)	19	0%	2.27%	0%	0%	3.77%	3.03%	1.56%
(wwwwws)	20	0%	0%	0%	0%	0%	0%	0%

Fuente: Prosodia.xls (Base de datos propia)

En la Tabla 3-69 podemos ver los valores totales y porcentuales de hombres y mujeres, además del total general. Ahí se puede observar que entre ambos sexos la única diferencia radica en el uso de los grupos 2 y 8, pues en las mujeres es más común el grupo 2 que el 8, mientras que entre los hombres se da el caso contrario. Además, en cuanto a la presencia o ausencia de determinados grupos rítmicos, entre las producciones de las mujeres no aparece ningún grupo 11, mientras que en las producciones de los hombres no hay ejemplos de 6 y 20. Así, estos tres grupos podrían considerarse muy poco comunes, mientras que los grupos 10, 15, 16 y 17 deben ser considerados extremadamente raros, si no es que imposibles.

Tabla 3-69: Número total y porcentaje de grupos rítmicos

Grupo rítmico		Mujeres		Hombres		Total	
(0)	0	115	14.22%	56	12.81%	171	13.72%
(s w)	1	239	29.54%	117	26.77%	356	28.57%
(w s)	2	94	11.62%	36	8.24%	130	10.43%
(s w w)	3	25	3.09%	6	1.37%	31	2.49%
(w s w)	4	120	14.83%	78	17.85%	198	15.89%
(w w s)	5	33	4.08%	24	5.49%	57	4.57%
(s w w w)	6	3	0.37%	0	0%	3	0.24%
(w s w w)	7	17	2.10%	10	2.29%	27	2.17%
(w w s w)	8	84	10.38%	55	12.59%	139	11.16%
(w w w s)	9	14	1.73%	6	1.37%	20	1.61%
(s w w w w)	10	0	0%	0	0%	0	0%
(w s w w w)	11	0	0%	2	0.46%	2	0.16%
(w w s w w)	12	10	1.24%	6	1.37%	16	1.28%
(w w w s w)	13	40	4.94%	29	6.64%	69	5.54%
(w w w w s)	14	3	0.37%	3	0.69%	6	0.48%
(s w w w w w)	15	0	0%	0	0%	0	0%
(w s w w w w)	16	0	0%	0	0%	0	0%
(w w s w w w)	17	0	0%	0	0%	0	0%
(w w w s w w)	18	4	0.49%	1	0.23%	5	0.40%
(w w w w s w)	19	7	0.87%	8	1.83%	15	1.20%
(w w w w w s)	20	1	0.12%	0	0%	1	0.08%
Total		809	100%	437	100%	1246	100%

Fuente: Prosodia.xls (Base de datos propia)

Entonces, el orden de importancia de los grupos rítmicos que utilizan los hablantes de Madrid sería el siguiente: **1**, **4**, **0**, **8**, **2**, **13**, **5**, 3, 7, **9**, 12, **19**, *14*, 18, 6, 11, *20*. Los distintos estilos tipográficos indican el tipo de acentuación: el estilo normal equivale a la acentuación nula, la letra cursiva indica la acentuación aguda (última sílaba), con la letra negrita se señala la acentuación grave (penúltima sílaba), el subrayado sencillo indica la acentuación esdrújula (antepenúltima sílaba) y el doble subrayado señala la acentuación sobresdrújula. Con esta marcación queda clara la supremacía de la acentuación grave (en total, 62.36%), seguida de la aguda (17.17%) y la nula (13.72%), lo que representa el 93.25% del total de grupos rítmicos. Ahora, si aplicamos el filtro para observar separadamente los distintos tipos de enunciados (véase Tabla 3-70), veremos que en los cuatro casos el grupo rítmico más común, con rangos

porcentuales entre 26.99% y 40.00%, es el 1. Entonces, en los enunciados interrogativos los grupos más comunes son **1, 4, 0, 2 y 8**; en los enunciados exclamativos, los más utilizados son **1, 2, 3, 4 y 0**; en los enunciados suspendidos son **1, 0, 4, 8, 2**, y por último, en los enunciados neutros, el orden de preferencia es **1, 4, 0, 8 y 2**. Aunque ninguna secuencia es idéntica, sobresale el caso de los enunciados exclamativos, en los que la presencia de grupos rítmicos con acentuación aguda y esdrújula es más alta que en los demás.

Tabla 3-70: Número total y porcentaje de grupos rítmicos por tipo de enunciado

Grupo rítmico		/+I -E -S/		/I +E -S/		/I -E +S/		/I -E -S/	
(0)	0	35	13.73%	2	8.00%	45	14.33%	89	13.65%
(s w)	1	82	32.16%	10	40.00%	88	28.03%	176	26.99%
(w s)	2	29	11.37%	4	16.00%	29	9.24%	68	10.43%
(s w w)	3	4	1.57%	3	12.00%	15	4.78%	9	1.38%
(w s w)	4	40	15.69%	3	12.00%	44	14.01%	111	17.02%
(w w s)	5	11	4.31%	1	4.00%	10	3.18%	35	5.37%
(s w w w)	6	0	0%	1	4.00%	1	0.32%	1	0.15%
(w s w w)	7	2	0.78%	0	0%	9	2.87%	16	2.45%
(w w s w)	8	27	10.59%	0	0%	35	11.15%	77	11.81%
(w w w s)	9	5	1.96%	0	0%	5	1.59%	10	1.53%
(s w w w w)	10	0	0%	0	0%	0	0%	0	0%
(w s w w w)	11	0	0%	0	0%	2	0.64%	0	0%
(w w s w w)	12	5	1.96%	0	0%	5	1.59%	6	0.92%
(w w w s w)	13	12	4.71%	0	0%	15	4.78%	42	6.44%
(w w w w s)	14	0	0%	0	0%	3	0.96%	3	0.46%
(s w w w w w)	15	0	0%	0	0%	0	0%	0	0%
(w s w w w w)	16	0	0%	0	0%	0	0%	0	0%
(w w s w w w)	17	0	0%	0	0%	0	0%	0	0%
(w w w s w w)	18	0	0%	1	4.00%	3	0.96%	1	0.15%
(w w w w s w)	19	3	1.18%	0	0%	5	1.59%	7	1.07%
(w w w w w s)	20	0	0%	0	0%	0	0%	1	0.15%
Total		255	100%	25	100%	314	100%	652	100%

Fuente: Prosodia.xls (Base de datos propia)

Finalmente, para dar por concluido el apartado sobre ritmo y, al mismo tiempo, clausurar la sección dedicada al *corpus* de Madrid, en la Tabla 3-71 se hace un recuento del número de casos en que los grupos rítmicos contiguos sean idénticos, en otras palabras, este análisis muestra de manera estadística los casos en los que se da el principio de isocronía que es percibido en el habla como el ritmo propiamente dicho.

Así es como podemos observar que, en general, el ritmo más común es el compuesto por secuencias de dos grupos del tipo 1, con un total de 15 casos, que es lo que en los estudios clásicos sobre métrica se denomina ritmo trocaico, teniendo en cuenta que estos términos se aplican a las secuencias de pies y, por lo tanto, en los casos de más de cuatro sílabas se tienen que concebir los grupos rítmicos como bípodos o bímetros. A continuación tenemos igual número de casos, 10, para la secuencia de dos grupos del tipo 4, ritmo anfibráquico, y tres grupos del tipo 1. De este último grupo rítmico también hay un caso excepcional donde hay

cuatro grupos contiguos idénticos. Otro grupo muy representado es el 8, que correspondería a un ritmo de tercer peón o dibraquio más troqueo, $(w w) + (s w)$, pues en 6 ocasiones se repite dos veces y en dos ocasiones se repite tres veces. Otros grupos rítmicos que se repiten dos veces son el 2, ritmo yámbico, el 13, combinación de dibraquio y anfibraco, $(w w) + (w s w)$, y finalmente los grupos 5, ritmo anapéstico, 7, ritmo de segundo peón, y 12, combinación de dibraquio con dáctilo, $(w w) + (s w w)$.

Tabla 3-71: Número de grupos rítmicos idénticos

Grupos rítmicos		2x	3x	4x	6x
(0)	0	4	4	0	1
(s w)	1	15	10	1	0
(w s)	2	3	0	0	0
(s w w)	3	0	0	0	0
(w s w)	4	10	0	0	0
(w w s)	5	1	0	0	0
(s w w w)	6	0	0	0	0
(w s w w)	7	1	0	0	0
(w w s w)	8	6	2	0	0
(w w w s)	9	0	0	0	0
(s w w w w)	10	0	0	0	0
(w s w w w)	11	0	0	0	0
(w w s w w)	12	1	0	0	0
(w w w s w)	13	2	0	0	0
(w w w w s)	14	0	0	0	0
(s w w w w w)	15	0	0	0	0
(w s w w w w)	16	0	0	0	0
(w w s w w w)	17	0	0	0	0
(w w w s w w)	18	0	0	0	0
(w w w w s w)	19	0	0	0	0
(w w w w w s)	20	0	0	0	0
Total		43	16	1	1
Real		39	12	1	0

Fuente: Prosodia.xls (Base de datos propia)

Se excluye del análisis las secuencias de dos, tres e incluso seis grupos rítmicos del tipo cero porque su contigüidad no siempre es percibida como una estructura rítmica. Esa es la razón por la cual hay un primer total en el que se incluye en el conteo y un segundo total más realista en el que se excluyen. Por tanto, en el *corpus* de Madrid se dan 39 casos de dos grupos idénticos contiguos, 12 casos de tres grupos idénticos y solamente un caso de cuatro grupos del mismo tipo en contigüidad, lo que nos da un total de 52 casos de isocronía.

Si tomamos en cuenta que en el caso de Madrid se analizaron 475 enunciados, nos daremos cuenta de que el número total de casos de isocronía se presenta únicamente en un 10.95% de enunciados. Seguramente hay situaciones en el habla cotidiana en las que el porcentaje de enunciados con algún caso de isocronía sea más elevado, sobre todo en situaciones lúdicas, pero en el presente *corpus* éste no fue el caso.

3.3 Corpus de la Ciudad de México

En la sección anterior, dedicada únicamente al análisis estadístico del *corpus* de grabaciones de Madrid, se fue justificando la metodología y las distintas decisiones arbitrarias que fueron tomadas con el fin de establecer unas bases propicias para el estudio comparativo. Todas estas implicaciones y observaciones aplican también para el *corpus* de la Ciudad de México, por lo que en esta sección se evitará profundizar en detalles metodológicos. En caso de ser necesario, se hará referencia al apartado homólogo en la sección previa. Además, en este punto, también se evitará hacer analogías con respecto a la información de Madrid, ya que el apartado 3.4 se dedicará exclusivamente al estudio comparativo entre las dos variedades.

3.3.1 Sílabas y enunciados

Para la creación del *corpus* de la Ciudad de México se analizaron 2101 sílabas producidas por mujeres (véase Tabla 3-72), repartidas entre 277 enunciados, lo que da un promedio general de 7.58 sílabas por enunciado. FAB es la mujer menos representada en el *corpus*, con 80 sílabas provenientes de 14 enunciados, mientras que MAR produjo un total de 727 sílabas en 80 enunciados. MIR es quien más enunciados aportó al *corpus*, 102. Las personas que más se alejan del promedio de sílabas por enunciado son CEC, con 4.93 sílabas por enunciado, y ADR, quien con 9.81 sílabas produce los enunciados más largos en promedio.

Tabla 3-72: Promedio de sílabas por enunciado (mujeres)

	Sílabas	Enunciados	Promedio
ADR	510	52	9.81
CEC	143	29	4.93
FAB	80	14	5.71
MAR	727	80	9.09
MIR	641	102	6.28
Mujeres	2101	277	7.58

Fuente: Prosodia.xls (Base de datos propia)

La contraparte masculina está representada con 4155 sílabas distribuidas a lo largo de 593 enunciados, lo que da un promedio general de 7.01 sílabas por enunciado (véase Tabla 3-73), que es ligeramente inferior al promedio de las mujeres. El hombre menos representado es ROD, con tan sólo 58 sílabas, mientras que MOI, quien participó en la misma conversación que ROD, es quien aportó la mayor cantidad de sílabas al análisis, 1209. También son ROD y MOI quienes representan los extremos con respecto a la cantidad de enunciados: 12 y 200, respectivamente. Las personas que más se alejan del promedio de sílabas por enunciado son ROD, quien con únicamente 4.83 sílabas produjo los enunciados más cortos del *corpus*, y FRA, quien aportó enunciados con un promedio de 8.47 sílabas.

Tabla 3-73: Promedio de sílabas por enunciado (hombres)

	Sílabas	Enunciados	Promedio
ALB	374	58	6.45
FRA	720	85	8.47
HUI	1128	151	7.47
IVA	666	87	7.66
MOI	1209	200	6.05
ROD	58	12	4.83
Hombres	4155	593	7.01

Fuente: Prosodia.xls (Base de datos propia)

Tomando en cuenta el número total de sílabas analizadas en el *corpus* de la Ciudad de México, 6256, distribuidas entre un total de 870 enunciados, nos arroja un promedio de 7.19 sílabas por enunciado (véase Tabla 3-74). Si los resultados se filtran según el tipo de enunciado, podremos observar claramente que los enunciados exclamativos son los menos representados, con 58 sílabas producidas a lo largo de 16 enunciados, pero también son, en promedio, los enunciados más cortos, con un promedio de 3.63 sílabas por enunciado. Por otra parte, los enunciados suspendidos son los más representados, con 3768 sílabas repartidas entre 499 enunciados, y, además, son los enunciados con el mayor número de sílabas, 7.55.

Tabla 3-74: Promedio de sílabas por tipo de enunciado

	Sílabas	Enunciados	Promedio
/+I -E -S/	659	104	6.34
/-I +E -S/	58	16	3.63
/-I -E +S/	3768	499	7.55
/-I -E -S/	1771	251	7.06
Total	6256	870	7.19

Fuente: Prosodia.xls (Base de datos propia)

Veamos ahora en la Tabla 3-75 la cantidad de enunciados de cada tipo que aportaron las mujeres. En relación directa con el número total de enunciados, FAB es quien menos enunciados de cada tipo produce, mientras que MIR es quien aporta la mayor cantidad de enunciados de cada tipo. Los enunciados suspendidos son los más comunes, 167, seguidos de los enunciados neutros, 78, y los interrogativos, 27, para terminar con los enunciados exclamativos, de los cuales las mujeres sólo aportan 5.

Tabla 3-75: Número de enunciados de cada tipo por persona (mujeres)

	ADR	CEC	FAB	MAR	MIR	Mujeres
/+I -E -S/	7	6	0	2	12	27
/-I +E -S/	1	1	0	1	2	5
/-I -E +S/	30	10	6	58	63	167
/-I -E -S/	14	12	8	19	25	78
Total	52	29	14	80	102	277

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres (véase Tabla 3-76), también es el número total de enunciados lo que determina que ROD sea quien menos enunciados de cada tipo produzca y que MOI sea

quien aporte la mayor cantidad de ellos. Además, el peso relativo que tiene cada uno de los tipos de enunciados es el mismo que en el caso de las mujeres, pues los de mayor presencia son los enunciados suspendidos, con 332, seguidos de los neutros, con 173, y los interrogativos, con 77 observaciones, mientras que los enunciados exclamativos son los menos representados, con sólo 11 casos.

Tabla 3-76: Número de enunciados de cada tipo por persona (hombres)

	ALB	FRA	HUI	IVA	MOI	ROD	Hombres
/+I -E -S/	12	9	19	9	26	2	77
/-I +E -S/	2	0	3	1	5	0	11
/-I -E +S/	17	49	85	62	112	7	332
/-I -E -S/	27	27	44	15	57	3	173
Total	58	85	151	87	200	12	593

Fuente: Prosodia.xls (Base de datos propia)

Como estos datos no se compararán en la sección 3.4 directamente con los de Madrid, cabe mencionar que en el *corpus* de Madrid son los enunciados neutros los más numerosos, seguidos de los enunciados suspendidos. Aunque, como se ha ido advirtiendo en distintos apartados, esta diferencia puede deberse, sobre todo, a diferencias de criterio a la hora de la transcripción, o bien, a factores tan abstractos como el grado de familiaridad entre los participantes de la conversación o las diferencias en la temática.

3.3.2 Duración

El número de sílabas producido por cada persona, pero sobre todo el número de enunciados, es el principal factor que influyó en la duración total de los segmentos analizados. Por esta razón, no es de sorprender que nuevamente sean FAB, con 12.712 segundos, y ROD, con 8.790 seg, quienes están menos representados en el *corpus*, mientras que MIR, con 112.290 seg, y MOI, con 192.957 seg, sean los que más aportan a la muestra (véase Tabla 3-77). La selección, como se comentó en la sección anterior, obedeció principalmente a criterios de calidad. Esta es la razón por la cual las mujeres, en total, contribuyeron con 328.309 seg, mientras que los hombres lo hicieron con 655.725 seg (una diferencia de 327.416 seg), para dar un total general de 984.03 segundos.

Tabla 3-77: Duración total de sílabas por persona (en segundos y minutos)

	ADR	CEC	FAB	MAR	MIR		Mujeres	Total
Duración total (s)	80.895	23.295	12.712	99.117	112.290		328.309	
Duración total (min)	01:20.9	00:23.3	00:12.7	01:39.1	01:52.3		05:28.3	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	
Duración total (s)	51.019	113.723	173.964	115.272	192.957	8.790	655.725	984.03
Duración total (min)	00:51.0	01:53.7	02:54.0	01:55.3	03:13.0	00:08.8	10:55.7	16:24.0

Fuente: Prosodia.xls (Base de datos propia)

Con respecto a la duración total presentada en la Tabla 3-2, 17:59.10 minutos, hay una diferencia de 1:35.10 minutos o 95.10 segundos debida a los silencios dentro de los fragmentos analizados.

Por otra parte, la duración máxima, mínima y promedio de las sílabas puede observarse en la Tabla 3-78. La duración máxima promedio entre las mujeres alcanza su valor máximo en MIR, con 0.299 seg, aunque ADR obtiene un valor muy cercano, 0.291 seg. MAR, por otro lado, es quien produce la duración mínima promedio de 0.078 seg. Quienes más se alejan del promedio femenino de 0.165 segundos son MIR, con 0.182 seg, y MAR, con 0.142 seg.

Tabla 3-78: Duración silábica máxima, mínima y promedio (segundos)

	ADR	CEC	FAB	MAR	MIR	Mujeres		Total
Dur. máx. promedio	0.291	0.249	0.237	0.233	0.299	0.270		
Dur. mín. promedio	0.089	0.114	0.111	0.078	0.106	0.096		
Duración promedio	0.165	0.168	0.164	0.142	0.182	0.165		
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	
Dur. máx. promedio	0.234	0.288	0.282	0.290	0.263	0.279	0.273	0.272
Dur. mín. promedio	0.091	0.088	0.088	0.101	0.099	0.103	0.094	0.095
Duración promedio	0.150	0.162	0.162	0.177	0.167	0.175	0.165	0.165

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres, la duración máxima promedio más alta es la de IVA, con 0.290 seg, aunque la diferencia con respecto a FRA, HUI y ROD es mínima. La duración mínima también se la disputan entre dos, HUI, con 0.086 seg, y FRA, con 0.088 seg. Sin embargo, los valores promediados sí son claros, pues quienes más se alejan del promedio general de 0.165 segundos son ALB, con 0.150 seg, e IVA, con 0.177 seg. El promedio de hombres y mujeres también es de 0.165 segundos.

Los tipos de enunciados muestran valores más diferenciados (véase Tabla 3-79), pues son los enunciados suspendidos los que alcanzan la máxima duración, 0.297 seg, y la duración promedio más alta, 0.175 seg, mientras que los enunciados interrogativos alcanzan la duración mínima, 0.085 seg, y el valor promedio más bajo, 0.147 seg.

Tabla 3-79: Duración silábica máxima, mínima y promedio por tipo de enunciado (segundos)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Dur. máx. promedio	0.231	0.229	0.297	0.243
Dur. mín. promedio	0.085	0.117	0.099	0.089
Duración promedio	0.147	0.170	0.175	0.151

Fuente: Prosodia.xls (Base de datos propia)

En cuanto a la velocidad, ésta puede ser medida indirectamente mediante la fórmula inversa, es decir, el análisis del número máximo, mínimo y promedio de sílabas por segundo (véase Tabla 3-80). Así, de entre las mujeres, MAR alcanza un valor máximo de 12.792 sílabas por segundo, mientras que MIR produjo un valor mínimo de 3.341 síl/seg. Además, con respecto

al promedio general femenino de 6.066 síl/seg, es MAR quien habla más rápido, con 7.063 síl/seg, y MIR es quien habla más lentamente, con 5.483 síl/seg.

De los hombres, FRA es quien alcanza el valor más alto, 11.408 síl/seg, aunque HUI lo sigue de cerca, mientras que IVA es quien alcanza la velocidad más baja, 3.445 síl/seg, y también es él quien alcanza el valor promedio más bajo, 5.655 síl/seg. Con respecto al promedio general masculino, ALB es quien habla más velozmente, con 6.673 síl/seg. De este modo, el promedio de la Ciudad de México es de 6.069 sílabas por segundo.

Tabla 3-80: Número máximo, mínimo y promedio de sílabas por segundo

	ADR	CEC	FAB	MAR	MIR	Mujeres		Total
Máx. sílabas x seg.	11.238	8.769	9.050	12.792	9.399	10.419		
Mín. sílabas x seg.	3.441	4.019	4.216	4.296	3.341	3.703		
Prom. sílabas x seg.	6.076	5.941	6.105	7.063	5.483	6.066		
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	Total
Máx. sílabas x seg.	10.964	11.408	11.326	9.892	10.130	9.717	10.618	10.554
Mín. sílabas x seg.	4.281	3.467	3.540	3.445	3.805	3.585	3.663	3.676
Prom. sílabas x seg.	6.673	6.166	6.186	5.655	5.999	5.705	6.069	6.068

Fuente: Prosodia.xls (Base de datos propia)

En la Tabla 3-81 podemos observar los valores alcanzados en cada uno de los tipos de enunciados. Por ejemplo, en los enunciados interrogativos se alcanza la velocidad máxima de 11.797 síl/seg, mientras que en los enunciados suspendidos se obtiene la velocidad mínima de 3.370 síl/seg. En promedio son otra vez los enunciados interrogativos los pronunciados con la mayor rapidez, mientras que los suspendidos son los que se pronuncian con la mayor lentitud, 6.818 y 5.700 sílabas por segundo respectivamente.

Tabla 3-81: Número máximo, mínimo y promedio de sílabas por segundo y por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Máx. sílabas x seg.	11.797	8.579	10.085	11.268
Mín. sílabas x seg.	4.323	4.364	3.370	4.121
Prom. sílabas x seg.	6.818	5.887	5.700	6.629

Fuente: Prosodia.xls (Base de datos propia)

Volviendo a la duración silábica, si se resta el porcentaje de variación mínima promedio al porcentaje de la variación máxima, obtenemos el rango promedio de variación de la duración silábica utilizada por cada uno de los hablantes de la Ciudad de México (véase Tabla 3-82). A saber, entre las mujeres, ADR es quien hace un juego más amplio con la duración, 197.81%, mientras que CEC es quien utiliza el rango más estrecho, 99.07%.

De los hombres, ROD es quien utiliza el rango de variación más amplio, 188.72%, mientras que ALB es quien hace el uso más estrecho del rango, 135.74%. Los promedios de hombres y mujeres son muy similares, lo que arroja un promedio general de 158.78% de oscilación entre la duración silábica.

Tabla 3-82: Variación máxima y mínima promedio de la duración silábica

	ADR	CEC	FAB	MAR	MIR	Mujeres	Total
Var. máxima promedio	160.93%	74.81%	79.75%	138.08%	116.84%	124.98%	
Var. mínima promedio	-36.88%	-24.26%	-34.96%	-44.00%	-31.18%	-35.42%	
Rango de var. prom.	197.81%	99.07%	114.71%	182.08%	148.02%	160.40%	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres
Var. máxima promedio	103.01%	137.52%	136.68%	123.73%	112.71%	157.54%	123.95%
Var. mínima promedio	-32.73%	-42.55%	-36.22%	-35.64%	-32.49%	-31.18%	-34.08%
Rango de var. prom.	135.74%	180.07%	172.90%	159.37%	145.20%	188.72%	158.03%

Fuente: Prosodia.xls (Base de datos propia)

En la Tabla 3-83 se aprecia que en los enunciados suspendidos es en los que se hace un uso más amplio del rango de variación, mientras que en los enunciados exclamativos el rango utilizado es extremadamente estrecho, casi un tercio con respecto al de los suspendidos.

Tabla 3-83: Variación máxima y mínima promedio de la duración silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Var. máxima promedio	110.13%	66.16%	133.16%	116.17%
Var. mínima promedio	-36.91%	10.68%	-33.83%	-37.73%
Rango de var. prom.	147.04%	55.48%	166.99%	153.90%

Fuente: Prosodia.xls (Base de datos propia)

Por último, se analiza también la desviación estándar máxima, mínima y promedio de la duración silábica (véase Tabla 3-84). De este modo, entre las mujeres, son las sílabas de ADR las que alcanzan una mayor desviación, 24.51%, mientras que la menor desviación aparece en las producciones de MAR, con 0.00%. En promedio es MIR quien alcanza una mayor desviación con respecto al promedio, con 5.74%, mientras que FAB es quien produce sus sílabas con la menor desviación estándar promedio, 4.33%.

Tabla 3-84: Desviación estándar máxima, mínima y promedio de la duración silábica

	ADR	CEC	FAB	MAR	MIR	Mujeres	Total
Desv. est. máxima	24.51%	15.14%	8.41%	12.83%	22.25%	24.51%	
Desv. est. mínima	0.25%	0.20%	0.15%	0.00%	0.05%	0.00%	
Desv. est. promedio	6.25%	5.01%	4.33%	4.79%	6.63%	5.74%	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres
Desv. est. máxima	11.87%	33.36%	29.64%	16.86%	31.42%	35.20%	35.20%
Desv. est. mínima	0.50%	0.10%	0.30%	0.55%	0.10%	0.45%	0.10%
Desv. est. promedio	5.13%	6.32%	6.51%	6.26%	5.85%	6.88%	6.10%

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres, las sílabas de ROD son las que más se desvían con respecto al promedio por enunciado, con 35.20%, mientras que la desviación mínima se la disputan entre dos, FRA y MOI, con 0.10%. Con respecto al promedio de 6.10%, ROD es quien más se desvía, 6.88%, mientras que ALB sólo lo hace un 5.13%, es decir, sus sílabas son las que tienen la duración más regular entre los hombres.

Por otra parte, si se compara los valores de la Tabla 3-85 con los de las Tablas 3-79 y 3-83, se podrá observar que los enunciados suspendidos son los que muestran una mayor irregularidad en cuanto a la duración de las sílabas que los conforman.

Tabla 3-85: Desv. est. máxima, mínima y promedio de la duración silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	15.96%	7.30%	35.20%	15.27%
Desv. est. mínima	0.25%	1.10%	0.00%	0.10%
Desv. est. promedio	5.03%	4.90%	6.65%	5.13%

Fuente: Prosodia.xls (Base de datos propia)

Con respecto a la desviación estándar, son los enunciados suspendidos los que alcanzan la desviación máxima más alta, pero también la desviación mínima más baja en comparación con los otros tipos de enunciados. Sin embargo, en promedio, son, de nuevo, los enunciados exclamativos los que muestran la menor variabilidad, con 4.90%, y los suspendidos tienen la mayor desviación, 6.73%.

3.3.3 Intensidad

En la primera parte de este apartado se hará un análisis de la intensidad silábica absoluta y más adelante se presentará los resultados de la estandarización de los valores. Así, en cuanto a la intensidad máxima promedio (véase Tabla 3-86), entre las mujeres, MAR es la hablante que alcanza el valor más alto, 81.30 dB, mientras que CEC es la persona que alcanzó el nivel más bajo, 63.41 dB. También en promedio, MAR y CEC representan los extremos, 76.15 dB y 65.94 dB respectivamente.

Tabla 3-86: Intensidad silábica máxima, mínima y promedio (decibeles)

	ADR	CEC	FAB	MAR	MIR	Mujeres		Total
Int. máxima prom.	73.60	68.59	79.14	81.30	72.67	75.23		
Int. mínima prom.	65.56	63.41	72.00	69.56	65.52	66.81		
Int. promedio	69.53	65.94	75.37	76.15	69.04	71.18		
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	
Int. máxima prom.	71.91	73.58	85.30	78.79	75.86	77.50	78.01	77.13
Int. mínima prom.	63.22	65.64	78.08	69.70	68.23	72.42	70.18	69.10
Int. promedio	67.71	69.62	81.88	74.15	71.92	74.95	74.10	73.17

Fuente: Prosodia.xls (Base de datos propia)

Al igual que con la mujeres, en el caso de los hombres hay dos claros extremos. Uno de ellos lo representa HUI, que alcanza una intensidad máxima promedio de 85.30 dB y el promedio más alto, 81.88 dB, mientras que ALB alcanza la intensidad mínima promedio de 63.22 y el promedio más bajo, 67.71 dB. Estas diferencias tan marcadas pueden deberse únicamente a las distancias relativas a las que se encontraban del micrófono al momento de la grabación.

En la Tabla 3-87 se puede observar que los enunciados suspendidos, con una intensidad máxima promedio de 77.41 dB, seguidos de cerca por los enunciados interrogativos, con 77.16 dB, son los que alcanzan los valores más altos, mientras que los enunciados neutros son los que alcanzan la intensidad más baja, 68.71 dB. En promedio, no hay grandes diferencias entre los distintos tipos de enunciados, pero los exclamativos son los que se producen con la mayor intensidad y los neutros con la menor intensidad.

Tabla 3-87: Intensidad silábica máxima, mínima y promedio por tipo de enunciado (dB)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Int. máxima prom.	77.16	76.81	77.41	76.57
Int. mínima prom.	69.21	71.25	69.21	68.71
Int. promedio	73.15	73.93	73.41	72.66

Fuente: Prosodia.xls (Base de datos propia)

En cuanto al rango de variación promedio de la intensidad utilizado por cada hablante, en la Tabla 3-88 podemos ver que, en el caso de las mujeres, MAR es quien utiliza el espectro más amplio, 19.24%, mientras que CEC utiliza el más estrecho, 8.28%. Además, de entre los hombres, IVA es quien hace un uso más amplio del rango de variación promedio, 5.63%, mientras ROD es quien usa el rango más estrecho, 7.33%, no sólo entre los hombres, sino en general, dentro del *corpus* de la Ciudad de México, pues el promedio general es de 12.89%.

Tabla 3-88: Variación máxima y mínima de la intensidad silábica

	ADR	CEC	FAB	MAR	MIR	Mujeres		Total
Var. máxima promedio	6.84%	3.74%	1.50%	9.23%	4.94%	6.24%		
Var. mínima promedio	-6.76%	-3.30%	-6.78%	-10.01%	-6.58%	-7.27%		
Rango de var. prom.	13.60%	7.04%	8.28%	19.24%	11.52%	13.51%		
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	
Var. máxima promedio	5.11%	6.69%	3.92%	7.66%	5.82%	3.78%	5.62%	
Var. mínima promedio	-7.91%	-8.01%	-6.21%	-7.97%	-6.62%	-3.55%	-6.98%	
Rango de var. prom.	13.02%	14.70%	10.13%	15.63%	12.44%	7.33%	12.60%	
								12.89%

Fuente: Prosodia.xls (Base de datos propia)

En cuanto a los tipos de enunciados (véase Tabla 3-89), el rango de variación promedio más amplio se da en los enunciados suspendidos, 13.49%, mientras que en los enunciados exclamativos, con un rango de 11.26%, varía la intensidad silábica muy poco.

Tabla 3-89: Variación máxima y mínima de la intensidad silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Var. máxima promedio	5.78%	6.64%	6.26%	5.37%
Var. mínima promedio	-7.17%	-4.62%	-7.23%	-6.87%
Rango de var. prom.	12.95%	11.26%	13.49%	12.24%

Fuente: Prosodia.xls (Base de datos propia)

Para continuar con el tema de la variación de la intensidad entre sílabas continuas, en la Tabla 3-90 se presentan los resultados de la desviación estándar máxima, mínima y promedio. De entre las mujeres, MAR es, también según este análisis, quien más se desvía del promedio,

con 694% de desviación estándar máxima y 373% de desviación estándar promedio. ADR, CEC, FAB y MIR producen al menos dos sílabas con justo la misma intensidad, 0%, aunque es CEC quien, en promedio, produce las sílabas con intensidad más regular, con 249%.

En el caso de los hombres, HUI es quien alcanza la desviación estándar más alta, 584%, aunque es IVA quien, en promedio hace un uso más amplio de la variación de la intensidad, con una desviación de 303%. La desviación mínima la alcanzaron todos los hablantes masculinos, pero es ROD quien alcanza la desviación más baja, 177%.

Tabla 3-90: Desviación estándar máxima, mínima y promedio de la intensidad silábica

	ADR	CEC	FAB	MAR	MIR	Mujeres	Total
Desv. est. máxima	618%	450%	550%	694%	457%	694%	
Desv. est. mínima	0%	0%	0%	50%	0%	0%	
Desv. est. promedio	261%	196%	258%	373%	249%	282%	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres
Desv. est. máxima	553%	465%	584%	529%	539%	469%	584%
Desv. est. mínima	0%	0%	0%	0%	0%	0%	0%
Desv. est. promedio	296%	260%	247%	303%	267%	177%	272%

Fuente: Prosodia.xls (Base de datos propia)

En los enunciados suspendidos es en los que hay una mayor desviación estándar absoluta y en promedio, 694% y 274%, mientras que la desviación estándar promedio más baja es la que se registra en los enunciados exclamativos, 239%.

Tabla 3-91: Desviación estándar máxima, mínima y promedio de la intensidad silábica por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	542%	600%	694%	638%
Desv. est. mínima	0%	50%	0%	0%
Desv. est. promedio	273%	239%	274%	268%

Fuente: Prosodia.xls (Base de datos propia)

Para facilitar los análisis comparativos entre los distintos hablantes, tomando en cuenta que el simple hecho de que la distancia con respecto al micrófono o que se haya utilizado micrófonos con distintos niveles de sensibilidad es determinante para la altura absoluta en decibeles, en las Tablas 3-92 y 3-93 se muestran los resultados tras estandarizar la intensidad a una altura inicial de 60 dB, adaptar los valores subsecuentes de la variación porcentual a una escala sobre 60 y sumar estos valores positivos o negativos al valor de la sílaba previa.

Según este análisis, los hablantes de la Ciudad de México que alcanzan la intensidad estandarizada más alta son MIR, 63.07 dB, e IVA, 64.01 dB. Quienes alcanzan las intensidades más bajas son MAR y ALB, con 53.78 dB y 55.18 dB respectivamente. Sin embargo, en cuanto al promedio general de 59.59 dB, son CEC, con 60.42 dB, e IVA, con 60.20 dB quienes se encuentran en el extremos más alto, aunque las diferencias con respecto a

otros hablantes son mínimas, mientras que FAB, con 58.38 dB, y HUI, con 58.69 dB se encuentran en el extremo más bajo.

Tabla 3-92: Intensidad silábica estandarizada máxima, mínima y promedio (dB inicial = 60)

	ADR	CEC	FAB	MAR	MIR		Mujeres	Total
Int. máxima prom.	62.66	62.87	61.26	62.85	63.07		62.82	
Int. mínima prom.	55.84	58.02	55.79	53.78	56.84		55.84	
Int. promedio	59.22	60.42	58.38	58.87	59.90		59.45	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	
Int. máxima prom.	62.72	63.54	61.14	64.01	63.40	62.17	62.84	62.84
Int. mínima prom.	55.18	56.67	55.97	56.58	56.99	58.02	56.47	56.27
Int. promedio	59.07	60.11	58.69	60.20	60.10	60.09	59.66	59.59

Fuente: Prosodia.xls (Base de datos propia)

Aplicando estos valores al filtro según los tipos de enunciados, nos encontramos con que en los enunciados suspendidos es donde se alcanza la intensidad estandarizada promedio más alta, 62.96 dB, y en los interrogativos la más baja, 56.21 dB. En promedio, los enunciados suspendidos son los pronunciados con mayor intensidad, mientras que los exclamativos se pronuncian con la intensidad estandarizada más baja.

Tabla 3-93: Intensidad silábica estandarizada máxima, mínima y promedio por tipo de enunciado (dB inicial = 60)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Int. máxima prom.	62.72	61.20	62.96	62.74
Int. mínima prom.	56.21	56.86	56.25	56.28
Int. promedio	59.44	58.94	59.68	59.52

Fuente: Prosodia.xls (Base de datos propia)

Si, tomando en cuenta estos nuevos valores, recalculamos la desviación estándar máxima, mínima y promedio (véase Tablas 3-94 y 3-95), podremos observar que los resultados de las Tablas 3-90 y 3-91 se confirman parcialmente: MAR no alcanza el nivel máximo de desviación estándar, sino CEC, 574%, pero sí el promedio más alto, 288%, mientras que CEC tiene el promedio general más bajo, 186%.

Tabla 3-94: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada

	ADR	CEC	FAB	MAR	MIR		Mujeres	Total
Desv. est. máxima	495%	574%	402%	508%	460%		574%	
Desv. est. mínima	0%	0%	0%	39%	0%		0%	
Desv. est. promedio	221%	186%	198%	288%	217%		234%	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	
Desv. est. máxima	519%	389%	422%	474%	475%	391%	519%	574%
Desv. est. mínima	0%	0%	0%	0%	0%	0%	0%	0%
Desv. est. promedio	257%	225%	176%	248%	224%	144%	217%	222%

Fuente: Prosodia.xls (Base de datos propia)

Las mayores diferencias se encuentran en que ya no es HUI quien alcanza el nivel máximo de desviación entre los hombres, sino ALB, con 519%, y ALB también tiene el promedio más

alto, 257%. El promedio más bajo de desviación estándar de la intensidad entre los hombres se confirma, ROD con 144%.

En la Tabla 3-95 se confirman los resultados de la Tabla 3-91: en los enunciados suspendidos es donde se alcanzan el mayor nivel de desviación.

Tabla 3-95: Desviación estándar máxima, mínima y promedio de la intensidad silábica estandarizada por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	485%	434%	574%	519%
Desv. est. mínima	0%	0%	0%	0%
Desv. est. promedio	224%	186%	224%	221%

Fuente: Prosodia.xls (Base de datos propia)

En promedio, es en los enunciados suspendidos y en los interrogativos en los que hay una mayor desviación estándar de la intensidad silábica estandarizada, 224%, y en los exclamativos es donde hay la menor desviación, 186%, es decir, en estos enunciados es donde hay la oscilación más baja entre la intensidad de las sílabas que los componen.

3.3.4 Frecuencia

En el apartado 3.2.4 se advirtió sobre el riesgo de analizar conjuntamente los datos sobre la frecuencia fundamental de mujeres y hombres, debido a las grandes diferencias entre los valores de unos y otros, e incluso entre hablantes del mismo sexo. Es por esta razón que en las Tablas 3-96 y 3-102 se ha marcado con fondo gris los valores promedio, pues, aparte de estos valores, las Tablas siguen conteniendo información importante para la caracterización frecuencial de los hablantes de la Ciudad de México.

De este modo, entre las mujeres, CEC es quien alcanza la altura tonal promedio más alta, 312.79 Hz, y también es ella quien, en promedio, habla más agudo, con 272.07 Hz, mientras que ADR es quien alcanza las notas más graves, 126.46 Hz (más graves incluso que las alcanzadas por ALB y FRA, por ejemplo), y, en general es ella quien habla más grave, con 226.53 Hz. Curiosamente, también son ADR y CEC quienes hacen un mayor y menor uso del rango de frecuencia, 160.62 Hz y 89.00 Hz respectivamente.

En el caso de los hombres, ALB es quien alcanza las notas más agudas, 246.88 Hz, mientras que IVA es quien hace uso de la frecuencia más baja, 105.57 Hz. También es IVA quien, en promedio, habla más grave, con 148.02 Hz, mientras que FRA es quien habla más agudo, con 201.13 Hz. Por otra parte, las diferencias en el uso del rango de frecuencia promedio es aún más extremo que en el caso de las mujeres, pues mientras ALB utiliza un rango de 113.88 Hz, la voz de ROD fluctúa dentro de un rango de 55.50 Hz.

Tabla 3-96: Frecuencia fundamental máxima, mínima y promedio (hertz)

	ADR	CEC	FAB	MAR	MIR		Mujeres	Total
Frec. máx. promedio	287.08	312.79	302.71	297.29	310.07		301.97	
Frec. mín. promedio	126.46	223.79	210.14	174.21	193.91		179.51	
Frecuencia promedio	226.53	272.07	259.09	235.17	258.25		247.12	
Rango frec. promedio	160.62	89.00	92.57	123.08	116.16		122.46	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	
Frec. máx. promedio	246.88	241.38	181.03	182.05	192.08	179.92	199.97	232.45
Frec. mín. promedio	133.00	142.06	108.82	105.57	119.17	124.42	119.28	138.46
Frecuencia promedio	198.42	201.13	148.22	148.02	158.48	159.44	164.37	190.72
Rango frec. promedio	113.88	99.32	72.21	76.48	72.91	55.50	80.69	93.99

Fuente: Prosodia.xls (Base de datos propia)

Según los valores inexactos provenientes del promedio general, es en los enunciados interrogativos en los que se alcanza la frecuencia más alta, pero también la más baja (véase Tabla 3-97). Por lo tanto, se podría afirmar de momento que es en estos enunciados en los que se hace un uso más amplio del rango de voz, mientras que en los enunciados exclamativos se utiliza el rango más estrecho. En promedio, los enunciados exclamativos son los más agudos y los neutros son los producidos de manera más grave.

Tabla 3-97: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (Hz)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. máx. promedio	242.38	221.94	234.55	224.83
Frec. mín. promedio	127.34	169.56	137.75	142.48
Frecuencia promedio	191.41	197.26	190.95	189.54
Rango frec. promedio	115.04	52.38	96.80	82.35

Fuente: Prosodia.xls (Base de datos propia)

Para comprobar los datos anteriores, se presenta en las Tablas 3-98 y 3-99 los mismos datos pero filtrados según si fueron producidos por mujeres u hombres. Así, en el caso de ellas, se confirma que, con 316.63 Hz, es en los enunciados interrogativos en los que se alcanzan las notas más altas, y además, con 168.67 Hz, es también en estos enunciados que se alcanzan las frecuencias más bajas. También el rango de frecuencia promedio en los enunciados interrogativos es el más amplio, con 147.96 Hz, y en los exclamativos se hace el uso más estrecho del rango frecuencial, con sólo 31.00 Hz.

Tabla 3-98: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (mujeres) (Hz)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. máx. promedio	316.63	264.80	303.81	295.36
Frec. mín. promedio	168.67	233.80	171.58	196.76
Frecuencia promedio	246.21	249.87	245.39	250.96
Rango frec. promedio	147.96	31.00	132.23	98.60

Fuente: Prosodia.xls (Base de datos propia)

La única diferencia con respecto a la Tabla 3-97 es que, en promedio, los enunciados suspendidos son los más graves, con 245.38 Hz, y los enunciados neutros son los más agudos, con 250.96 Hz.

En el caso de los hombres, excepto en la frecuencia promedio, se repite en mismo patrón: en los enunciados interrogativos se alcanza las frecuencias más altas y más bajas, 216.35 Hz y 112.84 Hz, y por lo tanto, son éstos los que hacen un uso más amplio del rango frecuencial, con 103.51 Hz, mientras los enunciados exclamativos son los que utilizan el rango más estrecho, con 62.09 Hz. Sin embargo, según el promedio, los enunciados suspendidos son los más graves (ligeramente más graves que los suspendidos), y los interrogativos son los más agudos (aunque el promedio de los exclamativos es muy parecido).

Tabla 3-99: Frecuencia fundamental máxima, mínima y promedio por tipo de enunciado (hombres) (Hz)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. máx. promedio	216.35	202.45	199.71	193.03
Frec. mín. promedio	112.84	140.36	120.73	118.01
Frecuencia promedio	172.19	173.35	163.57	161.85
Rango frec. promedio	103.51	62.09	78.98	75.02

Fuente: Prosodia.xls (Base de datos propia)

En la Tabla 3-100 se puede apreciar la variación porcentual máxima y mínima promedio de la frecuencia fundamental, de los cuales se obtiene el rango de variación promedio. En el caso de las mujeres, MAR es quien hace uso de un rango más amplio, 50.40%, mientras CEC es quien usa el rango de variación más estrecho, 27.47%.

De entre los hombres, HUI y ROD son quienes más se alejan del promedio de 39.24%, con rangos de variación de 47.24% y 21.63% respectivamente. El promedio de variación porcentual para la Ciudad de México es de 39.21%.

Tabla 3-100: Variación máxima y mínima promedio de la frecuencia fundamental

	ADR	CEC	FAB	MAR	MIR	Mujeres	Total
Var. máxima promedio	23.33%	12.67%	19.09%	29.64%	19.42%	22.38%	
Var. mínima promedio	-12.67%	-11.46%	-15.22%	-18.45%	-12.67%	-14.34%	
Rango de var. prom.	39.28%	27.47%	36.69%	50.40%	33.86%	39.12%	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres
Var. máxima promedio	25.11%	17.18%	28.59%	27.92%	21.46%	10.76%	23.75%
Var. mínima promedio	-11.88%	-13.64%	-16.94%	-15.16%	-10.30%	-4.64%	-13.22%
Rango de var. prom.	44.99%	32.18%	47.24%	44.36%	33.38%	21.63%	39.24%

Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, en la Tabla 3-101 se confirma la información de las Tablas 3-97, 3-98 y 3-99, pues en los enunciados interrogativos es en los que se hace uso del rango más amplio de variación promedio de la frecuencia, 41.86%, mientras los enunciados exclamativos son los que utilizan el rango más estrecho, 13.59%.

Tabla 3-101: Variación máxima y mínima promedio de la frecuencia fundamental por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Var. máxima promedio	30.27%	15.96%	23.90%	19.73%
Var. mínima promedio	-11.59%	2.37%	-14.34%	-13.91%
Rango de var. prom.	41.86%	13.59%	38.24%	33.64%

Fuente: Prosodia.xls (Base de datos propia)

Sin tomar en cuenta los promedios generales, podemos observar en la Tabla 3-102 que, de entre las mujeres, MIR es quien alcanza la mayor desviación estándar, 14009%, es decir, que sus sílabas son las que más se alejan del promedio individual de frecuencia fundamental de cada enunciado. También es MIR quien obtiene el valor más bajo de desviación estándar mínima, 100%. En promedio, ADR y CEC son quienes representan los extremos de la desviación con respecto al promedio femenino de 3830%, con valores de 4596% y 3132% respectivamente.

En el caso de los hombres, ALB es quien alcanza el valor más alto de desviación estándar máxima con 11461%, pero también, junto con ROD, es quien produjo al menos una vez dos sílabas contiguas con exactamente la misma frecuencia fundamental. Del mismo modo, son ALB y ROD quienes más se alejan del promedio masculino de 2606% con valores de 3781% y 2028% respectivamente.

Tabla 3-102: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental

	ADR	CEC	FAB	MAR	MIR	Mujeres	Total
Desv. est. máxima	10721%	9405%	10305%	9464%	14009%	14009%	
Desv. est. mínima	300%	500%	700%	759%	100%	100%	
Desv. est. promedio	4596%	3132%	3260%	3739%	3787%	3830%	
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres
Desv. est. máxima	11461%	9142%	6462%	7716%	8701%	6582%	11461%
Desv. est. mínima	0%	300%	50%	464%	47%	0%	0%
Desv. est. promedio	3781%	3023%	2352%	2440%	2385%	2028%	2606%
							2995%

Fuente: Prosodia.xls (Base de datos propia)

En cuanto a los tipos de enunciados, es en los suspendidos en los que se obtiene el valor más alto de desviación estándar máxima, pero también el valor más bajo de la mínima, 14009% y 0% (véase Tabla 3-103). En promedio, los enunciados en los que la frecuencia de las sílabas contiguas más se desvían del promedio es en los interrogativos, con 3636%, y en los que menos se desvían es en los enunciados exclamativos, con 2277%.

Tabla 3-103: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	10611%	8450%	14009%	12499%
Desv. est. mínima	200%	400%	0%	50%
Desv. est. promedio	3626%	2277%	3062%	2646%

Fuente: Prosodia.xls (Base de datos propia)

Sin embargo, estos últimos resultados se basan en el promedio general, lo que en la mayoría de las ocasiones podría considerarse inadecuado, por lo que en las Tablas 3-104 y 3-105 se presentan los resultados por separado para las mujeres y los hombres para confirmarlos.

En el caso de las mujeres, la mayor desviación se encuentra en los enunciados suspendidos, mientras que la menor se produce en los enunciados neutros. En promedio, las sílabas de los enunciados interrogativos son las que más se desvían, mientras que las de los exclamativos son las que menos se desvían. En otras palabras, únicamente el valor más bajo de la desviación estándar mínima se sale del esquema anterior.

Tabla 3-104: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (mujeres)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	10611%	2255%	14009%	12499%
Desv. est. mínima	200%	800%	150%	100%
Desv. est. promedio	4589%	1415%	4040%	3272%

Fuente: Prosodia.xls (Base de datos propia)

Por otro lado, entre los hombres, los valores máximo y mínimo se dan en los enunciados suspendidos, y en promedio, las sílabas de los que más se desvían del promedio de la F_0 son los interrogativos, mientras que la menor variación se dan en los enunciados neutros, y esa es la única diferencia con respecto a los resultados de la Tabla 3-103.

Tabla 3-105: Desviación estándar máxima, mínima y promedio de la frecuencia fundamental por tipo de enunciado (hombres)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	8419%	8450%	11461%	8425%
Desv. est. mínima	400%	400%	0%	50%
Desv. est. promedio	3289%	2669%	2571%	2364%

Fuente: Prosodia.xls (Base de datos propia)

Además, para evitar hacer análisis por separado de las producciones lingüísticas de mujeres y hombres, se puede recurrir al método de la frecuencia fundamental estandarizada, iniciando en 100 Hz, como se demuestra en las Tablas 3-106 a 3-109.

Según estos datos, las personas que alcanzan las frecuencias más altas son MAR, con 135.78 Hz, y HUI, con 135.41 Hz, mientras que las personas que alcanzan las frecuencias más bajas son ADR, con 83.96 Hz, y ALB, con 84.38 Hz. Con respecto al promedio general de 106.02 Hz, las mujeres tienen una frecuencia estandarizada promedio más baja que los hombres, 104.33 Hz frente a 106.81 Hz, y los casos extremos son ADR, con un promedio de 99.53 Hz y MAR, con un promedio de 108.26 Hz. Por último, la persona que utiliza el rango de frecuencia promedio más amplio es también MAR, con 48.68 Hz, mientras que ROD es quien hace uso del rango frecuencial más estrecho, de tan sólo 20.92 Hz.

Tabla 3-106: Frecuencia estandarizada máxima, mínima y promedio (F_0 inicial = 100)

	ADR	CEC	FAB	MAR	MIR	Mujeres	Total	
Frec. est. máx. prom.	119.92	116.38	118.79	135.78	123.07	125.23		
Frec. est. mín. prom.	83.96	89.00	90.43	87.10	90.59	88.16		
Frec. estándar prom.	99.53	102.03	102.79	108.26	104.56	104.33		
Rango frec. prom.	35.96	27.38	28.36	48.68	32.48	37.07		
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	Total
Frec. est. máx. prom.	126.76	116.76	135.41	126.89	127.81	115.75	127.68	126.90
Frec. est. mín. prom.	84.38	86.89	90.62	87.77	92.24	94.83	89.69	89.20
Frec. estándar prom.	104.56	99.71	112.21	105.23	107.23	104.55	106.81	106.02
Rango frec. prom.	42.38	29.87	44.79	39.12	35.57	20.92	37.99	37.70

Fuente: Prosodia.xls (Base de datos propia)

En cuanto a los tipos de enunciados, nuevamente, algunos de los resultados en las Tablas anteriores se confirman, pues es en los enunciados interrogativos en los que se produce la frecuencia estándar más alta, 133.78 Hz, mientras que en los enunciados neutros se produce la frecuencia estándar más baja, 88.75 Hz. En promedio, los enunciados interrogativos son los que se producen con la frecuencia más alta, 108.40 Hz, mientras que los neutros son los que se producen con la frecuencia más baja, 104.36 Hz. Además, en los enunciados interrogativos se utiliza el rango frecuencial más amplio, 44.22 Hz, mientras que en los enunciados exclamativos se hace uso del rango más estrecho, 23.18 Hz.

Tabla 3-107: Frecuencia estandarizada máxima, mínima y promedio por tipo de enunciado (F_0 inicial = 100)

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Frec. est. máx. prom.	133.78	119.06	128.10	122.16
Frec. est. mín. prom.	89.56	95.88	89.14	88.75
Frec. estándar prom.	108.40	106.86	106.33	104.36
Rango frec. promedio	44.22	23.18	38.96	33.41

Fuente: Prosodia.xls (Base de datos propia)

A partir de estos datos se puede, entonces, calcular nuevamente la desviación estándar en la que los valores de los distintos hablantes tienen una base común (véase Tabla 3-108).

Tabla 3-108: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada

	ADR	CEC	FAB	MAR	MIR	Mujeres	Total	
Desv. est. máxima	3518%	2566%	2258%	3543%	4619%	4619%		
Desv. est. mínima	200%	200%	300%	386%	50%	50%		
Desv. est. promedio	1188%	997%	998%	1537%	1113%	1232%		
	ALB	FRA	HUI	IVA	MOI	ROD	Hombres	Total
Desv. est. máxima	4217%	2825%	3805%	3100%	4602%	2250%	4602%	4619%
Desv. est. mínima	0%	150%	50%	236%	47%	0%	0%	0%
Desv. est. promedio	1524%	977%	1514%	1294%	1230%	826%	1296%	1276%

Fuente: Prosodia.xls (Base de datos propia)

Por lo tanto, de entre todos los hablantes, la frecuencia estandarizada de las sílabas de MIR es la que más se desvía del promedio por enunciado, 4619%, mientras que la frecuencia

estandarizada de las sílabas de ALB y ROD alcanzan la desviación de 0%. En promedio, las producciones lingüísticas de MAR y ALB, por un lado, con 1537% y 1524%, y las de ROD, por otro lado, con 826%, constituyen los extremos con respecto al promedio general de desviación estándar de 1276%.

Finalmente, con respecto a los tipos de enunciados, según estos datos, es en los enunciados neutros en donde se muestra una mayor desviación estándar, 4619%, mientras que en los enunciados suspendidos se da al menos un caso de desviación estándar de 0%. Sin embargo, en promedio, es en los enunciados interrogativos en los que la frecuencia estandarizada de las sílabas contiguas más se desvía del promedio de cada enunciado, con 1532%, mientras que en los exclamativos es en los que hay menos variación, con 999%.

Tabla 3-109: Desviación estándar máxima, mínima y promedio de la frecuencia estandarizada por tipo de enunciado

	/+I -E -S/	/-I +E -S/	/-I -E +S/	/-I -E -S/
Desv. est. máxima	4217%	3750%	4602%	4619%
Desv. est. mínima	100%	250%	0%	50%
Desv. est. promedio	1532%	999%	1298%	1142%

Fuente: Prosodia.xls (Base de datos propia)

Con esta información termina el análisis de los parámetros acústicos concretos de las producciones lingüísticas de los hablantes de la Ciudad de México. En los siguientes apartados, se analizarán los parámetros abstractos de la entonación y el ritmo.

3.3.5 Entonación

En el presente apartado se hará un recuento de los distintos tonos con los que fueron producidos los enunciados del *corpus* de la Ciudad de México. Además, por medio de las Tablas se podrá observar sus múltiples interrelaciones.

Para comenzar, en la Tabla 3-110 se puede apreciar que en 88.05% de los 870 enunciados que componen el *corpus* el tono inicial fue anotado como H%.

Tabla 3-110: Número y porcentaje de cada tipo de tono inicial

	H%	L%	Total
Tonos iniciales	766	104	870
Porcentaje	88.05%	11.95%	100%

Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, los tonos clasificados como prenucleares más comunes son *, con 27.83%, L+H*, con 21.02%, ¡L*+H, con 14.01%, y L*+H, con 13.37% (véase Tabla 3-111). Según la clasificación utilizada en la sección anterior, de acuerdo a la alineación del pico frecuencial con el acento, los tonos más comunes son los \acute{S} , con 37.47%, seguidos de $\prime S$, con 28.02%, S, con 27.83%, y S' , que son el grupo menos común con sólo 6.68%.

Tabla 3-111: Número y porcentaje de cada tipo de tono prenuclear

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
Tonos prenucleares	327	31	135	90	208	10	218	104	433	1556
Porcentaje	21.02%	1.99%	8.68%	5.78%	13.37%	0.64%	14.01%	6.68%	27.83%	100%

Fuente: Prosodia.xls (Base de datos propia)

El tono nuclear más común en el *corpus* de la Ciudad de México es L+H*, con 23.56%, seguido muy de cerca por *, con 22.99% (véase Tabla 3-112). Les siguen en importancia el tono ¡L+H*, con 13.33%, ¡L*+H, con 12.30%, y L*+H, con 12.18%. Entonces, al igual que en el caso de los tonos prenucleares, el grupo de tonos \acute{S} es el más abundante, con 44.13%, seguido de \acute{S} , con 24.71%, S, con 22.99%, y S', con 8.16%.

Tabla 3-112: Número y porcentaje de cada tipo de tono nuclear

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
Tonos nucleares	205	17	116	46	106	2	107	71	200	870
Porcentaje	23.56%	1.95%	13.33%	5.29%	12.18%	0.23%	12.30%	8.16%	22.99%	100%

Fuente: Prosodia.xls (Base de datos propia)

Para terminar con el recuento de los tonos aislados, en la Tabla 3-113 se puede observar que el tono final más común es M%, con 85.40%, seguido muy de lejos por H%, con 7.70% y L%, con 6.90%. Esta distribución, como se comentaba en el apartado 3.2.5, puede deberse simplemente a la clasificación como M% de toda variación final con un porcentaje menor a $\pm 25\%$, o bien a las características contextuales, temáticas o afectivas.

Tabla 3-113: Número y porcentaje de cada tipo de tono final

	H%	M%	L%	Total
Tonos finales	67	743	60	870
Porcentaje	7.70%	85.40%	6.90%	100%

Fuente: Prosodia.xls (Base de datos propia)

Ahora bien, aplicando el filtro según los tipos de enunciados (véase Tablas 3-114 y 3-115), podemos observar que, en todos los casos, el tono inicial más común es H%:

Tabla 3-114: Número de tonos iniciales por tipo de enunciado

	H%	L%	Total
/+I -E -S/	95	9	104
/-I +E -S/	16	0	16
/-I -E +S/	441	58	499
/-I -E -S/	214	37	251
Total	766	104	870

Fuente: Prosodia.xls (Base de datos propia)

Los porcentajes de preferencia por el tono H% van desde 85.26% en el caso de los enunciados neutros y 100% en el caso de los exclamativos, mientras que el promedio es de 88.05%.

Tabla 3-115: Porcentaje de tonos iniciales por tipo de enunciado

	H%	L%	Total
/+I -E -S/	91.35%	8.65%	100%
/-I +E -S/	100%	0%	100%
/-I -E +S/	88.38%	11.62%	100%
/-I -E -S/	85.26%	14.74%	100%
Total	88.05%	11.95%	100%

Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, tras filtrar los valores de los tonos nucleares según el tipo de enunciado, se obtienen los resultados de las Tablas 3-116 y 3-117. Los tonos nucleares más comunes en enunciados interrogativos son L+H*, L*+H, ¡L*+H, * y ¡L+H*; en los enunciados exclamativos son L*+H, * y L+H*; en los suspendidos son *, L+H*, ¡L*+H, ¡L+H*, L*+H y H+L*, mientras que en los neutros son *, L+H*, ¡L+H* y L*+H.

Tabla 3-116: Número de tonos nucleares por tipo de enunciado

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
/+I -E -S/	26	3	12	3	22	0	19	4	15	104
/-I +E -S/	3	0	1	0	7	0	1	0	4	16
/-I -E +S/	108	9	60	28	53	1	79	52	109	499
/-I -E -S/	68	5	43	15	24	1	8	15	72	251
Total	205	17	116	46	106	2	107	71	200	870

Fuente: Prosodia.xls (Base de datos propia)

Agrupados por su alineación acentual, veremos que los tonos más comunes en los enunciados interrogativos son los del tipo \acute{S} , con 42.30%, después los del tipo $\prime S$, con 39.42%, seguidos de S, con 14.42%, y finalmente los tonos S' , con sólo 3.85%. En los enunciados exclamativos los principales tonos son del tipo $\prime S$, con 50%, seguidos de \acute{S} y S, ambos con 25%. Los enunciados suspendidos muestran el mismo orden de preferencia que los interrogativos: \acute{S} , con 41.07%, $\prime S$, con 26.65%, S, con 21.84%, y S' , con 10.42%. Finalmente, en los enunciados neutros se hace uso principalmente de tonos del tipo \acute{S} , con 52.19%, seguidos de los del tipo S, con 28.69%, del tipo $\prime S$, con 13.15%, y los del tipo S' , con sólo 5.98%.

Tabla 3-117: Porcentaje de tonos nucleares por tipo de enunciado

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
/+I -E -S/	25.00%	2.88%	11.54%	2.88%	21.15%	0%	18.27%	3.85%	14.42%	100%
/-I +E -S/	18.75%	0%	6.25%	0%	43.75%	0%	6.25%	0%	25.00%	100%
/-I -E +S/	21.64%	1.80%	12.02%	5.61%	10.62%	0.20%	15.83%	10.42%	21.84%	100%
/-I -E -S/	27.09%	1.99%	17.13%	5.98%	9.56%	0.40%	3.19%	5.98%	28.69%	100%
Total	23.56%	1.95%	13.33%	5.29%	12.18%	0.23%	12.30%	8.16%	22.99%	100%

Fuente: Prosodia.xls (Base de datos propia)

En cuanto a los tonos finales utilizados en los distintos tipos de enunciados (véase Tablas 3-118 y 3-119), debido a su gran presencia, con 85.40% del total de tonos finales, M% es el que se usa en todos los tipos de enunciados con la mayor frecuencia. Los porcentajes varían entre 75% en el caso de los enunciados interrogativos y 93.75% en el caso de los exclamativos.

Tabla 3-118: Número de tonos finales por tipo de enunciado

	H%	M%	L%	Total
/+I -E -S/	20	78	6	104
/-I +E -S/	0	15	1	16
/-I -E +S/	43	422	34	499
/-I -E -S/	4	228	19	251
Total	67	743	60	870

Fuente: Prosodia.xls (Base de datos propia)

Sin embargo, es interesante observar que los tonos H% son los segundos en importancia en los enunciados interrogativos y suspendidos, mientras que los tonos L% tienen el segundo rango en los enunciados exclamativos y neutros.

Tabla 3-119: Porcentaje de tonos finales por tipo de enunciado

	H%	M%	L%	Total
/+I -E -S/	19.23%	75.00%	5.77%	100%
/-I +E -S/	0%	93.75%	6.25%	100%
/-I -E +S/	8.62%	84.57%	6.81%	100%
/-I -E -S/	1.59%	90.84%	7.57%	100%
Total	7.70%	85.40%	6.90%	100%

Fuente: Prosodia.xls (Base de datos propia)

Si ahora aplicamos al número total de tonos nucleares el filtro de los distintos tipos de tono final, obtendremos los datos que se muestran en las Tablas 3-120 y 3-121:

Tabla 3-120: Número de tonos nucleares combinados con cada tipo de tono final

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	15	0	7	1	13	0	31	0	0	67
M%	174	16	103	39	87	1	71	61	191	743
L%	16	1	6	6	6	1	5	10	9	60
Total	205	17	116	46	106	2	107	71	200	870

Fuente: Prosodia.xls (Base de datos propia)

Así, podemos observar que los tonos más comunes con terminación H% son: ¡L*+H, L+H*, L*+H y ¡L+H*, que equivalen a 65.67% de tonos del tipo 'S y el restante 34.33% del tipo S. Los tonos nucleares combinados con el tono final M% más comunes son: *, L+H*, ¡L+H*, L*+H y ¡L*+H, que da un total por grupos de 44.68% de S, 25.71% de S, 21.40% de 'S y sólo 8.21% de S'. Además, en el caso de los tonos nucleares combinados con tonos finales L%, los más frecuentes son L+H*, H+L* y *, aunque el total por grupos arroja un 48.34% de tonos S, 20% de 'S, 16.67% de S' y 15% de S.

Tabla 3-121: Porcentaje de tonos nucleares combinados con cada tipo de tono final

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	22.39%	0%	10.45%	1.49%	19.40%	0%	46.27%	0%	0%	100%
M%	23.42%	2.15%	13.86%	5.25%	11.71%	0.13%	9.56%	8.21%	25.71%	100%
L%	26.67%	1.67%	10.00%	10.00%	10.00%	1.67%	8.33%	16.67%	15.00%	100%
Total	23.56%	1.95%	13.33%	5.29%	12.18%	0.23%	12.30%	8.16%	22.99%	100%

Fuente: Prosodia.xls (Base de datos propia)

A estos datos se les puede aplicar también el filtro para el tipo de enunciado. En las Tablas 3-122 y 3-123 se presenta los resultados correspondientes únicamente a los enunciados interrogativos, cuyos únicos tonos nucleares combinados con H% son L*+H, ¡L+H* y L+H* y ¡L+H*, lo que equivale a 70% de tonos del tipo 'S y 30% del tipo Ś.

Tabla 3-122: Número de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	4	0	2	0	7	0	7	0	0	20
M%	21	3	10	3	13	0	11	2	15	78
L%	1	0	0	0	2	0	1	2	0	6
Total	26	3	12	3	22	0	19	4	15	104

Fuente: Prosodia.xls (Base de datos propia)

Los tonos nucleares que con mayor frecuencia se combinan con M% son L+H*, *, L*+H, ¡L*+H y ¡L+H*, aunque por grupos, el tipo Ś tiene un 47.44% de presencia, el 'S un 30.77%, el S un 19.23% y el S' únicamente un 2.56%. Por otro lado, los pocos tonos nucleares combinados con L% son L*+H y H+L*, ambos con dos observaciones, así como L+H* y ¡L*+H, cada uno con sólo una observación, lo que equivale a 50% de tonos 'S, 33.33% de S', y 16.67% de Ś.

Tabla 3-123: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados interrogativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	20.00%	0%	10.00%	0%	35.00%	0%	35.00%	0%	0%	100%
M%	26.92%	3.85%	12.82%	3.85%	16.67%	0%	14.10%	2.56%	19.23%	100%
L%	16.67%	0%	0%	0%	33.33%	0%	16.67%	33.33%	0%	100%
Total	25.00%	2.88%	11.54%	2.88%	21.15%	0%	18.27%	3.85%	14.42%	100%

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los enunciados exclamativos (véase Tablas 3-124 y 3-125), el panorama, debido a la poca presencia de este tipo de enunciados, es relativamente sencillo:

Tabla 3-124: Número de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0	0	0	0	0	0	0	0	0	0
M%	3	0	1	0	7	0	0	0	4	15
L%	0	0	0	0	0	0	1	0	0	1
Total	3	0	1	0	7	0	1	0	4	16

Fuente: Prosodia.xls (Base de datos propia)

Enunciados exclamativos con entonación final ascendente no hubo ninguno, mientras que con entonación descendente sólo hubo un caso, ¡L*+H (tipo 'S). Por otra parte, los únicos tonos nucleares en los enunciados terminados con tono final M% son L*+H, *, L+H* y ¡L+H*, lo que equivale a 46.67% de tonos del tipo 'S y 26.67% tanto del tipo Ś como del S.

Tabla 3-125: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados exclamativos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	0%	0%	0%	0%	0%	0%	0%	0%	0%	*100%
M%	20.00%	0%	6.67%	0%	46.67%	0%	0%	0%	26.67%	100%
L%	0%	0%	0%	0%	0%	0%	100%	0%	0%	100%
Total	18.75%	0%	6.25%	0%	43.75%	0%	6.25%	0%	25.00%	100%

Fuente: Prosodia.xls (Base de datos propia)

En el grupo más numeroso, el de los enunciados suspendidos, los principales tonos nucleares combinados con H% son ¡L*+H, L+H* y L*+H, es decir, 67.44% de 'S y 32.57% de Ś, mientras que los más combinados con L% son L+H*, *, H* y ¡L+H*, lo que equivale a 55.88% de Ś, 17.65% tanto de 'S como de S, y 8.82% de S' (véase Tablas 3-126 y 127).

Tabla 3-126: Número de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	10	0	3	1	6	0	23	0	0	43
M%	89	8	53	22	44	1	53	49	103	422
L%	9	1	4	5	3	0	3	3	6	34
Total	108	9	60	28	53	1	79	52	109	499

Fuente: Prosodia.xls (Base de datos propia)

En los enunciados suspendidos terminados en M%, los tonos nucleares más comunes son *, L+H*, ¡L+H* y ¡L*+H, H+L* y L*+H, lo que por grupos representa un 40.76% de Ś, 24.41% de S, 23.23% de 'S y 11.61% de S'.

Tabla 3-127: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados suspendidos

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	23.26%	0%	6.98%	2.33%	13.95%	0%	53.49%	0%	0%	100%
M%	21.09%	1.90%	12.56%	5.21%	10.43%	0.24%	12.56%	11.61%	24.41%	100%
L%	26.47%	2.94%	11.76%	14.71%	8.82%	0%	8.82%	8.82%	17.65%	100%
Total	21.64%	1.80%	12.02%	5.61%	10.62%	0.20%	15.83%	10.42%	21.84%	100%

Fuente: Prosodia.xls (Base de datos propia)

Finalmente, si se aplica el filtro para que aparezcan únicamente los datos relativos a los enunciados neutros, podremos observar en las Tablas 3-128 y 3-129 que en la combinación más escasa, H%, únicamente se utiliza los tonos ¡L+H*, L+H* y ¡L*+H, lo que equivale a un 75% de Ś por 25% de 'S.

Tabla 3-128: Número de tonos nucleares combinados con cada tipo de tono final en enunciados neutros

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	1	0	2	0	0	0	1	0	0	4
M%	61	5	39	14	23	0	7	10	69	228
L%	6	0	2	1	1	1	0	5	3	19
Total	68	5	43	15	24	1	8	15	72	251

Fuente: Prosodia.xls (Base de datos propia)

Los tonos nucleares que más se asocian con L% son L+H*, H+L*, * y ¡L+H*, es decir, 47.37% de \acute{S} , 26.35% de S' , 15.79% de S y 10.52% de \grave{S} . Por último, dentro del grupo más numeroso de enunciados neutros, los combinados con M%, los tonos nucleares más comunes son *, L+H*, ¡L+H* y L*+H, aunque la asociación por grupos da como resultado un 52.19% de \acute{S} , 30.26% de S, 13.16% de \grave{S} y sólo 4.39% de S' .

Tabla 3-129: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados neutros

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	25.00%	0%	50.00%	0%	0%	0%	25.00%	0%	0%	100%
M%	26.75%	2.19%	17.11%	6.14%	10.09%	0%	3.07%	4.39%	30.26%	100%
L%	31.58%	0%	10.53%	5.26%	5.26%	5.26%	0%	26.32%	15.79%	100%
Total	27.09%	1.99%	17.13%	5.98%	9.56%	0.40%	3.19%	5.98%	28.69%	100%

Fuente: Prosodia.xls (Base de datos propia)

Aplicando otro tipo de filtros al total general de tonos nucleares en combinación con los tres tonos finales se obtiene, por ejemplo, los valores respectivos a las producciones de mujeres (véase Tablas 3-130 y 3-131) y hombres (véase Tablas 3-132 y 3-133), lo que nos permitiría hacer comparaciones por sexos.

Tabla 3-130: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	3	0	2	0	8	0	14	0	0	27
M%	51	5	29	11	29	1	19	27	59	231
L%	4	0	3	2	1	0	1	4	4	19
Total	58	5	34	13	38	1	34	31	63	277

Fuente: Prosodia.xls (Base de datos propia)

Así, vemos que, en el caso de las mujeres, el tono final más común es M%, seguido de lejos por H%, y el menos común es L%, mientras que en el caso de los hombres, el tono final L% es ligeramente más común que H%.

Tabla 3-131: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por mujeres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	11.11%	0%	7.41%	0%	29.63%	0%	51.85%	0%	0%	100%
M%	22.08%	2.16%	12.55%	4.76%	12.55%	0.43%	8.23%	11.69%	25.54%	100%
L%	21.05%	0%	15.79%	10.53%	5.26%	0%	5.26%	21.05%	21.05%	100%
Total	20.94%	1.81%	12.27%	4.69%	13.72%	0.36%	12.27%	11.19%	22.74%	100%

Fuente: Prosodia.xls (Base de datos propia)

Entre las mujeres, los únicos cuatro tonos nucleares compatibles con H% son ¡L*+H, L*+H, L+H* y ¡L+H*, lo que equivale a 81.48% de \grave{S} y 18.52% de \acute{S} . Los principales tonos nucleares combinados con M% son *, L+H*, ¡L+H* y L*+H, así como H+L*, que por grupos arroja un resultado de 41.55% de \acute{S} , 25.54% de S, 21.21% de \grave{S} y 11.69% de S' . Por último, los tonos nucleares con terminación descendente más comunes son L+H*, H+L* y *, estos

tres con el mismo número de casos, así como ¡L+H* y H*, lo que equivale a 47.37% de \acute{S} , 21.05% tanto de S' como de S, y 10.52% de \grave{S} .

Tabla 3-132: Número de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	12	0	5	1	5	0	17	0	0	40
M%	123	11	74	28	58	0	52	34	132	512
L%	12	1	3	4	5	1	4	6	5	41
Total	147	12	82	33	68	1	73	40	137	593

Fuente: Prosodia.xls (Base de datos propia)

En el caso de los hombres, los tonos nucleares con terminación ascendente más comunes son ¡L*+H, L+H*, ¡L+H* y L*+H, o bien, por grupos, 55% de \acute{S} y 45% de \acute{S} . Los tonos nucleares compatibles con tono final M% más frecuentes son *, L+H*, ¡L+H*, L*+H (hasta ahí en el mismo orden que en el caso de las mujeres) y ¡L*+H, lo que equivale a 46.09% de \acute{S} , 25.78% de S, 21.49% de \acute{S} y 6.64% de S' (a pesar de las ligeras diferencias entre \acute{S} y S', muy similar a los resultados de las mujeres).

Tabla 3-133: Porcentaje de tonos nucleares combinados con cada tipo de tono final en enunciados producidos por hombres

	L+H*	L+!H*	¡L+H*	H*	L*+H	L*+!H	¡L*+H	H+L*	*	Total
H%	30.00%	0%	12.50%	2.50%	12.50%	0%	42.50%	0%	0%	100%
M%	24.02%	2.15%	14.45%	5.47%	11.33%	0%	10.16%	6.64%	25.78%	100%
L%	29.27%	2.44%	7.32%	9.76%	12.20%	2.44%	9.76%	14.63%	12.20%	100%
Total	24.79%	2.02%	13.83%	5.56%	11.47%	0.17%	12.31%	6.75%	23.10%	100%

Fuente: Prosodia.xls (Base de datos propia)

Por último, los tonos nucleares más comunes con tono final L% son L+H*, H+L*, * y L*+H, que por grupos representan un 48.79% de \acute{S} , 24.40% de \acute{S} , 14.63% de S' y 12.20% de S.

Este tipo de comparaciones intragrupalas son sólo una muestra de las posibilidades de combinación y discriminación de datos entre los distintos subgrupos de hablantes y que más adelante se aplicarán a comparaciones entre los hablantes de las variedades peninsular y mexicana del español.

3.3.6 Ritmo

La información sobre el peso silábico asignado a cada sílaba durante la transcripción detallada de los fragmentos seleccionados para cada persona, según se describió en el apartado 2.3.3.4, es la base del análisis de los grupos rítmicos que conforman las sílabas de cada enunciado.

En la presente sección se hará una presentación de los resultados del conteo de los grupos rítmicos respetando las convenciones descritas en 3.2.6, aplicando los filtros de género, tipo de enunciado y número de repeticiones de grupos rítmicos idénticos en posición contigua.

En la Tabla 3-134 se puede observar el número de cada uno de los grupos rítmicos que las mujeres produjeron en sus enunciados y en la Tabla 3-135 se presenta su equivalencia porcentual. De los valores en las Tablas se puede deducir el inventario de grupos rítmicos utilizados por cada hablante.

Tabla 3-134: Número de grupos rítmicos en enunciados producidos por mujeres

Grupo rítmico		ADR	CEC	FAB	MAR	MIR
(0)	0	18	20	16	21	37
(s w)	1	35	22	10	57	70
(w s)	2	14	8	1	39	26
(s w w)	3	0	5	1	6	8
(w s w)	4	24	8	6	56	37
(w w s)	5	15	1	0	24	13
(s w w w)	6	0	0	0	0	0
(w s w w)	7	1	0	2	2	6
(w w s w)	8	30	4	2	29	27
(w w w s)	9	7	0	0	7	6
(s w w w w)	10	0	0	0	0	0
(w s w w w)	11	1	0	0	0	0
(w w s w w)	12	1	0	0	0	1
(w w w s w)	13	12	1	1	17	9
(w w w w s)	14	1	0	0	0	1
(s w w w w w)	15	0	0	0	0	0
(w s w w w w)	16	0	0	0	0	0
(w w s w w w)	17	0	0	0	0	0
(w w w s w w)	18	2	0	0	0	0
(w w w w s w)	19	4	0	0	2	4
(w w w w w s)	20	0	0	0	0	0
Total		165	69	39	260	245

Fuente: Prosodia.xls (Base de datos propia)

El grupo rítmico más común entre las hablantes es el 1, excepto en el caso de FAB, que presenta más observaciones de 0 que de 1, aunque esto se debe seguramente a la pequeña cantidad de enunciados que aportó al *corpus*. Entre los grupos rítmicos que no están representados en el habla de la mujeres son 6, 10, 15, 16, 17 y 20. La característica común de casi todos estos grupos es la presencia de al menos tres sílabas átonas después de la tónica.

Atendiendo a los valores porcentuales, los grupos rítmicos más utilizados por ADR, en orden descendente son 1, 8, 4, 0, 5; CEC, por su parte, empleó con mayor frecuencia los grupos 1, 0, 2 y 4, 3, 8; FAB hizo mayoritariamente uso de los grupos 0, 1, 4, 7 y 8; MAR produjo principalmente los grupos 1, 4, 2, 8, 5, mientras que MIR optó por utilizar en especial los grupos 1, 0, 8, 2 y 5. Aunque ninguna de las combinaciones de grupos rítmicos es idéntico entre las hablantes, se puede observar una clara preferencia por los grupos 1 (21.21% - 31.88%), 4 (11.59% - 21.54%), 0 (8.08% - 41.03%), 8 (5.13% - 18.18%), 2 (2.56% - 15%), y, de éstos, 1, 4 y 8 tienen acentuación grave. Únicamente 2 tiene acentuación aguda, mientras que 0 es lo que se denomina 'pie libre'.

Tabla 3-135: Porcentaje de grupos rítmicos en enunciados producidos por mujeres

Grupo rítmico		ADR	CEC	FAB	MAR	MIR
(0)	0	10.91%	28.99%	41.03%	8.08%	15.10%
(s w)	1	21.21%	31.88%	25.64%	21.92%	28.57%
(w s)	2	8.48%	11.59%	2.56%	15.00%	10.61%
(s w w)	3	0%	7.25%	2.56%	2.31%	3.27%
(w s w)	4	14.55%	11.59%	15.38%	21.54%	15.10%
(w w s)	5	9.09%	1.45%	0%	9.23%	5.31%
(s w w w)	6	0%	0%	0%	0%	0%
(w s w w)	7	0.61%	0%	5.13%	0.77%	2.45%
(w w s w)	8	18.18%	5.80%	5.13%	11.15%	11.02%
(w w w s)	9	4.24%	0%	0%	2.69%	2.45%
(s w w w w)	10	0%	0%	0%	0%	0%
(w s w w w)	11	0.61%	0%	0%	0%	0%
(w w s w w)	12	0.61%	0%	0%	0%	0.41%
(w w w s w)	13	7.27%	1.45%	2.56%	6.54%	3.67%
(w w w w s)	14	0.61%	0%	0%	0%	0.41%
(s w w w w w)	15	0%	0%	0%	0%	0%
(w s w w w w)	16	0%	0%	0%	0%	0%
(w w s w w w)	17	0%	0%	0%	0%	0%
(w w w s w w)	18	1.21%	0%	0%	0%	0%
(w w w w s w)	19	2.42%	0%	0%	0.77%	1.63%
(w w w w w s)	20	0%	0%	0%	0%	0%

Fuente: Prosodia.xls (Base de datos propia)

De manera análoga al análisis de los grupos rítmicos en el habla de las mujeres, en la Tabla 3-136 se presenta el número de observaciones de cada grupo en el habla de los hombres, mientras que en la Tabla 3-137 se incluirá su equivalencia en porcentaje.

A diferencia de la mujeres, aquí sí hay una preferencia unánime por el grupo rítmico 1, y los grupos 10, 15, 16, 17 y 20 no están representados. El grupo 6 que no aparece en ninguno de los inventarios de las mujeres, aparece entre los hombres sólo 1 vez (IVA).

Según el número de observaciones y su porcentaje en orden descendente, ALB, hizo uso en especial de los grupos 1, 0, 4 y 8, 2, 3; FRA e IVA emplearon mayoritariamente los grupos 1, 0, 4, 8, 2; HUI y MOI utilizaron principalmente los grupos rítmicos 1, 0, 4, 2, 8, y ROD produjo especialmente los grupos 1, 4, 0, 8, 2 y 5.

Comparadas con las secuencias de las mujeres, no hay ningún caso idéntico. Sin embargo, entre los hombres sí hay coincidencias, y precisamente la secuencia utilizada por HUI y MOI corresponde a la del total general de los cinco grupos rítmicos más empleados por los hombres (véase Tabla 3-138): 1 (25.78% - 44%), 0 (16% - 23.65%), 4 (10.97% - 20%), 2 (4% - 11.78%) y 8 (7.01% - 12%).

En otras palabras, los grupos rítmicos más comunes en las producciones lingüísticas de hombres y mujeres son los mismos, únicamente se altera su orden: 1, 4, 0, 8, 2 (mujeres) vs. 1, 0, 4, 2, 8 (hombres). Como se mencionó anteriormente, los grupos 1, 4 y 8 tienen en común una acentuación grave, mientras que 2 tiene una acentuación aguda y 0 carece de ella.

Tabla 3-136: Número de grupos rítmicos en enunciados producidos por hombres

Grupo rítmico		ALB	FRA	HUI	IVA	MOI	ROD
(0)	0	28	61	91	60	118	4
(sw)	1	54	97	116	85	130	11
(ws)	2	16	22	53	18	53	1
(sww)	3	7	5	11	7	21	0
(wsw)	4	17	41	67	41	78	5
(wws)	5	6	10	22	14	14	1
(swww)	6	0	0	0	1	0	0
(wsww)	7	1	2	11	7	10	0
(wwsw)	8	17	29	44	28	35	3
(wwws)	9	2	8	3	2	9	0
(swwww)	10	0	0	0	0	0	0
(wswww)	11	0	0	0	0	1	0
(wwsww)	12	1	3	5	1	9	0
(wwwsww)	13	3	8	17	10	12	0
(wwwwws)	14	2	0	3	0	1	0
(swwwwww)	15	0	0	0	0	0	0
(wswwwww)	16	0	0	0	0	0	0
(wwswww)	17	0	0	0	0	0	0
(wwwsww)	18	0	0	2	0	4	0
(wwwwwsww)	19	1	7	5	0	4	0
(wwwwwws)	20	0	0	0	0	0	0
Total		155	293	450	274	499	25

Fuente: Prosodia.xls (Base de datos propia)

Tabla 3-137: Porcentaje de grupos rítmicos en enunciados producidos por hombres

Grupo rítmico		ALB	FRA	HUI	IVA	MOI	ROD
(0)	0	18.06%	20.82%	20.22%	21.90%	23.65%	16.00%
(sw)	1	34.84%	33.11%	25.78%	31.02%	26.05%	44.00%
(ws)	2	10.32%	7.51%	11.78%	6.57%	10.62%	4.00%
(sww)	3	4.52%	1.71%	2.44%	2.55%	4.21%	0%
(wsw)	4	10.97%	13.99%	14.89%	14.96%	15.63%	20.00%
(wws)	5	3.87%	3.41%	4.89%	5.11%	2.81%	4.00%
(swww)	6	0%	0%	0%	0.36%	0%	0%
(wsww)	7	0.65%	0.68%	2.44%	2.55%	2.00%	0%
(wwsw)	8	10.97%	9.90%	9.78%	10.22%	7.01%	12.00%
(wwws)	9	1.29%	2.73%	0.67%	0.73%	1.80%	0%
(swwww)	10	0%	0%	0%	0%	0%	0%
(wswww)	11	0%	0%	0%	0%	0.20%	0%
(wwsww)	12	0.65%	1.02%	1.11%	0.36%	1.80%	0%
(wwwsww)	13	1.94%	2.73%	3.78%	3.65%	2.40%	0%
(wwwwws)	14	1.29%	0%	0.67%	0%	0.20%	0%
(swwwwww)	15	0%	0%	0%	0%	0%	0%
(wswwwww)	16	0%	0%	0%	0%	0%	0%
(wwswww)	17	0%	0%	0%	0%	0%	0%
(wwwsww)	18	0%	0%	0.44%	0%	0.80%	0%
(wwwwwsww)	19	0.65%	2.39%	1.11%	0%	0.80%	0%
(wwwwwws)	20	0%	0%	0%	0%	0%	0%

Fuente: Prosodia.xls (Base de datos propia)

A manera de resumen, en la Tabla 3-138 se presenta el total de observaciones y su equivalencia en porcentaje de la distribución de grupos rítmicos en el habla de mujeres y

hombres, así como el total general. En la Tabla se aprecia que las diferencias entre las preferencias de mujeres y hombres radica únicamente en el orden de los grupos 0 y 4, por un lado, y el 2 y el 8, por el otro lado. Además, se nota que los grupos 10, 15, 16, 17 y 20 no existen en el *corpus* de la Ciudad de México, y el grupo 6, con una sola aparición, debe ser muy poco común.

Tabla 3-138: Número total y porcentaje de grupos rítmicos

Grupo rítmico		Mujeres		Hombres		Total	
(0)	0	112	14.40%	362	21.34%	474	19.16%
(s w)	1	194	24.94%	493	29.07%	687	27.77%
(w s)	2	88	11.31%	163	9.61%	251	10.15%
(s w w)	3	20	2.57%	51	3.01%	71	2.87%
(w s w)	4	131	16.84%	249	14.68%	380	15.36%
(w w s)	5	53	6.81%	67	3.95%	120	4.85%
(s w w w)	6	0	0%	1	0.06%	1	0.04%
(w s w w)	7	11	1.41%	31	1.83%	42	1.70%
(w w s w)	8	92	11.83%	156	9.20%	248	10.02%
(w w w s)	9	20	2.57%	24	1.42%	44	1.78%
(s w w w w)	10	0	0%	0	0%	0	0%
(w s w w w)	11	1	0.13%	1	0.06%	2	0.08%
(w w s w w)	12	2	0.26%	19	1.12%	21	0.85%
(w w w s w)	13	40	5.14%	50	2.95%	90	3.64%
(w w w w s)	14	2	0.26%	6	0.35%	8	0.32%
(s w w w w w)	15	0	0%	0	0%	0	0%
(w s w w w w)	16	0	0%	0	0%	0	0%
(w w s w w w)	17	0	0%	0	0%	0	0%
(w w w s w w)	18	2	0.26%	6	0.35%	8	0.32%
(w w w w s w)	19	10	1.29%	17	1.00%	27	1.09%
(w w w w w s)	20	0	0%	0	0%	0	0%
Total		778	100%	1696	100%	2474	100%

Fuente: Prosodia.xls (Base de datos propia)

En general, los grupos rítmicos que utilizan los hablantes de la Ciudad de México, según el orden de importancia, serían los siguientes: **1**, **0**, **4**, **2**, **8**, **5**, **13**, 3, 9, 7, **19**, 12, 18 y *14*, 11, 6. Por medio de estilos tipográficos se señala el tipo de acentuación bajo el cual se rige cada uno de los grupos: la letra regular señala la no acentuación, la cursiva significa acentuación aguda, la letra negrita indica la acentuación grave, el subrayado sencillo representa la acentuación esdrújula, y el subrayado doble indica acentuación sobreesdrújula.

Aunque la posición en la secuencia ya dice mucho sobre el peso porcentual de cada grupo, también la concentración de grupos con el mismo estilo tipográfico deja en claro la supremacía de la acentuación grave (en total, 57.88%), seguida de lejos por la acentuación nula (19.16%) y la acentuación aguda (17.10%), mientras que la acentuación esdrújula y la sobreesdrújula no representan más que el 5.86% del total, es decir, los grupos graves, nulos y agudos representan en conjunto el 94.14%.

A continuación, al total general se le aplicó el filtro para extraer por separado el número de observaciones de los distintos grupos rítmicos por tipo de enunciado y se obtuvieron los datos de la Tabla 3-139, que además incluye la equivalencia porcentual. En el caso de los enunciados interrogativos los grupos rítmicos más comunes son 0, 1, 4, 2, 8; en los exclamativos se encuentra más frecuentemente los grupos 1, 0 y 4, 2 y 3, 8; los enunciados suspendidos repiten la secuencia del total femenino: 1, 0, 4, 8, 2, mientras que en los enunciados neutros los grupos que más abundan son 1, 0 y 4, 8, 2, 13. En cuanto a la acentuación, en los enunciados interrogativos llama la atención que sea la única categoría que tiene una mayoría de unidades con acentuación nula. También es interesante que en los enunciados exclamativos el porcentaje de la acentuación esdrújula sea tan alto, posiblemente debido a las formas sufijadas de los pronombres de objeto en el imperativo.

Tabla 3-139: Número total y porcentaje de grupos rítmicos por tipo de enunciado

Grupo rítmico		/+I -E -S/		/-I +E -S/		/-I -E +S/		/-I -E -S/	
(0)	0	95	30.35%	3	11.11%	264	18.42%	112	15.98%
(s w)	1	80	25.56%	16	59.26%	385	26.87%	206	29.39%
(w s)	2	28	8.95%	2	7.41%	139	9.70%	82	11.70%
(s w w)	3	6	1.92%	2	7.41%	50	3.49%	13	1.85%
(w s w)	4	51	16.29%	3	11.11%	214	14.93%	112	15.98%
(w w s)	5	14	4.47%	0	0%	79	5.51%	27	3.85%
(s w w w)	6	0	0%	0	0%	1	0.07%	0	0%
(w s w w)	7	3	0.96%	0	0%	35	2.44%	4	0.57%
(w w s w)	8	17	5.43%	1	3.70%	144	10.05%	86	12.27%
(w w w s)	9	6	1.92%	0	0%	20	1.40%	18	2.57%
(s w w w w)	10	0	0%	0	0%	0	0%	0	0%
(w s w w w)	11	0	0%	0	0%	2	0.14%	0	0%
(w w s w w)	12	1	0.32%	0	0%	15	1.05%	5	0.71%
(w w w s w)	13	5	1.60%	0	0%	60	4.19%	25	3.57%
(w w w w s)	14	2	0.64%	0	0%	3	0.21%	3	0.43%
(s w w w w w)	15	0	0%	0	0%	0	0%	0	0%
(w s w w w w)	16	0	0%	0	0%	0	0%	0	0%
(w w s w w w)	17	0	0%	0	0%	0	0%	0	0%
(w w w s w w)	18	0	0%	0	0%	5	0.35%	3	0.43%
(w w w w s w)	19	5	1.60%	0	0%	17	1.19%	5	0.71%
(w w w w w s)	20	0	0%	0	0%	0	0%	0	0%
Total		313	100%	27	100%	1433	100%	701	100%

Fuente: Prosodia.xls (Base de datos propia)

Por último, en la Tabla 3-140 se hace un recuento del número de grupos rítmicos idéntico con el fin de observar cuáles son los grupos rítmicos más utilizados en la creación de configuraciones rítmicas mayores.

Como era de esperarse, mientras más pequeños sean los grupos rítmicos mayor será su propensión a ser utilizados en cadena. Además, se ha venido repitiendo que la acentuación más común es la grave. De este modo se explica que el grupo rítmico 1 (el más corto posible y con acentuación grave) sea el más común en la formación de secuencias rítmicas de dos,

tres y cuatro grupos consecutivos, con 40, 17 y 4 casos respectivamente, este es el ritmo tradicionalmente llamado trocaico. En importancia le sigue el grupo 4 con 13 casos de secuencia doble y 8 de secuencia triple, que dan lugar al ritmo llamado anfibráquico. Le sigue el grupo 8 con 6 casos de secuencia doble, 2 de secuencia triple y 1 de secuencia cuádruple, y que dan lugar al ritmo tradicionalmente llamado de tercer peón. Con el mismo número de observaciones, aunque 6 sean de secuencia doble y 3 de triple, está el grupo 2, que origina el ritmo yámbico. También cabe mencionar al grupo 13, con 3 casos de secuencia doble y 1 de triple, que da lugar a un ritmo combinado: dibráquico y anfibráquico, $(w w) + (w s w)$. Para terminar, mencionaremos al grupo 5, con 4 casos de secuencia doble (ritmo anapéstico), y a los grupos 3 y 19, ambos con una secuencia doble (ritmo dactílico y ritmo tribráquico-anfibráquico – $(w w w) + (w s w)$, respectivamente).

Tabla 3-140: Número de grupos rítmicos idénticos

Grupos rítmicos		2x	3x	4x
(0)	0	19	6	1
(s w)	1	40	17	4
(w s)	2	6	3	0
(s w w)	3	1	0	0
(w s w)	4	13	8	0
(w w s)	5	4	0	0
(s w w w)	6	0	0	0
(w s w w)	7	0	0	0
(w w s w)	8	6	2	1
(w w w s)	9	0	0	0
(s w w w w)	10	0	0	0
(w s w w w)	11	0	0	0
(w w s w w)	12	0	0	0
(w w w s w)	13	3	1	0
(w w w w s)	14	0	0	0
(s w w w w w)	15	0	0	0
(w s w w w w)	16	0	0	0
(w w s w w w)	17	0	0	0
(w w w s w w)	18	0	0	0
(w w w w s w)	19	1	0	0
(w w w w w s)	20	0	0	0
Total		93	37	6
Real		74	31	5

Fuente: Prosodia.xls (Base de datos propia)

En el *corpus* de la Ciudad de México se dan 74 casos reales de dos grupos rítmicos contiguos idénticos, 31 casos de tres grupos y 5 casos de cuatro grupos contiguos del mismo tipo, lo que nos arroja un total de 110 casos de isocronía, lo que, con respecto al número total de enunciados, 870, representa un 12.64% de enunciados con algún tipo de isocronía.

Con este análisis se da por concluido el análisis estadístico de los distintos parámetros fonéticos y fonológicos de la prosodia contenidos en el *corpus* de la Ciudad de México.

3.4 Análisis comparativo

A lo largo de cada una de las secciones anteriores se fue preparando el terreno para permitir un estudio sistemático de las similitudes y diferencias entre las variedades peninsular y mexicana más representativas en cuanto al número de hablantes y a la importancia que han ganado en cada uno de los países y en el resto del mundo por ser las capitales nacionales, por ser las urbes con el mayor número de habitantes, y por su consolidación como las variedades más difundidas por medios audiovisuales, así como su larga tradición cultural.

En la primera sección se fue presentando las distintas corrientes de estudio de la prosodia aplicadas al español y se eligió los modelos que más se adaptaban a un tratamiento computacional y que permitían una estandarización de los valores absolutos como condición indispensable para la comparación entre los datos de los dos *corpora*.

La metodología presentada en la segunda sección no solamente obedecía a criterios estéticos para la presentación de textos, grabaciones e imágenes de análisis de los fragmentos que componen las bases de datos, sino que los documentos resultantes cumplían con las funciones de clasificación y jerarquización de la información, presentación de los datos en distintos entornos, según la profundidad requerida por cada tipo de análisis, y sobre todo, la creación de la base de datos sobre la cuál se basó el análisis estadístico de esta sección.

En las siguientes páginas se presentará en cada gráfica un resumen de la información contenida paralelamente en las tablas de datos de las secciones correspondientes al *corpus* de Madrid y al *corpus* de la Ciudad de México.

Además, en las gráficas principales de las secciones sobre duración, intensidad y frecuencia se señala la significatividad de las diferencias entre los grupos por medio de la prueba estadística *t* de Student (Hernández Sampieri *et al.* 2003). La comparación se realizó sobre cada una de las variables, suponiendo que las desviaciones estándar de los dos grupos de hablantes son desiguales. Los niveles de significatividad serán indicados con *** (<.01), ** (<.05), * (<.1). Esto significa que en el caso, con una probabilidad de 99%, los valores medios obtenidos en cada grupo difieren significativamente, es decir, hay una posibilidad de error de 1%. Mientras menos observaciones haya, más difícil será alcanzar algún nivel de significatividad

3.4.1 Sílabas, unidades entonativas y enunciados

El recuento de sílabas y enunciados que se presentó en los apartados 3.2.1 y 3.3.1 aporta muy poco material para la comparación sistemática de las variedades lingüísticas del español que conforman la base de datos Prosodia.xls, pues, cuando mucho, nos permite ver las características de la muestra en sí. Por ejemplo, aunque se partió de la selección arbitraria de

3.4 Análisis comparativo

400 fragmentos de grabación producidos por hablantes de Madrid y 800 producidos por hablantes de la Ciudad de México, después de la transcripción detallada resultó que había varios casos en los que los silencios cortos dentro del fragmento podían ser interpretados como pausas entre enunciados, y, por esta razón, la base de datos aumentó a 475 enunciados de Madrid y 870 de la Ciudad de México.

El número de sílabas, por otra parte, dependió directamente de las características de los enunciados: 3,358 en el *corpus* de Madrid (véase Tablas 3-3, 3-4 y 3-5) y 6,256 en el *corpus* de la Ciudad de México (véase Tablas 3-72, 3-73 y 3-74), y, como se comentó en los apartados correspondientes, el criterio de calidad acústica fue el que se impuso por encima del de una representatividad estricta, con un número proporcional de hablantes de cada sexo y grupo de edad o social de acuerdo a la distribución poblacional, aunque esta haya sido la intención original.

Cabe mencionar que en el análisis estadístico se prescindió de la categoría de unidad entonativa porque se comprobó, en la mayoría de los casos, que éstas eran unidades mentales más que acústicas, pues no todas se delimitaban por pausas y el uso de alargamientos segmentales no era regular. Además, marcar cada una de esas unidades con tonos de frontera y nucleares hubiera saturado las rejillas de texto y se hubiera dificultado la identificación de los tonos relevantes.

Por este motivo, la sílaba fue considerada como la unidad prosódica mínima dentro de un fragmento de grabación, mientras que el enunciado fue considerado como la unidad máxima, ya que, incluso en los casos en los que había una interrupción o un límite de unidad entonativa antes de pausa, éstos también fueron tratados como límites de enunciado suspendido.

Finalmente, en los apartados 3.2.1 y 3.3.1 se puede observar la cantidad de enunciados de cada tipo que aportó cada uno de los hablantes a la base de datos y deducir el orden de sus preferencias (véase Tablas 3-5, 3-6, 3-7, y 3-74, 3-75, 3-76). Sin embargo, como muchas de estas diferencias se deben principalmente a criterios de selección según la calidad, las proporciones carecen de la fuerza necesaria para ser estudiadas comparativamente dentro del marco de esta investigación.

3.4.2 Duración

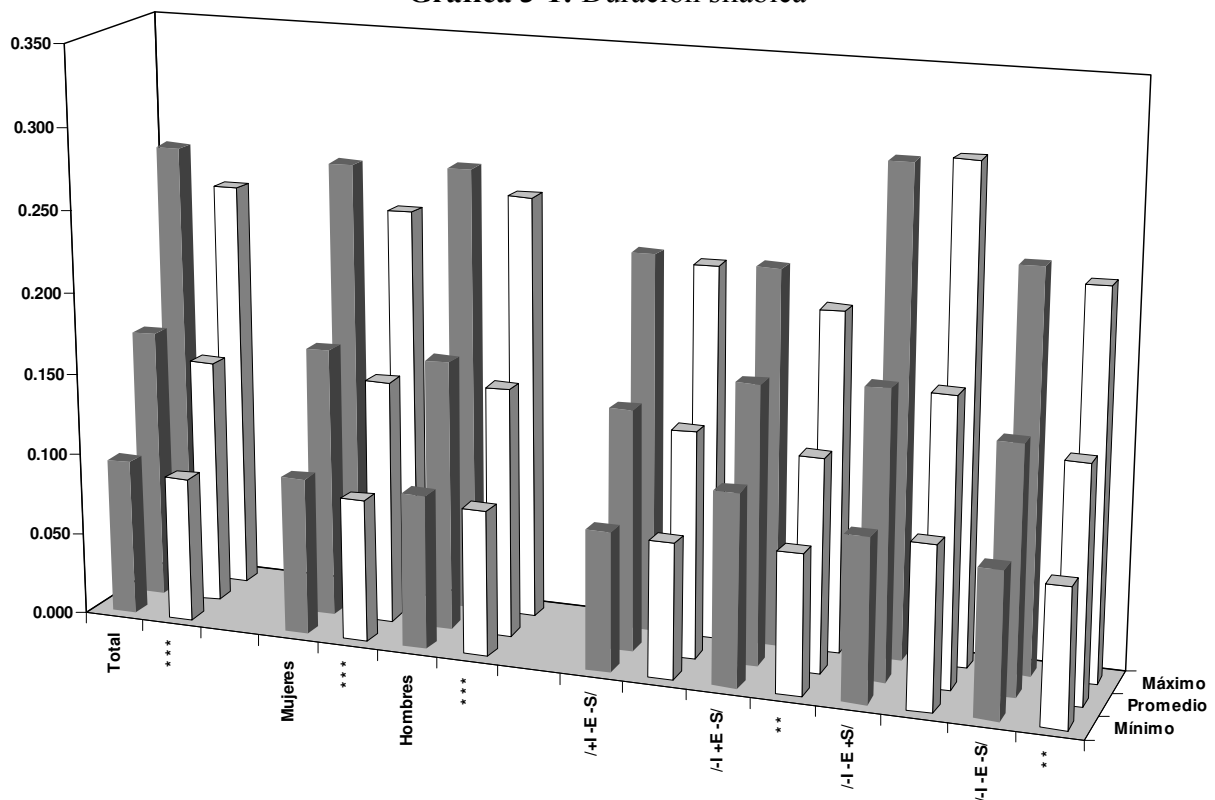
El primer fenómeno prosódico que se analizará en esta sección es el de la duración y, de manera indirecta, la velocidad del habla de los sujetos de estudio, pues este es un fenómeno abstracto que depende directamente de la duración silábica y su variación a lo largo del enunciado. Comentando acerca de su propia intuición, algunos de los hablantes mexicanos

dijeron en las entrevistas que tenían la impresión de que los españoles hablaban comparativamente más rápido que ellos. Si esta característica es percibida incluso por personas sin conocimientos especializados de fonética, debe de existir un correlato acústico que justifique esta percepción.

Para facilitar la lectura de la información, se optó por colorear con tonos grises más oscuros los resultados de las producciones lingüísticas de los hablantes de la Ciudad de México, mientras que las producciones de Madrid se dejaron en blanco o en tonos grises más claros.

Para comenzar, en la Gráfica 3-1 se presentan los valores mínimo, promedio y máximo de la duración silábica en las producciones lingüísticas de hombres y mujeres, así como en los distintos tipos de enunciados (véase Tablas 3-9 y 3-10, así como 3-78 y 3-79).

Gráfica 3-1: Duración silábica



Fuente: Prosodia.xls (Base de datos propia)

En el resultado total se puede observar que en las producciones de los mexicanos (a partir de aquí, por cuestiones prácticas, se designará como mexicanos no a los habitantes del país en general, sino únicamente a los originarios de la Ciudad de México) la duración silábica es siempre ligeramente mayor (entre 1 y 3 centésimas de segundo por sílaba) que en el caso de las producciones de los madrileños.

Igualmente, tanto en el caso de los hombres como en el de las mujeres, las producciones mexicanas tuvieron una duración silábica más larga que las madrileñas, con diferencias de hasta 2.6 centésimas de segundo por sílaba en el caso de los valores máximos en mujeres.

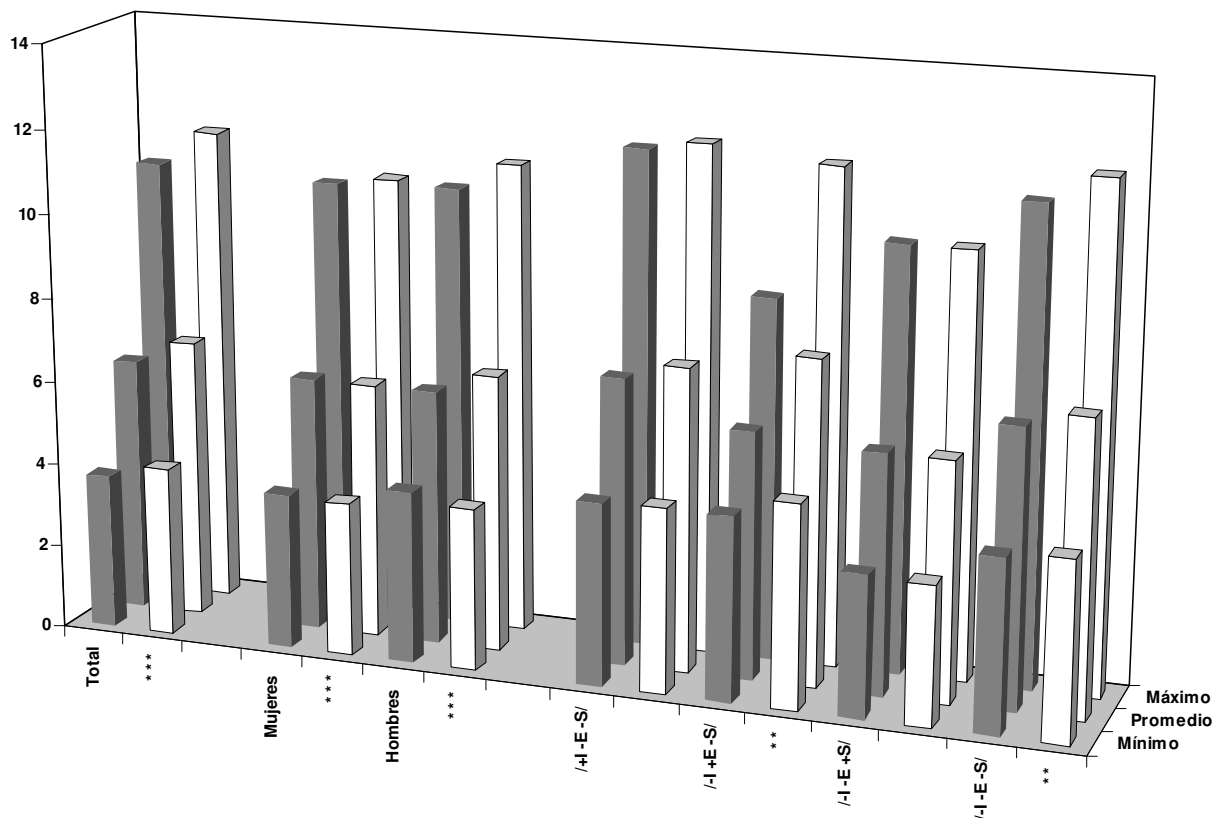
3.4 Análisis comparativo

Los enunciados interrogativos (/+I -E -S/) y los neutros (/I -E -S/) no presentan diferencias mayores a una centésima de segundo, mientras que en los exclamativos (/I +E -S/) la diferencia a favor de las producciones mexicanas llega a ser de hasta 3.2 centésimas en la duración mínima promedio y 3.4 en el promedio general. Por otra parte, en los enunciados suspendidos (/I -E +S/) la duración máxima promedio es el único valor en el que las sílabas de las producciones madrileñas son ligeramente más largas que las mexicanas, 0.4 centésimas de segundo, aunque estos enunciados, además de los interrogativos, no alcanzan ningún nivel de significatividad.

Hasta aquí se puede concluir que, en general, la duración silábica de las producciones mexicanas siempre es ligeramente mayor que la de las producciones madrileñas, sobre todo si se toma en cuenta los valores promediados.

Ahora, en la Gráfica 3-2 se presenta el resultado de la aplicación de la fórmula inversa, con la que se obtiene el número promedio de sílabas por segundo (véase Tablas 3-11 y 3-12, así como 3-80 y 3-81). Esto no significa que en realidad alguien haya producido alrededor de 12 sílabas por segundo, sino que, tomando en cuenta la duración silábica promedio, si todas las sílabas en un segundo tuvieran la misma longitud, arrojarían estos resultados.

Gráfica 3-2: Número de sílabas por segundo



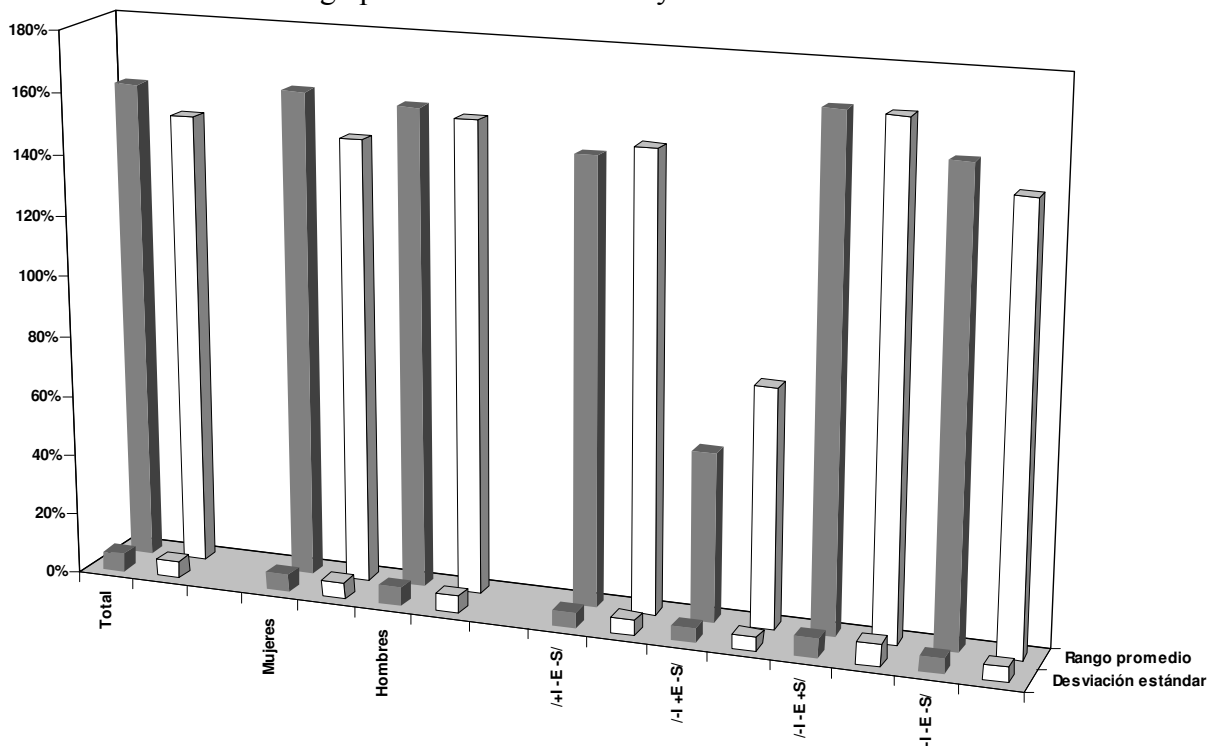
Fuente: Prosodia.xls (Base de datos propia)

Como podía preverse en la Gráfica 3-1, en general, las producciones lingüísticas de los madrileños son potencialmente más rápidas que las de los mexicanos (entre 0.33 y 0.85 síl/seg de diferencia). En el caso específico de las mujeres, la diferencia entre las producciones de las madrileñas y las mexicanas puede llegar a ser de hasta 1.99 síl/seg. Sin embargo, entre los distintos sexos las diferencias son mínimas: las mujeres madrileñas producen entre 0.199 y 0.258 síl/seg más que los hombres, mientras que en la Ciudad de México son los hombres quienes producen entre 0.005 y 0.388 síl/seg más que las mujeres mexicanas.

Excepto en el número mínimo promedio de sílabas en los enunciados suspendidos, donde los mexicanos producen 3.370 frente a 3.319 síl/seg de los madrileños, en los enunciados interrogativos, suspendidos y neutros, los madrileños producen un número ligeramente mayor de sílabas por segundo que los mexicanos. En el caso de los enunciados exclamativos, los madrileños producen entre 0.476 y 3.164 síl/seg más que los mexicanos.

Por último, en la Gráfica 3-3 se vacía la información sobre el rango promedio de variación entre los valores máximos y mínimos de variación porcentual en la duración de las sílabas en posición contigua, así como la desviación estándar promedio de la duración silábica.

Gráfica 3-3: Rango promedio de variación y desviación estándar de la duración



Fuente: Prosodia.xls (Base de datos propia)

En la desviación estándar (véase Tablas 3-15 y 3-16, así como 3-84 y 3-85), en total, las sílabas de los mexicanos se desvían 0.41% más del promedio que las de los madrileños. En ambas ciudades, las producciones de los hombres tienen una mayor desviación estándar de la duración promedio con respecto a las de las mujeres (0.36% en la Ciudad de México y 0.39%

3.4 Análisis comparativo

en Madrid). Además, excepto en los enunciados suspendidos, donde los madrileños tienen a favor una diferencia de 0.46% de desviación estándar, en el resto de los tipos de enunciados producidos por los mexicanos la desviación estándar es ligeramente mayor (0.03% a 0.59%).

En cuanto al rango promedio de variación (véase Tablas 3-13 y 3-14, así como 3-82 y 3-83), en total, los mexicanos utilizan un rango de variación intersilábica de la duración 9.06% mayor al de los madrileños. Esto se comprueba en el caso de las mujeres, pues las mexicanas utilizan un rango 13.98% mayor al de las madrileñas, aunque los mexicanos utilizan un rango 2.65% menor que los madrileños. Además, mientras entre hombres y mujeres mexicanas sólo hay una diferencia de 2.37% a favor de las mujeres, en Madrid son los hombres quienes hacen claramente un uso más amplio del rango de variación, con una diferencia de 18.96%.

Aparte de los enunciados neutros y los suspendidos, en los que los mexicanos utilizan rangos superiores a los de los madrileños por 9.27% y 0.74% respectivamente, los madrileños hacen claramente un uso más amplio del rango de variación de la duración en los demás tipos de enunciados: interrogativos por 3.72%, y 23.31% en el caso de los enunciados exclamativos.

En conclusión, la idea de que los mexicanos hablan más lento, con sílabas más largas y regulares con respecto al habla de los españoles, en general, sí tiene un sustento en la información acústica, al menos en las producciones lingüísticas de los hablantes que conforman la base de datos Prosodia.xsl. Cabe aclarar que estas diferencias normalmente no afectan la mutua comprensión entre los hablantes de las variedades peninsular y mexicana del español, más bien se perciben como rasgos característicos del habla regional.

3.4.3 Intensidad

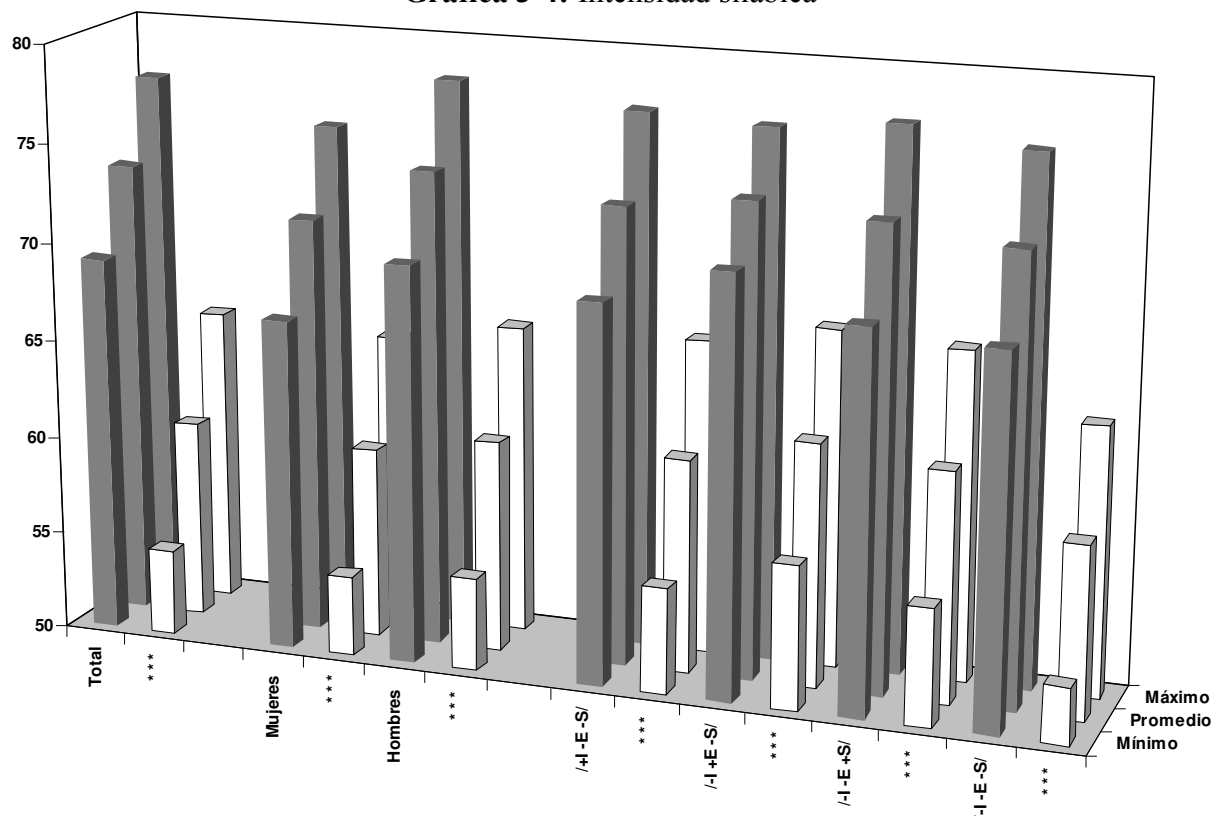
Otro de los aspectos prosódicos percibidos como diferentes entre los hablantes de variedades peninsular y mexicana del español es el volumen o intensidad de la voz. Los mexicanos consideran que los españoles hablan más ‘fuerte’, es decir, con una intensidad que en algunas ocasiones incluso puede parecer excesiva a los mexicanos. En este apartado se recopilarán los valores acústicos en los fragmentos de grabación para comprobar qué hay de cierto en esa percepción intuitiva de la diferencia de intensidad.

Primeramente, en la Gráfica 3-4 se concentran los valores absolutos de la intensidad silábica mínima, máxima y promedio de los hablantes de las dos variedades de español, también filtrados según el sexo y el tipo de enunciado (véase Tablas 3-17, 3-18, y 3-86, 3-87).

Lo primero que salta a la vista en esta Gráfica es que las barras de color gris oscuro, correspondientes a los hablantes de la Ciudad de México, son los que muestran los valores más altos, en promedio entre 12.73 y 15.43 dB por encima de las producciones de los

hablantes de Madrid. Estas diferencias se repiten a lo largo de todas las categorías de clasificación. Sin embargo, como se ha venido mencionando en apartados anteriores, estas diferencias, especialmente debido a su regularidad dentro de cada grupo de hablantes, son debidas principalmente a la distancia de los hablantes con respecto al micrófono (además de que en algunas de las grabaciones de Madrid la grabadora estaba escondida), y las diferencias en la sensibilidad de los micrófonos utilizados para la captura de las conversaciones.

Gráfica 3-4: Intensidad silábica



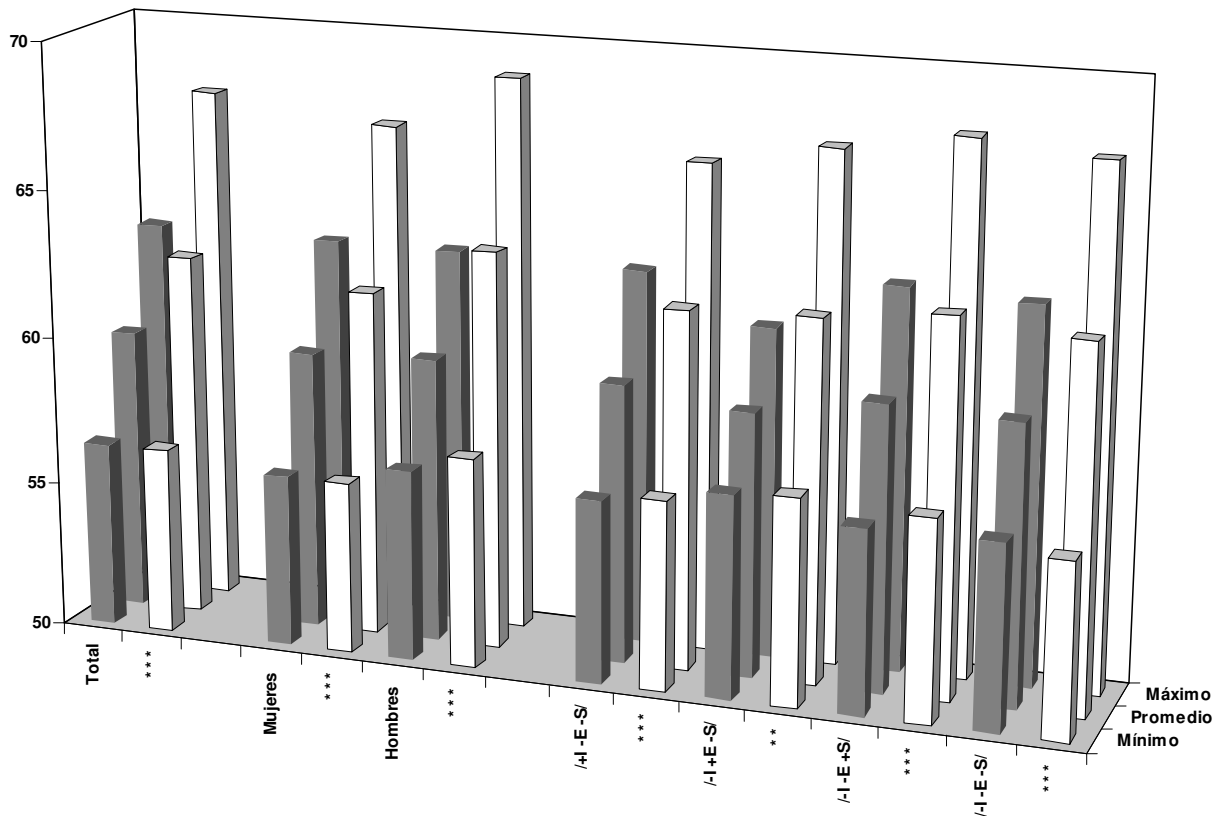
Fuente: Prosodia.xls (Base de datos propia)

Por todas estas razones, es preferible hacer uso de los valores estandarizados de la intensidad silábica según la fórmula introducida en 1.4, como se muestra en la Gráfica 3-5 (véase Tablas 3-23, 3-24, y también 3-92, 3-93). De este modo observamos en la Gráfica 3-5 los resultados de la variación de la intensidad dentro de cada enunciado. Así, la intensidad estandarizada mínima de las producciones lingüísticas de los mexicanos sigue siendo más alta, con una diferencia promedio de 0.02 dB. Sin embargo, todos los valores máximos y promedio son más altos en las producciones madrileñas, 4.77 dB y 2.79 dB en el total general respectivamente. Esto quiere decir que los mexicanos utilizan un rango de variación más estrecho que los madrileños, aunque este aspecto se estudiará en la Gráfica 3-6.

Comparando las producciones masculinas y femeninas vemos que en ambos grupos de hablantes los hombres hacen uso de una intensidad ligeramente mayor, en promedio 0.21 dB

más entre los mexicanos y 1.74 dB entre los madrileños, y entre los hombres son los madrileños quienes hablan con una intensidad promedio de 3.82 dB por encima de las producciones de su contraparte mexicana.

Gráfica 3-5: Intensidad silábica estandarizada



Fuente: Prosodia.xls (Base de datos propia)

En cuanto a los tipos de enunciados, en los cuatro casos los madrileños hacen uso de una intensidad promedio más alta, con diferencias de 2.69 dB en interrogaciones, 2.79 dB en los enunciados neutros, 3.05 dB en los enunciados suspendidos, y 3.33 dB en exclamaciones.

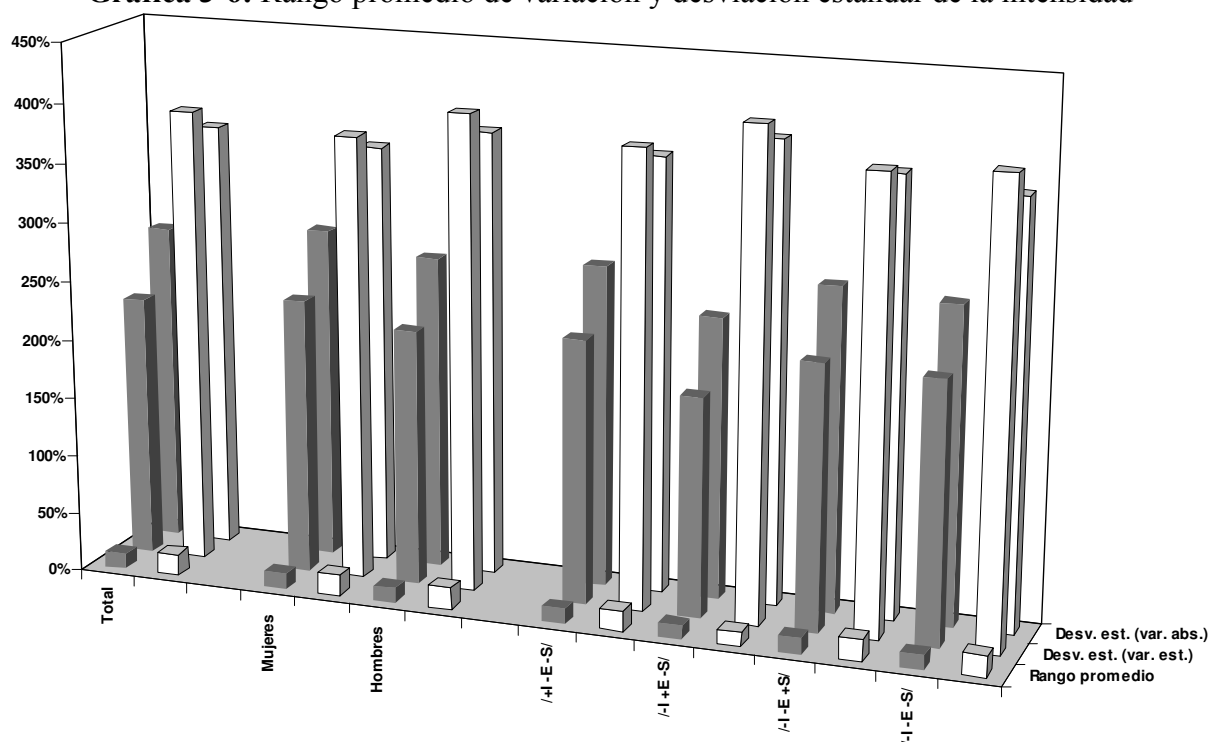
El rango de variación (véase Tablas 3-19 y 3-20, así como 3-88 y 3-89) que ya podía apreciarse en la Gráfica anterior se puede observar en la Gráfica 3-6 junto con los valores de la desviación estándar, partiendo de los valores absolutos (véase Tablas 3-21 y 3-22, también 3-90 y 3-91) y los estandarizados (véase Tablas 3-25, 3-26, y 3-94 y 3-95).

El rango de variación promedio de la intensidad, que, por cuestiones prácticas, se incluyó en la misma Gráfica que la desviación estándar a pesar de ser mucho más bajo que estas dos últimas, es, en todos los casos, más estrecho en las producciones de los mexicanos que en las de los madrileños, en promedio 5.33% menor, en mujeres 4.10%, y en hombres hasta 6.56%. En los enunciados exclamativos los madrileños utilizan un rango 1.09% más amplio, los interrogativos 4.78%, los suspendidos 5.51% y los neutros 6.02%.

En cuanto a la desviación estándar, que en todos los casos los valores estandarizados son mayores que en los valores absolutos, ambas series de valores muestran valores claramente

mayores en las producciones madrileñas, lo que también refuerza los resultados del rango promedio. Por ejemplo, tomando en cuenta únicamente los valores estandarizados, en total, los producciones madrileñas se desvían 164% más del promedio que las mexicanas, y las diferencias en los tipos de enunciados varían entre 166% y 225%, en neutros y exclamativos respectivamente. También las producciones de mujeres se desvían 142% más en Madrid que en la Ciudad de México, y las de hombres se desvían 185%. Lo que llama la atención es que en Madrid son las producciones de los hombres las que más se desvían del promedio, 26% con respecto a las producciones de las madrileñas, mientras que en la Ciudad de México se da el caso contrario, pues las producciones lingüísticas de las mujeres se desvían 17% más del promedio que las de los hombres.

Gráfica 3-6: Rango promedio de variación y desviación estándar de la intensidad



Fuente: Prosodia.xls (Base de datos propia)

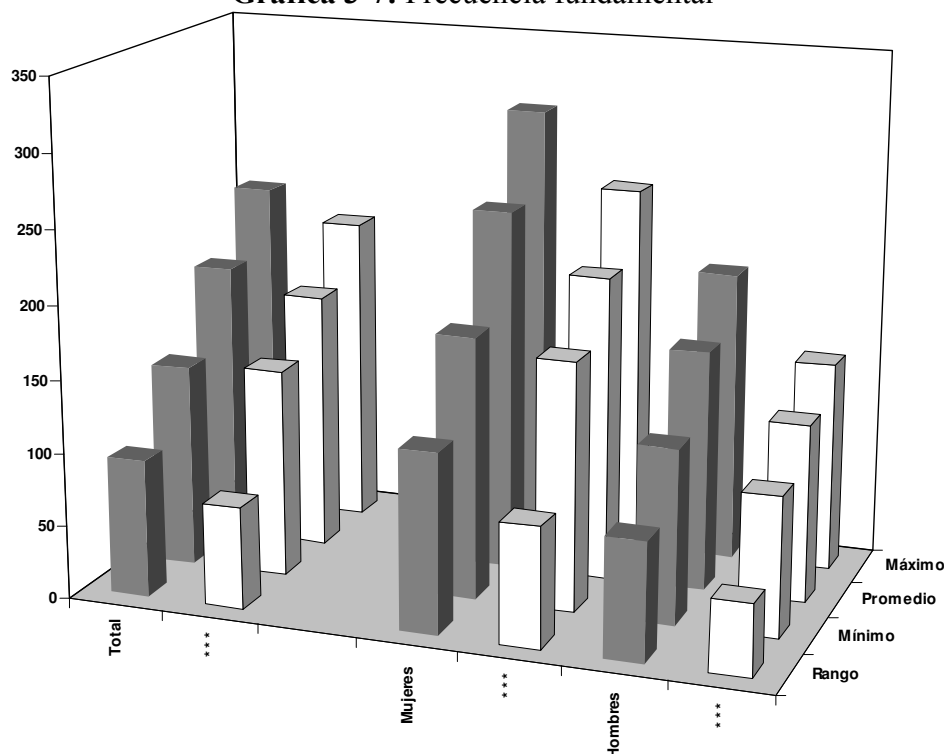
Resumiendo, también en el caso de la intensidad, partiendo de valores estandarizados que sientan las bases para un estudio comparativo independiente de factores extralingüísticos, los datos confirman la concepción generalizada de que los madrileños hablan a un volumen más alto, es decir, con una intensidad más alta (todos los grupos de datos con los niveles de significatividad más altos), y además hacen un uso más amplio del rango de variación, lo que se percibe como aumentos y disminuciones de la intensidad más marcados dentro de los enunciados que en el habla de los mexicanos.

3.4.4 Frecuencia

Pasando al tema de la frecuencia, aunque pocos hablantes son conscientes de las diferencias en el uso de los registros, altura de la voz o rango de la frecuencia fundamental utilizado por los hablantes de distintas regiones geográficas, intuitivamente, cuando los mexicanos intentan imitar el habla de los españoles, cambian automáticamente a un registro más grave. Esta conducta nos revela que el rango frecuencial es una realidad percibida por los hablantes como un factor característico de su habla personal que puede contrastar con el rango frecuencial de los hablantes de un grupo lingüístico diferente. Además, estas diferencias, al igual que el uso distinto de la duración silábica, la velocidad del habla o la intensidad, parece ser percibido por los hablantes como idiosincrático o característico de una comunidad lingüística sin que produzca diferencias funcionales o que lleve a malentendidos entre unos hablantes y otros.

A continuación veremos si en los datos acústicos se encuentra la información necesaria para afirmar que los hablantes de una u otra variedad del español hablan más grave o más agudo. Para esto recurriremos, primeramente, a la Gráfica 3-7 que contiene la información básica sobre la frecuencia fundamental máxima, mínima, promedio, y el rango frecuencial de hombres y mujeres de la Ciudad de México y Madrid (véase Tablas 3-27 y 3-96).

Gráfica 3-7: Frecuencia fundamental



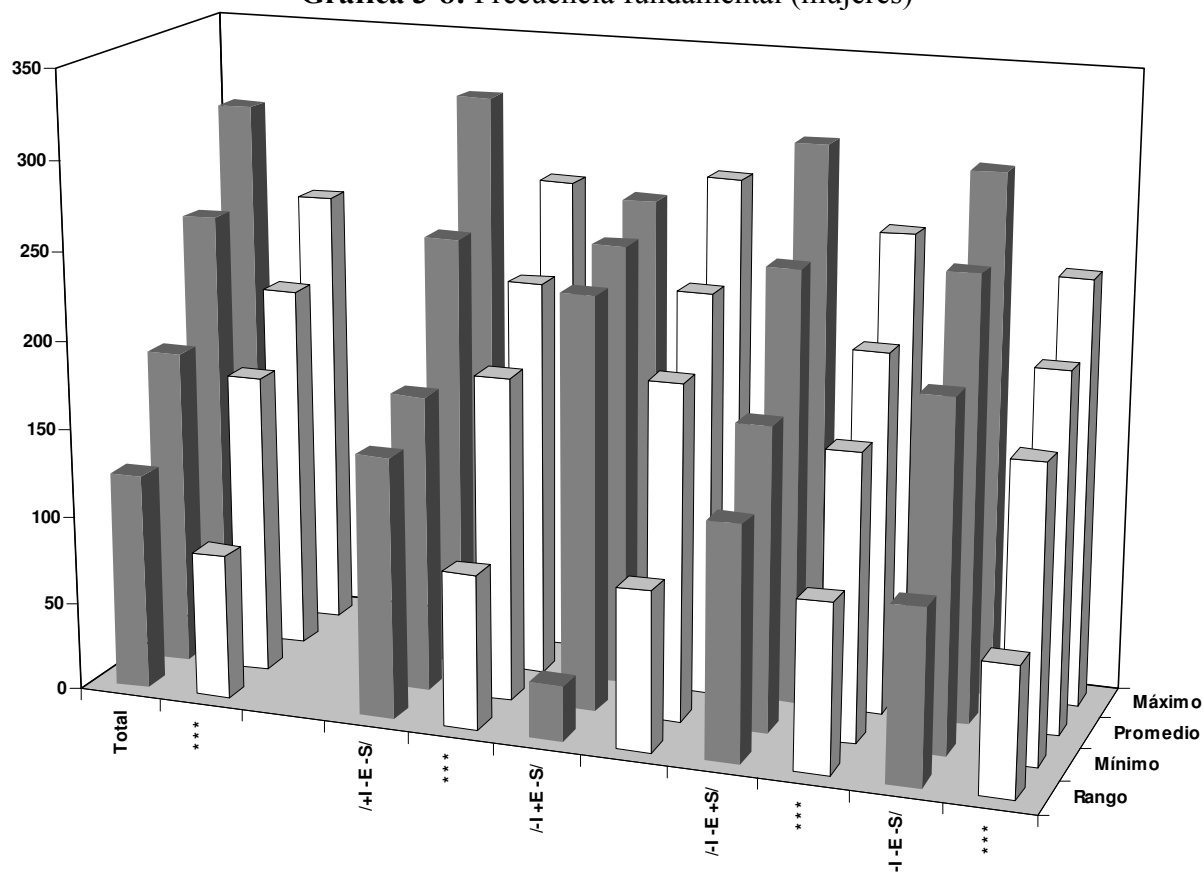
Fuente: Prosodia.xls (Base de datos propia)

A pesar de que se ha advertido en múltiples ocasiones de las dificultades que implica utilizar valores totales en los que se promedie la frecuencia fundamental de hombres y mujeres, en la presente Gráfica se aprecia claramente que en las cuatro series de valores, las producciones

lingüísticas mexicanas obtuvieron valores más altos que las de los madrileños y eso se observa también en los valores totales.

No obstante, estos valores se analizarán en detalle en la Gráfica 3-8, únicamente para los datos de las mujeres (véase Tablas 3-29 y 3-98), y en la Gráfica 3-9, para los datos de los hombres (véase Tablas 3-30 y 3-99).

Gráfica 3-8: Frecuencia fundamental (mujeres)



Fuente: Prosodia.xls (Base de datos propia)

En el caso de las mujeres, con excepción del rango de variación y la frecuencia fundamental máxima en los enunciados exclamativos, así como la mínima en los interrogativos, las producciones lingüísticas de las mexicanas siempre tienen una frecuencia fundamental mayor que las de las madrileñas.

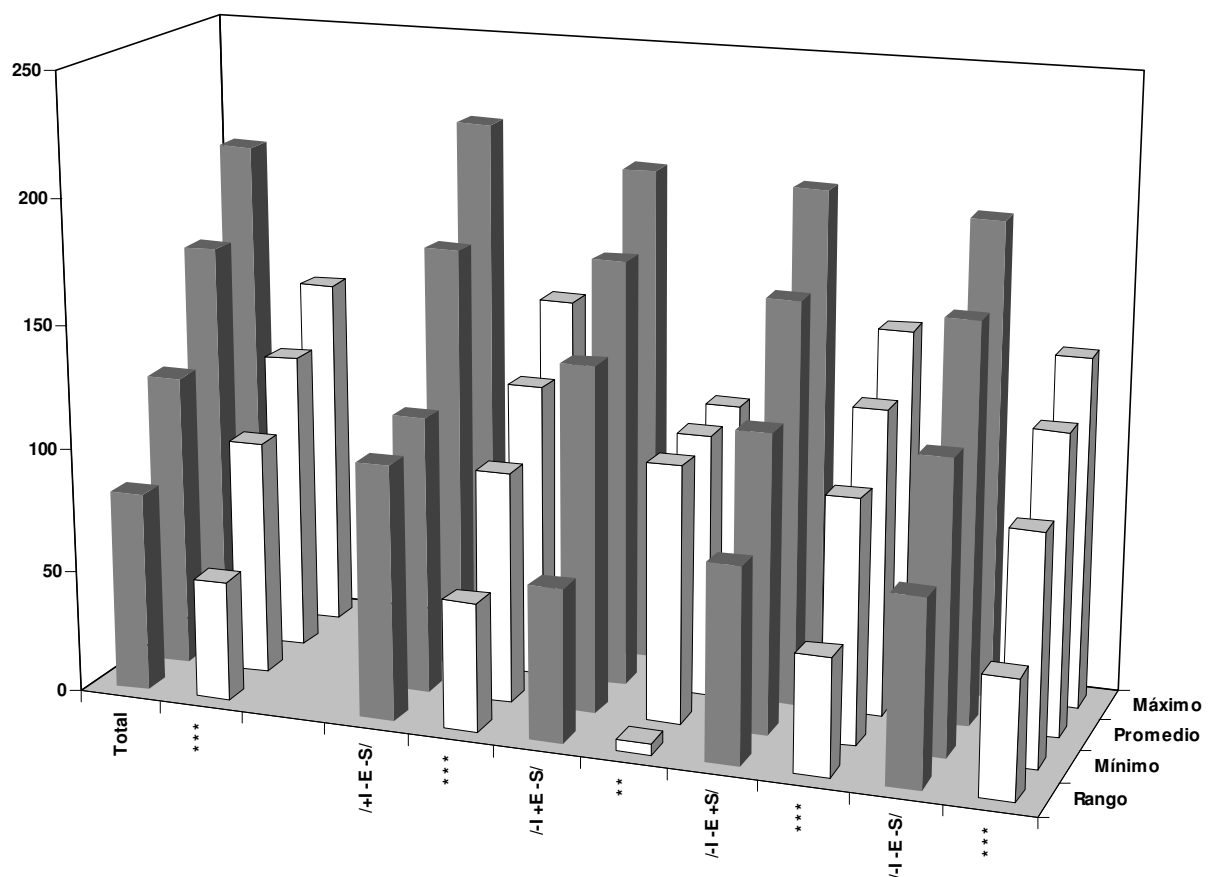
La frecuencia fundamental promedio de las mexicanas es 38.87 Hz mayor que la de las madrileñas, además de que la diferencia de la frecuencia fundamental en los distintos tipos de enunciados va desde 18.00 Hz en los exclamativos hasta 49.58 Hz en los neutros.

También el rango frecuencial general es 40.56 Hz más amplio en las producciones de las mexicanas, y en los distintos tipos de enunciados la diferencia oscila entre 25.16 Hz en los neutros y 71.82 Hz en los interrogativos. Únicamente el rango frecuencial en los enunciados exclamativos es 65.82 Hz mayor en las producciones de las madrileñas.

3.4 Análisis comparativo

Por otro lado, en el caso de los hombres (véase Gráfica 3-9), no hay ninguna excepción, pues todos los valores son más altos en las producciones lingüísticas de los hablantes de la Ciudad de México. Por ejemplo, la frecuencia fundamental promedio difiere por 42.51 Hz entre las dos variantes, y en los distintos tipos de enunciados la brecha varía entre 35.96 Hz en los suspendidos y 71.36 Hz en los enunciados exclamativos. Además, con respecto al rango frecuencial, en promedio, los mexicanos utilizan un espectro de frecuencias 32.12 Hz más amplio que el de los madrileños, mientras que, dependiendo del tipo de enunciado, la diferencia oscila entre 27.07 Hz en los enunciados neutros y 56.09 Hz en los exclamativos.

Gráfica 3-9: Frecuencia fundamental (hombres)



Fuente: Prosodia.xls (Base de datos propia)

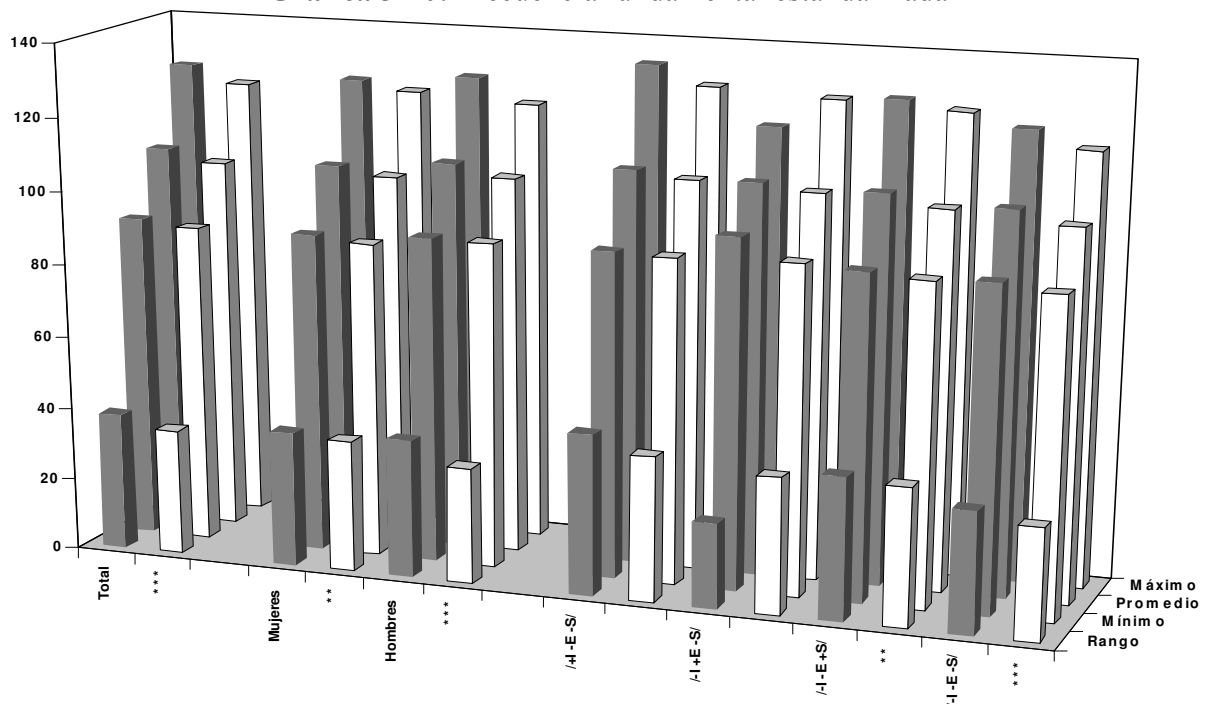
Sin embargo, para poder hacer realmente un estudio comparativo entre todas las producciones lingüísticas de hombres y mujeres de las dos ciudades, tendremos que recurrir a los valores de la frecuencia fundamental estandarizada (véase Tablas 3-37 y 3-38, así como 3-106 y 3-107), como se muestran en la Gráfica 3-10.

La sustitución de la frecuencia fundamental de la primera sílaba de cada enunciado por 100 y la adición de la variación porcentual de cada sílaba subsiguiente nos permite comparar las series de datos partiendo de una base común (véase 1.1.1.2). Así, por ejemplo, independientemente de que las voces de los hablantes sean graves o agudas, vemos que la

frecuencia estándar promedio de los mexicanos es 3.20 Hz más alta que la de los madrileños, y que los hombres, tanto mexicanos como madrileños hacen un uso mayor de la variación frecuencial que las mujeres, 2.48 y 1.71 Hz respectivamente. Además, las producciones de las mexicanas son 2.14 Hz más altas que las de las madrileñas, mientras que las producciones lingüísticas de los mexicanos son 2.91 Hz más altas que las de los madrileños.

Si tomamos en cuenta los distintos tipos de enunciados, las diferencias, todas ellas a favor de los mexicanos, van de 1.60 Hz en el caso de los exclamativos hasta 3.46 Hz en los neutros, a pesar de que los valores máximos de la F_0 estandarizada sean hasta 8.25 Hz más bajos. Sin embargo, el uso de valores estandarizados hace que los enunciados interrogativos y exclamativos ya no alcancen ningún nivel de significatividad.

Gráfica 3-10: Frecuencia fundamental estandarizada



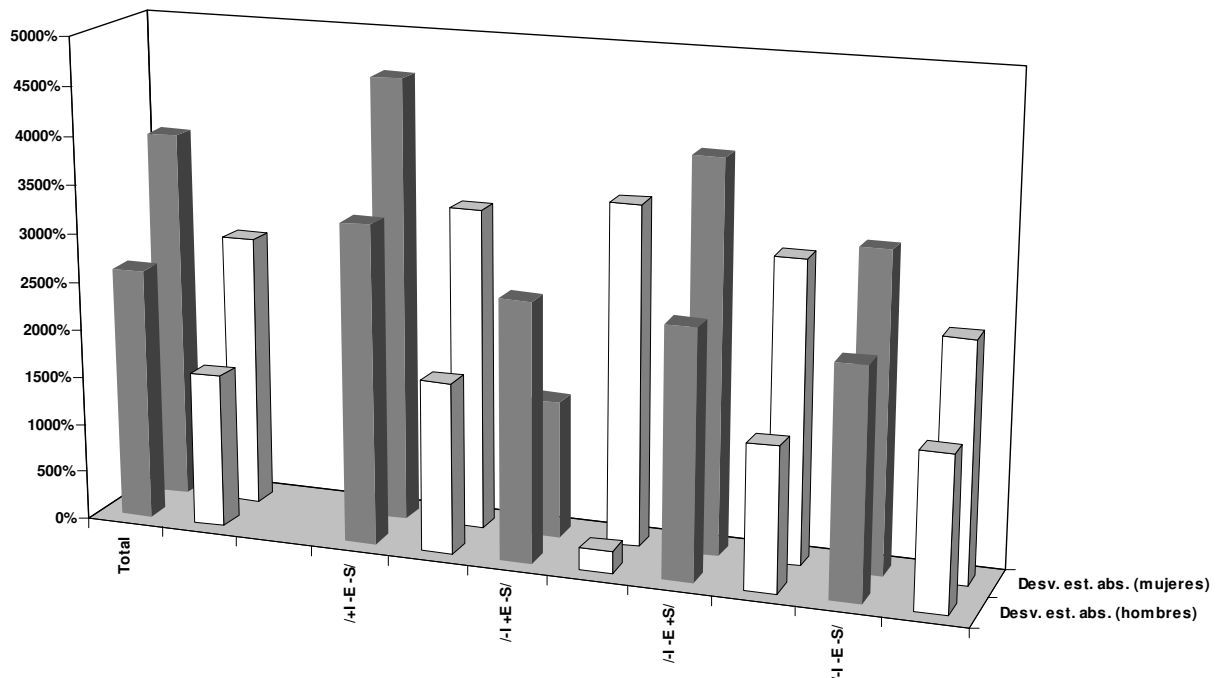
Fuente: Prosodia.xls (Base de datos propia)

En cuanto al rango frecuencial promedio, los mexicanos utilizan una gama de frecuencias estandarizadas 3.09 Hz mayor que los madrileños. Los hombres en la Ciudad de México sólo usan un rango 0.92 Hz mayor que las mujeres, mientras que las madrileñas usan un rango 4.44 Hz más amplio que los hombres. Además, mientras que el rango frecuencial de las mexicanas es 0.82 Hz mayor que el de las madrileñas, los mexicanos hacen uso de un rango 6.18 Hz más amplio que los madrileños.

En los distintos tipos de enunciados los mexicanos hacen uso de un rango frecuencial más amplio (1.13 Hz en los suspendidos, 2.76 Hz en los neutros y 4.43 en los interrogativos), excepto en el caso de los enunciados exclamativos, pues la diferencia con respecto al rango utilizado por los mexicanos asciende a 14.13 Hz.

En la Gráfica 3-11 se observa en detalle las diferencias del uso que hacen hombres y mujeres de la desviación estándar de la frecuencia fundamental (véase Tablas 3-33 y 3-102 para los totales, 3-35 y 3-104 para los valores femeninos, así como 3-36 y 3-105 para los masculinos).

Gráfica 3-11: Desviación estándar de la frecuencia fundamental



Fuente: Prosodia.xls (Base de datos propia)

En el caso de las mujeres, únicamente en las sílabas de los enunciados exclamativos las frecuencias fundamentales de las madrileñas se desvían más que las de las mexicanas, con una diferencia de 2097%, mientras que en el resto de los tipos de enunciados, las diferencias que alcanzan las mujeres van de 802% en los neutros a 1275% en los interrogativos.

Por otro lado, en el caso de los hombres, en cada uno de los tipos de enunciados, los mexicanos alcanzan las desviaciones estándar más altas en comparación con las producciones de los madrileños, es decir, los mexicanos hacen un uso menos regular de la frecuencia fundamental: 773% en los enunciados neutros, 1087% en los suspendidos, 1531% en los interrogativos, e incluso 2444% en los exclamativos. Sin embargo, debemos recordar que, en parte, la altura de esta desviación estándar depende de los valores absolutos de la frecuencia fundamental que, de por sí, ya eran más altos en las producciones lingüísticas mexicanas.

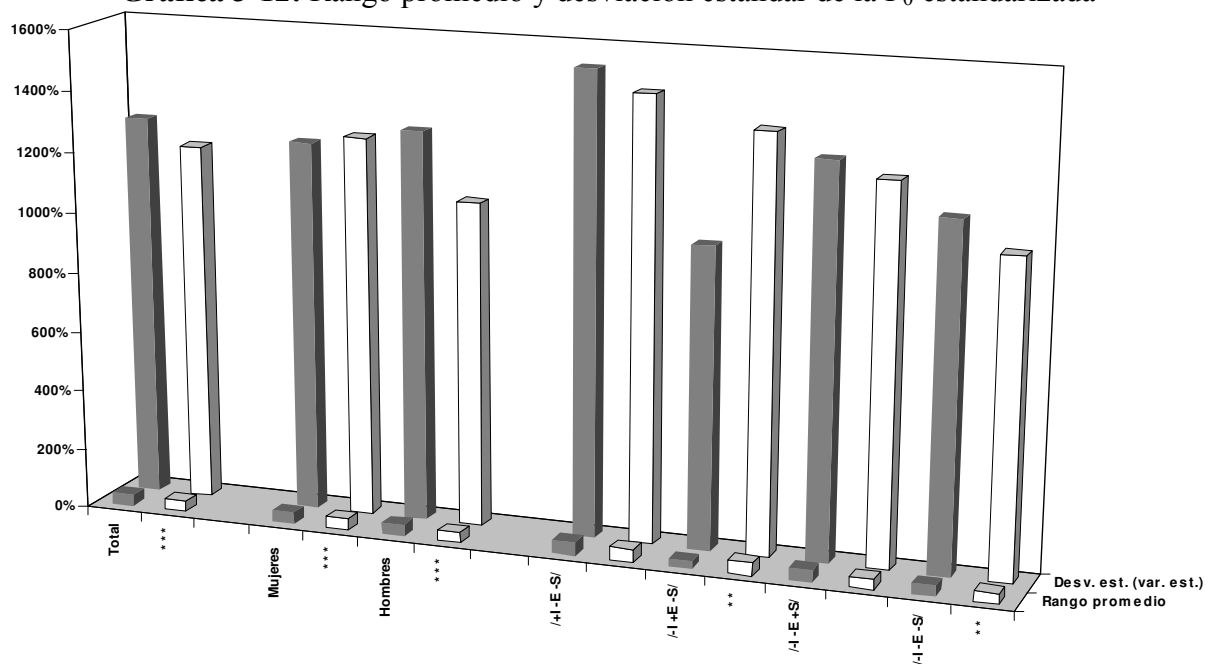
Para comprobar estos resultados, en la Gráfica 3-12 se emplea los valores de la frecuencia fundamental estandarizada para calcular las diferencias en la desviación estándar (véase Tablas 3-39, 3-40, y también 3-108, 3-109) y el rango promedio de variación porcentual de la frecuencia (véase Tablas 3-31 y 3-32, así como 3-100 y 3-101).

En lo referente a la desviación estándar, en total, los mexicanos se desvían 83% más del promedio que los madrileños, aunque en realidad esto sólo sucede en el caso de los hombres,

donde los mexicanos alcanzan una diferencia de 219% (64% con respecto a las mexicanas), mientras que, entre las mujeres, las madrileñas alcanzan una diferencia de 30%, y también con respecto a los hombres de Madrid alcanzan una diferencia de 185%.

En los resultados de la desviación estándar según los tipos de enunciados vemos repetirse el mismo patrón: los madrileños alcanzan una desviación 373% mayor que los mexicanos en los enunciados exclamativos, pero en los demás enunciados los mexicanos obtienen valores más altos: 51% de diferencia en los suspendidos, 67% en los interrogativos y 99% en los neutros.

Gráfica 3-12: Rango promedio y desviación estándar de la F₀ estandarizada



Fuente: Prosodia.xls (Base de datos propia)

En cuanto al rango promedio de variación frecuencial, en general, los resultados anteriores se confirman: los mexicanos hacen uso de un rango 3.52% mayor que los madrileños, los hombres 6.45% y las mujeres 1.74%; los hombres mexicanos utilizan un rango tan sólo 0.12% mayor que las mexicanas, mientras que las madrileñas alcanzan una diferencia de 4.59% con respecto a los hombres de Madrid; en los enunciados exclamativos los madrileños hacen uso de un rango frecuencial 16.25% mayor que los mexicanos, mientras que estos últimos alcanzan una diferencia de 1.52% en los enunciados interrogativos, 2.83% en los neutros y 5.98% en los suspendidos.

En conclusión, con excepción de los enunciados exclamativos, que, en realidad, debido a su escasa presencia en la base de datos, no arrojan resultados representativos, en general se ve una tendencia clara que confirma la percepción intuitiva de registros de voz más agudos para los mexicanos desde el punto de vista madrileño y viceversa. Además, mientras que entre los mexicanos no hay gran diferencia en cuanto al uso del rango de frecuencias, en Madrid parece que las mujeres hablan de manera más expresiva que los hombres.

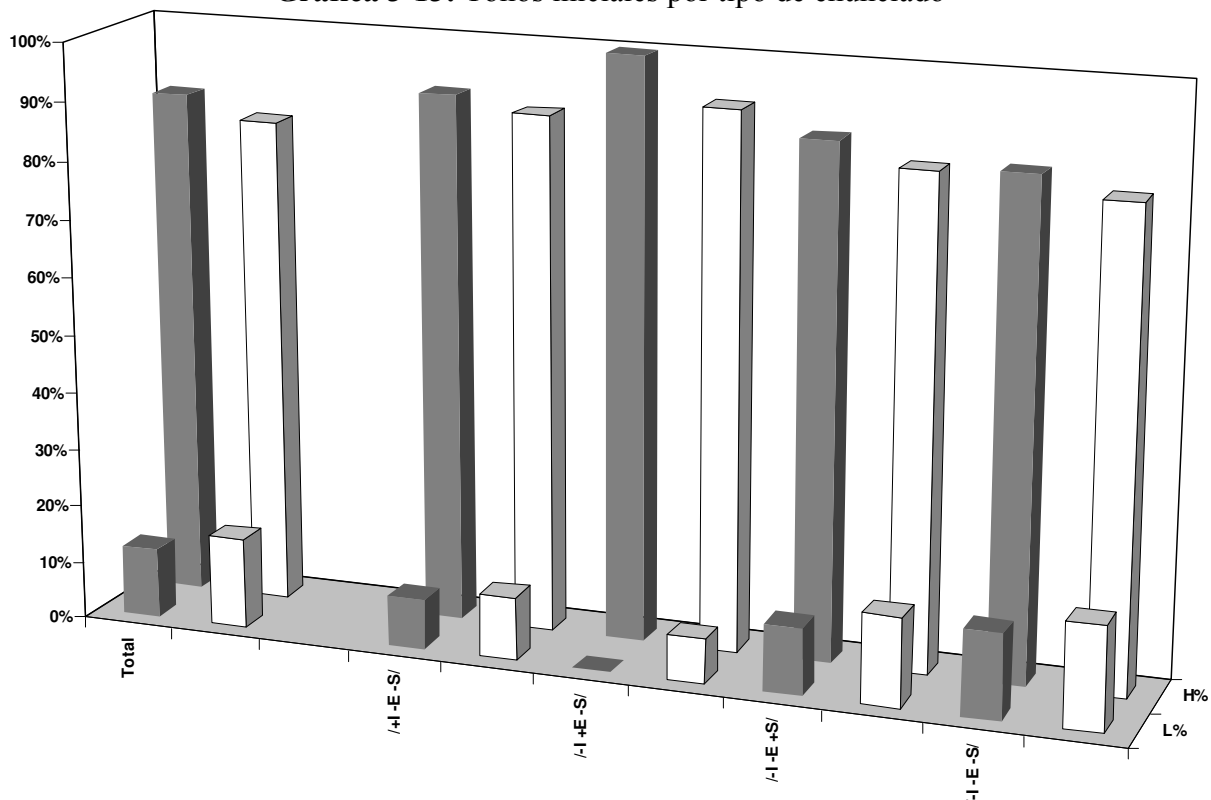
3.4.5 Entonación

La entonación es un fenómeno prosódico que muy pocas personas perciben. Los hablantes mexicanos a quienes se pidió que intentaran imitar el habla de un español, además de hablar más rápido, con mayor intensidad y con un registro de voz más grave en comparación con su manera habitual de hablar, algunas veces también emplearon algunas configuraciones tonales ligeramente distintas a las empleadas por ellos mismos, pero fueron incapaces de describirlas. La razón de esto, como se ha venido comentando, es que entre las dos variedades no existen realmente solapamientos de funciones comunicativas entre el significado conferido a los enunciados con tal o cual contorno entonativo, sino que las diferencias son interpretadas como melodías características del habla de otros grupos hispanohablantes.

A continuación se comparará de manera sistemática los datos correspondientes a la frecuencia de uso de los distintos tonos en posición inicial, prenuclear, nuclear y final en los distintos tipos de enunciados producidos por mexicanos y madrileños.

Comenzamos con la Gráfica 3-13, en la que se presentan los datos sobre los tonos iniciales (véase Tablas 3-41 y 3-46, así como 3-110 y 3-115), donde se puede ver que, como la suma de ambos tonos es 100%, en todos los casos, los mexicanos produjeron más tonos H%: en total 3.63%, en enunciados interrogativos 2.35%, en neutros 3.25%, en suspendidos 3.83, y en los enunciados exclamativos incluso 7.69%.

Gráfica 3-13: Tonos iniciales por tipo de enunciado

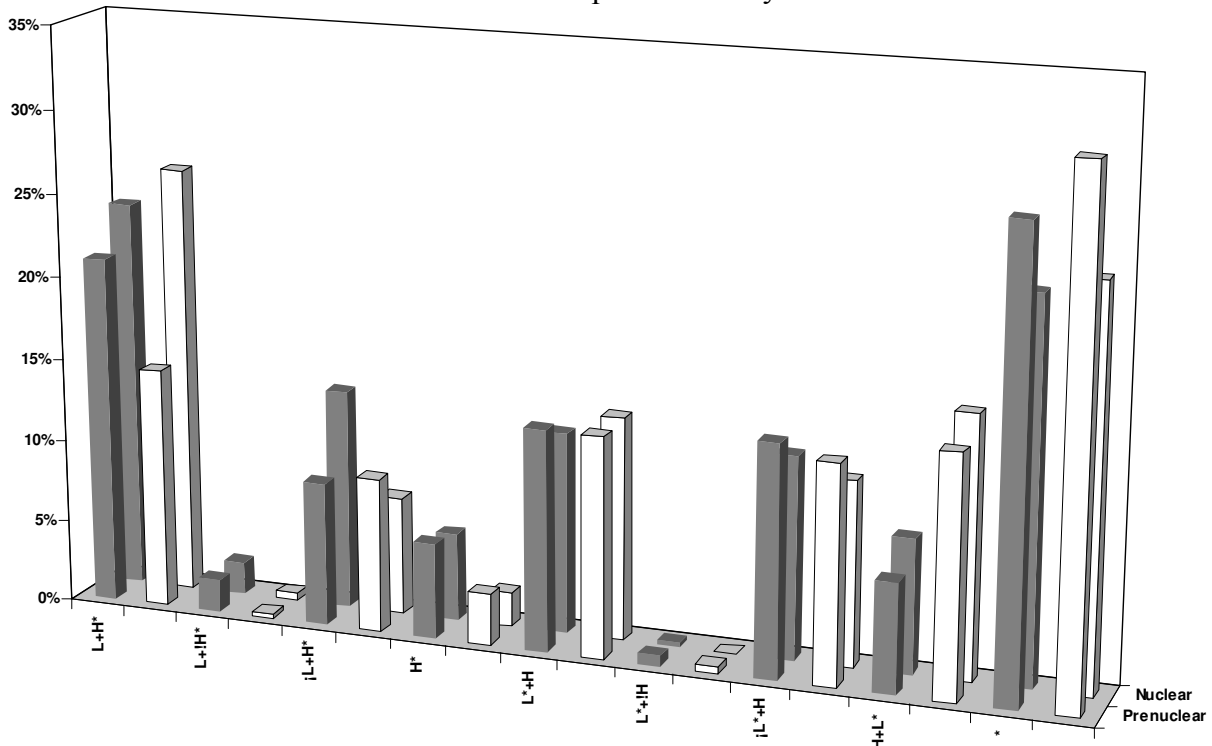


Fuente: Prosodia.xls (Base de datos propia)

Sin embargo, las proporciones son muy similares en ambos casos, pues los enunciados para los cuales los hablantes de ambos grupos hacen mayor uso del tono L%, en orden descendente, son los enunciados neutros, los suspendidos, los interrogativos, y finalmente los exclamativos, de los cuales no se encontró ningún caso en la Ciudad de México.

En cuanto a los tonos intermedios, es decir, los prenucleares y nucleares (véase Tablas 3-42 y 3-111, así como 3-43 y 3-112), en la Gráficas 3-14 y 3-15 se puede apreciar su distribución:

Gráfica 3-14: Tonos prenucleares y nucleares

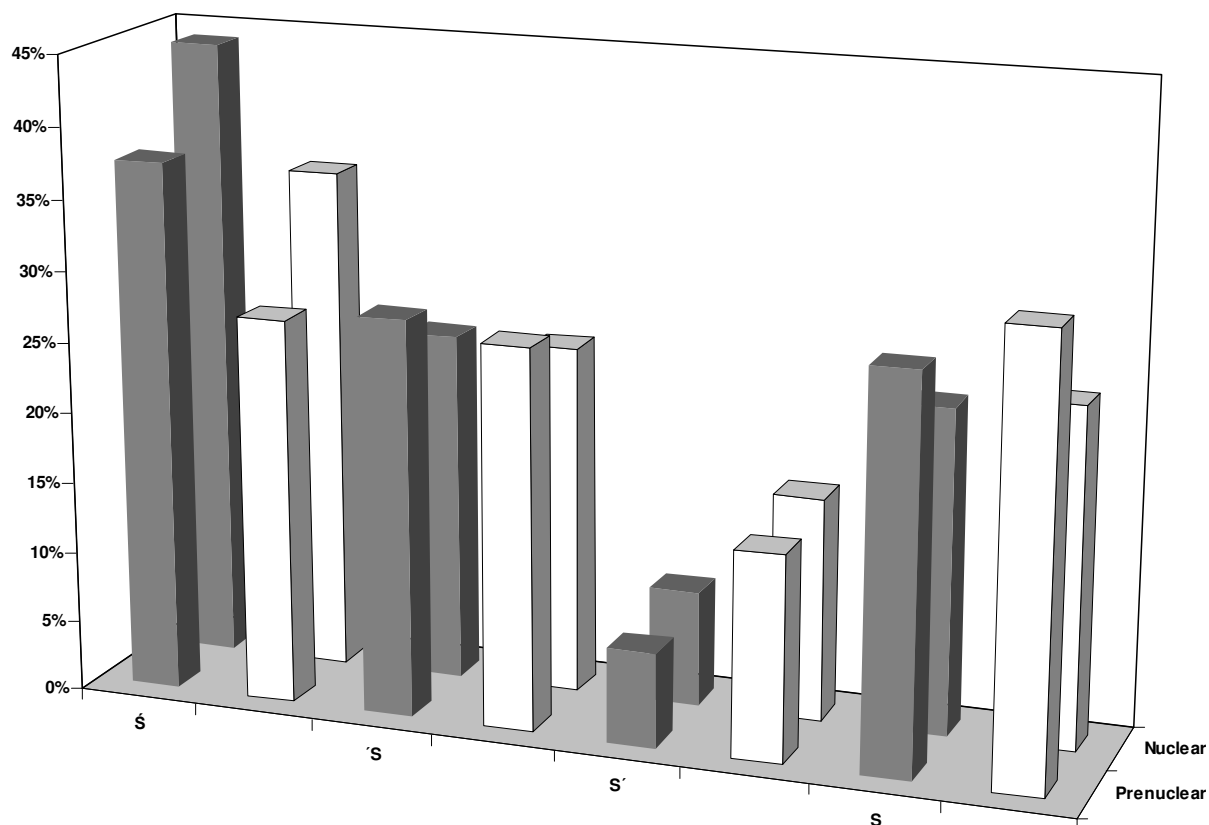


Fuente: Prosodia.xls (Base de datos propia)

A simple vista, se puede ver que, en el caso de los tonos prenucleares, en orden descendente, los tonos más comunes entre los mexicanos son *, L+H*, ¡L*+H, L*+H, ¡L+H*, H+L* y H*, mientras que los tonos de escalonamiento descendente L+!H* y L*+!H son muy poco frecuentes. En las producciones de los madrileños, por otra parte, el orden de importancia es *, L+H*, H+L*, L*+H, ¡L*+H, ¡L+H* y H*, mientras que los tonos de escalonamiento descendente son incluso más bajos que en el caso de los mexicanos. Las principales diferencias en el orden de importancia de los tonos es la alternancia del tono ¡L*+H de los mexicanos y el H+L* de los madrileños.

En la Gráfica 3-15 se presenta los tonos prenucleares agrupados según la alineación del pico frecuencial con respecto a la sílaba tónica, lo que nos permite ver que entre los mexicanos las preferencias son \acute{S} , 'S, S y S', mientras que entre los madrileños son S, \acute{S} , 'S, y S'. En los architonemas \acute{S} y 'S las diferencias a favor de los mexicanos son de 10.27 y 1.11%, mientras que en S y S' las diferencias son a favor de los madrileños, 3.53 y 7.86% respectivamente.

Gráfica 3-15: Tonos prenucleares y nucleares agrupados



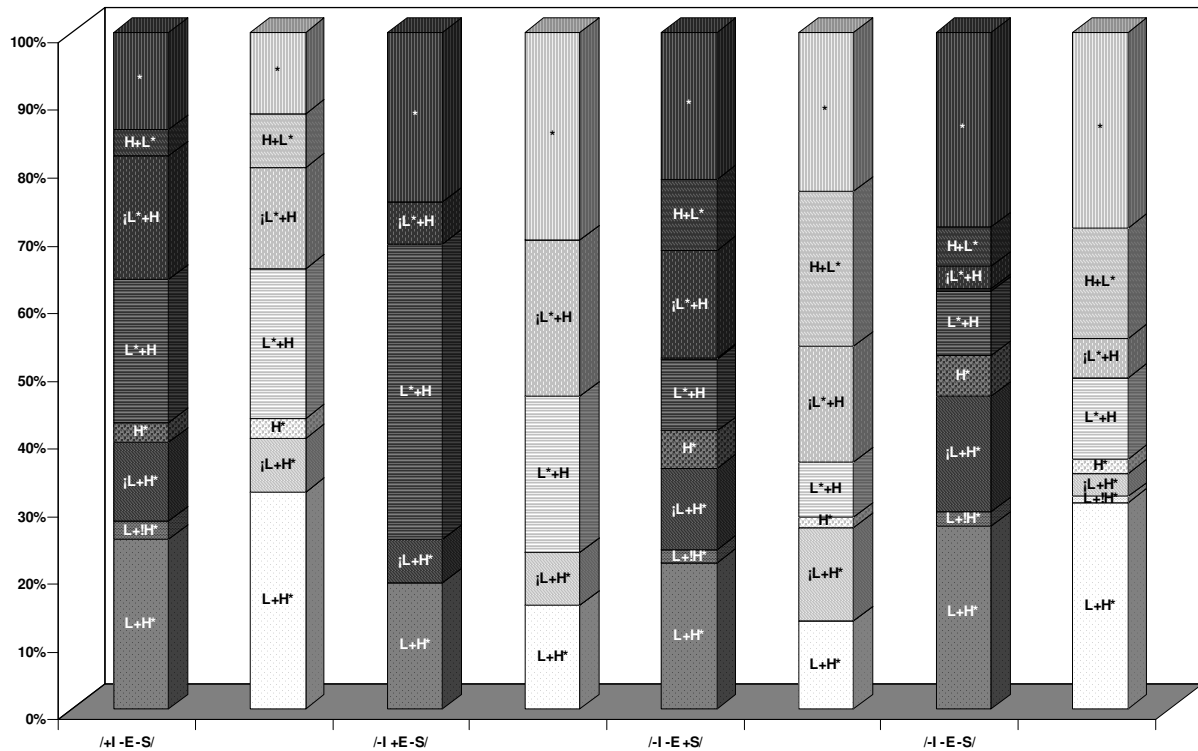
Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, en cuanto a los tonos nucleares, el orden de preferencia de los mexicanos es $L+H^*$, $*$, $\downarrow L+H^*$, $\downarrow L^*+H$, L^*+H , $H+L^*$ y H^* (los tonos con escalonamiento descendente son irrelevantes en todos los casos), mientras que los tonos preferidos por los madrileños son $L+H^*$, $*$, $H+L^*$, L^*+H , $\downarrow L^*+H$, $\downarrow L+H^*$ y H^* . Entre estas dos secuencias la principal diferencia es que los tonos $\downarrow L+H^*$, $\downarrow L^*+H$, L^*+H y $H+L^*$ se utilizan exactamente en el orden inverso. Con respecto a los tonos prenucleares, en los dos casos, la diferencia más notoria es la inversión de los tonos $L+H^*$ y $*$.

Sin embargo, en la Gráfica 3-15 se ve con mayor claridad que, tanto entre los mexicanos como entre los madrileños, las preferencias son \acute{S} , \acute{S} , S y S' , aunque las diferencias a favor de los mexicanos en los architonemas \acute{S} y \acute{S} son de 8.55 y 0.08%, mientras que en S y S' las diferencias son a favor de los madrileños, 1.01 y 7.63% respectivamente. En otras palabras, tanto en el uso de tonos prenucleares como nucleares, los mexicanos se distinguen por un mayor uso de los tonos \acute{S} y los madrileños usan comparativamente más los tonos S' .

Ahora, si dejamos de lado los tonos prenucleares, en la Gráfica 3-16 se ilustra los valores porcentuales de los tonos nucleares filtrados según el tipo de enunciado (véase Tablas 3-48 y 3-123). Las barras de tonos grises más oscuros corresponden a los resultados de los hablantes de la Ciudad de México y las barras de tonos más claros a los enunciados de Madrid.

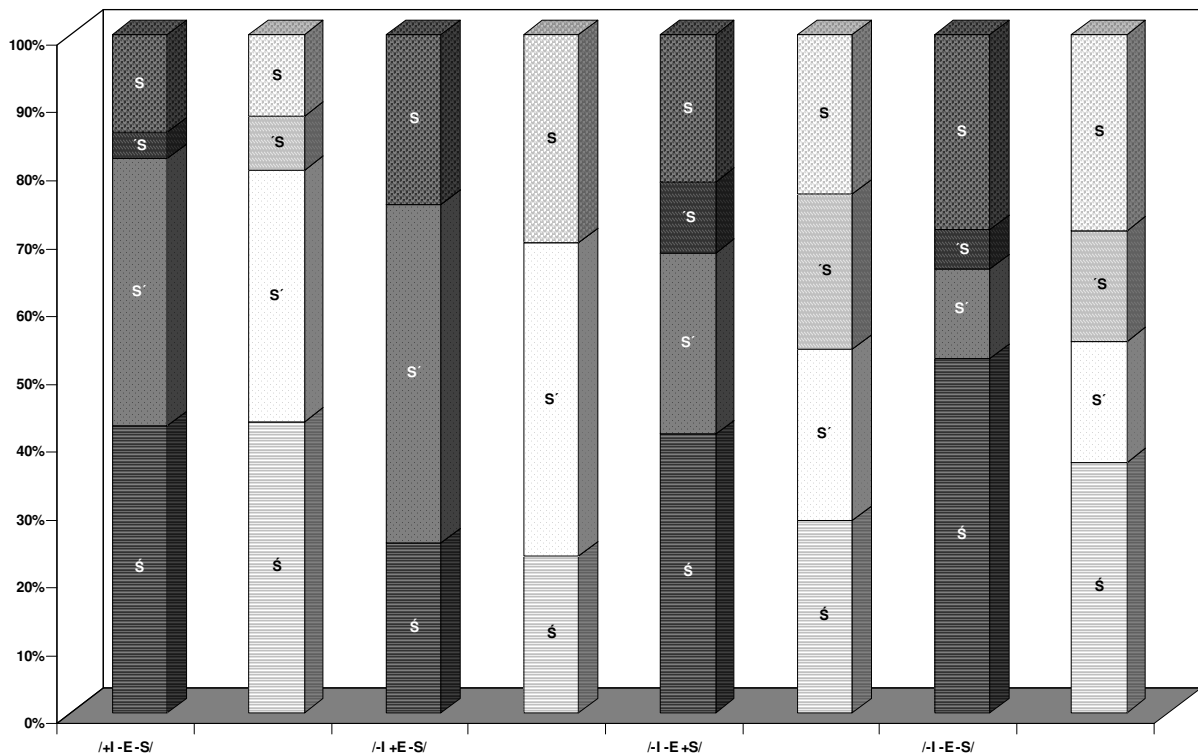
Gráfica 3-16: Tonos nucleares por tipo de enunciado



Fuente: Prosodia.xls (Base de datos propia)

Paralelamente, en la Gráfica 3-17 se representa la suma de los tonos individuales en los cuatro architonemas, según la alineación del pico frecuencial con la sílaba tónica, para permitir una comparación de los tonos como categorías fonológicas independientes del contexto.

Gráfica 3-17: Tonos nucleares agrupados por tipo de enunciado



Fuente: Prosodia.xls (Base de datos propia)

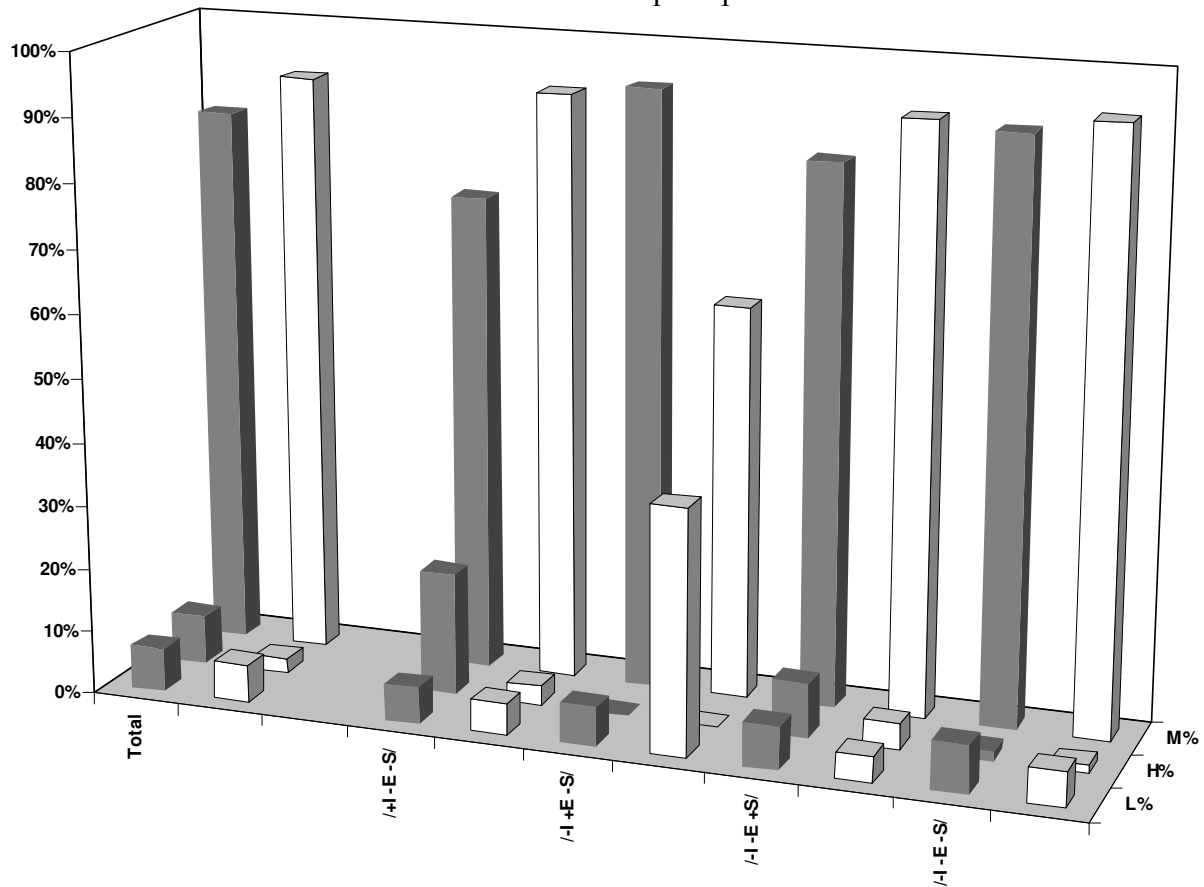
De aquí en adelante, cuando se haga referencia a los distintos tonos nucleares, se presentarán dos gráficas, una con los tonos individuales y otra con los tonos agrupados. Los tonos individuales deben ser considerados ‘alotonos’ o variantes en distribución complementaria de un ‘architonema’, pues, por ejemplo, la diferencia entre L+H*, ¡L+H*, L+!H* y H* es únicamente el contexto en el que aparecen (pico después de un valle, pico después de un valle más alto que el valle anterior, pico más bajo que el anterior después de un valle y pico a la misma altura que un pico anterior, respectivamente), mientras que todas comparten la característica de que el pico frecuencial coincide con la sílaba tónica. Por esta razón, las diferencias entre el uso de los tonos por mexicanos y madrileños se hará únicamente tomando en cuenta los cuatro architonemas: \acute{S} , ‘S, S’ y S.

En el caso de los enunciados interrogativos, las diferencias en el uso de los tonos son mínimas. Por ejemplo, la diferencia entre el uso de tonos nucleares del grupo \acute{S} es de sólo 0.70% a favor de los madrileños, quienes también hacen un uso 4.15% mayor de tonos S’. Los mexicanos, por su parte, hacen mayor uso de tonos de los tipos ‘S y S, en ambos casos con una diferencia de 2.42%. En los enunciados exclamativos las similitudes entre las dos variedades también son muchas, pues las diferencias a favor de los mexicanos en los tonos nucleares de los tipos \acute{S} y ‘S son de sólo 1.93 y 3.84% respectivamente. Como en ninguno de los dos casos se produjo tonos del tipo S’, la diferencia porcentual entre el uso de S entre las dos variedades es igual a la suma de los tipos \acute{S} y ‘S: 5.77%. En los enunciados suspendidos, los mexicanos hacen un uso 12.61 y 1.45% mayor de los tipos de tonos nucleares \acute{S} y ‘S, respectivamente, mientras que los madrileños hacen mayor uso de S’, con una diferencia de 12.34%, y S, por sólo 1.74%. Finalmente, en cuanto a los enunciados neutros, los mexicanos muestran una gran preferencia por los tonos nucleares del tipo \acute{S} , con una diferencia de 15.37%, mientras que los madrileños usan más los tipos S’, ‘S y S, con 10.34, 4.84 y 0.18% más que los mexicanos.

Para concluir con la presentación de los tonos aislados más comunes en las variedades mexicana y madrileña, en la Gráfica 3-18 podemos observar la distribución de los tonos finales L%, M% y H% por grupos de hablantes y tipos de enunciado (véase Tablas 3-44 y 3-50, así como 3-113 y 3-119).

En todos los casos, los tonos de frontera final más comunes son los M%, aunque su distribución en las dos variedades de español son muy variadas. Por ejemplo, en total, los madrileños utilizan este tono 6.39% más que los mexicanos, y también en los enunciados interrogativos, 17%, en los suspendidos, 7.30%, y en los neutros, 2.47%, pero los mexicanos hacen un uso muy superior del tono en enunciados exclamativos, en comparación, 32.21%.

Gráfica 3-18: Tonos finales por tipo de enunciado



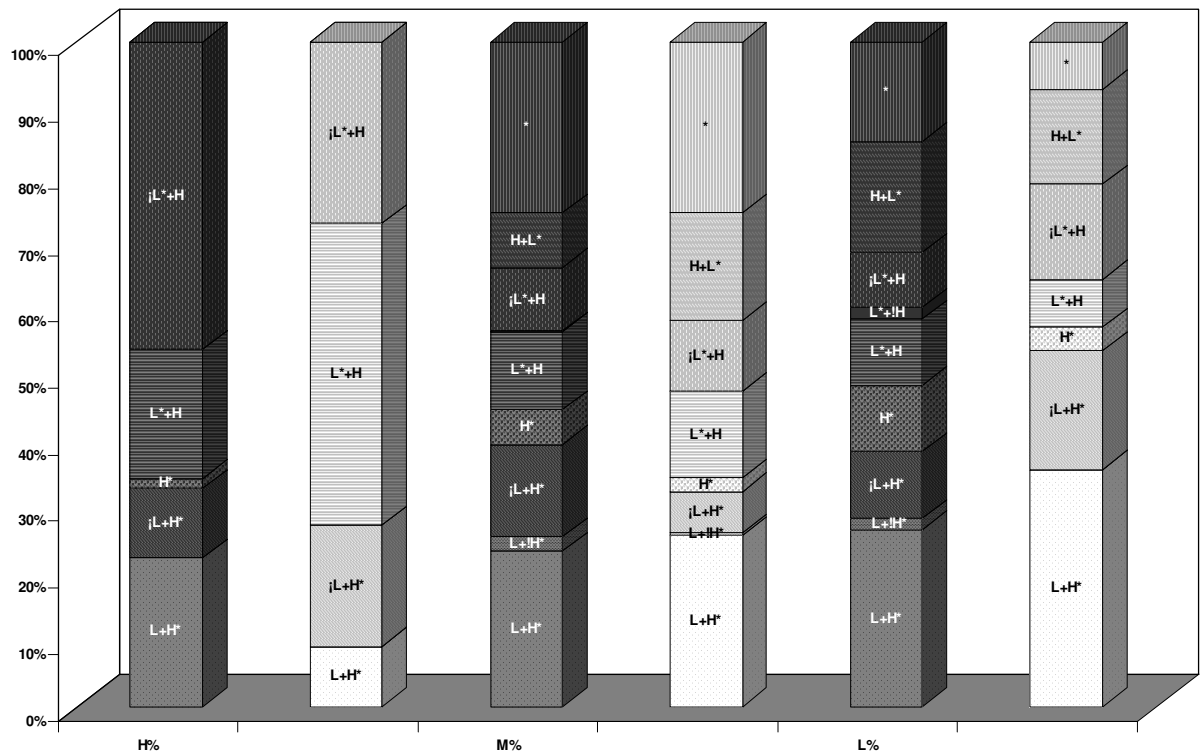
Fuente: Prosodia.xls (Base de datos propia)

Por el contrario, en cuanto al uso del tono L%, excepto en los enunciados exclamativos, con una diferencia a favor de los madrileños de 32.21%, los mexicanos lo utilizan más frecuentemente en todos los demás contextos: en total 1.01%, en los enunciados suspendidos 2.74%, en los neutros 2.13% y en los interrogativos sólo 0.77% más. También el tono H% es utilizado con mayor frecuencia por los mexicanos: en total 5.38% más que los madrileños, en los enunciados interrogativos 16.23%, en los suspendidos 4.55% y en los neutros sólo 0.33%. Ahora, analizando la combinación de los tonos nucleares con los tonos finales, obtenemos los resultados que se muestran en las Gráficas 3-19 y 3-20 (véase Tablas 3-52 y 3-121).

Como se había mencionado anteriormente, en la Gráfica 3-19 se puede ver la distribución de todos los tonos nucleares individuales, y la presentación en forma de columnas porcentuales permite una comparación detallada de la presencia de tal o cual tono en combinación con cada uno de los tres tonos finales posibles. Sin embargo, una descripción de este tipo complicaría el análisis innecesariamente.

En la Gráfica 3-20 vemos, entonces, que en ambas variedades los tonos nucleares combinados con el tono final H% son únicamente de dos tipos: \acute{S} y \acute{S} . Por lo tanto, la diferencia de 7.06% entre las dos columnas es a favor de los mexicanos en \acute{S} y de los madrileños en \acute{S} .

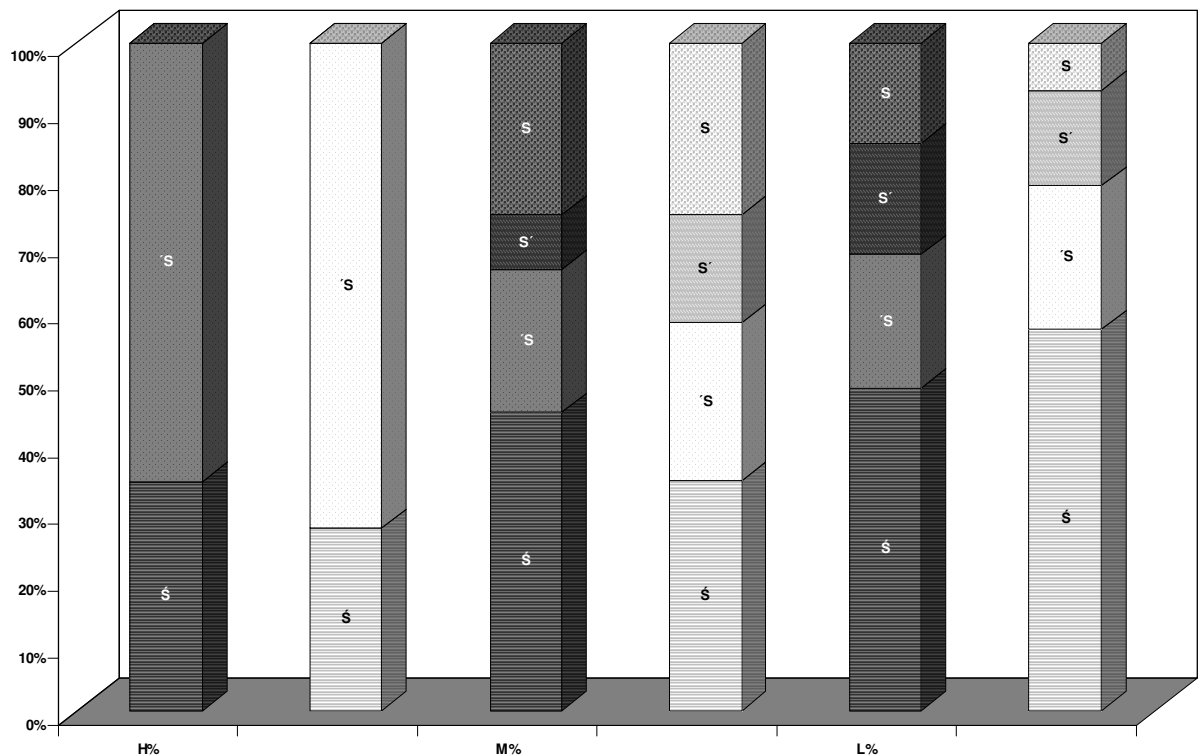
Gráfica 3-19: Tonos nucleares por cada tono final



Fuente: Prosodia.xls (Base de datos propia)

En el caso de los tonos nucleares combinados con M%, los mexicanos utilizan 10.28% más de tonos del tipo \acute{S} , mientras que los madrileños producen 2.22% más de tonos del tipo 'S y 8.07% del tipo S. La diferencia en el uso de tonos S es casi nula, 0.02%.

Gráfica 3-20: Tonos nucleares agrupados por cada tono final

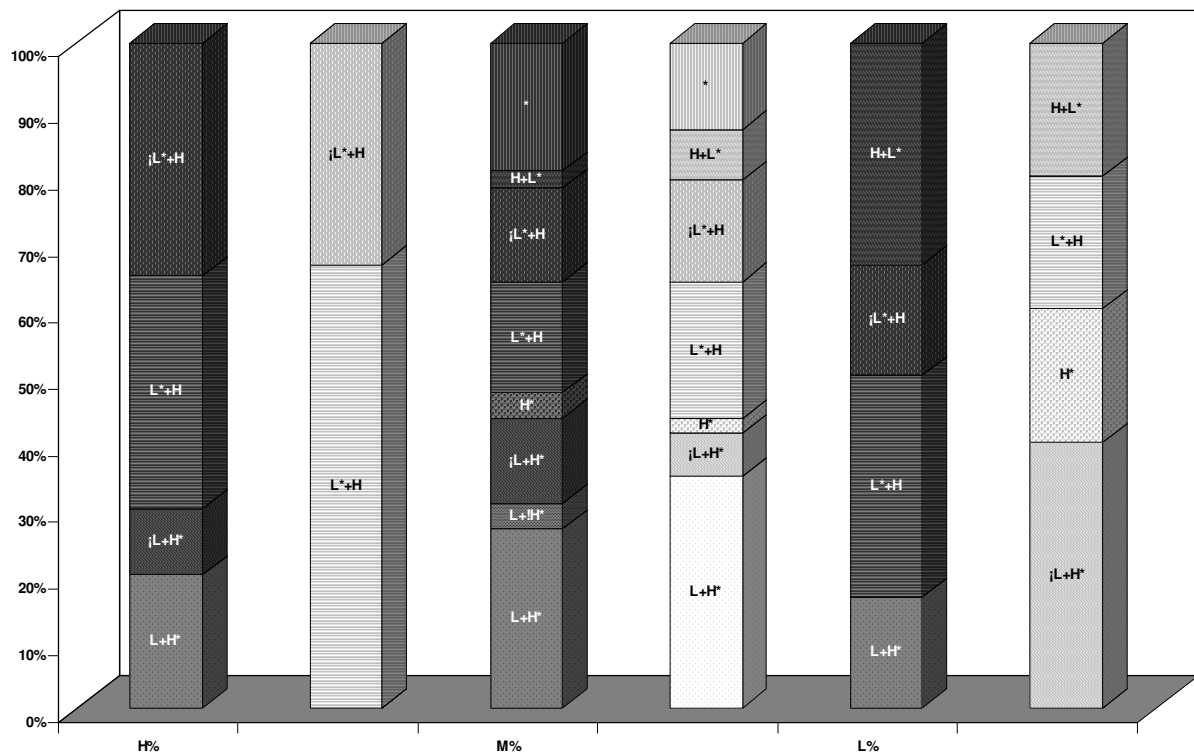


Fuente: Prosodia.xls (Base de datos propia)

Por otra parte, los tonos nucleares combinados con L% también presentan tonos de los cuatro grupos. De hecho, la distribución de tonos que hacen los mexicanos en este caso es muy similar a la distribución con el tono M%, al menos en los dos grupos de mayor presencia (Ś y 'S). Sin embargo, son precisamente estas dos combinaciones de tonos nucleares con finales los que aparecen con más frecuencia en Madrid, con una diferencia de 8.80% para Ś y 1.43% para 'S. Los mexicanos sobresalen en el uso de las otras dos combinaciones, S' con 2.38% y S con una diferencia de 7.86%.

Ahora, si tomamos la información sobre los tonos nucleares combinados con tonos finales únicamente en el caso de los enunciados interrogativos (véase Tablas 3-54 y 3-125), obtenemos la distribución presentada en las Gráficas 3-21 y 3-22.

Gráfica 3-21: Tonos nucleares por cada tono final en enunciados interrogativos



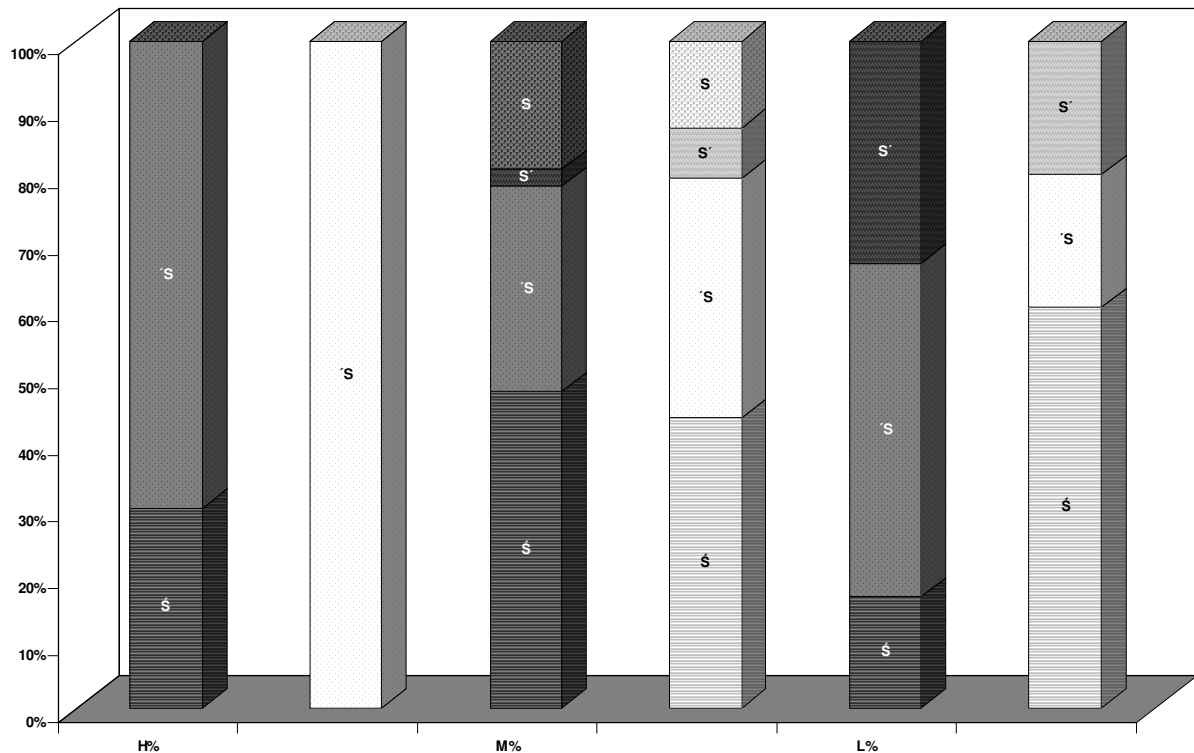
Fuente: Prosodia.xls (Base de datos propia)

En cuanto a los tonos nucleares combinados con el tono final H%, en la Gráfica 3-22 se puede observar una clara diferencia entre las producciones mexicanas y las madrileñas, pues el 100% de los enunciados fueron producidos con tonos del grupo 'S, mientras que los mexicanos produjeron 30% de enunciados con tonos del grupo Ś.

Los tonos nucleares combinados con M% pertenecen a los cuatro distintos grupos y las diferencias entre las dos variedades de español no fueron tan pronunciadas: 6.19 y 3.97% a favor de los mexicanos en los grupos S y Ś respectivamente, mientras que los madrileños utilizaron 5.10% más de tonos tipo 'Sy 5.05% de tonos S'.

El tono L% se combinó sólo con tres grupos de tonos en las dos variedades de español, pues no se produjo ningún caso de tonos S. Además, en este caso, las diferencias entre la preferencia por los distintos grupos fueron muy marcadas: 43.33% a favor de los madrileños en cuanto al uso de tonos del grupo \acute{S} , mientras que los mexicanos alcanzaron diferencias de 30 y 13.33% en el uso de tonos de los tipos 'S y S', respectivamente.

Gráfica 3-22: Tonos nucleares agrupados por cada tono final en enunciados interrogativos

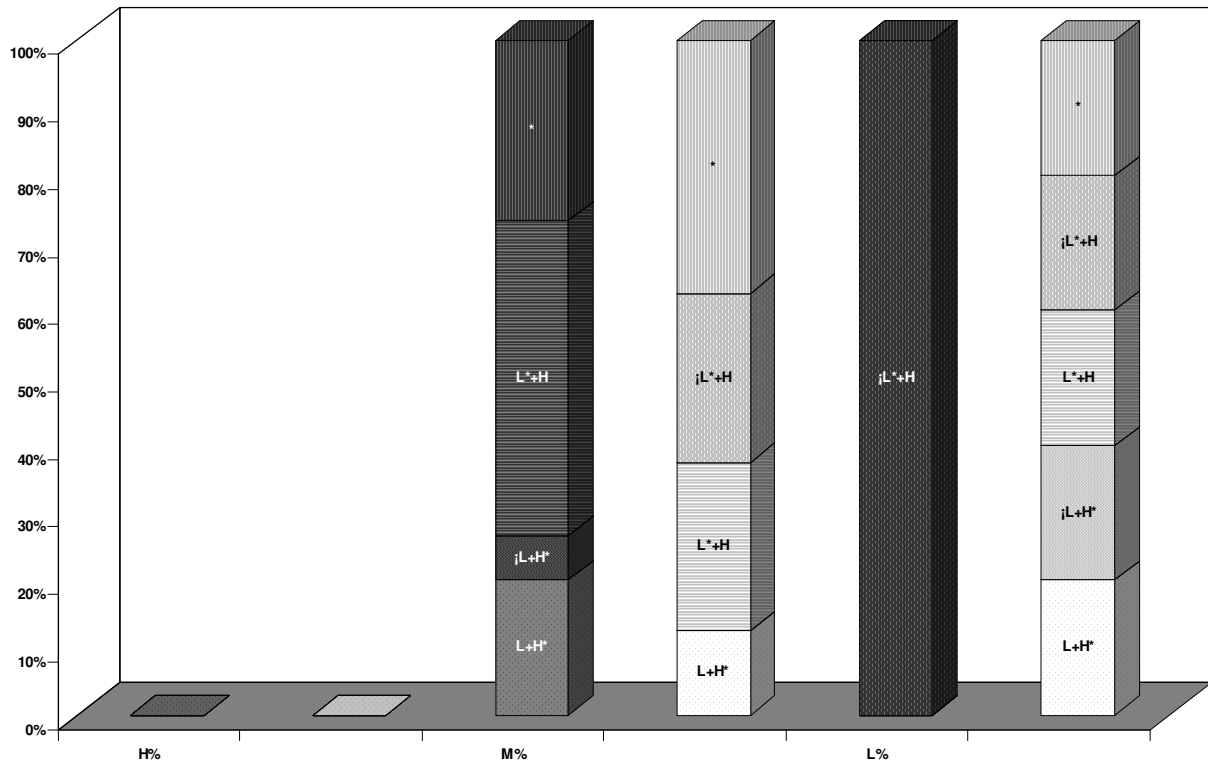


Fuente: Prosodia.xls (Base de datos propia)

Por lo tanto, en los contornos entonativos ascendentes (H%) de los enunciados interrogativos parece favorecerse el uso de tonos del tipo 'S, sobre todo por parte de los madrileños. En los contornos suspendidos (M%) hay un mayor equilibrio entre el uso de tonos de los tipos \acute{S} y 'S, pues se observan preferencias muy similares entre los dos grupos de hablantes. En contornos descendentes (L%), aparecen también los tonos del tipo S' en el habla de ambos grupos, aunque la proporción de \acute{S} aumenta entre los madrileños.

En el caso de los enunciados exclamativos, como puede observarse en las Gráficas 3-23 y 3-24 (véase Tablas 3-56 y 3-127), posiblemente debido a su bajo nivel de presencia dentro del *corpus*, no se dio ningún caso de entonación ascendente en ninguno de los dos grupos de hablantes y, aparte, en la entonación suspendida y descendente tampoco se dio ningún caso de tonos del tipo S'. Por lo tanto, entre los tonos nucleares combinados con M%, los mexicanos producen 14.17% más de tonos \acute{S} , mientras que esta diferencia se reparte en 10.83 y 3.33% a favor de los madrileños en los tonos 'S y S.

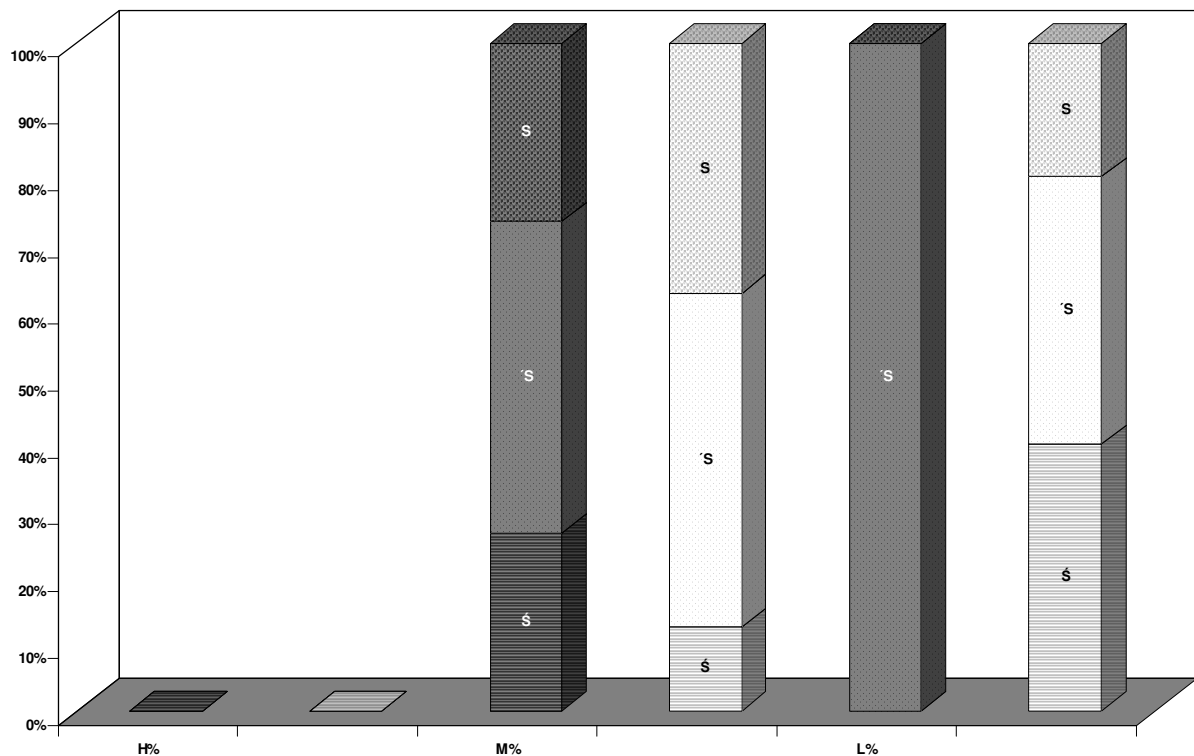
Gráfica 3-23: Tonos nucleares por cada tono final en enunciados exclamativos



Fuente: Prosodia.xls (Base de datos propia)

En cuanto a los tonos nucleares combinados con L%, los mexicanos producen en todos los casos tonos del tipo 'S, mientras que este tipo representa sólo el 40% entre los madrileños, los restantes 40 y 20% corresponden a los tonos Ṡ y S, respectivamente.

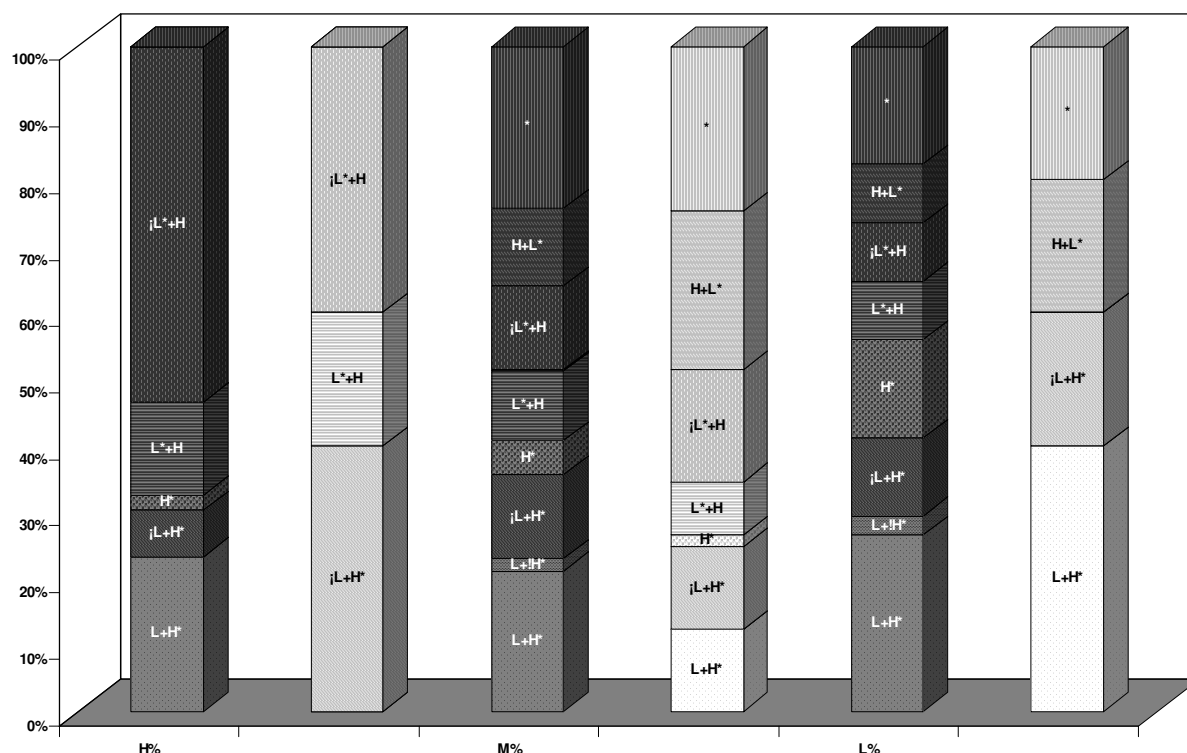
Gráfica 3-24: Tonos nucleares agrupados por cada tono final en enunciados exclamativos



Fuente: Prosodia.xls (Base de datos propia)

Entonces, las tendencias en los enunciados exclamativos son que los casos de entonación ascendente son pocos o nulos, mientras que en la entonación suspendida y descendente hay una preferencia general por los tonos 'S, \acute{S} y S, excepto en el caso de los enunciados con contornos descendentes de los mexicanos, en los que se utiliza únicamente tonos del tipo 'S. En las Gráficas 3-25 y 3-26 (véase Tablas 3-58 y 3-129) puede apreciarse los resultados del filtramiento de las combinaciones de tonos nucleares y finales según si se daban en enunciados suspendidos.

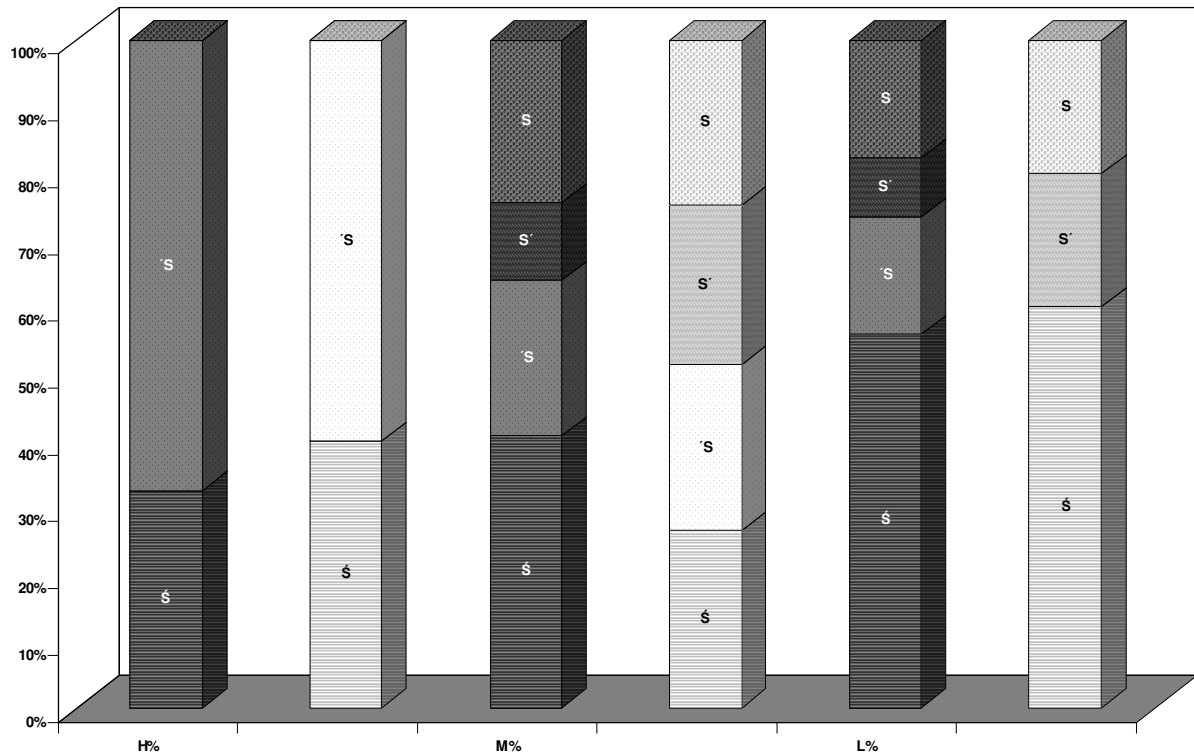
Gráfica 3-25: Tonos nucleares por cada tono final en enunciados suspendidos



Fuente: Prosodia.xls (Base de datos propia)

En los enunciados con entonación ascendente únicamente se dieron, en los dos grupos de hablantes, dos tipos de tonos, \acute{S} y 'S. La diferencia de 7.44% entre las dos columnas se da en la preferencia del tipo \acute{S} por parte de los madrileños y del tipo 'Sen el caso de los mexicanos. En cuanto a los enunciados con entonación suspendida, las principales diferencias se dan entre los tipos \acute{S} , 14.21% a favor de los mexicanos, y S', 12.28% a favor de los madrileños. Las diferencias en los tipos 'S y S son extremadamente bajas, 1.54% y 0.37% a favor de los madrileños. Por último, en la entonación descendente, sólo en el caso de los mexicanos se utiliza tonos del tipo 'S, lo que representa 17.64% del total de la Ciudad de México. Esta diferencia se proyecta en porcentajes mayores de los tres tipos de tonos restantes a favor de los madrileños, 11.18% más de S', 4.12% de \acute{S} , y 2.35% de S. Es decir, en general se da preferencia a los tonos del tipo \acute{S} , pero en los contornos ascendentes se prefiere tonos 'S.

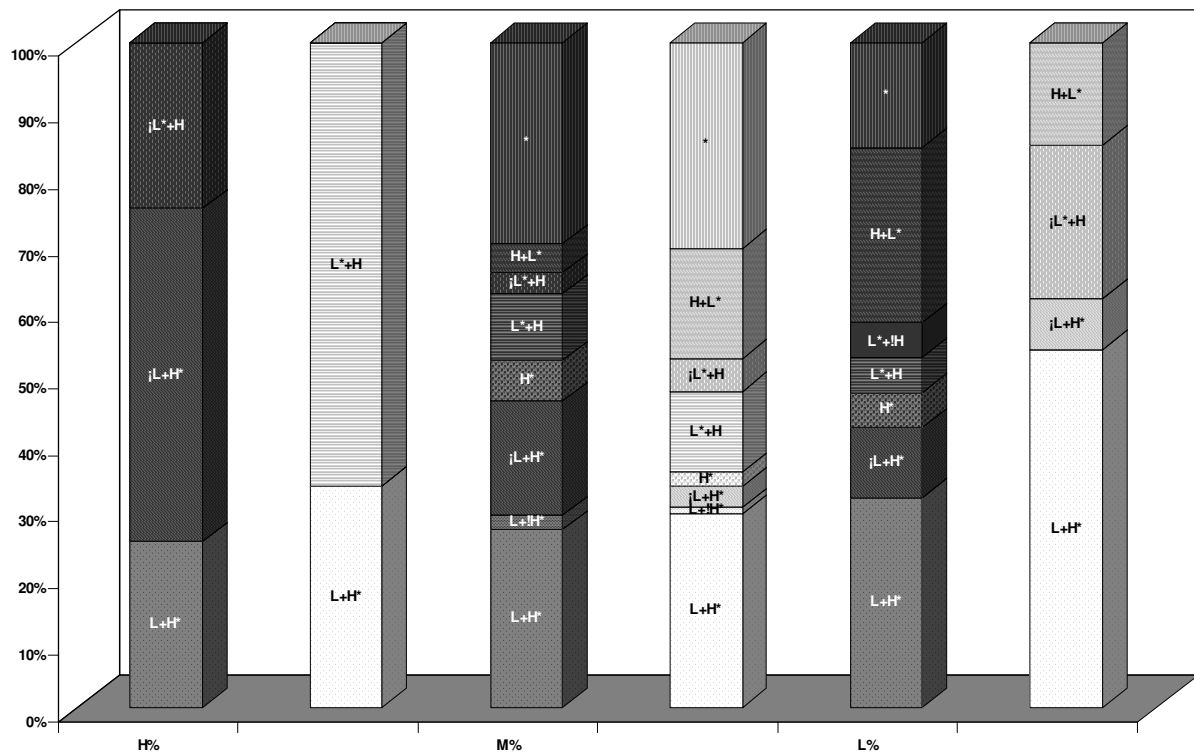
Gráfica 3-26: Tonos nucleares agrupados por cada tono final en enunciados suspendidos



Fuente: Prosodia.xls (Base de datos propia)

Para terminar con la entonación de los tipos de enunciados, en las Gráficas 3-27 y 3-28 se presenta las series de valores correspondientes a los enunciados neutros (véase Tablas 3-60 y 3-131):

Gráfica 3-27: Tonos nucleares por cada tono final en enunciados neutros



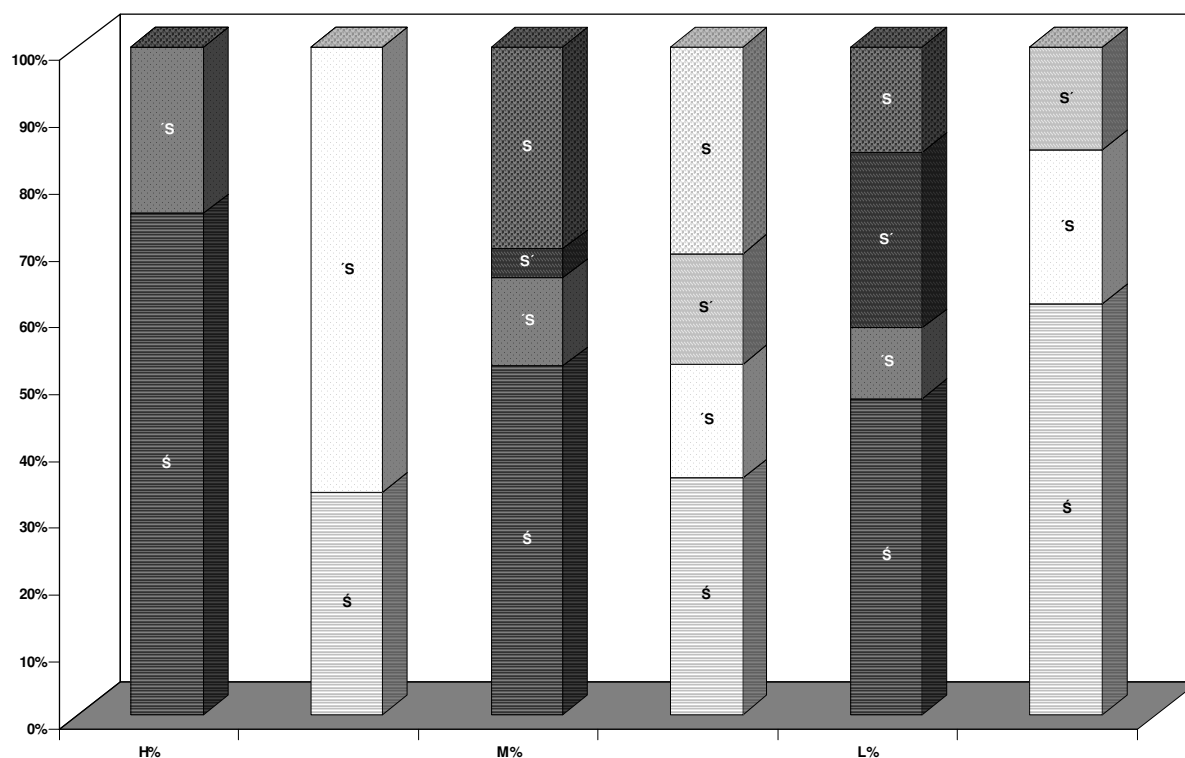
Fuente: Prosodia.xls (Base de datos propia)

En los enunciados neutros con entonación ascendente no hubo casos, en ninguno de los dos grupos de hablantes, en los que se haya utilizado tonos de los tipos S' y S. Por lo tanto, la diferencia de 45% entre las dos columnas se traduce como una preferencia de los mexicanos por los tonos del tipo \acute{S} , y una preferencia por tonos del tipo S' por parte de los madrileños.

En cuanto a la entonación suspendida, los mexicanos muestran una clara preferencia por tonos del tipo \acute{S} , 16.76% mayor que los madrileños. Por lo tanto, el resto de valores porcentuales son mayores en las preferencias de los madrileños: 12.20% en tonos del tipo S', 3.88% en tonos 'S y tan sólo 0.68% en tonos del tipo S.

En el caso de la entonación descendente, sólo los mexicanos utilizan tonos del tipo S, 15.79%. Además, este grupo de hablantes hace un uso 10.94% mayor de tonos del tipo S'. Es por eso que los madrileños, por el contrario, muestran valores más elevados de tonos de los tipos \acute{S} y 'S, de 14.17 y 12.56% respectivamente.

Gráfica 3-28: Tonos nucleares agrupados por cada tono final en enunciados neutros

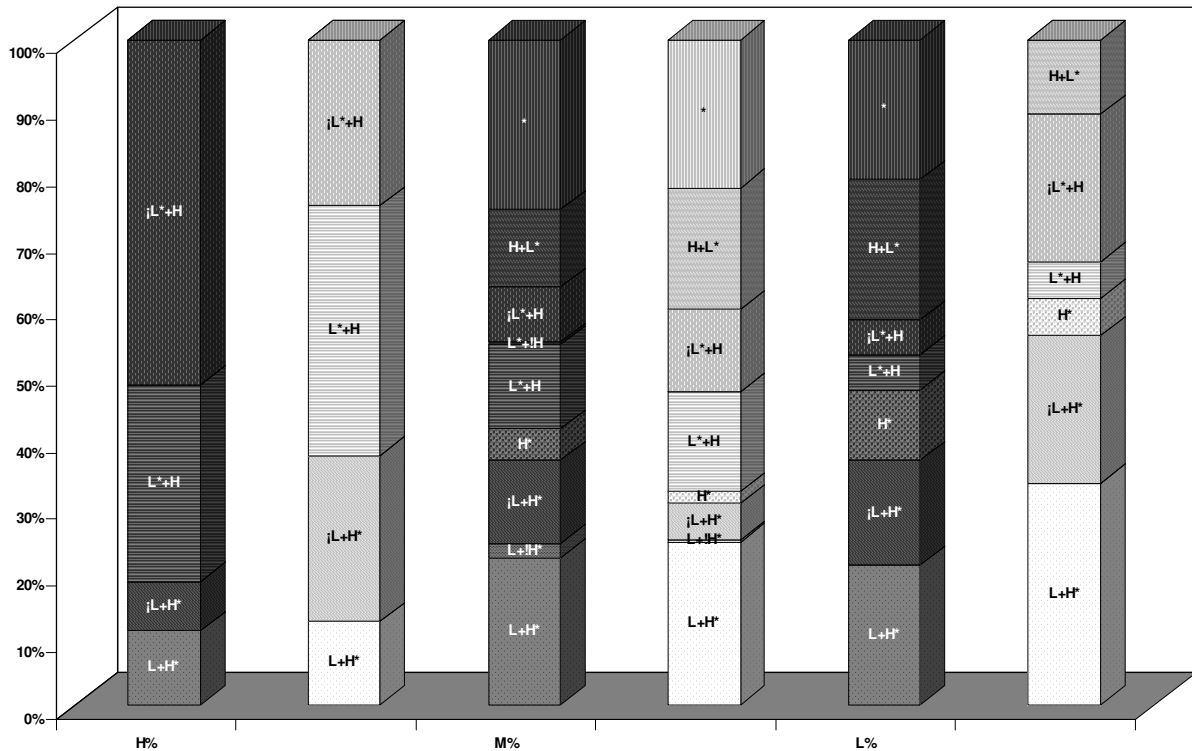


Fuente: Prosodia.xls (Base de datos propia)

Resumiendo, en los enunciados neutros hay una clara preferencia por los tonos del tipo \acute{S} , excepto en los casos de entonación ascendente en las producciones de Madrid, donde el tipo de tono predominante es 'S.

Por otra parte, en las Gráficas 3-29 y 3-30 se presentan las combinaciones de tonos nucleares y finales en el caso específico de las mujeres (véase Tablas 3-62 y 3-131), y en las Gráficas 3-31 y 3-32 la distribución de tonos en el caso de los hombres (véase Tablas 3-64 y 3-133).

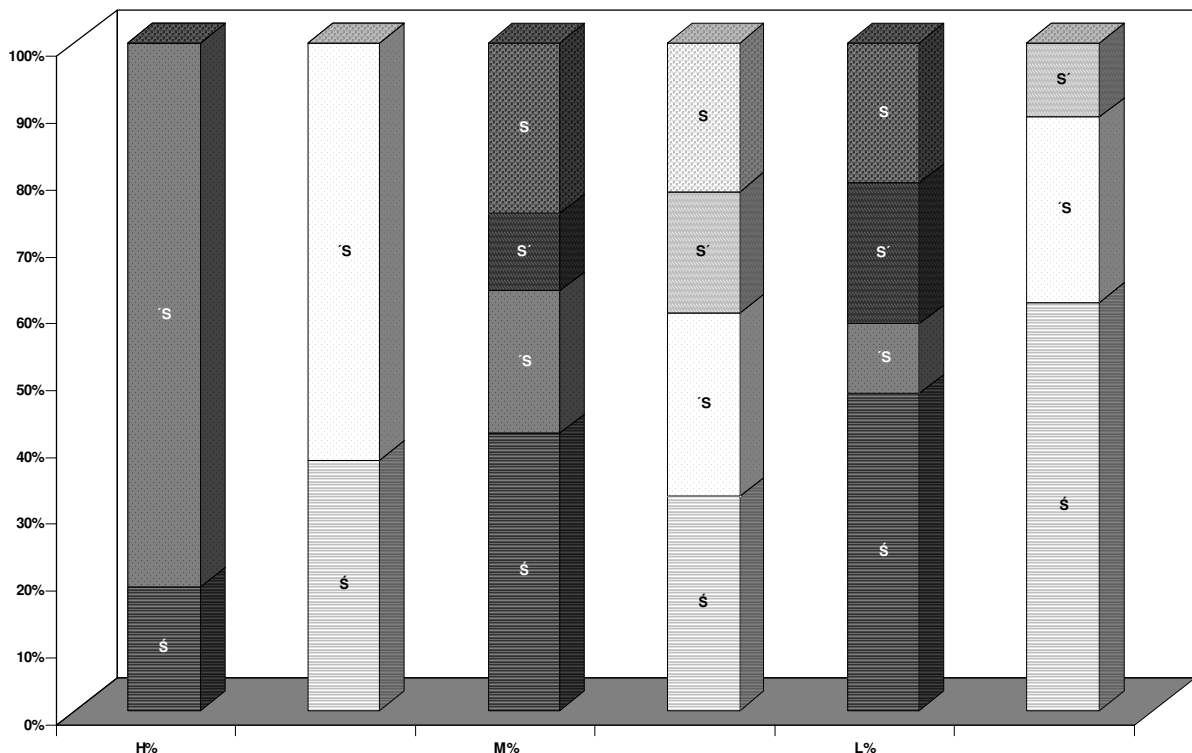
Gráfica 3-29: Tonos nucleares por cada tono final en enunciados producidos por mujeres



Fuente: Prosodia.xls (Base de datos propia)

También en estas Gráficas se presenta primeramente todos los valores arrojados por los tonos individuales y en una segunda Gráfica se muestran los tonos agrupados por tipos:

Gráfica 3-30: Tonos nucleares agrupados por cada tono final en enunciados producidos por mujeres

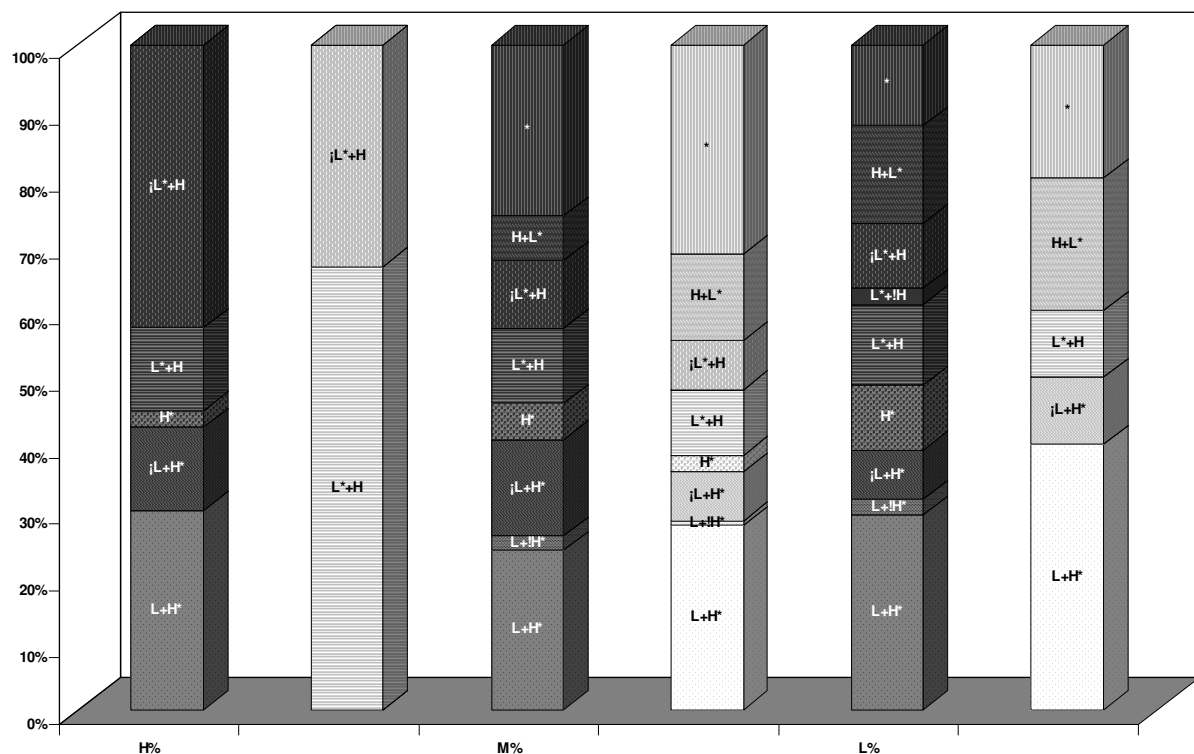


Fuente: Prosodia.xls (Base de datos propia)

Así, en el caso de las mujeres, de todos los enunciados producidos con entonación ascendente en ambos grupos de hablantes, solamente se dan casos de tonos de los tipos \acute{S} y \acute{S} , aunque en distinta proporción. La diferencia de 18.98% entre las dos columnas representa una mayoría de \acute{S} en las producciones de las madrileñas, y una mayoría de \acute{S} en las de las mexicanas. En cuanto a la entonación suspendida, se observa una mayor preferencia de las mexicanas por tonos de los tipos \acute{S} y S , con una diferencia de 9.45 y 3.28%, mientras que las madrileñas muestran una mayor predilección por tonos de los tipos S' y \acute{S} , con diferencias de 6.56 y 6.16% respectivamente. Por otra parte, a diferencia de las madrileñas, las mexicanas utilizan los tonos del tipo S , lo que representa un 13.74% del total. Además, las mexicanas usan 9.94% más de tonos del tipo S' que las madrileñas, quienes, por el contrario, hacen mayor uso de tonos de los tipos \acute{S} y \acute{S} , con diferencias de 17.26 y 13.74% respectivamente.

En el caso de los hombres, además de que no existen los tonos S' y S en la entonación ascendente, los madrileños tampoco hacen usos de tonos del tipo \acute{S} en conjunto con el tono final $H\%$; a diferencia de los mexicanos, que los usan en 45% de los casos.

Gráfica 3-31: Tonos nucleares por cada tono final en enunciados producidos por hombres

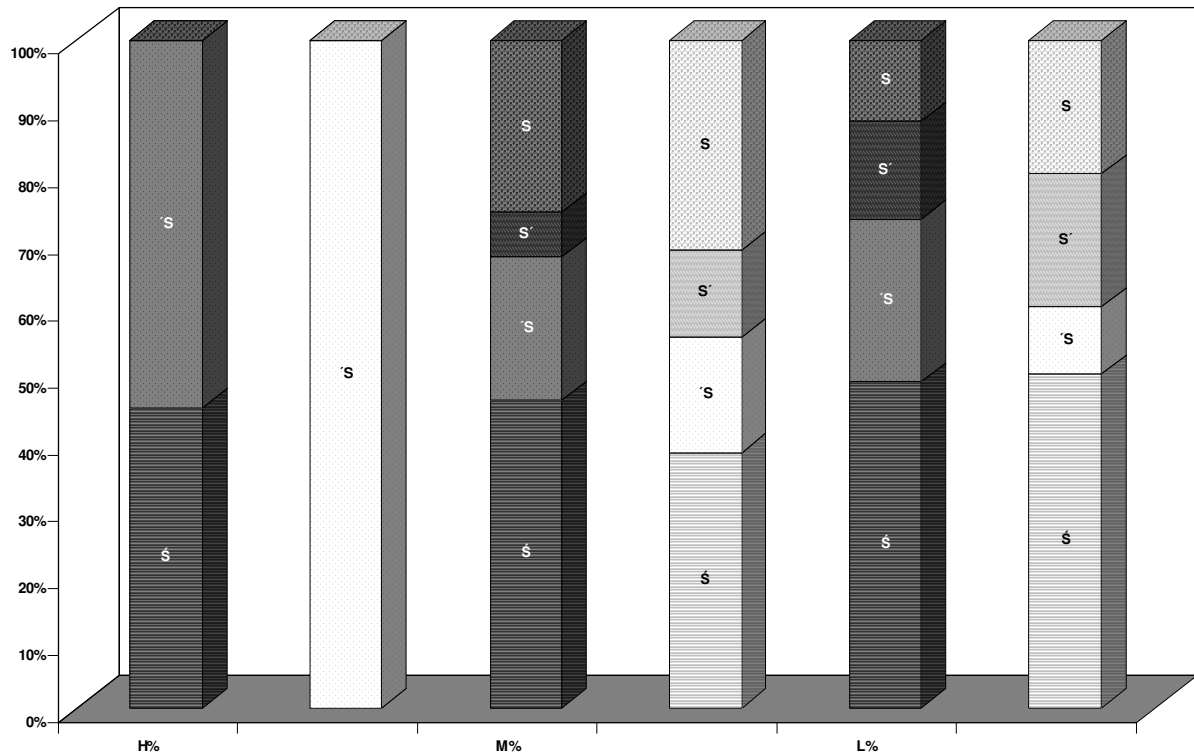


Fuente: Prosodia.xls (Base de datos propia)

En cuanto a la entonación suspendida, los mexicanos hacen un mayor uso de los tonos de los tipos \acute{S} y \acute{S} , con diferencias de 7.81 y 4.20%, mientras que los madrileños usan más frecuentemente los tonos S' y S , con diferencias de 6.32 y 5.70% respectivamente.

Finalmente, la entonación descendente de los mexicanos se caracteriza por su preferencia por tonos del tipo 'S, con una diferencia de 14.40%, mientras que los madrileños recuperan esta diferencia de porcentaje haciendo, sobre todo, un mayor uso de tonos de los tipos S y S', con 7.80 y 5.37%, pues la diferencia en el uso de tonos del tipo \acute{S} es muy pequeña: 1.21% a favor de los madrileños.

Gráfica 3-32: Tonos nucleares agrupados por cada tono final en enunciados producidos por hombres



Fuente: Prosodia.xls (Base de datos propia)

En general, se puede observar entre los valores de las producciones de mujeres y hombres que en la entonación ascendente hay una clara preferencia de 'S sobre \acute{S} , caso contrario a lo que sucede en la entonación suspendida y descendente. En el caso específico de la entonación suspendida, para la cual se contó con el mayor número de ejemplos, si se comparan las columnas de hombres y mujeres se notará que las proporciones son muy similares. Tal vez cuando se amplíe la base de datos y se disponga de un mayor número de enunciados con tonos finales H% y L% los valores entre mujeres y hombres se asemejen más.

Entonces, en este apartado sobre la entonación, a pesar de la influencia que puede tener la diferencia natural entre el número de observaciones de cada tipo, se ha podido observar ciertas tendencias: 1) los mexicanos inician una mayor cantidad de enunciados con tonos altos que los madrileños; 2) los tonos nucleares (y prenucleares), a pesar de sus variadas formas dependientes del contexto, pueden agruparse en sólo cuatro categorías, dependiendo de si el pico frecuencial recae en la sílaba tónica, antes, después o es muy poco perceptible; 3) según

esta clasificación, el tono prenuclear más común en Madrid no sobrepasa el 5% de variación intersilábica, que corresponde al tono *, tipo S, mientras que en la Ciudad de México los tonos prenucleares más comunes son del tipo \acute{S} ; 4) los tonos nucleares más comunes en ambas variedades lingüísticas son del tipo \acute{S} ; 5) la mayoría de tonos finales empleados más frecuentemente por los madrileños y mexicanos son del tipo M%, es decir, entonación suspendida, aunque los mexicanos hacen un uso ligeramente mayor de una variación de F0 mayor a $\pm 25\%$; 6) según los valores de la base de datos Prosodia.xls, en ningún tipo de enunciado se dan casos de entonación ascendente después de tonos de los tipos S' y S; 7) en los distintos tipos de enunciados las cuatro (o dos) categorías se combinan de las maneras más variadas, sin embargo, en todos los casos hay tipos de tonos que contrastan y otros que permanecen relativamente estables; y 8) la mayor frecuencia de combinaciones contrastantes de tipos de tonos nucleares con los tres tipos de tono final son los que permiten a los hablantes identificar la melodía característica de las distintas variedades regionales de una lengua, en nuestro caso, del español de Madrid y del de la Ciudad de México.

3.4.6 Ritmo

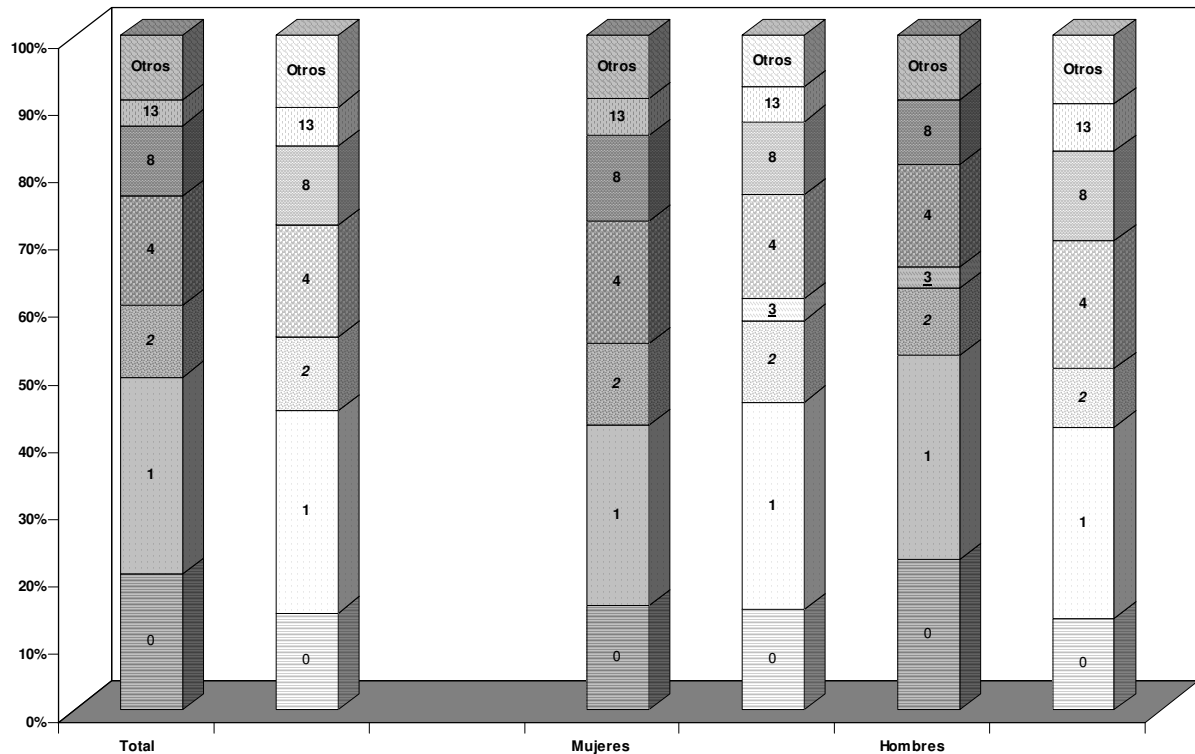
El ritmo en el habla espontánea es posiblemente el fenómeno prosódico más difícil de percibir y controlar conscientemente. Los hablantes normalmente no se dan cuenta de que ellos mismos producen secuencias rítmicas cuando hablan rápido, cuando recitan alguna fórmula humorística o dramática, cuando imitan a algún personaje o cuando colorean alguna frase con un matiz expresivo o retórico, en fin, cuando juegan a ser poetas.

Sin embargo, independientemente del grado de escolaridad, a la mayoría de los hispanohablantes se les dificulta la identificación de sílabas tónicas y átonas y, por lo tanto, la localización de estructuras rítmicas en el flujo del habla.

Es por esta razón que el ritmo cumple un papel secundario a la hora de la codificación de mensajes. Si se le compara con el control activo de otros recursos como la duración silábica, la intensidad de la voz, el registro de voz y la melodía de la voz, nos daremos cuenta de que, en la mayoría de los casos, la creación de ritmo se da de manera automática y natural.

También es por estos motivos que en el presente apartado no se intentará calificar el uso de los distintos grupos rítmicos como un rasgo utilizado de manera distintiva y consciente por los hablantes de Madrid y de la Ciudad de México, aunque los datos se seguirán separando del mismo modo. Por lo tanto, en la Gráficas 3-33 se hará únicamente un recuento de los grupos rítmicos utilizados más frecuentemente por mujeres y hombres de las dos variedades de español, así como su valor porcentual (véase Tablas 3-69 y 3-138).

Gráfica 3-33: Grupos rítmicos por género



Fuente: Prosodia.xls (Base de datos propia)

Así, vemos que, en total, los mexicanos hacen un mayor uso del grupo rítmico 0, (0), con una diferencia de 5.44%, y del grupo 5, (w w s), con una frecuencia de aparición tan sólo 0.28% mayor. El resto de los grupos rítmicos aparecen con mayor frecuencia en el habla de los madrileños, con diferencias que van de 1.90% en el grupo 13, (w w w s w), hasta 0.28% en el grupo 2, (w s). En el caso específico de las mujeres, las madrileñas hacen un uso 4.60% mayor del grupo 1, (s w), y 0.31% del grupo 2, mientras que las diferencias que alcanzan las mexicanas en el resto de los grupos rítmicos son ligeramente superiores, entre 2.73% en el grupo 5, (w w s), y 0.18% en el grupo 0. Entre los hombres, las producciones de los mexicanos destacan por la frecuencia en que aparecen los grupos rítmicos 0, 1 y 2, con diferencias de 8.53, 2.30 y 1.37% con respecto a los madrileños, quienes, a su vez, producen el resto de grupos con mayor frecuencia, con diferencias entre 3.39% en el grupo 8 y 1.54% en el grupo 5. En todos los casos, el grupo más común es el 1, con acentuación grave, y le siguen los grupos 4 y 8, también con acentuación grave, (w s w) y (w w s w) respectivamente (véase 1.3.4).

Con el estilo de letra cursiva se marcan los grupos con acentuación aguda, con letra negrita los grupos con acentuación grave, con subrayado sencillo la acentuación esdrújula, con subrayado doble la sobreesdrújula y con el estilo normal se indica la no acentuación. Con esta ayuda visual es más sencillo ubicar los grupos según sus propiedades acentuales.

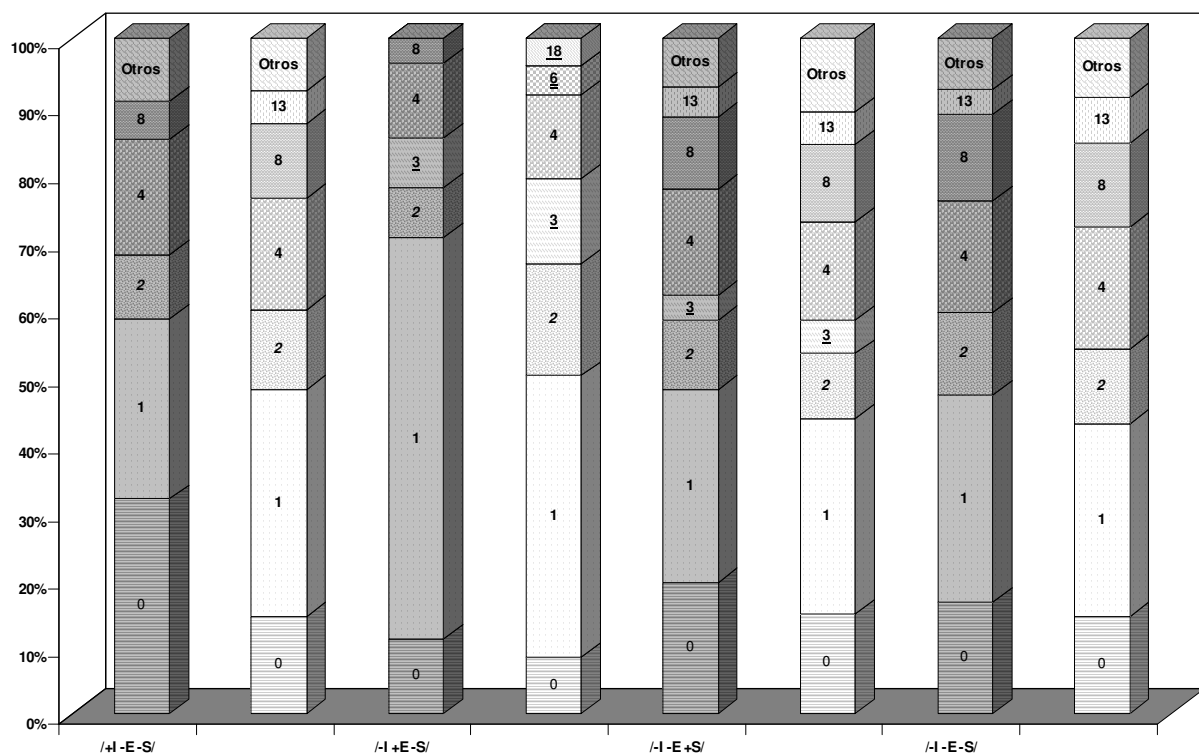
Por otro lado, en la Gráfica 3-34 se presenta el valor porcentual de los grupos rítmicos en los distintos tipos de enunciados (véase Tablas 3-70 y 3-139). Por ejemplo, en los enunciados interrogativos, los mexicanos producen 16.62% más grupos 0 que los madrileños, y, en contraste, la diferencia en los grupos 4 y 5 sólo es 0.60 y 0.16% superior, mientras que los madrileños compensan estas diferencias con valores más altos de los grupos 1, 8, 13 y 2, cuyas diferencias porcentuales son de 6.60, 5.16, 4.71 y 2.42%.

En los enunciados exclamativos, los mexicanos producen un mayor porcentaje de los grupos 1, 8 y 0, con diferencias de 19.26, 3.70 y 3.11%, mientras que los madrileños emplean con mayor frecuencia los grupos 2, 3, 6 y 4, con diferencias de 8.59, 4.59, 4.00 y 0.89%. En este caso sobresale el uso de grupos rítmicos con acentuación esdrújula y sobreesdrújula.

En cuanto a los enunciados suspendidos, excepto en el empleo del grupo 0, las diferencias son mínimas, ya que los mexicanos utilizan 4.09, 2.33, 0.92 y 0.46% más los grupos 0, 5, 4 y 2, respectivamente, pero los madrileños hacen un mayor uso de los grupos rítmicos 3, 1, 8 y 13, con diferencias de 1.29, 1.16, 1.10 y 0.59%.

Por último, en el caso de los enunciados neutros, también comparados con los valores de los enunciados suspendidos, las proporciones son muy similares entre ambos grupos de hablantes, a no ser por la mayor utilización por parte de los mexicanos de los grupos 1, 0, 2 y 8, con 2.40, 2.33, 1.27 y 0.46%, así como el empleo más frecuente de los grupos rítmicos 13, 5 y 4, con diferencias porcentuales de 2.87, 1.52 y 1.04% por parte de los madrileños.

Gráfica 3-34: Grupos rítmicos por tipo de enunciado

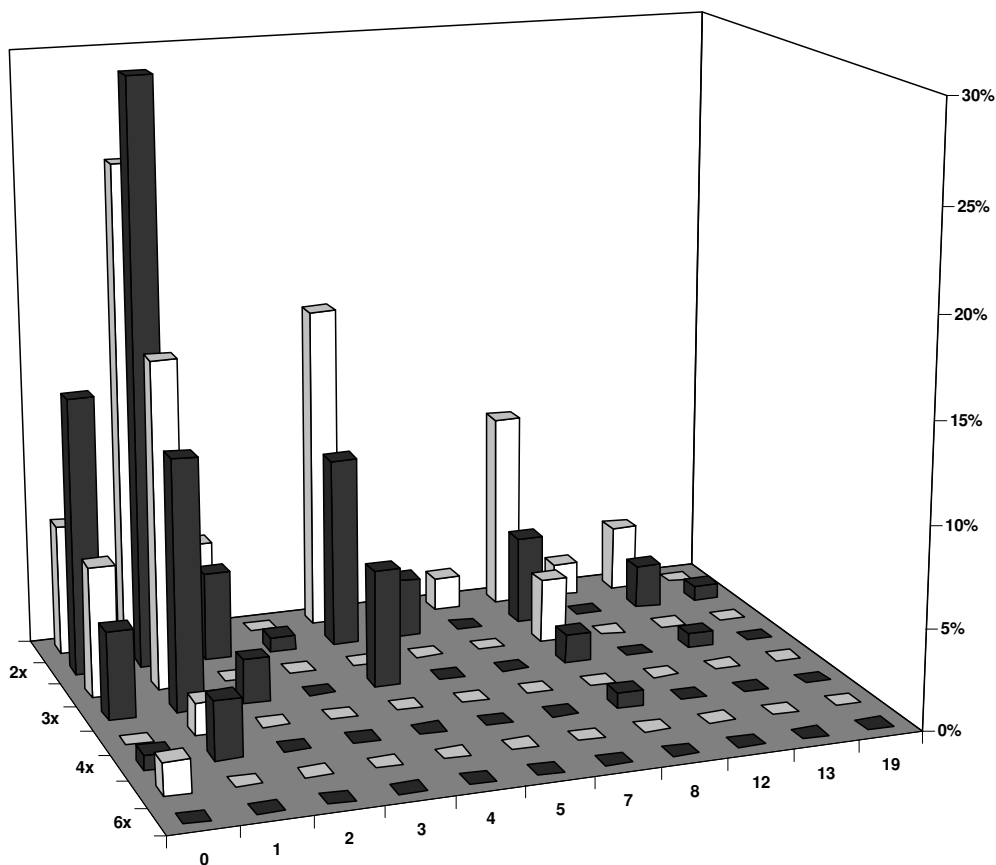


Fuente: Prosodia.xls (Base de datos propia)

Con la única excepción de los enunciados interrogativos de los mexicanos, donde el grupo rítmico más común es el 0, en todos los demás casos, el grupo 1, seguido del 4, 2 y 8 son los que aparecen con mayor frecuencia. Además, el uso de los grupos rítmicos en los enunciados suspendidos y los neutros (que son los que cuentan con el mayor número de observaciones), son muy similares entre sí y entre los distintos grupos de hablantes. En cuanto al tipo de acentuación más común, podemos ver por medio de los estilos tipográficos que la acentuación grave es la predominante en todos los casos, seguida de lejos por la aguda, y en el caso especial de los enunciados exclamativos, también por la acentuación esdrújula.

Pasando ahora realmente al análisis del ritmo en las producciones de los hablantes de Madrid y la Ciudad de México, en la Gráfica 3-35 se muestra la cantidad de veces que aparecen grupos rítmicos idénticos en una secuencia continua, que es la única manera en que pueden ser percibidos como estructuras rítmicas (véase Tablas 3-71 y 3-140).

Gráfica 3-35: Grupos rítmicos idénticos incluyendo el grupo 0



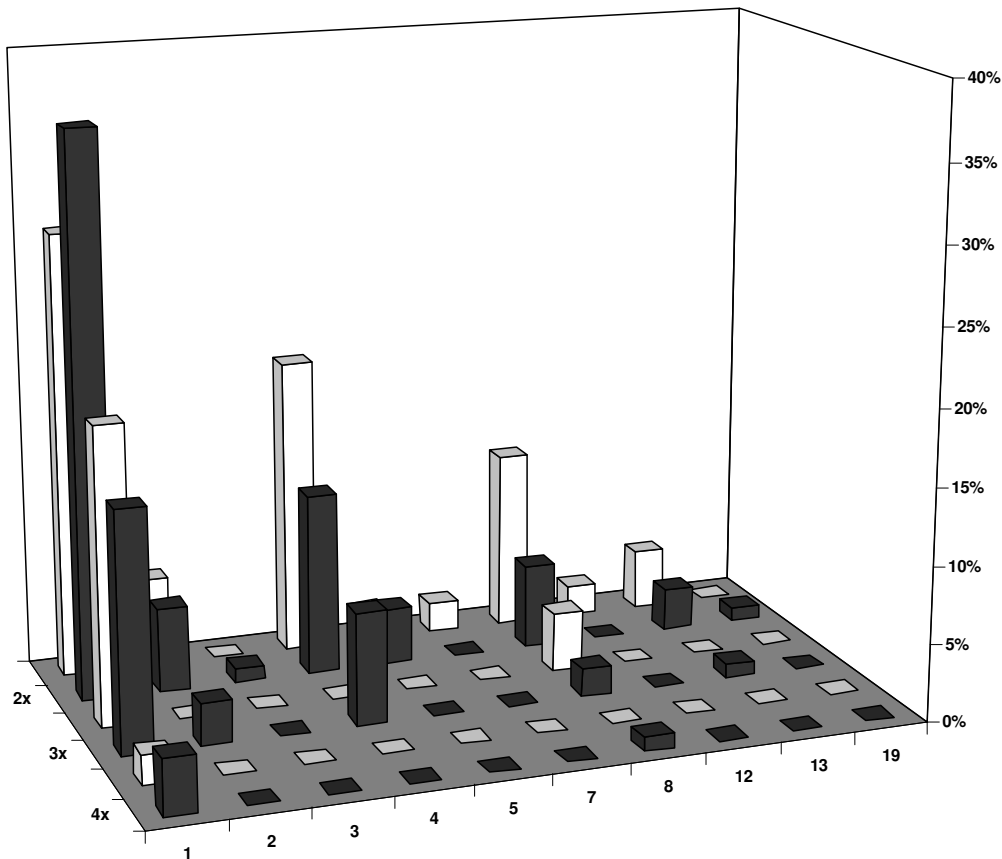
Fuente: Prosodia.xls (Base de datos propia)

A pesar la frecuencia con la que aparecen las secuencias de pies libres, en apartados anteriores se mencionó que las secuencias de grupos rítmicos 0 son poco fiables a la hora de analizar el ritmo, pues una secuencia de este tipo no implica que pueda ser percibido como una estructura

rítmica, sobre todo por las diferencias en la duración silábica, ya que esta es una de las principales condiciones para que se cumpla el principio de isocronía (véase 1.3.1 y 1.3.3).

Por esta razón, se presenta también la Gráfica 3-36, en la que se ha eliminado todos los resultados correspondientes a las secuencias del grupo rítmico 0:

Gráfica 3-36: Grupos rítmicos idénticos



Fuente: Prosodia.xls (Base de datos propia)

Tomando en cuenta estos nuevos porcentajes, para empezar, podemos comparar la serie más numerosa de grupos consecutivos idénticos, el **1**, el ritmo trocaico (*s w*). En el caso de la aparición de únicamente dos grupos rítmicos **1**, los mexicanos alcanzan un porcentaje de 36.36% de todos los casos de ritmicidad, lo que representa un 7.52% de diferencia en comparación con las producciones madrileñas, que, con 28.85% también constituye el grupo más empleado. En los casos que se dio una secuencia de tres grupos rítmicos **1**, los madrileños le sacan a los mexicanos una ventaja de 3.78%, aunque en las secuencias de cuatro grupos idénticos los mexicanos vuelven a sacar una diferencia a su favor de 1.71%.

En cuanto al ritmo yámbico (*w s*), conformado por secuencias de grupos rítmicos del tipo **2**, así como el ritmo anfibráquico (*w s w*), **4**, el ritmo de tercer peón (*w w s w*), **8**, y el ritmo a base de díbracos y anfibracos (*w w + s w*), **13**, al menos en las secuencias de dos grupos, los madrileños son quienes los emplean más a menudo, con diferencias de 0.31, 7.41,

6.08 y 1.12% respectivamente. Los mexicanos, por su parte, emplean la secuencia de dos grupos 5, ritmo anapéstico (w w s), 1.71% más que los madrileños.

Además, para terminar con las secuencias de dos grupos rítmicos, hay algunos ritmos que sólo aparecen en uno u otro de los *corpora*: en el *corpus* de Madrid se dan casos de ritmo de segundo peón (w s w w), 7, y de combinación de díbracos con dáctilos (w w + s w w), 12, mientras que en el *corpus* de la Ciudad de México se dan casos aislados de ritmo dactílico (s w w), 3, y de ritmo compuesto por tríbracos y anfíbracos (w w w + w s w), 19.

Dentro de las secuencias de tres grupos rítmicos, aparte de la ya comentada sobre el grupo 1, los madrileños emplean 2.03% más el ritmo de tercer peón, 8, mientras que los mexicanos producen tres estructuras rítmicas que no aparecen en tríadas en el habla de los madrileños, a saber, los ritmos compuestos por los grupos 2, 4 y 13. También en las secuencias de cuatro grupos, sólo en el habla de los mexicanos se da un caso de ritmo de tercer peón, 8.

En conclusión, al menos con la información de la base de datos, se puede afirmar que los mexicanos hacen un mayor uso del ritmo trocaico que los españoles, aunque para ambos grupos este es el ritmo más común. Los madrileños, por su parte, hacen un mayor uso de ritmos con acentuación grave pero con una, dos o tres sílabas átonas precediendo a la tónica, como el anfibráquico, el de tercer peón y el combinado de díbracos y anfíbracos. A no ser por el ritmo yámbico, de acentuación aguda, los casos en los que se crean estructuras rítmicas con base en una acentuación aguda o esdrújula ocurren únicamente como eventos aislados.

3.5 Resumen

En esta investigación se partió del supuesto que, si los hablantes son capaces de identificar de manera más o menos activa y consciente las características prosódicas de un habla distinta de la suya, estas características deben estar presentes en el plano acústico. Dentro de las mismas grabaciones no fue posible pedir a los hablantes madrileños que intentaran imitar o describir el habla de los mexicanos, aunque este sí ha sido el tema de conversación con investigadores procedentes de la Península Ibérica. En el caso de las grabaciones de las entrevistas y conversaciones con hablantes de la Ciudad de México, la mayoría de ellos mostraron interés por esta investigación y el tema de las diferencias entre su habla propia y la de los españoles en general les pareció divertido. Esa habla típica de los españoles que los mexicanos describen e imitan es la mejor prueba de que, independientemente de las diferencias que presenta cada dialecto en el plano segmental y de las características específicamente idiosincrásicas, hay rasgos prosódicos que pueden ser abstraídos y que en la mayoría de los casos demuestran tener correlatos acústicos.

La tecnología ha avanzado en tal medida que ahora es posible que cualquier investigador tenga acceso a una multitud de programas gratuitos de análisis acústico con una calidad comparable a los programas de alto costo. El programa Praat es un ejemplo de estos programas gratuitos de alta calidad que son utilizados por una inmensidad de investigadores en todo el planeta que sugieren modificaciones y ampliaciones del programa, lo que conlleva continuas mejoras del producto con cada actualización.

Fue con este programa que se analizó y extrajo toda la gama de valores acústicos de duración, intensidad y frecuencia fundamental. Estos datos fueron sometidos a una estructura predefinida del lenguaje XML, gracias a la cual se pudo hacer la selección específica de entre la gran multitud de datos para obtener páginas *web* que desplegaran los datos con distintos formatos y gracias a las cuales se pudo consolidar la base de datos en Excel.

Las tablas de datos presentadas en los apartados 3.2 y 3.3 son tan sólo los resultados de unas de las muchas posibilidades de selección, procesamiento, filtrado y presentación de los subconjuntos de datos. En esos dos apartados se pudo ya ir previendo las posibilidades de análisis comparativos, pues los datos de los distintos hablantes y los valores promediados para mujeres y hombres ya dejaban entrever las principales similitudes y diferencias. Sin embargo, se intentó únicamente presentar los datos, mostrando los valores extremos con respecto a los promedios y, sobre todo, justificar el uso de valores estandarizados de la intensidad y la frecuencia fundamental que más tarde permitirían hacer comparaciones de la variación de tales parámetros sin que los factores extralingüísticos afectaran los resultados.

Una vez que se hubo presentado todos los componentes individuales para el análisis, en el apartado 3.4 se procedió a combinar los valores promedio correspondientes a los hablantes de Madrid y a los de la Ciudad de México en gráficas que permitieran identificar visualmente las principales diferencias entre las características de las producciones lingüísticas de cada uno de los grupos de hablantes.

Los resultados fueron muy satisfactorios, pues se pudo obtener pruebas concretas e identificables acústicamente de las diferencias percibidas por los hablantes, a saber, que, por ejemplo, desde el punto de vista de los madrileños, 1) los mexicanos producen las sílabas de manera más lenta y regular, lo que hace parecer su habla más lenta, 2) hablan a un volumen o intensidad más baja y regular, es decir, tienden menos a alzar y bajar la voz dentro de un mismo enunciado, 3) utilizan un rango de voz o registro más alto, lo que también repercute en una mayor variación de la frecuencia, es decir, que hablan de manera menos monótona, 4) inician más a menudo los enunciados con una entonación alta, pronuncian más frecuentemente las sílabas tónicas con un pico frecuencial simultáneo y más alto que ellos, y,

comparativamente, terminan los enunciados con entonación ascendente usando una frecuencia más alta y los contornos descendentes con una frecuencia porcentualmente más baja, 5) por lo general, aunque la distribución de tonos nucleares y finales en los distintos tipos de enunciados es muy variable, en los enunciados terminados con un contorno suspendido (menor a $\pm 25\%$) los mexicanos utilizan más frecuentemente tonos del tipo \acute{S} , es decir, tonos cuyo pico frecuencial coincide con la sílaba tónica ($L+H^*$, $\acute{L}+H^*$, $L+!H^*$ o H^*), y 6) tienden a formar estructuras rítmicas más cortas, preferentemente de acentuación grave.

Sin embargo, lo que parece funcionar muy bien en la comparación de sólo dos variedades dialectales de la misma lengua tendrá que ser probado en el futuro con conjuntos de datos más amplios y provenientes de una mayor cantidad de dialectos. Sólo así se podrá identificar y solucionar nuevos problemas que conlleven nuevas implicaciones metodológicas.

Lo que debe destacarse del presente método, aparte de la aplicación en su totalidad a *corpora* lingüísticos lo suficientemente robustos con datos provenientes de la lengua oral y grabados con distintas calidades acústicas, es el hecho de analizar de manera sistemática y conjunta distintos parámetros acústicos, en lo que podríamos definir como prosodia, un área de estudio única y diversa, es decir, una vertiente de la fonología que puede estudiarse como un todo, pero que se compone de distintos fenómenos que dan cuenta de las características físicas de la materia fónica.

4. Conclusiones

Trabajar con habla espontánea implica renunciar a muchas de las comodidades de estudiar el habla de laboratorio, por ejemplo, la calidad y nivel de grabación no siempre es la que uno desearía; los hablantes pocas veces dirán justo el tipo de enunciados que uno necesita con la intensidad, velocidad, entonación y ritmo que uno espera; la situación, el contexto y el ruido de ambiente pocas veces pueden ser controlados, e incluso las temáticas algunas veces pueden ser poco aptas para la investigación y su consiguiente divulgación. Sin embargo, desarrollar toda una metodología con la que, incluso con grabaciones de conversaciones espontáneas de tales características, también se pueda obtener resultados satisfactorios, es la mejor prueba de que el método funciona y que podrá aplicarse a cualquier otro tipo de grabaciones.

En la Introducción se hizo mención de las preguntas, objetivos e hipótesis que sirvieron de guías para la planeación, el desarrollo y la conformación de la presente investigación. Las cinco hipótesis propuestas son las siguientes:

H₁: Los factores prosódicos que influyen en la comunicación son la entonación, el ritmo y el tempo, la intensidad y la calidad de la voz.

H₂: Los modelos teóricos que mejor se aplican al estudio y representación electrónica de la entonación del español son el métrico autosegmental, Sp-ToBI, y el de los contornos melódicos con frecuencia fundamental estandarizada; para el ritmo, el modelo más apto es el de los grupos rítmicos; la mejor manera de analizar el tempo y la intensidad es por medio de la variación porcentual de una sílaba con respecto a la anterior; la calidad de la voz no es fácilmente analizable en grabaciones espontáneas.

H₃: Dichos modelos son suficientes para dar cuenta de las similitudes funcionales y las diferencias regionales de la prosodia del español oral de la Ciudad de México y Madrid.

H₄: Haciendo uso del lenguaje XML, es posible representar jerárquicamente los distintos fenómenos prosódicos que aparecen en el habla espontánea y especificar por medio de variables sus características acústicas.

H₅: El programa Praat es el más adaptado para realizar los análisis acústicos y crear automáticamente documentos de XML con la información necesaria para permitir análisis detallados de la entonación, el ritmo, el tempo y la intensidad de las producciones lingüísticas de los hablantes castellanos y mexicanos.

Con respecto a la hipótesis H₁, H₂ y H₃, en la sección 1 de este trabajo se presentó los conceptos básicos sobre la prosodia, pues mi intención original era contemplar no sólo el estudio de la entonación, el ritmo o la intensidad por separado, sino ofrecer una visión panorámica de la complejidad de los fenómenos suprasegmentales como una unidad diversa.

La prosodia no es solamente el análisis de la entonación o el ritmo o los rasgos acústicos por separado, sino que los modelos de análisis deben incluirlos a todos ellos e intentar dar cuenta de sus interacciones e interdependencia.

En el caso del español han sido pocos los estudios que han abarcado más de uno de los aspectos de la prosodia, con una predilección especial por la entonación. El ritmo ha sido considerado poco útil para la producción de contrastes semánticos o pragmáticos, pero ¿quién no sabe que la misma frase dicha melodiosa y rítmicamente no tiene una diferente función pragmática en el contexto preciso?

Aquí se intentó abordar los principales modelos teóricos que ya han aplicados al español por algún autor y que ofreciera unas ideas clave que valiera la pena rescatar y aplicar en la propia investigación. Es por este motivo que al final se incluyó teorías que son consideradas incluso por sus mismos autores como incompatibles con los supuestos de las principales corrientes, como en el caso de la teoría de los patrones melódicos del español en el habla espontánea de Cantero (2002 y 2005) con respecto a la teoría métrica autosegmental, en especial la teoría de los tonos e índices de disyunción o Sp-ToBI (Sosa 1999 y 2003, Beckman *et al.* 2002, Hualde 2003 y Prieto 2003). Sin embargo, al menos para la realización de esta investigación, se consideró que algunos aspectos de la teoría de los patrones melódicos facilitaban el análisis de los contornos entonativos de enunciados como unidades a la vez autónomas e interdependientes, y la métrica autosegmental permitía un mejor análisis de los tonos como componentes de esos contornos entonativos. Es más, al menos dentro del marco de esta investigación, la utilización de una de las dos teorías sin tomar en cuenta la otra no hubiera permitido realizar el presente estudio con el nivel de profundidad e interdependencia que se esperaba.

Por otro lado, la teoría de los patrones melódicos desató nuevas ideas que aún no habían sido desarrolladas por otros investigadores. Dentro del área del ritmo, la descripción de los grupos rítmicos que hace Cantero (2002), se aplicó aquí en el sentido más amplio del término, pues incluso se creó un sistema de transcripción parecido a la representación del peso silábico en la mayoría de las corrientes fonológicas actuales, por ejemplo, la representación de la base del ritmo trocaico con la secuencia ($s w$), donde s representa la sílaba tónica, w la sílaba átona, y la longitud y los límites del pie o grupo acentual se delimitan por medio de los paréntesis. Esta representación no sólo probó ser útil e indispensable para el análisis del ritmo para el investigador como ser humano, sino que era fácilmente detectable y procesable por los programas de computadora. De este modo se detectó, por ejemplo, cuáles eran las secuencias imposibles o muy poco comunes en el habla cotidiana, a saber, todos los grupos rítmicos con tres o más sílabas átonas después de la tónica, lo que reduce el total de combinaciones posibles de veintiuna a sólo quince o dieciséis.

Además, el método de la estandarización de la frecuencia fundamental probó ser útil no sólo para el análisis de la variación frecuencial, sino también para el análisis, sobre todo comparativo, de la variación de la intensidad. Aunque es muy probable que alguien más ya hubiera llegado a una idea parecida, hasta el momento parece no haber ninguna obra que haya utilizado la intensidad estandarizada.

También el inventario de tonos del Sp-ToBI, que además yo aumenté con la inclusión del tono ¡L*+H, no sólo por cuestiones de sistematicidad y simetría, sino para que hubiera una congruencia teórica, resultó ser más fácilmente comparable si se consideraba estos tonos individuales como alotonos o variantes fonéticas en distribución complementaria de cuatro architonemas: \acute{S} (tonos cuyo pico frecuencial coincide con la sílaba acentuada: L+H*, L+!H*, ¡L+H* y H*), S' (tonos cuyo pico frecuencial se encuentra después de la sílaba tónica: L*+H, L*+!H y ¡L*+H), \grave{S} (tonos con pico frecuencial antes de la sílaba acentuada: H+L*), y S (tonos sin pico frecuencial o muy poco perceptible: *).

Además, dentro del área de la duración, se utilizó la información sobre la duración silábica y su variación intersilábica para obtener un índice aproximado del tempo o velocidad promedio a la que los hablantes se comunican normalmente.

Así, en el apartado 3.5, se da por confirmado que la selección de parámetros acústicos y modelos teóricos permiten dar cuenta de las principales diferencias acústicas y perceptivas de la prosodia de las variedades del español de Madrid y la Ciudad de México fue la correcta.

Por último, con respecto a las hipótesis H₄ y H₅, en el marco de esta investigación se planeó y desarrolló un sistema de transcripción basado en el lenguaje XML, que permite la introducción de texto y etiquetas con información estructural. El inventario de etiquetas y su gramática subyacente se diseñaron con la finalidad de que representaran las relaciones jerárquicas y funcionales de los distintos componentes de la prosodia.

De manera paralela se seleccionó las grabaciones de donde se extraerían los fragmentos idóneos para el análisis acústico de las grabaciones de los hablantes de Madrid y la Ciudad de México. Además, se determinaron normas de transcripción para asegurar el más alto nivel de consistencia entre los textos de las transcripciones.

A manera de adaptación entre los dos sistemas, se ideó que, por medio de subprogramas o *scripts* del programa Praat, se automatizara el análisis acústico de las transcripciones segmentadas por sílabas y que esa información se imprimiera en un archivo de texto rodeada de secuencias de signos que, aunque para Praat fueran sólo texto, en el archivo de texto resultante serían los patrones sintácticos del etiquetado de XML.

Una vez que se obtuvo los documentos de XML correspondientes a cada fragmento de grabación seleccionado, se procedió a dar distintos formatos a la información y al texto dentro y fuera de las etiquetas de elementos y atributos. Las páginas *web* resultantes de la transformación de los datos por medio del lenguaje XSLT fueron utilizadas dentro de esta investigación con distintas finalidades: la presentación en forma de página *web* interactiva con imágenes realizadas automáticamente en Praat sobre el análisis de la frecuencia fundamental y enlaces a los fragmentos de grabación correspondientes (todo lo cuál sirvió para una mejor administración y selección manual de los fragmentos idóneos); la presentación selectiva, en la que se permitía ver un conjunto filtrado de datos, es decir, que muestra únicamente los fragmentos producidos por determinadas personas o tipos de fenómenos prosódicos que no dificultaran el análisis acústico), y las presentaciones con fines estadísticos, para las cuales se seleccionó y ordenó la información de tal manera que pudiera ser simplemente copiada a hojas de cálculo de Excel para proceder con los análisis estadísticos de las series de datos.

En las hojas de cálculo se obtuvo por medio de una gran variedad de métodos y fórmulas la información que se vació en las tablas de datos de las secciones 3.2 y 3.3. Finalmente, esos datos correspondientes a los *corpora* de Madrid y la Ciudad de México se integraron en Gráficas con las que se sustentó también visualmente las similitudes y diferencias en el uso que los hablantes de los dos grupos hicieron de los distintos fenómenos prosódicos.

Los resultados obtenidos superaron las expectativas planteadas inicialmente, aunque este estudio, por muy amplia que haya sido la muestra, no deja de ser un estudio exploratorio que deberá de ser puesto a prueba con mayores cantidades de datos correspondientes a un mayor número de orígenes dialectales.

Con respecto a las posibles aplicaciones que se haga de los índices de variación dialectal obtenidos por medio de esta metodología, se prevé que podría ser útil para sistemas de reconocimiento de voz, conversión de texto en habla o síntesis de voz, pero también simplemente para la creación de inventarios prosódicos de distintas lenguas y sus variedades dialectales, aunque será el tiempo el que decida finalmente cuáles de ellas se harán realidad.

Sólo resta por decir que, a pesar del poco contacto que tuve con otros investigadores durante la realización de la presente investigación, o tal vez con mayor razón, la experiencia de adentrarse en el tema de la prosodia fue mucho más enriquecedora y desafiante de lo que alguna vez imaginé. Gracias a las exigencias de la propia investigación, muchas áreas de conocimiento como la informática, la estadística o la física acústica pasaron a jugar un papel más importante dentro de mi formación y cada una de ellas aportó una pieza indispensable del rompecabezas para la creación de esta obra.

5. Bibliografía

- 'T HART, Johan, Collier, René y Cohen, Antonie (1990): *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge: CUP.
- ABERCROMBIE, David (1965): Forgotten phoneticians. En *Studies in phonetics and linguistics*. Londres: Oxford University Press. Pp. 45-75
- ALARCOS LLORACH, Emilio (1950): *Fonología española*. Madrid: Gredos.
- ALCOBA, Santiago y Murillo, Julio (1998): Intonation in Spanish. En Hirst, Daniel y Di Cristo, Albert (Eds.): *Intonation systems: a survey of twenty languages*. Cambridge: CUP. Pp. 152-166.
- ALMEIDA, Manuel (1999): *Tiempo y ritmo en el español canario. Un estudio acústico*. Madrid: Iberoamericana; Frankfurt a. M.: Vervuert. (Lingüística Iberoamericana, 8)
- ARMSTRONG, Liliás E. y Ward, Ida C. (1931) [1926]: *A handbook of English intonation*. Cambridge: Heffer and Sons.
- AUER, Peter; Couper-Kuhlen, Elizabeth y Müller, Frank (1999): *Language in time. The rhythm and tempo of spoken interaction*. New York, Oxford: OUP (Oxford Studies in Sociolinguistics).
- BAQUÉ, Lorraine y Estruch, Mònica (2003): Modelo Aix-en-Provence. En Prieto, Pilar (Ed.): *Teorías de la entonación*. Barcelona: Ariel. Pp. 123-153.
- BARNARD, David T. y Ide, Nancy M. (1995): The Text Encoding Initiative: Flexible and Extensible Document Encoding. Technical Report 96-396. Department of Computing and Information Science, Queen's University Kingston, Ontario: 12/1995. [<http://www.cs.queensu.ca/TechReports/Reports/1996-396.pdf>]
- BECKMAN, Mary E. y Ayers Elam, Gayle (1997): Guidelines for ToBI Labelling. Version 3. [[http://ling.ohio-state.edu/~tobi/ Guidelines for ToBI Labelling.htm](http://ling.ohio-state.edu/~tobi/Guidelines%20for%20ToBI%20Labelling.htm)]. Fecha de consulta: 30.09.02.
- BECKMAN, Mary E.; Díaz-Campos Manuel; Tevis McGory, Julia y Morgan, Terrell A. (2002): Intonation across Spanish, in the Tones and Break Indices framework. En: *Probus*, 14: 9-36.
- BENKWITZ, Annaliese (2003): *Kontrastive phonetische Untersuchungen zum Rhythmus. Britisches Englisch als Ausgangssprache - Deutsch als Zielsprache*. Frankfurt a. M. et al.: Peter Lang. (Hallesche Schriften zur Sprechwissenschaft und Phonetik, 14)
- BERNERS-LEE, Tim; Hender, James y Lassila, Ora (2001): The Semantic Web [en línea]. En *Scientific American*: 17/05/2001 [Fecha de consulta: 21/09/2005] [<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>]
- BIRD, Steven y Liberman, Mark (2001): A formal framework for linguistic annotation. In: *Speech Communication* 33(1,2), 23-60.

- BIRON, Paul V. y Malhotra, Ashok (Eds.) (2004): XML Schema Part 2: Datatypes Second Edition [en línea]. W3C (World Wide Web Consortium) Recommendation. 28/10/2004 [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/>]
- BLEVINS, Juliette (1995): 6. The syllable in phonological theory. En Goldsmith, John A. (Ed.) *The Handbook of Phonological Theory*. Cambridge, Oxford: Blackwell. Pp. 206-244.
- BLOOMFIELD, Leonard (1993): *El lenguaje*. Lima: Universidad Nacional Mayor de San Marcos, 1964.
- BOERSMA, Paul (2001): Praat, a system for doing phonetics by computer. En *Glott International* 5:9/10, 341-345.
- BOERSMA, Paul y Weenink, David (2008): Praat: doing phonetics by computer (Versión 5.0.23) [Programa computacional]. [Actualización: 09/05/2008] [<http://www.praat.org/>]
- BOS, Bert; Wium Lie, Håkon; Lilley, Chris y Jacobs, Ian (Eds.) (1998): Cascading Style Sheets, level 2. CSS2 Specification [en línea]. W3C (World Wide Web Consortium) Recommendation: 12/05/1998. [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/1998/REC-CSS2-19980512>]
- BRAY, Tim; Paoli, Jean; Sperberg-McQueen, C. M., Maler, Eve y Yergeau, François (Eds.) (2004): Extensible Markup Language (XML) 1.0 Second Edition [en línea]. W3C (World Wide Web Consortium) Recommendation: 04/02/2000 [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/REC-xml>]
- BROSELOW, Ellen (1995): 5. Skeletal positions and moras. En Goldsmith, John A. (Ed.) *The Handbook of Phonological Theory*. Cambridge, Oxford: Blackwell. Pp. 175-205.
- BROWN, Allen; Fuchs, Matthew; Robie, Jonathan y Wadler, Philip (Eds.) (2001): XML Schema: Formal Description [en línea]. W3C (World Wide Web Consortium) Recommendation: 25/09/2001. [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/2001/WD-xmlschema-formal-20010925/>]
- BRUCE, Gösta (1977): *Swedish word accents in sentence perspective*. Lund: CWK Gleerup.
- BURNARD, Lou (1995): Text Encoding for Information Interchange. An Introduction to the Text Encoding Initiative [en línea]. Oxford University Computing Services: 07/1995. [<http://www.tei-c.org/Vault/SC/J31/>]
- BURNETT, Daniel C.; Walker, Mark R. y Hunt, Andrew (Eds.) (2004): Speech Synthesis Markup Language (SSML) Version 1.0 [en línea]. W3C (World Wide Web Consortium) Recommendation: 07/09/2004 [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/2004/REC-speech-synthesis-20040907/>]
- CANTERO SERENA, Francisco José (2002): *Teoría y análisis de la entonación*. Barcelona: Universitat de Barcelona.
- CANTERO SERENA, Francisco José (2005): Patrones melódicos del español en habla espontánea. En *Actas del III Congreso de Fonética Experimental*. Santiago de Compostela: Universidad de Santiago de Compostela.

- CARRASCO González, María y Celis Sánchez, Cruz (Revisado por Brian MacWhinney) (2004): Sistema de Transcripción CHAT [en línea]. CHILDES Project: 20/04/2004 [Fecha de consulta: 21/09/2005] [<http://childes.psy.cmu.edu/intro/spanish.pdf>]
- CLARK, James (Ed.) (1999): XSL Transformations (XSLT) Version 1.0 [en línea]. W3C (World Wide Web Consortium) Recommendation: 16/11/1999 [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/1999/REC-xslt-19991116>]
- CLARK, James (Ed.) (2002): RELAX NG Compact Syntax [en línea]. OASIS Committee Specification: 21/11/2001. [Fecha de consulta: 21/09/2005] [<http://www.relaxng.org/compact-20021121.html>]
- CLARK, James y DeRose, Steve (Eds.) (1999): XML Path Language (XPath) Version 1.0 [en línea]. W3C (World Wide Web Consortium) Recommendation: 16/11/1999 [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/1999/REC-xpath-19991116>]
- CLARK, James y Murata, Makoto (Eds.) (2001): RELAX NG Specification [en línea]. OASIS Committee Specification: 03/12/2001. [Fecha de consulta: 21/09/2005] [<http://www.oasis-open.org/committees/relax-ng/spec.html>]
- COUPER-KUHLEN, Elizabeth (1986): *An introduction to English prosody*. Londres: Edward Arnold.
- CRESTI, Emanuela y Moneglia, Massimo (Eds.) (2005): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.
- CRYSTAL, David (1969): *Prosodic systems and intonation in English*. Cambridge: CUP. (Cambridge Studies in Linguistics, 1)
- CRYSTAL, David (1997): *The Cambridge Encyclopedia of Language*. 2. ed. Cambridge: CUP.
- CUBER, Ulrich (1997): *VBA-Programmierung mit Word 97*. München y Düsseldorf: ECON.
- DI CRISTO, Albert (2003): De la métrique et du rythme de la parole ordinaire: l'exemple du français. En *SEMEN. Revue de sémiolinguistique des textes et discours*, 16. Besançon: Presses Universitaires Franc-Comtoises. Pp. 25-43
- DITTMAR, Norbert (2002): *Transkription*. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien. Opladen: Leske + Budrich. (Qualitative Sozialforschung, Band 10)
- DRISCOLL, M. J. (2004): The Text Encoding Initiative. En: Преглед НИЦД (Review of the National Center for Digitization) 5 (2004), pp. 5-18 [<http://www.komunikacija.org.yu/komunikacija/casopisi/ncd/5/d003/download>]
- EDWARDS, Jane (1993): 1. Principles and Contrasting Systems of Discourse Transcription. En: Edwards, Jane A. y Lampert, Martin D.: *Talking Data: Transcription and Coding in Discourse Research*. London: Lawrence Erlbaum. Pp. 3-31
- EDWARDS, Jane A. (1995): Principles and alternative systems in the transcription, coding and mark-up of spoken discourse. En: Leech, Geoffrey; Myers, Greg y Thomas, Jenny (Eds.): *Spoken English on computer*. New York: Longman. Pp. 19-34.

- EDWARDS, Jane A. y Lampert, Martin D. (1993): *Talking Data: Transcription and Coding in Discourse Research*. London: Lawrence Erlbaum.
- EHLICH, Konrad (1993): 5. HIAT - a Transcription System for Discourse Data. En: Edwards, Jane A. y Lampert, Martin D.: *Talking Data: Transcription and Coding in Discourse Research*. London: Lawrence Erlbaum. Pp. 123-148.
- ENRÍQUEZ, Emilia V.; Casado, Celia y Santos, Andrés (1989): La percepción del acento en español. En *Lingüística Española Actual*, 11. Pp. 241-269
- FALLSIDE, David y Walmsley, Priscilla (Eds.) (2004): XML Schema Part 0: Primer Second Edition [en línea]. W3C (World Wide Web Consortium) Recommendation: 28/10/2004 [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/2004/REC-xmlschema-0-20041028/>]
- FONT ROTCHÉS, Dolors (2005a): *L'entonació del català*. Tesis doctoral, Laboratori de Fonètica Aplicada, Universitat de Barcelona.
- FOX, Anthony (2000): *Prosodic features and prosodic structures. The phonology of suprasegmentals*. Oxford: OUP.
- GARCÍA-LECUMBERRI, María Luisa (2003): Análisis por configuraciones: la escuela británica. En Prieto, Pilar (Ed.): *Teorías de la entonación*. Barcelona: Ariel. Pp. 35-61.
- GARRIDO Almiñana, Juan María (2003): La escuela holandesa: el modelo IPO. En Prieto, Pilar (Ed.): *Teorías de la entonación*. Barcelona: Ariel. Pp. 97-122.
- GIL FERNÁNDEZ, Juana (1988): *Los sonidos del lenguaje*. Madrid: Síntesis.
- GILI GAYA, Samuel (1950): *Elementos de fonética general*. Madrid: Gredos.
- GOLDFARB, Charles F. y Prescod, Paul (2000): *The XML Handbook*, 2a. ed. London et al.: Pentice Hall.
- GOLDSMITH, John (1976): *Autosegmental phonology*. Tesis doctoral, MIT. (Publ. 1979, Nueva York: Garland)
- GUMPERZ, John Joseph (1982): *Discourse strategies*. Cambridge: CUP (Studies in Interactional Sociolinguistics 1).
- GÜNTHER, Carsten (1999): *Prosodie und Sprachproduktion*. Tübingen: Niemeyer (Linguistische Arbeiten 401).
- HALLE, Morris e Idsardi, William (1995): 11. General properties of stress and metrical structure. En Goldsmith, John A. (Ed.) *The Handbook of Phonological Theory*. Cambridge, Oxford: Blackwell. Pp. 403-443
- HALLIDAY, Michael A. K. (1970): *A course in spoken English: intonation*. Oxford: Oxford University Press.
- HANKE, Johann-Christian (2000): *HTML/XML & JavaScript*. Düsseldorf: Data Becker.

- HAYES, Bruce (1999) [1989]: 18. Compensatory lengthening in moraic phonology. En Goldsmith, John A. *Phonological theory. The essential readings*. Oxford: Blackwell. Pp. 352-369
- HERNÁNDEZ Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar (2003): *Metodología de la investigación*. 3a. ed. México et al.: McGraw-Hill Interamericana.
- HIRST, Daniel y Di Cristo, Albert (Eds.) (1998): *Intonation systems: a survey of twenty languages*. Cambridge: CUP.
- HOCKETT, Charles F. (1971) [1958]: *Curso de lingüística moderna*. Buenos Aires: EUDEBA, 1971.
- HUALDE, José Ignacio (2003): El modelo métrico y autosegmental. En Prieto, Pilar (Ed.): *Teorías de la entonación*. Barcelona: Ariel. Pp. 155-184.
- HYMAN, Larry M. (2003) [1985]: *A theory of phonological weight*. Stanford: CSLI.
- KAGER, René (1995): 10. The metrical theory of word stress. En Goldsmith, John A. (Ed.) *The Handbook of Phonological Theory*. Cambridge, Oxford: Blackwell. Pp. 367-402.
- KEHREIN, Roland (2002): *Prosodie und Emotionen*. Tübingen: Niemeyer. (Reihe Germanistische Linguistik, 231)
- KLASMEYER, Gudrun y Sendlmeier, Walter F. (2000): 15. Voice and emotional states. En Kent, Raymond D. y Ball, Martin J. (Eds.): *Voice quality measurement*. San Diego: Singular. Pp. 339-357
- KRANICH, Wieland (2003): *Phonetische Untersuchungen zur Prosodie emotionaler Sprechausdrucksweisen*. Frankfurt am Main: Lang. (Hallesche Schriften zur Sprechwissenschaft und Phonetik, 11)
- LADD, D. Robert (1980): *The structure of intonational meaning*. Bloomington: Indiana University Press.
- LADD, Robert (1996): *Intonational phonology*. Cambridge: CUP.
- LADEFOGED, Peter (2003): *Phonetic data analysis. An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.
- LAVER, John (2000): 4. Phonetic evaluation of voice quality. En Kent, Raymond D. y Ball, Martin J. (Eds.): *Voice quality measurement*. San Diego: Singular. Pp. 37-48
- LAWLER, John und Aristar Dry, Helen (Eds.) (1998): *Using Computers in Linguistics. A Practical Guide*. London/New York: Routledge.
- LIBERMAN, Mark (1975): *The intonational system of English*. Tesis doctoral, MIT. (Publ. 1978, Nueva York: Garland)
- LIBERMAN, Mark (1979): *The intonational system of English*. Nueva York, Londres: Garland.

- LIBERMAN, Mark y Prince, Alan (1999) [1977]: 20. On stress and linguistic rhythm. En Goldsmith, John A. *Phonological theory. The essential readings*. Oxford: Blackwell. Pp. 392-404
- MARTÍNEZ CELDRÁN, Eugenio (2003): Análisis por niveles: la escuela americana. En Prieto, Pilar (Ed.): *Teorías de la entonación*. Barcelona: Ariel. Pp. 63-95.
- MATHIESON, Lesley (2000): 1. Normal-disordered continuum. En Kent, Raymond D. y Ball, Martin J. (Eds.): *Voice quality measurement*. San Diego: Singular. Pp. 3-12
- MCCARTHY, John y Prince, Alan (1999) [1986]: 13. Prosodic morphology. En Goldsmith, John A. (Ed.) *Phonological theory. The essential readings*. Oxford: Blackwell. Pp. 238-288
- MCGLASHAN, Scott; Burnett, Daniel C.; Carter, Jerry; Danielsen, Peter; Ferrans, Jim; Hunt, Andrew; Lucas, Bruce; Porter, Brad; Rehor, Ken y Tryphonas, Steph (2004): Voice Extensible Markup Language (VoiceXML) Version 2.0 [en línea]. W3C (World Wide Web Consortium) Recommendation: 16/03/2004. [Fecha de consulta: 21/09/2005] [<http://www.w3.org/TR/2004/REC-voicexml20-20040316/>]
- MCWHINNEY, Brian (1999): 14. The CHILDES System. En: Ritchie, William C. y Bhatia, Tej K. (Eds.): *Handbook of Child Language Acquisition*. San Diego et al.: Academic Press.
- MCWHINNEY, Brian (2000): *The CHILDES Project. Tools for Analyzing Talk, Third Edition. Volume I: Transcription Format and Programs*. New Jersey, London: Erlbaum.
- NAVARRO TOMÁS, Tomás (1948): *Manual de entonación española*. Madrid: Guadarrama.
- NESPOR, Marina y Vogel, Irene (1986): *Prosodic Phonology*. Dordrecht: Foris Publications.
- NEUBER, Baldur (2002): *Prosodische Formen in Funktion. Leistungen der Suprasegmentalia für das Verstehen, Behalten und die Bedeutungs(re)konstruktion*. Frankfurt am Main: Lang. (Hallesche Schriften zur Sprechwissenschaft und Phonetik, 7)
- NIEDERMAIR, Elke y Niedermair, Michael (2001): *XML. Das große Buch*. Düsseldorf: Data Becker.
- NÚÑEZ CEDEÑO, Rafael y Morales-Font, Alfonso (1999): *Fonología generativa contemporánea de la lengua española*. Georgetown: University Press.
- O'CONNOR, J. D. y Arnold, G. F. (1973): *Intonation of colloquial English*. Londres: Longman.
- PAESCHKE, Astrid (2003): *Prosodische Analyse emotionaler Sprechweise*. Berlin: Logos.
- PERLMUTTER, David (1995): 8. Phonological quantity and multiple association. En Goldsmith, John A. (Ed.) *The Handbook of Phonological Theory*. Cambridge, Oxford: Blackwell. Pp. 307-317.
- PIERREHUMBERT, Janet (1980): *The phonetics and phonology of English intonation*. Tesis doctoral, MIT.
- PIERREHUMBERT, Janet y Hirschberg, Julia (1990): Chapter 14. The meaning of intonational contours in the interpretation of discourse. En Cohen, Philip R.; Morgan, Jerry y Pollack, Martha (Eds.): *Intentions in communication*. Massachusetts: MIT. Pp. 271-311.

- PIKE, Kenneth L. (1945): *The intonation of American English*. Ann Arbor: University of Michigan Press.
- PRIETO, Pilar (Ed.) (2003): *Teorías de la entonación*. Barcelona: Ariel.
- QUILIS, Antonio (1981): *Fonética acústica de la lengua española*. Madrid: Gredos.
- QUILIS, Antonio (1985): Entonación dialectal hispánica. *Lingüística Española Actual (LEA)*, VII, 2: 145-190.
- QUILIS, Antonio (1992): 362. Spanisch: Intonationsforschung und Prosodie. En Holtus, Günther; Metzeltin, Michael y Schmitt, Christian. *Lexikon der Romanistischen Linguistik (LRL), Band VI, 1. Aragonesisch/Navarresisch, Spanisch, Asturianisch/Leonesisch*. Tübingen: Niemeyer. Pp. 62-68.
- QUILIS, Antonio y Fernández, Joseph A. (1985): *Curso de fonética y fonología españolas*. Madrid: CSIC.
- RABANUS, Stefan (2001): *Intonatorische Verfahren im Deutschen und Italienischen. Gesprächsanalyse und autosegmentale Phonologie*. Tübingen: Niemeyer (Linguistische Arbeiten 439).
- SCHMIDT, Thomas (2001): The transcription system EXMARaLDA: An application of the annotation graph formalism as the Basis of a Database of Multilingual Spoken Discourse. En: Bird, Steven; Buneman, Peter; Liberman, Mark (Eds.): *Proceedings of the IRCS Workshop On Linguistic Databases*, 11-13 December 2001. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania. Pp. 219-227. [http://www.rrz.uni-hamburg.de/exmaralda/Daten/4D-Literatur/Vortraege-Dokumente/IRCS_Paper.pdf]
- SCHMIDT, Thomas (2002): EXMARaLDA - ein System zur Diskurstranskription auf dem Computer. *Arbeiten zur Mehrsprachigkeit, Serie B (34)*. Hamburg. [<http://www.rrz.uni-hamburg.de/exmaralda/Daten/4D-Literatur/AZM.pdf>]
- SCHMIDT, Thomas (2004): EXMARaLDA - ein Modellierungs- und Visualisierungs-verfahren für die computergestützte Transkription gesprochener Sprache. En: Buchberger, Ernst (Ed.): *Proceedings of Konvens 2004, Schriftenreihe der Österreichischen Gesellschaft für Artificial Intelligence, Band 5*, Wien, 2004. [http://www.rrz.uni-hamburg.de/exmaralda/Daten/4D-Literatur/Paper_LREC.pdf]
- SCHMIDT, Thomas (2005): Time-based data models and the Text Encoding Initiative's guidelines for transcription of speech. En: *Arbeiten zur Mehrsprachigkeit (Working Papers in Multilingualism), Serie B (62)*. [Hamburg. http://www.rrz.uni-hamburg.de/exmaralda/Daten/4D-Literatur/SFB_AzM62.pdf]
- SELKIRK, Elizabeth O. (1984): *Phonology and Syntax. The relation between sound and structure*. Cambridge, MA: MIT
- SELTING, Margret (1995): *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer. (Linguistische Arbeiten, 329)
- SELTING, Margret; Auer, Peter; Barden, Birgit; Bergmann, Jörg; Couper Kuhlen, Elizabeth; Günther, Susanne; Quasthoff, Uta; Meier, Christoph; Schlobinski, Peter y Uhmann,

- Susanne (1998): Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173: 91-122.
- SILVA-FUENZALIDA, I. (1956-57): La entonación en el español y su morfología. En *Boletín de Filología*, IX, Santiago de Chile. Pp. 177-187.
- SOSA, Juan Manuel (1999): *La entonación del español. Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra.
- SOSA, Juan Manuel (2003): La notación tonal del español en el modelo Sp-ToBI. En Prieto, Pilar (Ed.): *Teorías de la entonación*. Barcelona: Ariel. Pp. 185-208.
- SPERBERG-McQUEEN, C. M. y Burnard, Lou (Eds.) (1994): Text Encoding Initiative. The XML Version of the TEI Guidelines [en línea] [Fecha de consulta: 21/09/2005] [<http://www.tei-c.org/Vault/GL/teip3doc.tar.gz>]
- SPERBERG-McQUEEN, C. M. y Burnard, Lou (Eds.) (1999): Guidelines for Electronic Text Encoding and Interchange [en línea]. Revisión final, publicada electrónicamente como "P4-beta", 05/1999: [Fecha de consulta: 21/09/2005] [<http://www.tei-c.org/Vault/GL/P3/>], [<http://www.tei-c.org/Vault/GL/p4beta.pdf>]
- SPERBERG-McQUEEN, C. M. y Burnard, Lou (Eds.) (Revisado y reeditado por Syd Bauman y Lou Burnard: 14/01/2005) (2004): TEI P5. Guidelines for Electronic Text Encoding and Interchange [en línea]. The TEI Consortium. [Fecha de consulta: 21/09/2005] [<http://www.tei-c.org/P5/Guidelines/>]
- STEYER, Ralph (2003): *JavaScript... in 21 Tagen*. München: Markt + Technik Verlag.
- STOCKWELL, Robert P.; Bowen, J. Donald y Silva-Fuenzalida, I. (1956): Spanish juncture and intonation. En *Language*, 32, núm. 4. Pp. 641-665.
- TOLEDO, Guillermo Andrés (1988): *El ritmo en el español. Estudio fonético con base computacional*. Madrid: Gredos.
- TRAGER, George L. y Smith, Henry Lee Jr. (1951): *An outline of English structure*. Washington: American Council of Learned Societies, 1957.
- WELLS, John (1989): Computer coded phonemic notation of the individual languages of the European Community. En *Journal of the International Phonetic Association*, 19: 32-54.
- WELLS, Rulon S. (1945): The pitch phonemes of English. En *Language*, 21: 27-39.
- WENNERSTROM, Ann (2001): *The Music of Everyday Speech. Prosody and Discourse Analysis*. Oxford: OUP.

6. Anexos

Anexo 1: Normas de transcripción en rejillas de texto (*text grids*) de Praat

Enunciados:	□//	Terminación declarativa
	□?	Terminación interrogativa
	□...	Terminación suspendida
Unidades tonales:	□/□	Límite de unidad tonal
Interrupciones:	□+	Autointerrupción a final de turno
	□+//	Interrupción por parte del interlocutor
Reinicios:	[/]	Repetición
	[//]	Reformulación parcial
	[///]	Reformulación total
Pausa:	#	
Comentarios:	[%act:□...]	Acciones de los participantes
	[%sit:□...]	Situación o evento relevante
	[%add:□...]	Interlocutor a quien se dirige el texto
	[%par:□...]	Gestos o eventos paralingüísticos
	[%com:□...]	Comentarios del transcriptor
	[%lan:□...]	Idioma alternativo
	[%rec:□-...-]	Reconstrucción de palabra
	(%alt:□...)	Propuestas normativas para errores de producción
	[%qlt:□...]	□(□...□) Calidad de voz del hablante
	[%amb:]	Ambiente
	[%sce:]	Escena
Producción ininteligible:	xxx	
Fragmento:	..-	
Solapamiento:	<□...□>	Texto solapado
	<[<]□<□...□>	Texto causante del solapamiento
Delimitador de sílabas:	.	Por ejemplo: pro.so.dia

Anexo 2: Definición de Tipo de Documento PROSODY_XML.DTD

```

<!-- PROSODY_XML.DTD Version 6 -->
<!-- Declaration of XML doc type -->
<!-- Declaration of DTD document -->

<!-- Date: 25th November 2004 -->
<!-- Last Update: 15th June 2007 -->
<!-- Written by Eduardo Velazquez -->

<!-- Root element -->
<!ELEMENT Transcript (Header?,Text)>

<!-- Declaration of Header-relevant tags -->
<!ELEMENT Header (Title|File|Data|Participants|Date|Place|Situation|Topic|Source|Class+|
Length|Words|Acoustic_quality|Transcriber|Revisor|Comments)*>

<!ELEMENT Title (#PCDATA)>
<!ELEMENT File (#PCDATA)>
<!ELEMENT Data (#PCDATA)>
<!ELEMENT Participants (Speaker+)>
  <!ELEMENT Speaker (#PCDATA|Name|Sex|Age|Education|Occupation|Role|Origin)*>
    <!ELEMENT Name (#PCDATA)>
    <!ELEMENT Sex (#PCDATA)>
    <!ATTLIST Sex Type (woman|man|x) #REQUIRED >
    <!ELEMENT Age (#PCDATA)>
    <!ATTLIST Age Type (A|B|C|D|x) #REQUIRED >
    <!ELEMENT Education (#PCDATA)>
    <!ATTLIST Education Type (1|2|3|x) #REQUIRED >
    <!ELEMENT Occupation (#PCDATA)>
    <!ELEMENT Role (#PCDATA)>
    <!ELEMENT Origin (#PCDATA)>
<!ELEMENT Date (#PCDATA)>
<!ELEMENT Place (#PCDATA)>
<!ELEMENT Situation (#PCDATA)>
<!ELEMENT Topic (#PCDATA)>
<!ELEMENT Source (#PCDATA)>
<!ELEMENT Class (#PCDATA)>
<!ATTLIST Class Type1 (CDATA|informal|formal) #REQUIRED >
<!ATTLIST Class Type2 (familiar-private|public|formal_in_natural_context|
media|telephone) #IMPLIED >
<!ATTLIST Class Type3 (monologue|dialogue|conversation) #IMPLIED >
<!ELEMENT Length (#PCDATA)>
<!ELEMENT Words (#PCDATA)>
<!ELEMENT Acoustic_quality (#PCDATA)>
<!ATTLIST Acoustic_quality Type (A|B|C) #REQUIRED >
<!ELEMENT Transcriber (#PCDATA)>
<!ELEMENT Revisor (#PCDATA)>
<!ELEMENT Comments (#PCDATA)>

<!-- Declaration of Text-relevant tags -->
<!ELEMENT Text (Turn+)>
<!ELEMENT Turn (#PCDATA|U|Pause|Comment)*>
<!ATTLIST Turn Name CDATA #REQUIRED>

<!ELEMENT Pause (#PCDATA)>
<!ATTLIST Pause Sec CDATA #IMPLIED
Type (s|l|x|1) #IMPLIED>

<!ELEMENT Comment (#PCDATA)>
<!-- Utterance-level comments -->
<!ATTLIST Comment act CDATA #IMPLIED
qlt CDATA #IMPLIED
sit CDATA #IMPLIED
add CDATA #IMPLIED
par CDATA #IMPLIED
com CDATA #IMPLIED
lan CDATA #IMPLIED
amb CDATA #IMPLIED
sce CDATA #IMPLIED>

```

```

<!-- U = Utterance -->
<!ELEMENT U      (#PCDATA|IU|Pause|Restart)*>
<!ATTLIST U      Type (nt|qst|emp|ssp|slf|ext|none) #IMPLIED
                Subtype CDATA #IMPLIED
                Quest (Y|N) #REQUIRED
                Emph (Y|N) #REQUIRED
                Susp (Y|N) #REQUIRED>

<!-- IU = Intonation Unit -->
<!ELEMENT IU      (#PCDATA|F|S|Overlap|Restart|Break|Pause|Fragment|
                Unintelligible|Interruption|Comment|Border)*>
<!ATTLIST IU      img      CDATA #IMPLIED
                id      CDATA #IMPLIED>

<!ELEMENT Overlap (#PCDATA|F|S|Fragment|Mora|Comment|Unintelligible)*>
<!ATTLIST Overlap Type (pas|act|cont) #IMPLIED
                Ref      CDATA #IMPLIED>

<!ELEMENT Restart      (#PCDATA)>
<!ATTLIST Restart      Type (rep|prt|tot) #REQUIRED>
<!ELEMENT Fragment      (#PCDATA)>
<!ELEMENT Unintelligible (#PCDATA)>
<!ELEMENT Interruption      EMPTY>
<!ATTLIST Interruption Type (slf|ext) #REQUIRED>
<!-- Tone-Unit-level comments -->
<!ATTLIST Comment      alt      CDATA #IMPLIED
                pho      CDATA #IMPLIED>

<!ELEMENT Border      (#PCDATA)>
<!ATTLIST Border      Type (slf|ext|iu|nt|qst|emp|ssp|rep|prt|tot|none) #REQUIRED
                Sec      CDATA #IMPLIED>

<!-- F = Foot -->
<!ELEMENT F      (#PCDATA|S|Break|Overlap)*>
<!ATTLIST F      Wt      CDATA #REQUIRED>

<!ELEMENT Break      EMPTY>
<!ATTLIST Break      Type (0|1|2|3|4) #REQUIRED>

<!-- S = Syllable -->
<!ELEMENT S      (#PCDATA|Mora|Break|Comment|Fragment|Unintelligible|Overlap)*>
<!ATTLIST S      Beg      CDATA #REQUIRED
                End      CDATA #REQUIRED
                Dur      CDATA #REQUIRED
                Durvar CDATA #REQUIRED
                Tmp (nt|all|l|accel|rall) #REQUIRED
                F0min CDATA #REQUIRED
                F0max CDATA #REQUIRED
                F0dBmax CDATA #REQUIRED
                F0      CDATA #REQUIRED
                F0var CDATA #REQUIRED
                F0vst CDATA #REQUIRED
                F0std CDATA #REQUIRED
                dBmin CDATA #REQUIRED
                dBmax CDATA #REQUIRED
                tdBmax CDATA #REQUIRED
                dB      CDATA #REQUIRED
                dBvar CDATA #REQUIRED
                Vol (nt|f|ff|p|pp|cresc|dim) #REQUIRED
                Phon CDATA #REQUIRED
                Orth CDATA #REQUIRED
                Stress (Y|N) #REQUIRED
                Focus (Y|N) #REQUIRED
                Tone CDATA #REQUIRED>
<!-- Spanish-ToBI: L*+H, L+H*, H+L*, H*, *, L*+!H, L+!H*, !L*+H, !L+H*,
L%, M%, H% - upstep must be marked with upside-down exclamation mark! -->

<!ELEMENT Mora      EMPTY>
<!ATTLIST Break      Type (0) #IMPLIED>
<!-- Syllable-level comments -->
<!ATTLIST Comment      rec      CDATA #IMPLIED>

<!-- End of PROSODY_XML.DTD, created by Eduardo Velazquez -->

```

Anexo 3: Definición de Tipo de Documento PROSODY_XML_SET.DTD

```

<!-- PROSODY_XML_SET.DTD Version 6 -->
<!-- Declaration of XML doc type -->
<!-- Declaration of DTD document -->

<!-- Date: 25th November 2004 -->
<!-- Last Update: 15th June 2007 -->
<!-- Written by Eduardo Velazquez -->

<!-- Root element -->
<ELEMENT Transcript (Header?,Set)>

<!-- Declaration of Header-relevant tags -->
<ELEMENT Header (Title|File|Data|Date|Place|Source|Transcriber|Revisor|Comments)*>

<ELEMENT Title (#PCDATA)>
<ELEMENT File (#PCDATA)>
<ELEMENT Data (#PCDATA)>
<ELEMENT Date (#PCDATA)>
<ELEMENT Source (#PCDATA)>
<ELEMENT Transcriber (#PCDATA)>
<ELEMENT Revisor (#PCDATA)>
<ELEMENT Comments (#PCDATA)>

<!-- Declaration of Sample Set-relevant tags -->
<ELEMENT Set (Sample+)>
<ELEMENT Sample (#PCDATA|U|Comment)*>
<ATTLIST Sample Name CDATA #REQUIRED>

<ELEMENT Comment (#PCDATA)>
<!-- Utterance-level comments -->
<ATTLIST Comment act CDATA #IMPLIED
                qlt CDATA #IMPLIED
                sit CDATA #IMPLIED
                add CDATA #IMPLIED
                par CDATA #IMPLIED
                com CDATA #IMPLIED
                lan CDATA #IMPLIED
                amb CDATA #IMPLIED
                sce CDATA #IMPLIED>

<!-- U = Utterance -->
<ELEMENT U (#PCDATA|IU|Pause|Restart)*>
<ATTLIST U Type (nt|qst|emp|ssp|slf|ext|none) #IMPLIED
           Subtype CDATA #IMPLIED
           Quest (Y|N) #REQUIRED
           Emph (Y|N) #REQUIRED
           Susp (Y|N) #REQUIRED>

<ELEMENT Pause (#PCDATA)>
<ATTLIST Pause Sec CDATA #IMPLIED
               Type (s|l|x|1) #IMPLIED>

<ELEMENT Restart (#PCDATA)>
<ATTLIST Restart Type (rep|prt|tot) #REQUIRED>

<!-- IU = Intonation Unit -->
<ELEMENT IU (#PCDATA|F|Pause|Restart|Fragment|Unintelligible|
            Interruption|Comment|Border)*>
<ATTLIST IU img CDATA #REQUIRED
            id CDATA #REQUIRED>

<ELEMENT Fragment (#PCDATA)>
<ELEMENT Unintelligible (#PCDATA)>
<ELEMENT Interruption EMPTY>
<ATTLIST Interruption Type (slf|ext) #REQUIRED>
<!-- Tone-Unit-level comments -->
<ATTLIST Comment alt CDATA #IMPLIED
                 pho CDATA #IMPLIED>

<ELEMENT Border (#PCDATA)>
<ATTLIST Border Type (iu|nt|qst|emp|ssp|rep|prt|tot|slf|ext|none) #REQUIRED
               Sec CDATA #IMPLIED>

```



```

<!-- F = Foot -->
<!ELEMENT F      (#PCDATA|S|Break)*>
<!ATTLIST F      Wt          CDATA          #REQUIRED>

<!ELEMENT Break          EMPTY>
<!ATTLIST Break          Type      (0|1|2|3|4) #REQUIRED>

<!-- S = Syllable -->
<!ELEMENT S      (#PCDATA|Mora|Break|Comment|Fragment|Unintelligible)*>
<!ATTLIST S      Beg          CDATA          #REQUIRED
                End          CDATA          #REQUIRED
                Dur          CDATA          #REQUIRED
                Durvar       CDATA          #REQUIRED
                Tmp (nt|all|l|accel|rall) #REQUIRED
                F0min        CDATA          #REQUIRED
                F0max        CDATA          #REQUIRED
                F0dBmax      CDATA          #REQUIRED
                F0           CDATA          #REQUIRED
                F0var        CDATA          #REQUIRED
                F0vst        CDATA          #REQUIRED
                F0std        CDATA          #REQUIRED
                dBmin        CDATA          #REQUIRED
                dBmax        CDATA          #REQUIRED
                tdBmax       CDATA          #REQUIRED
                dB           CDATA          #REQUIRED
                dBvar        CDATA          #REQUIRED
                Vol (nt|f|ff|p|pp|cresc|dim) #REQUIRED
                Phon        CDATA          #REQUIRED
                Orth        CDATA          #REQUIRED
                Stress (Y|N) #REQUIRED
                Focus (Y|N) #REQUIRED
                Tone        CDATA          #REQUIRED>
<!-- Spanish-ToBI: L*+H, L+H*, H+L*, H*, *, L*+!H, L+!H*, !L*+H, !L+H*,
L%, M%, H% - upstep must be marked with upside-down exclamation mark! -->

<!ELEMENT Mora          EMPTY>
<!ATTLIST Break          Type      (0)          #REQUIRED>
<!-- Syllable-level comments -->
<!ATTLIST Comment          rec      CDATA          #IMPLIED>

<!-- End of PROSODY_XML_SET.DTD, created by Eduardo Velazquez -->

```

Anexo 4: Glosario de elementos de PROSODY_XML.DTD y PROSODY_XML_SET.DTD

<Acoustic_quality>

```
<Acoustic_quality Type='B' />
<Acoustic_quality Type='B'>B</Acoustic_quality>
```

<Age>

```
<Age Type="B" />
<Age Type="B">B</Age> ,
```

<Border>

```
<Border Type='iu' Sec='0.725'> //</Border>
<Border Type='qst'> ?</Border>
```

<Break>

```
<Break Type='4' />
y.<Break Type='0' />e1
```

<Class>

```
<Class Type1='informal, public, conversation' />
<Class Type1='informal' Type2='public' Type3='conversation' />
<Class Type1='informal' Type2='public' Type3='conversation'>informal, public,
  conversation</Class>
```

<Comment>

```
p<Comment rec='-ue-' />s
<Comment sit='away from microphone' />
```

<Comments>

```
<Comments>Eventual microphone interferences.</Comments>
```

<Data>

```
<Data>Hotel_1.wav, df1-04-1.doc, df1-04-1.mp3</Data>
```

<Date>

```
<Date>05/08/2004</Date>
```

<Education>

```
<Education Type="2" />
<Education Type="2">2</Education> ,
```

<F>

```
<F Wt='s'></F>
<F Wt='W )'></F>
```

<File>

```
<File>df1-04-1</File>
```

<Fragment>

```
m-<Fragment />
```

<Header>

```
<Header></Header>
```

<Interruption>

```
<U><IU><Interruption Type='slf' /></IU></U>
  <Border Type='slf'> +</Border>
<U><IU><Interruption Type='ext' /></IU></U>
  <Border Type='ext'> +</Border>
```

<IU>

```
<IU img="df92_1973.86-1974.69_HUI_qst" id="df92_1973.86-1974.69_HUI_qst"></IU>
```

<Length>

```
<Length>2&#039;00&#034;</Length>
```

<Mora>

```
<S Beg="0.050" End="0.128" Dur="0.078" Durvar="0" Tmp="nt" F0min="361" F0max="361"
  F0dBmax="361" F0="361" F0var="0" F0vst="0" F0std="100" dBmin="72" dBmax="78" tdBmax="0.123"
  dB="74" dBvar="0" Vol="nt" Stress="Y" Focus="N" Tone="H%" Phon="sí" Orth="sí~">
  sí~<Mora/>
</S>
<Break Type="0"/>
```

<Name>

```
<Name>Miriam</Name>
```

<Occupation>

```
<Occupation>hotel receptionist,</Occupation>
```

<Origin>

```
<Origin>Mexico City&#041;</Origin>
```

<Overlap>

```
<Overlap Type="pas" Ref="1">&lt; digital &gt;</Overlap>
<Overlap Type="act" Ref="1">[&lt;] &lt; de la &gt;</Overlap>
```

<Participants>

```
<Participants><Speaker></Speaker></Participants>
```

<Pause>

```
<Pause/>
<Pause Sec='1.90' />
<Pause Sec='1.90' Type='l' />
```

<Place>

```
<Place>Hotel Isabel la Católica, Mexico City</Place>
```

<Restart>

```
<F Wt="w )">
  <S Beg="1.720" End="1.894" Dur="0.174" Durvar="-43.93" Tmp="all" F0min="171" F0max="172"
    F0dBmax="172" F0="170" F0var="-8.11" F0vst="-0.97" F0std="80" dBmin="70" dBmax="78"
    tdBmax="1.889" dB="74" dBvar="3.97" Vol="cresc" Stress="N" Focus="N" Tone="M%" Phon="el"
    Orth="el [ / ]">el [ / ]</S>
  <Break Type="3" />
</F>
<Restart Type="rep" />
<Border Type="rep" />
```

<Revisor>

```
<Revisor>Eduardo; Eduardo (prosody)</Revisor>
```

<Role>

```
<Role>participant</Role>,</Role>
```

<S>

```
<S Beg="0.168" End="0.274" Dur="0.106" Durvar="0" Tmp="nt" F0min="129" F0max="154"
  F0dBmax="153" F0="129" F0var="0" F0vst="0" F0std="100" dBmin="77" dBmax="88" tdBmax="0.268"
  dB="84" dBvar="0" Vol="nt" Stress="Y" Focus="N" Tone="H% L*+H" Phon="kó" Orth="có">có</S>
```

<Sex>

```
<Sex Type="woman" /> (
<Sex Type="woman">woman</Sex>,</Sex>
```

<Situation>

```
<Situation>Receptionists at Hotel Isabel talking alternatively with Interviewer when not
  attending hotel guests. Not-hidden, researcher present.</Situation>
```

<Source>

<Source>DF 2004 (own material)</Source>

<Speaker>

<Speaker>ADR</Speaker>,

<Text>

<Text><Turn></Turn></Text>

<Title>

<Title>Las dos recepcionistas</Title>

<Topic>

<Topic>Touristic places in Mexico City and daily life</Topic>

<Transcriber>

<Transcriber>Eduardo</Transcriber>

<Transcription>

<Transcription><Text></Text></Transcription>

<Turn>

<Turn Name='ADR'>

<U>

<U Quest="Y" Emph="N" Susp="N">

<Unintelligible>

<Unintelligible>xxx</Unintelligible>

<Words>

<Words>302</Words>

Anexo 5: Subprograma `ExtractTextGrid+Sound+XML.praat`

Líneas	Secciones
20 - 38	Definición del formulario
42 - 50	Selección y asociación de TextGrid y LongSound
54 - 97	Impresión del encabezado del documento de XML
101 - 104	Inicio del ciclo de extracción de información de cada una de las hileras
109 - 115	Inicio del ciclo de inspección y asociación de cada uno de los intervalos en la hilera
111 - 112	Asignación de los puntos inicial y final del intervalo actual [<code>st</code> y <code>et</code>]
123 - 187	Asignación de variable para el tipo de frontera del último intervalo que no esté vacío [<code>border0\$</code>]
191 - 256	Asignación de variable para el tipo de frontera del penúltimo intervalo que no esté vacío [<code>border00\$</code>]
260 - 338	Asignación de variable para el tipo de frontera del intervalo actual que no esté vacío [<code>border1\$</code>]
342 - 378	Inspección de hasta ocho intervalos consecutivos para localizar intervalos vacíos y transcritos
386 - 462	Asociación de intervalos consecutivos sin vacíos intermedios
386 - 462	Asignación de variable para los puntos inicial y final del grupo asociado de intervalos [<code>t1</code> y <code>t2</code>]
386 - 462	Asignación de variable para los puntos inicial y final del intervalo vacío después del grupo [<code>t4</code> y <code>t5</code>]
467 - 468	Adición de 0.05 segundos antes y después de los límites del grupo [<code>t0</code> y <code>t3</code>]
472 - 548	Asignación de variable para el tipo de frontera del grupo [<code>border\$</code>]
553 - 612	Estandarización de formato de las variables de tiempo con ceros a la izquierda
617 - 620	Extracción del fragmento de TextGrid entre <code>t0</code> y <code>t3</code> , registro con nombre especial
625 - 631	Extracción la hilera actual del fragmento de TextGrid y registro con nombre especial
636 - 643	Extracción del fragmento correspondiente del archivo de audio y registro con nombre especial
647 - 653	Extracción de la frecuencia (Pitch) del archivo de audio actual y cálculo de la frecuencia mínima y máxima
657 - 674	Redondeo de la frecuencia mínima para establecer una base más holgada para las imágenes
678 - 702	Redondeo de la frecuencia máxima para establecer una cima más holgada para las imágenes
706 - 707	Extracción del espectrograma (Spectrogram) del archivo de audio actual

- 711 - 791 Creación de imágenes
- 795 - 800 Escritura de la imagen a archivos
- 804 - 805 Limpieza de la ventana de imágenes antes de pasar al siguiente intervalo
- 809 - 1216 Extracción de información del TextGrid para crear un archivo compatible con XML
- 815 - 837 Impresión del inicio de la unidad entonativa, enunciado o turno de palabra, tomando en cuenta la frontera de los intervalos previos
- 843 - 904 Reemplazo de signos al inicio y dentro del intervalo actual por código compatible con XML
- 912 - 1148 Reemplazo de signos al final del intervalo actual por código compatible con XML (tipo de frontera, final de unidad entonativa, final de enunciado, final de turno de palabra - seguido inmediatamente de pausa)
- 912 - 1148 Etiquetado explícito de pausas únicamente cuando sean mayores a 1 segundo
- 1153 - 1215 Reemplazo de signos al final del intervalo actual por código compatible con XML (para intervalos que no estén seguidos inmediatamente de pausa)
- 1221 - 1242 Eliminación de archivos generados por los análisis automáticos según las tareas seleccionadas (Extract TextGrids, Extract Sounds, Create Pictures, Create XML)
- 1246 - 1247 Selección de los archivos TextGrid y LongSound para analizar el siguiente intervalo
- 1249 Fin del ciclo de intervalos
- 1253 Fin del ciclo de hileras
- 1259 - 1261 Impresión del final del documento de XML
- 1265 - 1268 Escritura del contenido de la Info Window a un archivo de texto con terminación .XML
- 1272 Fin del subprograma

```

01 #+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----#
02 #
03 # This program was last updated on 26th February, 2007 by Eduardo Velázquez
04 #
05 # ExtractTextGrid+Sound+XML.praat will display a form where the user should
06 # 1) provide a unique ID to be used as a prefix and 2) the directory where
07 # files will be saved, 3) check the desired output files (TextGrids, Sound
08 # files, pictures and/or XML file), 4) set a minimum of metainformation
09 # about the conversation.
10 # After this, one long sound file and its corresponding text grid file
11 # (basic transcripts where intervals correspond to utterances and pauses
12 # and each speaker's interventions are kept in different tiers) must be
13 # open and selected in order to allow the program to extract the information
14 # into the desired output files.
15 #
16 #+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----#
17
18 # -> Definition of Form
19
20 form Extract transcribed intervals as TextGrid and Sound files
21   comment Enter unique ID to be used as a prefix:
22     sentence prefix df
23   comment Enter directory where files will be saved:
24     sentence dir C:\
25     boolean extract_TextGrids no
26     boolean extract_Sounds no
27     boolean create_pictures no
28     boolean create_XML_file yes

```

```

29     sentence source_file df-2004
30     comment Title:
31     text title
32     sentence source DF-2004
33     sentence transcriber Eduardo
34     sentence revisor Eduardo; Eduardo (prosody)
35     comment Comments:
36     text comments
37     boolean kill_octave_jumps no
38 endform
39
40 # -> Selection and association of TextGrid and LongSound
41
42 Scale times
43
44 name$ = selected$ ("TextGrid")
45 sound$ = selected$ ("LongSound")
46
47 select TextGrid 'name$'
48 n_tier = Get number of tiers
49 total_duration$ = Get total duration
50 total_length$ = total_duration$ - " seconds"
51
52 # -> Print Header of XML document
53
54 if 'create_XML_file' = 1
55 clearinfo
56
57 printline <?xml version='1.0' encoding='ISO-8859-1' standalone='no' ?>
58 ...'newline$'<!-- Written by Eduardo Velázquez -->
59 ...'newline$'<?xml-stylesheet type='text/xsl' href='Prosody.xsl'?>
60 ...'newline$'<!DOCTYPE Transcript SYSTEM 'PROSODY_XML.DTD'>
61 ...'newline$'<Transcript>
62 ...'newline$'<Header>
63 ...'newline$'<Title>'title$'</Title>
64 ...'newline$'<File>'source_file$'</File>
65 ...'newline$'<Data>Other files</Data>
66 ...'newline$'<Participants>
67 for i_tier to n_tier
68 tier_name$ = Get tier name... i_tier
69 printline <Speaker>'tier_name$',
70 ...'newline$'<Name>'tier_name$'</Name>
71 ...'newline$'(<Sex Type='x'>x</Sex>,
72 ...'newline$'<Age Type='x'>x</Age>,
73 ...'newline$'<Education Type='x'>x</Education>,
74 ...'newline$'<Occupation>x</Occupation>,
75 ...'newline$'<Role>x</Role>,
76 ...'newline$'<Origin>x</Origin>)
77 ...'newline$'</Speaker>
78 ...'newline$'</Participants>
79 endfor
80 printline </Participants>
81 ...'newline$'<Date>Date</Date>
82 ...'newline$'<Place>Place</Place>
83 ...'newline$'<Situation>Situation</Situation>
84 ...'newline$'<Topic>Topic</Topic>
85 ...'newline$'<Source>'source$'</Source>
86 ...'newline$'<Class Type1='informal' Type2='public'
87 ...Type3='conversation'>informal, public, conversation</Class>
88 ...'newline$'<Length>'total_length$'</Length>
89 ...'newline$'<Words>Words</Words>
90 ...'newline$'<Acoustic_quality Type='A'>A</Acoustic_quality>
91 ...'newline$'<Transcriber>'transcriber$'</Transcriber>
92 ...'newline$'<Revisor>'revisor$'</Revisor>
93 ...'newline$'<Comments>'comments$'</Comments>
94 ...'newline$'</Header>
95 ...'newline$'<Text>
96
97 endif
98
99 # -> Loop to extract the desired information from all tiers
100
101 for i_tier to n_tier
102 select TextGrid 'name$'
103 tier_name$ = Get tier name... i_tier
104 n = Get number of intervals... i_tier
105
106 # -> Loop to inspect and associate all intervals in tier
107 # -> Set starting and end point of current interval [st and et]

```

```

108
109   for i from 2 to n-1
110       select TextGrid 'name$'
111         st = Get starting point... i_tier i
112         et = Get end point... i_tier i
113         i_label$ = Get label of interval... i_tier i
114         label0$ = Get label of interval... i_tier i-1
115         label2$ = Get label of interval... i_tier i+1
116
117   # -> Set last interval's type of border if not void [border0$]
118   # -> (nt = neutral, assertive, declarative; ext = externally interrupted;
119   # -> slf = self-interrupted; qst = question, exclamative; emp = emphatic,
120   # -> exclamative; ssp = suspended; iu = intonation unit; rep = repetition;
121   # -> prt = partial restart; tot = total restart)
122
123       if label0$ <> ""
124           nt$ = " //"
125           when = endsWith (label0$, nt$)
126           if when = 1
127               border0$ = "nt"
128           endif
129
130           ext$ = " +/"
131           when = endsWith (label0$, ext$)
132           if when = 1
133               border0$ = "ext"
134           endif
135
136           slf$ = " +"
137           when = endsWith (label0$, slf$)
138           if when = 1
139               border0$ = "slf"
140           endif
141
142           qst$ = " ?"
143           when = endsWith (label0$, qst$)
144           if when = 1
145               border0$ = "qst"
146           endif
147
148           emp$ = " !"
149           when = endsWith (label0$, emp$)
150           if when = 1
151               border0$ = "emp"
152           endif
153
154           ssp$ = " ..."
155           when = endsWith (label0$, ssp$)
156           if when = 1
157               border0$ = "ssp"
158           endif
159
160           iu$ = " /"
161           when = endsWith (label0$, iu$)
162           if when = 1
163               border0$ = "iu"
164           endif
165
166           rep$ = " [/"
167           when = endsWith (label0$, rep$)
168           if when = 1
169               border0$ = "rep"
170           endif
171
172           prt$ = " [//]"
173           when = endsWith (label0$, prt$)
174           if when = 1
175               border0$ = "prt"
176           endif
177
178           tot$ = " [///]"
179           when = endsWith (label0$, tot$)
180           if when = 1
181               border0$ = "tot"
182           endif
183       else
184           border0$ = ""
185       border00$ = ""
186

```



```

187     endif
188
189 # -> Set previous-to-last interval's type of border if not void [border00$]
190
191     if i >= 3 and i < n-1
192         label00$ = Get label of interval... i_tier i-2
193
194         if label0$ = "" and label00$ <> ""
195             when = endsWith (label00$, nt$)
196             if when = 1
197                 border00$ = "nt"
198             endif
199
200             ext$ = " +/"
201             when = endsWith (label00$, ext$)
202             if when = 1
203                 border00$ = "ext"
204             endif
205
206             slf$ = " +"
207             when = endsWith (label00$, slf$)
208             if when = 1
209                 border00$ = "slf"
210             endif
211
212             qst$ = " ?"
213             when = endsWith (label00$, qst$)
214             if when = 1
215                 border00$ = "qst"
216             endif
217
218             emp$ = "!"
219             when = endsWith (label00$, emp$)
220             if when = 1
221                 border00$ = "emp"
222             endif
223
224             ssp$ = " .."
225             when = endsWith (label00$, ssp$)
226             if when = 1
227                 border00$ = "ssp"
228             endif
229
230             iu$ = " /"
231             when = endsWith (label00$, iu$)
232             if when = 1
233                 border00$ = "iu"
234             endif
235
236             rep$ = " [/]"
237             when = endsWith (label00$, rep$)
238             if when = 1
239                 border00$ = "rep"
240             endif
241
242             prt$ = " [//]"
243             when = endsWith (label00$, prt$)
244             if when = 1
245                 border00$ = "prt"
246             endif
247
248             tot$ = " [///]"
249             when = endsWith (label00$, tot$)
250             if when = 1
251                 border00$ = "tot"
252             endif
253         else
254             border00$ = ""
255         endif
256     endif
257
258 # -> Set current interval's type of border if not void [border1$]
259
260     if i_label$ <> ""
261         nt$ = " //"
262         when = endsWith (i_label$, nt$)
263         if when = 1
264             border1$ = "nt"

```

```

265     u_quest1$ = "N"
266     u_emph1$ = "N"
267     u_suspl$ = "N"
268   endif
269
270
271   ext$ = " +/"
272   when = endsWith (i_label$, ext$)
273   if when = 1
274     border1$ = "ext"
275     u_quest1$ = "N"
276     u_emph1$ = "N"
277     u_suspl$ = "Y"
278   endif
279
280   slf$ = " +"
281   when = endsWith (i_label$, slf$)
282   if when = 1
283     border1$ = "slf"
284     u_quest1$ = "N"
285     u_emph1$ = "N"
286     u_suspl$ = "Y"
287   endif
288
289   qst$ = " ?"
290   when = endsWith (i_label$, qst$)
291   if when = 1
292     border1$ = "qst"
293     u_quest1$ = "Y"
294     u_emph1$ = "N"
295     u_suspl$ = "N"
296   endif
297
298   emp$ = "!"
299   when = endsWith (i_label$, emp$)
300   if when = 1
301     border1$ = "emp"
302     u_quest1$ = "N"
303     u_emph1$ = "Y"
304     u_suspl$ = "N"
305   endif
306
307   ssp$ = " ..."
308   when = endsWith (i_label$, ssp$)
309   if when = 1
310     border1$ = "ssp"
311     u_quest1$ = "N"
312     u_emph1$ = "N"
313     u_suspl$ = "Y"
314   endif
315
316   iu$ = " /"
317   when = endsWith (i_label$, iu$)
318   if when = 1
319     border1$ = "iu"
320   endif
321
322   rep$ = " [/"
323   when = endsWith (i_label$, rep$)
324   if when = 1
325     border1$ = "rep"
326   endif
327
328   prt$ = " [/"
329   when = endsWith (i_label$, prt$)
330   if when = 1
331     border1$ = "prt"
332   endif
333
334   tot$ = " [/"
335   when = endsWith (i_label$, tot$)
336   if when = 1
337     border1$ = "tot"
338   endif
339
340 # -> Inspect up to 8 following intervals to locate void and unvoid labels
341
342   if n - i >= 8
343     label9$ = Get label of interval... i_tier i+8

```

```

344     label8$ = Get label of interval... i_tier i+7
345     label7$ = Get label of interval... i_tier i+6
346     label6$ = Get label of interval... i_tier i+5
347     label5$ = Get label of interval... i_tier i+4
348     label4$ = Get label of interval... i_tier i+3
349     label3$ = Get label of interval... i_tier i+2
350   elseif n - i >= 7
351     label8$ = Get label of interval... i_tier i+7
352     label7$ = Get label of interval... i_tier i+6
353     label6$ = Get label of interval... i_tier i+5
354     label5$ = Get label of interval... i_tier i+4
355     label4$ = Get label of interval... i_tier i+3
356     label3$ = Get label of interval... i_tier i+2
357   elseif n - i >= 6
358     label7$ = Get label of interval... i_tier i+6
359     label6$ = Get label of interval... i_tier i+5
360     label5$ = Get label of interval... i_tier i+4
361     label4$ = Get label of interval... i_tier i+3
362     label3$ = Get label of interval... i_tier i+2
363   elseif n - i >= 5
364     label6$ = Get label of interval... i_tier i+5
365     label5$ = Get label of interval... i_tier i+4
366     label4$ = Get label of interval... i_tier i+3
367     label3$ = Get label of interval... i_tier i+2
368   elseif n - i >= 4
369     label5$ = Get label of interval... i_tier i+4
370     label4$ = Get label of interval... i_tier i+3
371     label3$ = Get label of interval... i_tier i+2
372   elseif n - i >= 3
373     label4$ = Get label of interval... i_tier i+3
374     label3$ = Get label of interval... i_tier i+2
375   elseif n - i >= 2
376     label3$ = Get label of interval... i_tier i+2
377   endif
378
379
380 # -> Associate intervals without intermediary void intervals (this is the
381 # -> reason why pauses must not be marked extra in text grid)
382
383 # -> Set starting and end point of associated groups of intervals [t1 and t2]
384 # -> Set starting and end point of void interval after group [t4 and t5]
385
386   if n - i >= 8 and label0$ = "" and label2$ <> "" and label3$ <> ""
387     ...and label4$ <> "" and label5$ <> "" and label6$ <> ""
388     ...and label7$ <> "" and label8$ = ""
389     t1 = Get starting point... i_tier i
390     t2 = Get end point... i_tier i+7
391     t4 = Get starting point... i_tier i+8
392     t5 = Get end point... i_tier i+8
393     durnext = t5 - t4
394     label$ = label8$
395     labels$ = "'i_label$' 'label2$' 'label3$' 'label4$' 'label5$'
396     ...'label6$' 'label7$'"
397   elseif n - i >= 7 and label0$ = "" and label2$ <> "" and label3$ <> ""
398     ...and label4$ <> "" and label5$ <> "" and label6$ <> ""
399     ...and label7$ <> "" and label8$ = ""
400     t1 = Get starting point... i_tier i
401     t2 = Get end point... i_tier i+6
402     t4 = Get starting point... i_tier i+7
403     t5 = Get end point... i_tier i+7
404     durnext = t5 - t4
405     label$ = label7$
406     labels$ = "'i_label$' 'label2$' 'label3$' 'label4$' 'label5$'
407     ...'label6$' 'label7$'"
408   elseif n - i >= 6 and label0$ = "" and label2$ <> "" and label3$ <> ""
409     ...and label4$ <> "" and label5$ <> "" and label6$ <> ""
410     ...and label7$ = ""
411     t1 = Get starting point... i_tier i
412     t2 = Get end point... i_tier i+5
413     t4 = Get starting point... i_tier i+6
414     t5 = Get end point... i_tier i+6
415     durnext = t5 - t4
416     label$ = label6$
417     labels$ = "'i_label$' 'label2$' 'label3$' 'label4$' 'label5$'
418     ...'label6$'"
419   elseif n - i >= 5 and label0$ = "" and label2$ <> "" and label3$ <> ""
420     ...and label4$ <> "" and label5$ <> "" and label6$ = ""
421     t1 = Get starting point... i_tier i
422     t2 = Get end point... i_tier i+4

```

```

423     t4 = Get starting point... i_tier i+5
424     t5 = Get end point... i_tier i+5
425     durnext = t5 - t4
426     label$ = label5$
427     labels$ = "'i_label$' 'label2$' 'label3$' 'label4$' 'label5$'"
428     elsif n - i >= 4 and label0$ = "" and label2$ <> "" and label3$ <> ""
429     ...and label4$ <> "" and label5$ = ""
430         t1 = Get starting point... i_tier i
431         t2 = Get end point... i_tier i+3
432         t4 = Get starting point... i_tier i+4
433         t5 = Get end point... i_tier i+4
434         durnext = t5 - t4
435         label$ = label4$
436         labels$ = "'i_label$' 'label2$' 'label3$' 'label4$'"
437     elsif n - i >= 3 and label0$ = "" and label2$ <> "" and label3$ <> ""
438     ...and label4$ = ""
439         t1 = Get starting point... i_tier i
440         t2 = Get end point... i_tier i+2
441         t4 = Get starting point... i_tier i+3
442         t5 = Get end point... i_tier i+3
443         durnext = t5 - t4
444         label$ = label3$
445         labels$ = "'i_label$' 'label2$' 'label3$'"
446     elsif n - i >= 2 and label0$ = "" and label2$ <> "" and label3$ = ""
447         t1 = Get starting point... i_tier i
448         t2 = Get end point... i_tier i+1
449         t4 = Get starting point... i_tier i+2
450         t5 = Get end point... i_tier i+2
451         durnext = t5 - t4
452         label$ = label2$
453         labels$ = "'i_label$' 'label2$'"
454     elsif n - i >= 1 and label0$ = "" and label2$ = ""
455         t1 = Get starting point... i_tier i
456         t2 = Get end point... i_tier i
457         t4 = Get starting point... i_tier i+1
458         t5 = Get end point... i_tier i+1
459         durnext = t5 - t4
460         label$ = i_label$
461         labels$ = "'i_label$'"
462     endif
463
464 # -> Add 0.05 seconds before and after group boundaries [t0 and t3]
465
466     t0 = t1 - 0.05
467     t3 = t2 + 0.05
468
469
470 # -> Set current group's type of border [border$]
471
472     nt$ = " //"
473     when = endsWith (label$, nt$)
474     if when = 1
475         border$ = "nt"
476         u_quest$ = "N"
477         u_emph$ = "N"
478         u_susp$ = "N"
479     endif
480
481     ext$ = " +/"
482     when = endsWith (label$, ext$)
483     if when = 1
484         border$ = "ext"
485         u_quest$ = "N"
486         u_emph$ = "N"
487         u_susp$ = "Y"
488     endif
489
490     slf$ = " +"
491     when = endsWith (label$, slf$)
492     if when = 1
493         border$ = "slf"
494         u_quest$ = "N"
495         u_emph$ = "N"
496         u_susp$ = "Y"
497     endif
498
499     qst$ = " ?"
500     when = endsWith (label$, qst$)
501     if when = 1

```

```

502     border$ = "qst"
503     u_quest$ = "Y"
504     u_emph$ = "N"
505     u_susp$ = "N"
506   endif
507
508   emp$ = "!"
509   when = endsWith (label$, emp$)
510   if when = 1
511     border$ = "emp"
512     u_quest$ = "N"
513     u_emph$ = "Y"
514     u_susp$ = "N"
515   endif
516
517   ssp$ = " ..."
518   when = endsWith (label$, ssp$)
519   if when = 1
520     border$ = "ssp"
521     u_quest$ = "N"
522     u_emph$ = "N"
523     u_susp$ = "Y"
524   endif
525
526   iu$ = "/"
527   when = endsWith (label$, iu$)
528   if when = 1
529     border$ = "iu"
530   endif
531
532   rep$ = "[/]"
533   when = endsWith (label$, rep$)
534   if when = 1
535     border$ = "rep"
536   endif
537
538   prt$ = "[//]"
539   when = endsWith (label$, prt$)
540   if when = 1
541     border$ = "prt"
542   endif
543
544   tot$ = "[///]"
545   when = endsWith (label$, tot$)
546   if when = 1
547     border$ = "tot"
548   endif
549
550 # -> Fill time variables with zeros to the left in order to normalize
551 # -> figures to four-digit whole numbers and two decimals (####.##)
552
553   if t1 < 10
554     t1$ = "000't1:2'"
555   elseif t1 > 10 and t1 < 100
556     t1$ = "00't1:2'"
557   elseif t1 > 100 and t1 < 1000
558     t1$ = "0't1:2'"
559   elseif t1 > 1000
560     t1$ = "'t1:2'"
561   endif
562
563   if t2 < 10
564     t2$ = "000't2:2'"
565   elseif t2 > 10 and t2 < 100
566     t2$ = "00't2:2'"
567   elseif t2 > 100 and t2 < 1000
568     t2$ = "0't2:2'"
569   elseif t2 > 1000
570     t2$ = "'t2:2'"
571   endif
572
573
574   if t4 < 10
575     t4$ = "000't4:2'"
576   elseif t4 > 10 and t4 < 100
577     t4$ = "00't4:2'"
578   elseif t4 > 100 and t4 < 1000
579     t4$ = "0't4:2'"
580   elseif t4 > 1000

```

```

581     t4$ = "'t4:2'"
582     endif
583
584     if t5 < 10
585         t5$ = "000't5:2'"
586     elseif t5 > 10 and t5 < 100
587         t5$ = "00't5:2'"
588     elseif t5 > 100 and t5 < 1000
589         t5$ = "0't5:2'"
590     elseif t5 > 1000
591         t5$ = "'t5:2'"
592     endif
593
594     if st < 10
595         st$ = "000'st:2'"
596     elseif st > 10 and st < 100
597         st$ = "00'st:2'"
598     elseif st > 100 and st < 1000
599         st$ = "0'st:2'"
600     elseif st > 1000
601         st$ = "'st:2'"
602     endif
603
604     if et < 10
605         et$ = "000'et:2'"
606     elseif et > 10 and et < 100
607         et$ = "00'et:2'"
608     elseif et > 100 and et < 1000
609         et$ = "0'et:2'"
610     elseif et > 1000
611         et$ = "'et:2'"
612     endif
613
614 # -> Extract TextGrid fragment between t0 and t3 and save with special name
615 # -> (prefix_normalizedStartingPoint-normalizedEndPoind_NUMBERofTiers_border)
616
617     if 'extract_TextGrids' = 1
618         Extract part... t0 t3 no
619         Write to text file... 'dir$'\HTML\'prefix$'_t1$'-
620         't2$'_n_tier'tiers_'border$'.TextGrid
621
622 # -> Extract current tier from TextGrid fragment and save with special name
623 # -> (prefix_normalizedStartingPoint-normalizedEndPoind_NAMEofTier_border)
624
625     select TextGrid 'name$'_part
626     Extract tier... i_tier
627     Into TextGrid
628
629     Write to text file... 'dir$'\HTML\'prefix$'_t1$'-
630     't2$'_tier_name$'_border$'.TextGrid
631     endif
632
633 # -> Extract corresponding Sound file fragment and save with special name
634 # -> (prefix_normalizedStartingPoint-normalizedEndPoind_NAMEofTier_border)
635
636     if 'extract_Sounds' = 1
637     select LongSound 'sound$'
638     Extract part... t0 t3 no
639     Write to WAV file... 'dir$'\HTML\'prefix$'_t1$'-
640     't2$'_tier_name$'_border$'.wav
641     Write to WAV file... 'dir$'\HTML\Sounds\'prefix$'_t1$'-
642     't2$'_tier_name$'_border$'.wav
643     endif
644
645 # -> Extract Pitch from current Sound file and get Pitch minimum and maximum
646
647     if 'create_pictures' = 1
648     To Pitch... 0 50 600
649     if 'kill_octave_jumps' = 1
650         Kill octave jumps
651     endif
652     min = Get minimum... 0 0 Hertz Parabolic
653     max = Get maximum... 0 0 Hertz Parabolic
654
655 # -> Round down Pitch minimum to set a looser base for pictures
656
657     if min > 0 and min < 75
658         base = 50

```

```

659     elsif min > 75 and min < 110
660         base = 75
661     elsif min > 110 and min < 160
662         base = 100
663     elsif min > 160 and min < 210
664         base = 150
665     elsif min > 210 and min < 260
666         base = 200
667     elsif min > 260 and min < 310
668         base = 250
669     elsif min > 310 and min < 360
670         base = 300
671     else
672         base = 75
673     endif
674
675
676 # -> Round up Pitch maximum to set a looser top for pictures
677
678     if max > 0 and max < 75
679         top = 75
680     elsif max > 75 and max < 90
681         top = 100
682     elsif max > 90 and max < 140
683         top = 150
684     elsif max > 140 and max < 190
685         top = 200
686     elsif max > 190 and max < 240
687         top = 250
688     elsif max > 240 and max < 290
689         top = 300
690     elsif max > 290 and max < 340
691         top = 350
692     elsif max > 340 and max < 390
693         top = 400
694     elsif max > 390 and max < 440
695         top = 450
696     elsif max > 440 and max < 490
697         top = 500
698     elsif max > 490 and max < 540
699         top = 550
700     elsif max > 540
701         top = 600
702     endif
703
704 # -> Extract Spectrogram from current Sound file
705
706     select Sound 'name$'
707     To Spectrogram... 0.05 600 0.002 20 Gaussian
708
709 # -> Create Pictures
710
711     # -> Specify font type size, color
712
713     Times
714     Font size... 12
715     Black
716
717     # -> Define size and position of Waveform
718     # -> (by specifying grid coordinates)
719
720     Viewport... 0 9 0 2
721
722     # -> Draw Waveform
723
724     select Sound 'name$'
725     Draw... 0 0 0 0 no curve
726
727     # -> Define size and position of Spectrogram
728
729     Viewport... 0 9 1 4
730
731     # -> Draw Spectrogram
732
733     select Spectrogram 'name$'
734     Paint... 0 0 'base' 'top' 100 yes 50 5 0 no
735
736     # -> Draw Pitch curve
737     # -> First as a thick white line

```

```

738         select Pitch 'name$'
739         Line width... 10
740         White
741         Draw... 0 0 'base' 'top' no
742
743     # -> then as a thin black line
744
745         Line width... 3
746         Black
747         Draw... 0 0 'base' 'top' no
748
749     # -> Label y axis
750
751         Line width... 1
752         One mark left... 'top' no no yes
753         One mark left... 'base' no no yes
754         Marks left every... 1 50 yes yes no
755         Font size... 15
756         Text left... yes Fundamental frequency (Hz)
757
758     # -> Define size and position of TextGrid
759
760         Viewport... 0 9 1 5
761         Font size... 12
762
763     # -> Draw TextGrid
764
765         select TextGrid grid
766         Draw... 0 0 yes yes no
767
768     # -> Define size and position of inner box
769
770         Viewport... 0 9 0 5
771
772     # -> Draw inner box
773
774         Black
775         Draw inner box
776
777     # -> Label x axis
778
779         Viewport... 0 9 4 5
780         Font size... 15
781         Text bottom... yes Time (s)
782         Marks bottom every... 1 0.1 no yes no
783         Marks bottom every... 1 0.5 yes yes no
784
785         Font size... 15
786         Select outer viewport... 8 9 4 4.5
787         Text... 0 Left 0 Top 'tier_name$'
788
789         Viewport... 0 9 0 5
790
791     # -> Write picture to a file (Windows MetaFile / Praat Picture File)
792
793     #
794         Write to Windows metafile... 'dir$\HTML\Images\'prefix$'_t1$'-
795         't2$'_tier_name$'_border$'.wmf
796         Write to Windows metafile... 'dir$\HTML\'prefix$'_t1$'-
797         't2$'_tier_name$'_border$'.wmf
798         Write to praat picture file... 'dir$\HTML\'prefix$'_t1$'-
799         't2$'_tier_name$'_border$'.praapic
800
801     # -> Erase picture before going onto next interval
802
803         Erase all
804     endif
805
806     # -> Extract information from TextGrid to create XML-like file
807
808     if 'create_XML_file' = 1
809         select TextGrid 'name$'
810
811     # -> Print beginning of intonation unit, utterance or turn, taking the
812     # -> borders of previous intervals into account
813
814         if border0$ = "iu" or border00$ = "iu"
815             printline <IU img='prefix$'_t1$'-t2$'_tier_name$'_border$'
816

```



```

817 ... id='prefix$'_st$'-et$'_tier_name$'_border1$'>
818 elseif border0$ = "rep" or border00$ = "rep"
819   printline <IU img='prefix$'_t1$'-t2$'_tier_name$'_border$'
820   ... id='prefix$'_st$'-et$'_tier_name$'_border1$'>
821 elseif border0$ = "prt" or border00$ = "prt"
822   printline <IU img='prefix$'_t1$'-t2$'_tier_name$'_border$'
823   ... id='prefix$'_st$'-et$'_tier_name$'_border1$'>
824 elseif border0$ = "tot" or border00$ = "tot"
825   printline <IU img='prefix$'_t1$'-t2$'_tier_name$'_border$'
826   ... id='prefix$'_st$'-et$'_tier_name$'_border1$'>
827 elseif border0$ = "nt" or border0$ = "qst" or border0$ = "emp" or
828   ..border0$ = "ssp" or border0$ = "ext" or border0$ = "slf"
829   printline <U Quest='u_quest1$' Emph='u_emph1$' Susp='u_susp1$'>
830   ...'newline$'<IU img='prefix$'_t1$'-t2$'_tier_name$'_border$'
831   ... id='prefix$'_st$'-et$'_tier_name$'_border1$'>
832 elseif border0$ = ""
833   printline <Turn Name='tier_name$'>
834   ...'newline$'<U Quest='u_quest$' Emph='u_emph$' Susp='u_susp$'>
835   ...'newline$'<IU img='prefix$'_t1$'-t2$'_tier_name$'_border$'
836   ... id='prefix$'_st$'-et$'_tier_name$'_border1$'>
837 endif
838
839 # -> Replace signs at the beginning and in the middle of current interval
840 # -> by XML-like code (Overlaps, Comments, Unintelligible productions,
841 # -> Fragments, Restarts, and intermediary Intonation Unit boundaries)
842
843   length = length (i_label$)
844
845   ltlt$ = "\[<\] \< "
846   lt$ = "\< "
847   gt$ = "\>"
848   com$ = "([\%](\{3})(:\s)([a-zA-ZáéíóúñÁÉÍÓÚÑ_])*)(\)(\s\()?"
849   c$ = "\)"
850
851   if index_regex (i_label$, "xxx")
852     right = rindex_regex (i_label$, "xxx")
853     left$ = mid$ (i_label$, 1, right - 1)
854     right$ = mid$ (i_label$, right + 3, length)
855     i_label$ = left$ + "<Unintelligible>xxx</Unintelligible>" + right$
856   endif
857
858   if index_regex (i_label$, ltlt$)
859     i_label$ = replace_regex$ (i_label$, ltlt$, "<Overlap Type='act'/'>
860     ...\&lt;] \&lt; ", 0)
861   endif
862
863   if index_regex (i_label$, lt$)
864     i_label$ = replace_regex$ (i_label$, lt$, "<Overlap Type='pas'/'>
865     ...\&lt; ", 0)
866   endif
867
868
869   if index_regex (i_label$, com$)
870     i_label$ = replace_regex$ (i_label$, com$, "<Comment\2=
871     ...'\4'/'>\1\2\3\4\5\6", 0)
872   endif
873
874   length = length (i_label$)
875
876   if index_regex (i_label$, c$)
877     i_label$ = replace_regex$ (i_label$, c$, c$+"<Comment/'>", 0)
878   endif
879
880   if index_regex (i_label$, gt$)
881     i_label$ = replace_regex$ (i_label$, gt$, " \&gt;<Overlap/'>", 0)
882   endif
883
884
885   if index_regex (i_label$, "- ")
886     i_label$ = replace$ (i_label$, "- ", "-<Fragment/'> ", 0)
887   endif
888   if index (i_label$, " / ")
889     i_label$ = replace$ (i_label$, " / ", " /<Border Type='iu'/'></IU>
890     ...<IU> ", 0)
891   endif
892   if index (i_label$, " [/ ")
893     i_label$ = replace$ (i_label$, " [/ ", " [/<Restart Type='rep'/'>
894     ...<Border Type='rep'/'></IU><IU> ", 0)
895   endif

```

```

896   if index (i_label$, " [//] ")
897     i_label$ = replace$ (i_label$, " [//] ", " [//]<Restart Type='prt'/">
898     ...<Border Type='prt'/"></IU><IU> ", 0)
899   endif
900   if index (i_label$, " [///] ")
901     i_label$ = replace$ (i_label$, " [///] ", " [///]<Restart Type=
902     ...'tot'/"><Border Type='tot'/"></IU><IU> ", 0)
903   endif
904   print 'i_label$'
905
906 # -> Replace signs at the end of current interval by XML-like code (Border
907 # -> type, end of Intonation Unit, end of Utterance, end of Turn - with
908 # -> immediately following pause)
909
910 # -> Pauses will be explicitly marked only when longer than 1 second.
911
912   if i < n - 1
913     nt$ = " //"
914     end = endsWith (i_label$, nt$)
915     if end = 1
916       print <Border Type=''border1$'"/>
917       ...'newline$'</IU>
918       ...'newline$'</U>
919       if label2$ = ""
920         if durnext < 0.50
921           pause$ = "s"
922         elsif durnext > 0.50 and durnext < 1.00
923           pause$ = "l"
924         elsif durnext > 1.00
925           pause$ = "xl"
926         endif
927         if pause$ = "xl"
928           printline <Pause Sec=''durnext:2'' Type=''pause$'"/>
929           ... #</Pause>
930         else
931           printline <Pause Sec=''durnext:2'' Type=''pause$'"/>
932         endif
933         printline </Turn>
934       endif
935     endif
936
937     ext$ = " +/"
938     end = endsWith (i_label$, ext$)
939     if end = 1
940       print <Border Type=''border1$'"/>
941       ...'newline$'</IU>
942       ...'newline$'</U>
943       if label2$ = ""
944         if durnext < 0.50
945           pause$ = "s"
946         elsif durnext > 0.50 and durnext < 1.00
947           pause$ = "l"
948         elsif durnext > 1.00
949           pause$ = "xl"
950         endif
951         if pause$ = "xl"
952           printline <Pause Sec=''durnext:2'' Type=''pause$'"/>
953           ... #</Pause>
954         else
955           printline <Pause Sec=''durnext:2'' Type=''pause$'"/>
956         endif
957         printline </Turn>
958       endif
959     endif
960
961     slf$ = " +"
962     end = endsWith (i_label$, slf$)
963     if end = 1
964       print <Border Type=''border1$'"/>
965       ...'newline$'</IU>
966       ...'newline$'</U>
967       if label2$ = ""
968         if durnext < 0.50
969           pause$ = "s"
970         elsif durnext > 0.50 and durnext < 1.00
971           pause$ = "l"
972         elsif durnext > 1.00
973           pause$ = "xl"
974         endif

```

```

975         endif
976         if pause$ = "xl"
977             printline <Pause Sec=''durnext:2'' Type=''pause$''>
978             ... #</Pause>
979         else
980             printline <Pause Sec=''durnext:2'' Type=''pause$''/>
981         endif
982         printline </Turn>
983     endif
984 endif
985
986 qst$ = " ?"
987 end = endsWith (i_label$, qst$)
988 if end = 1
989     print <Border Type=''border1$''/>
990     ...'newline$'</IU>
991     ...'newline$'</U>
992 if label2$ = ""
993     if durnext < 0.50
994         pause$ = "s"
995     elsif durnext > 0.50 and durnext < 1.00
996         pause$ = "l"
997     elsif durnext > 1.00
998         pause$ = "xl"
999     endif
1000     if pause$ = "xl"
1001         printline <Pause Sec=''durnext:2'' Type=''pause$''>
1002         ... #</Pause>
1003     else
1004         printline <Pause Sec=''durnext:2'' Type=''pause$''/>
1005     endif
1006     printline </Turn>
1007 endif
1008 endif
1009
1010 emp$ = " !"
1011 end = endsWith (i_label$, emp$)
1012 if end = 1
1013     print <Border Type=''border1$''/>
1014     ...'newline$'</IU>
1015     ...'newline$'</U>
1016 if label2$ = ""
1017     if durnext < 0.50
1018         pause$ = "s"
1019     elsif durnext > 0.50 and durnext < 1.00
1020         pause$ = "l"
1021     elsif durnext > 1.00
1022         pause$ = "xl"
1023     endif
1024     if pause$ = "xl"
1025         printline <Pause Sec=''durnext:2'' Type=''pause$''>
1026         ... #</Pause>
1027     else
1028         printline <Pause Sec=''durnext:2'' Type=''pause$''/>
1029     endif
1030     printline </Turn>
1031 endif
1032 endif
1033
1034 ssp$ = " ..."
1035 end = endsWith (i_label$, ssp$)
1036 if end = 1
1037     print <Border Type=''border1$''/>
1038     ...'newline$'</IU>
1039     ...'newline$'</U>
1040 if label2$ = ""
1041     if durnext < 0.50
1042         pause$ = "s"
1043     elsif durnext > 0.50 and durnext < 1.00
1044         pause$ = "l"
1045     elsif durnext > 1.00
1046         pause$ = "xl"
1047     endif
1048     if pause$ = "xl"
1049         printline <Pause Sec=''durnext:2'' Type=''pause$''>
1050         ... #</Pause>
1051     else
1052         printline <Pause Sec=''durnext:2'' Type=''pause$''/>

```

```

1053         endif
1054         printline </Turn>
1055     endif
1056 endif
1057
1058 iu$ = " /"
1059 end = endsWith (i_label$, iu$)
1060 if end = 1
1061     print <Border Type='border1$' />
1062     ...'newline$'</IU>
1063     if label2$ = ""
1064         if durnext < 0.50
1065             pause$ = "s"
1066         elsif durnext > 0.50 and durnext < 1.00
1067             pause$ = "l"
1068         elsif durnext > 1.00
1069             pause$ = "xl"
1070         endif
1071     endif
1072     if pause$ = "xl"
1073         printline <Pause Sec='durnext:2' Type='pause$'>
1074         ... #</Pause>
1075     else
1076         printline <Pause Sec='durnext:2' Type='pause$' />
1077     endif
1078 endif
1079 endif
1080
1081 rep$ = " [/"
1082 when = endsWith (i_label$, rep$)
1083 if when = 1
1084     print <Restart Type='rep' />
1085     ...'newline$'<Border Type='border1$' />
1086     ...'newline$'</IU>
1087     if label2$ = ""
1088         if durnext < 0.50
1089             pause$ = "s"
1090         elsif durnext > 0.50 and durnext < 1.00
1091             pause$ = "l"
1092         elsif durnext > 1.00
1093             pause$ = "xl"
1094         endif
1095     endif
1096     if pause$ = "xl"
1097         printline <Pause Sec='durnext:2' Type='pause$'>
1098         ... #</Pause>
1099     else
1100         printline <Pause Sec='durnext:2' Type='pause$' />
1101     endif
1102 endif
1103
1104 prt$ = " [//]"
1105 when = endsWith (i_label$, prt$)
1106 if when = 1
1107     print <Restart Type='prt' />
1108     ...'newline$'<Border Type='border1$' />
1109     ...'newline$'</IU>
1110     if label2$ = ""
1111         if durnext < 0.50
1112             pause$ = "s"
1113         elsif durnext > 0.50 and durnext < 1.00
1114             pause$ = "l"
1115         elsif durnext > 1.00
1116             pause$ = "xl"
1117         endif
1118     endif
1119     if pause$ = "xl"
1120         printline <Pause Sec='durnext:2' Type='pause$'>
1121         ... #</Pause>
1122     else
1123         printline <Pause Sec='durnext:2' Type='pause$' />
1124     endif
1125 endif
1126
1127 tot$ = " [///]"
1128 when = endsWith (i_label$, tot$)
1129 if when = 1
1130     print <Restart Type='tot' />
1131     ...'newline$'<Border Type='border1$' />

```

```

1132     ...'newline$'</IU>
1133     if label2$ = ""
1134         if durnext < 0.50
1135             pause$ = "s"
1136         elseif durnext > 0.50 and durnext < 1.00
1137             pause$ = "l"
1138         elseif durnext > 1.00
1139             pause$ = "xl"
1140         endif
1141         if pause$ = "xl"
1142             printline <Pause Sec=''durnext:2'' Type=''pause$''>
1143             ... #</Pause>
1144         else
1145             printline <Pause Sec=''durnext:2'' Type=''pause$''/>
1146         endif
1147     endif
1148 endif
1149
1150 # -> Replace signs at the end of current interval by XML-like code (for
1151 # -> intervals without immediately following pause)
1152
1153 else
1154     nt$ = " //"
1155     end = endsWith (i_label$, nt$)
1156     if end = 1
1157         print <Border Type=''border1$''/>
1158         ...'newline$'</IU>
1159         ...'newline$'</U>
1160         ...'newline$'</Turn>
1161     endif
1162
1163     ext$ = " +/"
1164     end = endsWith (i_label$, ext$)
1165     if end = 1
1166         print <Border Type=''border1$''/>
1167         ...'newline$'</IU>
1168         ...'newline$'</U>
1169         ...'newline$'</Turn>
1170     endif
1171
1172     slf$ = " +"
1173     end = endsWith (i_label$, slf$)
1174     if end = 1
1175         print <Border Type=''border1$''/>
1176         ...'newline$'</IU>
1177         ...'newline$'</U>
1178         ...'newline$'</Turn>
1179     endif
1180
1181     qst$ = " ?"
1182     end = endsWith (i_label$, qst$)
1183     if end = 1
1184         print <Border Type=''border1$''/>
1185         ...'newline$'</IU>
1186         ...'newline$'</U>
1187         ...'newline$'</Turn>
1188     endif
1189
1190     emp$ = " !"
1191     end = endsWith (i_label$, emp$)
1192     if end = 1
1193         print <Border Type=''border1$''/>
1194         ...'newline$'</IU>
1195         ...'newline$'</U>
1196         ...'newline$'</Turn>
1197     endif
1198
1199     ssp$ = " ..."
1200     end = endsWith (i_label$, ssp$)
1201     if end = 1
1202         print <Border Type=''border1$''/>
1203         ...'newline$'</IU>
1204         ...'newline$'</U>
1205         ...'newline$'</Turn>
1206     endif
1207
1208     iu$ = " /"
1209     end = endsWith (i_label$, iu$)
1210

```

```

1211     if end = 1
1212         print <Border Type='border1$' />
1213         printline </IU>
1214     endif
1215 endif
1216 endif
1217
1218 # -> Remove files generated by automatic analyses according to the selected
1219 # -> tasks (Extract TextGrids, Extract Sounds, Create Pictures, Create XML)
1220
1221     if 'extract_TextGrids' = 1
1222         select TextGrid 'name$_part
1223         plus IntervalTier 'tier_name$'
1224         plus TextGrid grid
1225         Remove
1226     endif
1227     if 'extract_Sounds' = 1
1228         select Sound 'name$'
1229         Remove
1230     endif
1231     if 'create_pictures' = 1
1232         select Spectrogram 'name$'
1233         plus Pitch 'name$'
1234         Remove
1235
1236     if 'kill_octave_jumps' = 1
1237         select Pitch 'name$'
1238         Remove
1239
1240     endif
1241     endif
1242 endif
1243
1244 # -> Select TextGrid and Long Sound files again to analyse following interval
1245
1246     select TextGrid 'name$'
1247     plus LongSound 'sound$'
1248
1249     endfor
1250
1251 # -> End of interval loop
1252
1253 endfor
1254
1255 # -> End of tier loop
1256
1257 # -> Print end of Transcript
1258
1259 if 'create_XML_file' = 1
1260     printline </Text>
1261     ...'newline$'</Transcript>
1262
1263 # -> Write contents of Info Window into a Text file with XML-ending
1264
1265     filedelete 'dir$'\prefix$.xml
1266     fappendinfo 'dir$'\prefix$.xml
1267 endif
1268
1269
1270 #+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
1271 #
1272 # -> End of program ExtractTextGrid+Sound+XML.praat by Eduardo Velázquez
1273 #
1274 #+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

Anexo 6: Subprograma `Create_XML(U+).praat`

Líneas	Secciones
18 - 32	Definición del formulario
36	Limpieza de la Info Window
42 - 47	Impresión del encabezado de documento de XML
51 - 52	Creación de lista de archivos de sonido en el directorio especificado
57 - 61	Inicio del ciclo para procesar todos los archivos especificados
65 - 68	Lectura de los archivos TextGrid y de sonido con el mismo nombre de base
72 - 92	Ajuste de la opción de segmentación manual a las características de la TextGrid
96 - 361	Segmentación de sílabas y asignación de valores paso a paso
138 - 347	Cálculo del porcentaje de variación y valores estándar de los valores dados de f_0
150 - 176	Localización y extracción de la frecuencia del intervalo actual, el siguiente y los dos anteriores
185 - 347	Extracción de valores para las variables de frecuencia
365 - 369	Escritura del archivo de TextGrid completado
374 - 384	Extracción de las hileras <code>phon</code> , <code>f0</code> , <code>rhythm</code> , <code>tones</code> y la ortográfica
385 - 390	Conversión de las hileras en una nueva TextGrid para utilizarse en las imágenes
394 - 399	Elaboración de los objetos Pitch, Intensity y Spectrogram
403 - 419	Redondeo de la frecuencia mínima para establecer una base más holgada para las imágenes
423 - 453	Redondeo de la frecuencia máxima para establecer una cima más holgada para las imágenes
457 - 550	Creación de imágenes
554 - 555	Escritura de la imagen a archivos
559 - 561	Limpieza de la ventana de imágenes antes de pasar al siguiente intervalo
565 - 897	Extracción de información de la TextGrid
572 - 575	Predefinición de variables en posición 0
580	Inicio del ciclo que procesa todos los intervalos de la hilera ortográfica silábica, <code>syll</code> , (1)
584 - 677	Interpretación de los rasgos de frontera y de enunciado en el intervalo actual
681 - 705	Obtención de los valores de la TextGrid para el intervalo actual, el previo y los dos posteriores
709 - 761	Extracción e interpretación de la hilera rítmica, <code>rhythm</code> , (6)

765 - 809	Extracción de los puntos en la hilera tonal, <i>tones (7)</i> , durante el intervalo actual
813 - 835	Extracción e interpretación de los valores de las variables de duración
839 - 871	Extracción e interpretación de los valores de las variables de intensidad
875 - 897	Extracción y simplificación de los valores de las variables de frecuencia
901 - 906	Inicio de la impresión de las muestras de grabación
910 - 912	Inicio de la impresión de código compatible con XML por cada sílaba
916 - 938	Impresión de los valores de las variables en la Info Window con una sintaxis compatible con el XML
942 - 989	Interpretación de la puntuación en la hilera ortográfica por sílabas para obtener los valores de sus disyunciones (<i>breaks</i>)
993 - 1190	Interpretación e impresión de la frontera de unidad entonativa o enunciado
1194 - 1198	Selección del archivo TextGrid para analizar el siguiente intervalo
1202	Fin del ciclo de intervalos (<i>syll</i>)
1207	Impresión del final de una muestra de grabación individual
1211 - 1222	Selección y eliminación de todos los objetos generados por Praat para la presente muestra de grabación
1226 - 1230	Fin del ciclo de archivos (<i>i_file</i>)
1234 - 1235	Impresión del final del documento de XML
1239 - 1241	Escritura del contenido de la Info Window a un archivo de texto con terminación <i>.XML</i>
1245	Fin del subprograma

```

01  ##+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----#
02  #
03  # This program was last updated on 5th June, 2007 by Eduardo Velázquez
04  #
05  # Create_XML(U+).praat will display a form asking for information which
06  # could be divided into four areas: 1) directory location, 2) check
07  # buttons to allow or skip the creation of pictures, killing of octave
08  # jumps or manual segmentation of syllables, 3) location where the re-
09  # sulting text grid, picture and XML files will be saved.
10  # After this, all sound and text grid files in the given directory will
11  # be read in and analysed before extracting a whole palette of acoustic
12  # values and then printed with an XML-like syntax on the Info Window.
13  #
14  ##+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----#
15
16  # -> Definition of Form
17
18  form Create XML document of selected files
19    comment Enter directory where Sound and TextGrid files are kept:
20    sentence dir C:\
21    boolean Create_pictures no
22    boolean Kill_octave_jumps no
23    boolean Segment_syllables_manually no
24    comment Enter directory where TextGrid files will be saved:
25    sentence grid_dir C:\
26    comment Enter directory where Picture files will be saved:
27    sentence pic_dir C:\
28    comment Name of output XML document:

```



```

29     sentence xml_file
30     comment Enter directory where XML file will be saved:
31     sentence xml_dir C:\
32 endform
33
34 # -> Clear Info Window
35
36 clearinfo
37
38 # -> Print Header of XML document
39
40 # -> Header for compilation of Fragments
41
42 printline <?xml version='1.0' encoding='ISO-8859-1' standalone='no' ?>
43 ...'newline$'<!-- Written by Eduardo Velázquez -->
44 ...'newline$'<?xml-stylesheet type='text/xsl' href='Prosody-set.xsl'?>
45 ...'newline$'<!DOCTYPE Transcript SYSTEM 'PROSODY_XML_SET.DTD'>
46 ...'newline$'<Transcript>
47 ...'newline$'<Set>
48
49 # -> Create list of Sound files
50
51 Create Strings as file list... list 'dir$'\*.wav
52 Sort
53
54 # -> Loop that goes through all the specified files
55
56 #-#-#-#-#-#-#(for i_file to n_files)
57 n_files = Get number of strings
58 for i_file to n_files
59     select Strings list
60     file_name$ = Get string... i_file
61     file$ = file_name$ - ".wav"
62
63 # -> Read in the TextGrid and Sound files with that base name
64
65     Read from file... 'dir$'\file$.wav
66     Read from file... 'dir$'\file$.TextGrid
67     praatname$ = selected$ ("TextGrid")
68
69
70 # -> Adjust the manual segmentation to TextGrid characteristics
71
72     tiers = Get number of tiers
73     if tiers = 1
74         segment_syllables_manually = 1
75         n_ortho = Get number of intervals... 1
76         tier_name$ = Get tier name... 1
77     elseif tiers > 1 and tiers < 8
78         segment_syllables_manually = 0
79         select TextGrid 'praatname$'
80         Extract tier... 'tiers'
81         Into TextGrid
82         Write to text file... 'grid_dir$'\file$.TextGrid
83         n_ortho = Get number of intervals... 1
84         tier_name$ = Get tier name... 1
85     elseif tiers = 8
86         segment_syllables_manually = 0
87         n_ortho = Get number of intervals... 8
88         tier_name$ = Get tier name... 8
89     endif
90
91 select Sound 'praatname$'
92 noprogess To Pitch... 0 50 600
93
94 # -> Step-by-step segmentation of syllables and assignation of values
95
96 if segment_syllables_manually = 1
97     select TextGrid 'praatname$'
98     ortho1$ = Get label of interval... 1 2
99     t1_1 = Get starting point... 1 2
100    t2_1 = Get end point... 1 2
101
102    select Sound 'praatname$'
103    plus TextGrid 'praatname$'
104
105    Edit
106        editor TextGrid 'praatname$'
107        Duplicate tier... 1 syll

```

```

108     pause Segment and label each syllable. Then, press 'Continue'.
109     endeditor
110
111     select TextGrid 'praatname$'
112 else
113     select TextGrid 'praatname$'
114     ortho1$ = Get label of interval... 8 2
115     t1_1 = Get starting point... 8 2
116     t2_1 = Get end point... 8 2
117 endif
118
119 if segment_syllables_manually = 1
120     editor TextGrid 'praatname$'
121     Duplicate tier... 2 phon
122     pause Transcribe each syllable phonetically.
123     ...Then, press 'Continue'.
124
125     Duplicate tier... 3 f0
126     Remove all text from tier
127     pause Write valid f0 values for each syllable.
128     ...Then, press 'Continue'.
129
130     Duplicate tier... 4 f0var
131     Remove all text from tier
132
133     Duplicate tier... 5 f0std
134     endeditor
135
136 # -> Calculate percent variation and standard values of given f0 values
137
138     select TextGrid 'praatname$'
139     n = Get number of intervals... 3
140     f0prev = 0
141     f0std = 100
142     f0var = 0
143     f0vst = 0
144
145     for i to n
146
147 # -> Localize and extract Pitch of current, next, and two previous
148 # -> intervals (first and last intervals will always be empty)
149
150     select TextGrid 'praatname$'
151     f0 = Get label of interval... 3 i
152     ti = Get starting point... 3 i
153     f0$ = "'f0'"
154     if f0$ = ""
155         Set interval text... 3 i 0
156     endif
157     f0 = Get label of interval... 3 i
158     f0$ = "'f0'"
159
160     if i < n - 1
161         f0next = Get label of interval... 3 i+1
162     endif
163
164     if i > 2
165         f0prev = Get label of interval... 3 i-1
166         f0prev2 = Get label of interval... 3 i-2
167         f0prev$ = "'f0prev'"
168         f0prev2$ = "'f0prev2'"
169         if f0prev$ = ""
170             f0prev$ = "0"
171         endif
172         if f0prev2$ = ""
173             f0prev2$ = "0"
174         endif
175     endif
176     endif
177
178 # -> Extract values for the Pitch variables
179 # -> (f0 value of annotated intervals without perceivable pitch last
180 # -> must be marked with 0, standardized f0 will be the average of pre-
181 # -> vious and next intervals and will be displayed between parenthesis.)
182 # -> (In TextGrids containing more than one utterance the values for
183 # -> f0var and f0std will be reset to 0% and 100.)
184
185     if n_ortho < 4
186         tortho2_1 = Get starting point... 6 3

```

```

187   tortho2_2 = Get end point... 6 3
188
189   if i > 2 and f0 <> 0 and ti <> tortho2_1
190     f0var = 100 * (f0 - f0prev) / f0prev
191     f0vst = f0var / 100 * 12
192     f0std = (1 + (f0var / 100)) * f0stdprev
193     f0varnext = 100 * (f0next - f0) / f0
194     f0stdnext = (1 + (f0varnext / 100)) * f0std
195     if f0prev$ = "0"
196       f0var = 100 * (f0 - f0prev2) / f0prev2
197       f0vst = f0var / 100 * 12
198       if i = 3
199         f0std = 100
200       elseif i >= 4
201         f0std = (1 + (f0var / 100)) * f0stdprev2
202       endif
203     endif
204
205   elseif i > 2 and f0 = 0 and ti <> tortho2_1
206     f0var = 0
207     f0vst = 0
208     f0varnext = 100 * (f0next - f0prev) / f0prev
209     f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
210     f0std = (f0stdnext + f0stdprev) / 2
211     f0var$ = "'f0var'"
212     f0vst$ = "'f0vst'"
213     f0std$ = "'f0std'"
214
215     elseif ti = tortho2_1
216       f0var = 0
217       f0vst = 0
218       f0varnext = 100 * (f0next - f0prev) / f0prev
219       f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
220       f0std = 100
221       f0var$ = "'f0var'"
222       f0vst$ = "'f0vst'"
223       f0std$ = "'f0std'"
224     endif
225
226   elseif n_ortho = 4
227     tortho2_1 = Get starting point... 6 3
228     tortho2_2 = Get end point... 6 3
229     tortho3_1 = Get starting point... 6 4
230     tortho3_2 = Get end point... 6 4
231
232     if i > 2 and f0 <> 0 and ti <> tortho2_1 and ti <> tortho3_1
233       f0var = 100 * (f0 - f0prev) / f0prev
234       f0vst = f0var / 100 * 12
235       f0std = (1 + (f0var / 100)) * f0stdprev
236       f0varnext = 100 * (f0next - f0) / f0
237       f0stdnext = (1 + (f0varnext / 100)) * f0std
238       if f0prev$ = "0"
239         f0var = 100 * (f0 - f0prev2) / f0prev2
240         f0vst = f0var / 100 * 12
241         if i = 3
242           f0std = 100
243         elseif i >= 4
244           f0std = (1 + (f0var / 100)) * f0stdprev2
245         endif
246       endif
247     endif
248
249     elseif i > 2 and f0 = 0 and ti <> tortho2_1 and ti <> tortho3_1
250       f0var = 0
251       f0vst = 0
252       f0varnext = 100 * (f0next - f0prev) / f0prev
253       f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
254       f0std = (f0stdnext + f0stdprev) / 2
255       f0var$ = "'f0var'"
256       f0vst$ = "'f0vst'"
257       f0std$ = "'f0std'"
258
259     elseif ti = tortho2_1 or ti = tortho3_1
260       f0var = 0
261       f0vst = 0
262       f0varnext = 100 * (f0next - f0prev) / f0prev
263       f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
264       f0std = 100
265       f0var$ = "'f0var'"

```

```

265     f0vst$ = "'f0vst'"
266     f0std$ = "'f0std'"
267     endif
268
269
270   elseif n_ortho = 5
271     tortho2_1 = Get starting point... 6 3
272     tortho2_2 = Get end point... 6 3
273     tortho3_1 = Get starting point... 6 4
274     tortho3_2 = Get end point... 6 4
275     tortho4_1 = Get starting point... 6 5
276     tortho4_2 = Get end point... 6 5
277
278     if i > 2 and f0 <> 0 and ti <> tortho2_1 and ti <> tortho3_1
279     ...and ti <> tortho4_1
280       f0var = 100 * (f0 - f0prev) / f0prev
281       f0vst = f0var / 100 * 12
282       f0std = (1 + (f0var / 100)) * f0stdprev
283       f0varnext = 100 * (f0next - f0) / f0
284       f0stdnext = (1 + (f0varnext / 100)) * f0std
285       if f0prev$ = "0"
286         f0var = 100 * (f0 - f0prev2) / f0prev2
287         f0vst = f0var / 100 * 12
288         if i = 3
289           f0std = 100
290         elseif i >= 4
291           f0std = (1 + (f0var / 100)) * f0stdprev2
292         endif
293       endif
294
295   elseif i > 2 and f0 = 0 and ti <> tortho2_1 and ti <> tortho3_1
296     ...and ti <> tortho4_1
297     f0var = 0
298     f0vst = 0
299     f0varnext = 100 * (f0next - f0prev) / f0prev
300     f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
301     f0std = (f0stdnext + f0stdprev) / 2
302     f0var$ = "'f0var'"
303     f0vst$ = "'f0vst'"
304     f0std$ = "'f0std'"
305
306     elseif ti = tortho2_1 or ti = tortho3_1 or ti = tortho4_1
307     f0var = 0
308     f0vst = 0
309     f0varnext = 100 * (f0next - f0prev) / f0prev
310     f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
311     f0std = 100
312     f0var$ = "'f0var'"
313     f0vst$ = "'f0vst'"
314     f0std$ = "'f0std'"
315     endif
316
317   endif
318
319   if f0var = undefined
320     f0var = 0
321   endif
322   if f0vst = undefined
323     f0vst = 0
324   endif
325   if f0std = undefined
326     f0std = 0
327   endif
328
329   if i >= 4
330     f0stdprev2 = f0stdprev
331   endif
332
333   f0stdprev = f0std
334
335   if i >= 2 and i <= n - 1
336   if f0var > 0
337     Set interval text... 4 i +'f0var:2'\%
338   else
339     Set interval text... 4 i 'f0var:2'\%
340   endif
341   if f0$ <> "0"
342     Set interval text... 5 i 'f0std:0'
343   else

```

```

344     Set interval text... 5 i ('f0std:0')
345     endif
346   endif
347 endfor
348
349   select Sound 'praatname$'
350   plus TextGrid 'praatname$'
351   Edit
352     editor TextGrid 'praatname$'
353
354     Duplicate tier... 6 rhythm
355     Remove all text from tier
356     pause Write rhythm structure. Then, press 'Continue'.
357
358     Add point tier... 7 tones
359     pause Write tone structure. Then, press 'Continue'.
360
361     endeditor
362
363 # -> Save filled-up TextGrid file
364
365   select TextGrid 'praatname$'
366   Write to text file... 'grid_dir$'\file$.TextGrid
367
368   endif
369
370
371 # -> Extract Phonetic, f0, Rhythm, Tones, and Orthographic Tiers
372 # -> Convert them into a new TextGrid to be used in pictures
373
374   select TextGrid 'praatname$'
375   tier_name$ = Get tier name... 8
376   Extract tier... 2
377   select TextGrid 'praatname$'
378   Extract tier... 3
379   select TextGrid 'praatname$'
380   Extract tier... 6
381   select TextGrid 'praatname$'
382   Extract tier... 7
383   select TextGrid 'praatname$'
384   Extract tier... 8
385   select IntervalTier phon
386   plus IntervalTier f0
387   plus IntervalTier rhythm
388   plus TextTier tones
389   plus IntervalTier 'tier_name$'
390   Into TextGrid
391
392 # -> Make Pitch, Intensity and Spectrogram objects
393
394   select Pitch 'praatname$'
395   if 'kill_octave_jumps' = 1
396     Kill octave jumps
397   endif
398   min = Get minimum... 0 0 Hertz Parabolic
399   max = Get maximum... 0 0 Hertz Parabolic
400
401 # -> Round down Pitch minimum to set a looser base for Pictures
402
403   if min > 0 and min < 75
404     base = 50
405   elsif min > 75 and min < 110
406     base = 75
407   elsif min > 110 and min < 160
408     base = 100
409   elsif min > 160 and min < 210
410     base = 150
411   elsif min > 210 and min < 260
412     base = 200
413   elsif min > 260 and min < 310
414     base = 250
415   elsif min > 310 and min < 360
416     base = 300
417   else
418     base = 75
419   endif
420
421 # -> Round up Pitch maximum to set a looser top for Pictures
422

```

```

423     if max > 0 and max < 75
424         top = 75
425     elsif max > 75 and max < 90
426         top = 100
427     elsif max > 90 and max < 140
428         top = 150
429     elsif max > 140 and max < 190
430         top = 200
431     elsif max > 190 and max < 240
432         top = 250
433     elsif max > 240 and max < 290
434         top = 300
435     elsif max > 290 and max < 340
436         top = 350
437     elsif max > 340 and max < 390
438         top = 400
439     elsif max > 390 and max < 440
440         top = 450
441     elsif max > 440 and max < 490
442         top = 500
443     elsif max > 490 and max < 540
444         top = 550
445     elsif max > 540
446         top = 600
447     endif
448
449     select Sound 'praatname$'
450     noprogess To Intensity... 75 0 yes
451
452     select Sound 'praatname$'
453     noprogess To Spectrogram... 0.05 1200 0.002 20 Gaussian
454
455 # -> Create Pictures
456
457 if 'create_pictures' = 1
458     Erase all
459
460     # -> Specify font type size, color
461
462     Times
463     Font size... 12
464     Black
465
466     # -> Define size and position of Waveform
467     # -> (by specifying grid coordinates)
468
469     Viewport... 0 9 0 2
470
471
472     # -> Draw Waveform
473
474     select Sound 'praatname$'
475     Draw... 0 0 0 0 no curve
476
477     # -> Define size and position of Spectrogram
478
479     Viewport... 0 9 1 4
480
481     # -> Draw Spectrogram
482
483     top2 = top + 2
484
485     select Spectrogram 'praatname$'
486     Paint... 0 0 'base' 'top2' 100 yes 50 5 0 no
487
488     # -> Draw Pitch curve
489     # -> First as a thick white line
490
491     select Pitch 'praatname$'
492     Line width... 10
493     White
494     Draw... 0 0 'base' 'top' no
495
496     # -> then as a thin black line
497
498     Line width... 3
499     Black
500     Draw... 0 0 'base' 'top' no
501

```

```

502 # -> Label y axis
503
504     Line width... 1
505     One mark left... 'top' no no yes
506     One mark left... 'base' no no yes
507     Marks left every... 1 50 yes yes no
508     Font size... 15
509     Text left... yes Fundamental frequency (Hz)
510
511 # -> Define size and position of TextGrid
512
513     Viewport... 0 9 1 6.5
514     Font size... 12
515
516 # -> Draw TextGrid
517
518     select TextGrid grid
519     Draw... 0 0 yes no no
520
521 # -> Define size and position of inner box
522
523     Viewport... 0 9 0 6.5
524
525 # -> Draw inner box
526
527     Black
528     Draw inner box
529
530 # -> Label x axis
531
532     Viewport... 0 9 4 6.5
533     Font size... 15
534     Text bottom... yes Time (s)
535     Font size... 10
536     Marks bottom every... 1 0.1 no yes no
537     Marks bottom every... 1 0.5 yes yes no
538
539     Font size... 15
540     Select outer viewport... 8 9 3.5 4
541     Text... 0 Left 0 Half phon
542     Select outer viewport... 8 9 4 4.5
543     Text... 0 Left 0 Half F_0
544     Select outer viewport... 8 9 4.5 5
545     Text... 0 Left 0 Half rhythm
546     Select outer viewport... 8 9 5 5.5
547     Text... 0 Left 0 Half tones
548     Select outer viewport... 8 9 5.5 6
549     Text... 0 Left 0 Half 'tier_name$'
550     Viewport... 0 9 0 6.5
551
552 # -> Write picture to a file (Windows MetaFile / Praat Picture File)
553
554     Write to Windows metafile... 'pic_dir$'\file$.wmf
555     Write to praat picture file... 'pic_dir$'\file$.praapic
556
557 # -> Erase picture before going onto next interval
558
559     Erase all
560
561 endif
562
563 # -> Extract information from TextGrid
564
565     select TextGrid 'praatname$'
566     n = Get number of intervals... 1
567     name$ = Get tier name... 8
568
569 # -> Predefinition of variables in position 0
570
571     dbprev = 0
572     dbvar = 0
573     durprev = 0
574     durvar = 0
575
576 # -> Loop that goes through all intervals of Syllables Tier (1)
577
578 #-#-#-#(for i from 2 to n-1)
579     for i from 2 to n-1
580

```

```

581
582 # -> Interpret border and utterance features of current interval
583
584 select TextGrid 'praatname$'
585   t1 = Get starting point... 1 i
586   t2 = Get end point... 1 i
587
588   nt$ = " //"
589   when = endsWith (orthol$, nt$)
590   if when = 1
591     border$ = "nt"
592     u_quest$ = "N"
593     u_emph$ = "N"
594     u_susp$ = "N"
595   endif
596
597   ext$ = " +/"
598   when = endsWith (orthol$, ext$)
599   if when = 1
600     border$ = "ext"
601     u_quest$ = "N"
602     u_emph$ = "N"
603     u_susp$ = "Y"
604   endif
605
606   slf$ = " +"
607   when = endsWith (orthol$, slf$)
608   if when = 1
609     border$ = "slf"
610     u_quest$ = "N"
611     u_emph$ = "N"
612     u_susp$ = "Y"
613   endif
614
615   qst$ = " ?"
616   when = endsWith (orthol$, qst$)
617   if when = 1
618     border$ = "qst"
619     u_quest$ = "Y"
620     u_emph$ = "N"
621     u_susp$ = "N"
622   endif
623
624   emp$ = "!"
625   when = endsWith (orthol$, emp$)
626   if when = 1
627     border$ = "emp"
628     u_quest$ = "N"
629     u_emph$ = "Y"
630     u_susp$ = "N"
631   endif
632
633   ssp$ = " ..."
634   when = endsWith (orthol$, ssp$)
635   if when = 1
636     border$ = "ssp"
637     u_quest$ = "N"
638     u_emph$ = "N"
639     u_susp$ = "Y"
640   endif
641
642   iu$ = " /"
643   when = endsWith (orthol$, iu$)
644   if when = 1
645     border$ = "iu"
646     u_quest$ = "N"
647     u_emph$ = "N"
648     u_susp$ = "Y"
649   endif
650
651   rep$ = " [/"
652   when = endsWith (orthol$, rep$)
653   if when = 1
654     border$ = "rep"
655     u_quest$ = "N"
656     u_emph$ = "N"
657     u_susp$ = "Y"
658   endif

```



```

659
660     prt$ = " [//]"
661     when = endsWith (orthol$, prt$)
662     if when = 1
663         border$ = "prt"
664         u_quest$ = "N"
665         u_emph$ = "N"
666         u_susp$ = "Y"
667     endif
668
669
670     tot$ = " [///]"
671     when = endsWith (orthol$, tot$)
672     if when = 1
673         border$ = "tot"
674         u_quest$ = "N"
675         u_emph$ = "N"
676         u_susp$ = "Y"
677     endif
678
679 # -> Get TextGrid values for present, next and two previos intervals
680
681     syll$ = Get label of interval... 1 i
682     phon$ = Get label of interval... 2 i
683     f0 = Get label of interval... 3 i
684     f0$ = "'f0'"
685     f0var$ = Get label of interval... 4 i
686     v1$ = replace$ (f0var$, "+", "", 0)
687     v2$ = replace$ (v1$, "% ", "", 0)
688     f0var = 'v2$'
689     f0var$ = "'f0var'"
690     f0vst = f0var / 100 * 12
691     f0vst$ = "'f0vst'"
692     f0std = Get label of interval... 5 i
693     f0std$ = "'f0std'"
694
695     if i <= n - 1
696         syllnext$ = Get label of interval... 1 i+1
697         t3 = Get end point... 1 i+1
698     endif
699
700     if i > 2
701         syllprev$ = Get label of interval... 1 i-1
702         syll_prev$ = Get label of interval... 1 i-2
703         t0 = Get starting point... 1 i-1
704         t_0 = Get starting point... 1 i-2
705     endif
706
707 # -> Extract and interpret Rhythm Tier (6)
708
709     rhythm$ = Get label of interval... 6 i
710     if rhythm$ = "( W"
711         stress$ = "Y"
712         focus$ = "N"
713     endif
714     if rhythm$ = "( S"
715         stress$ = "Y"
716         focus$ = "Y"
717     endif
718     if rhythm$ = "( w"
719         stress$ = "N"
720         focus$ = "N"
721     endif
722     if rhythm$ = "( s"
723         stress$ = "Y"
724         focus$ = "N"
725     endif
726     if rhythm$ = "W )"
727         stress$ = "Y"
728         focus$ = "N"
729     endif
730     if rhythm$ = "S )"
731         stress$ = "Y"
732         focus$ = "Y"
733     endif
734     if rhythm$ = "w )"
735         stress$ = "N"
736         focus$ = "N"
737     endif

```

```

738     if rhythm$ = "s )"
739         stress$ = "Y"
740     focus$ = "N"
741     endif
742     if rhythm$ = "W"
743         stress$ = "Y"
744     focus$ = "N"
745     endif
746     if rhythm$ = "S"
747         stress$ = "Y"
748     focus$ = "Y"
749     endif
750     if rhythm$ = "w"
751         stress$ = "N"
752     focus$ = "N"
753     endif
754     if rhythm$ = "s"
755         stress$ = "Y"
756     focus$ = "N"
757     endif
758     if rhythm$ = "( 0 )"
759         stress$ = "N"
760     focus$ = "N"
761     endif
762
763 # -> Extract points in current interval - Tone Tier (7)
764
765     Extract part... t1 t2 yes
766     npoint = Get number of points... 7
767     if npoint = 1
768         tone$ = Get label of point... 7 1
769         when = endsWith (tone$, "% ")
770         if when = 1
771             tone$ = replace$ (tone$, "% ", "%", 1)
772         endif
773     endif
774     if npoint = 2
775         tone1$ = Get label of point... 7 1
776         when = endsWith (tone1$, "% ")
777         if when = 1
778             tone1$ = replace$ (tone1$, "% ", "%", 1)
779         endif
780         tone2$ = Get label of point... 7 2
781         when = endsWith (tone2$, "% ")
782         if when = 1
783             tone2$ = replace$ (tone2$, "% ", "%", 1)
784         endif
785         tone$ = tone1$ + " " + tone2$
786     endif
787     if npoint = 3
788         tone1$ = Get label of point... 7 1
789         when = endsWith (tone1$, "% ")
790         if when = 1
791             tone1$ = replace$ (tone1$, "% ", "%", 1)
792         endif
793         tone2$ = Get label of point... 7 2
794         when = endsWith (tone2$, "% ")
795         if when = 1
796             tone2$ = replace$ (tone2$, "% ", "%", 1)
797         endif
798         tone3$ = Get label of point... 7 3
799         when = endsWith (tone3$, "% ")
800         if when = 1
801             tone3$ = replace$ (tone3$, "% ", "%", 1)
802         endif
803         tone$ = tone1$ + " " + tone2$ + " " + tone3$
804     endif
805     if npoint = 0
806         tone$ = ""
807     endif
808     Remove
809
810 # -> Extract and interpret values for the Duration variables
811
812     if durprev > 0
813         dur = t2 - t1
814         durvar = 100 * (dur - durprev) / durprev
815         durprev = dur

```

```

817     else
818         dur = t2 - t1
819         durprev = dur
820         durvar = 0
821     endif
822
823     durnext = t3 - t2
824
825     if durvar > 75
826         tmp$ = "l"
827     elsif durvar > 25 and durvar < 74.99
828         tmp$ = "rall"
829     elsif durvar > -24.99 and durvar < 24.99
830         tmp$ = "nt"
831     elsif durvar > -25 and durvar < -74.99
832         tmp$ = "accel"
833     elsif durvar > -75
834         tmp$ = "all"
835     endif
836
837 # -> Extract and interpret values for the Intensity variables
838
839     select Intensity 'praatname$'
840     if dbprev > 50
841         db = Get mean... t1 t2 dB
842         dbmin = Get minimum... t1 t2 None
843         dbmax = Get maximum... t1 t2 None
844         tdbmax = Get time of maximum... t1 t2 Parabolic
845         dbvar = 100 * (db - dbprev) / dbprev
846         dbprev = db
847     else
848         db = Get mean... t1 t2 dB
849         dbmin = Get minimum... t1 t2 None
850         dbmax = Get maximum... t1 t2 None
851         tdbmax = Get time of maximum... t1 t2 Parabolic
852         dbvar = 0
853         dbprev = db
854     endif
855
856     if dbvar > 7.50
857         vol$ = "ff"
858     elsif dbvar > 5.00 and dbvar < 7.49
859         vol$ = "f"
860     elsif dbvar > 2.50 and dbvar < 4.99
861         vol$ = "cresc"
862     elsif dbvar > -2.49 and dbvar < 2.49
863         vol$ = "nt"
864     elsif dbvar > -2.50 and dbvar < -4.99
865         vol$ = "dim"
866     elsif dbvar > -5.00 and dbvar < -7.49
867         vol$ = "p"
868     elsif dbvar > -7.50
869         vol$ = "pp"
870     endif
871
872 # -> Extract and simplify values for the Pitch variables
873
874     select Pitch 'praatname$'
875     f0min = Get minimum... t1 t2 Hertz None
876     f0max = Get maximum... t1 t2 Hertz None
877     f0tdbmax = Get value at time... tdbmax Hertz Linear
878
879     if f0min = undefined
880         f0min = 0
881     endif
882     if f0max = undefined
883         f0max = 0
884     endif
885     if f0tdbmax = undefined
886         f0tdbmax = 0
887     endif
888     if f0var = undefined
889         f0var = 0
890     endif
891     if f0vst = undefined
892         f0vst = 0
893     endif
894     if f0std = undefined

```

```

896         f0std = 100
897     endif
898
899 # -> Begin printing series of Samples
900
901     if i = 2
902         printline <Sample Name='tier_name$'>
903             ...'newline$'<U Quest='u_quest$' Emph='u_emph$'
904             ...Susp='u_susp$'>
905             ...'newline$'<IU img='file$' id='file$'>
906         endif
907
908 # -> Begin printing XML-like code for each syllable
909
910     if syll$ <> "" and syll$ <> "#"
911
912         print <F Wt='rhythm$'>'newline$'
913
914 # -> Print variables' values on the Info Window with an XML-like syntax
915
916         print <S Beg='t1:3'
917             ...End='t2:3'
918             ...Dur='dur:3'
919             ...Durvar='durvar:2'
920             ...Tmp='tmp$'
921             ...F0min='f0min:0'
922             ...F0max='f0max:0'
923             ...F0dBmax='f0tdbmax:0'
924             ...F0='f0:0'
925             ...F0var='f0var:2'
926             ...F0vst='f0vst:2'
927             ...F0std='f0std:0'
928             ...dBmin='dbmin:0'
929             ...dBmax='dbmax:0'
930             ...tdBmax='tdbmax:3'
931             ...dB='db:0'
932             ...dBvar='dbvar:2'
933             ...Vol='vol$'
934             ...Stress='stress$'
935             ...Focus='focus$'
936             ...Tone='tone$'
937             ...Phon='phon$'
938             ...Orth='syll$'>
939
940 # -> Interpret punctuation in orthographical tier to get breaks' values
941
942     length = length (syll$)
943     break$ = "[a-zA-ZáéíóúñÁĒÍÓÚÑ]\.[a-zA-ZáéíóúñÁĒÍÓÚÑ]"
944     break0$ = "[a-zA-ZáéíóúñÁĒÍÓÚÑ]"
945     break1$ = "\."
946
947     if index_regex (syll$, break$)
948         right = rindex_regex (syll$, break$)
949         left = index_regex (syll$, break$)
950         if right = left and left > 0
951             first$ = left$ (syll$, left + 1)
952             second$ = right$ (syll$, length - (left + 1))
953             print 'first$'<Break Type='0'/'>'second$'
954         elseif right > left
955             first$ = left$ (syll$, left + 1)
956             second$ = mid$ (syll$, left + 2, length - (right + 1))
957             third$ = right$ (syll$, length - (right + 1))
958             print 'first$'<Break Type='0'/'>'second$'
959             ...<Break Type='0'/'>'third$'
960         endif
961     elseif index_regex (syll$, "xxx")
962         print <Unintelligible>'syll$'</Unintelligible>
963     elseif index_regex (syll$, ":")
964         locate = index_regex (syll$, ":")
965         first$ = left$ (syll$, locate)
966         second$ = right$ (syll$, length - locate)
967         print 'first$'<Elongation/>'second$'
968     elseif index_regex (syll$, "-")
969         print 'syll$'<Fragment/>
970     else
971         print 'syll$'
972     endif
973
974

```

```

975     if index_regex (syll$, "~")
976         print <Mora/></S><Break Type='0'/></F>'newline$'
977     endif
978
979     right0 = rindex_regex (syll$, break0$)
980     if right0 = length
981         print </S><Break Type='0'/></F>'newline$'
982     endif
983
984     syll$ = replace$ (syll$, " ...", " ..", 1)
985
986     right1 = rindex_regex (syll$, break1$)
987     if right1 = length
988         print </S><Break Type='1'/></F>'newline$'
989     endif
990
991 # -> Interpret and print intonation unit or utterance boundaries
992
993     if i <= n - 1
994         nt$ = " //"
995         when = endsWith (syll$, nt$)
996         if when = 1
997             print </S><Break Type='4'/></F>
998             ...'newline$'<Border Type='nt'/>
999             ...'newline$'</IU>
1000             ...'newline$'</U>'newline$'
1001             if syllnext$ <> ""
1002                 printline <U Quest='u_quest$' Emph='u_emph$'
1003                     ...Susp='u_susp$'>
1004                     ...'newline$'<IU img='file$' id='file$'>
1005             endif
1006         endif
1007
1008         ext$ = " +/"
1009         when = endsWith (syll$, ext$)
1010         if when = 1
1011             print </S><Break Type='4'/></F>
1012             ...'newline$'<Border Type='ext'/>
1013             ...'newline$'</IU>
1014             ...'newline$'</U>'newline$'
1015             if syllnext$ <> ""
1016                 printline <U Quest='u_quest$' Emph='u_emph$'
1017                     ...Susp='u_susp$'>
1018                     ...'newline$'<IU img='file$' id='file$'>
1019             endif
1020         endif
1021
1022         slf$ = " +"
1023         when = endsWith (syll$, slf$)
1024         if when = 1
1025             print </S><Break Type='4'/></F>
1026             ...'newline$'<Border Type='slf'/>
1027             ...'newline$'</IU>
1028             ...'newline$'</U>'newline$'
1029             if syllnext$ <> ""
1030                 printline <U Quest='u_quest$' Emph='u_emph$'
1031                     ...Susp='u_susp$'>
1032                     ...'newline$'<IU img='file$' id='file$'>
1033             endif
1034         endif
1035
1036         qst$ = " ?"
1037         when = endsWith (syll$, qst$)
1038         if when = 1
1039             print </S><Break Type='4'/></F>
1040             ...'newline$'<Border Type='qst'/>
1041             ...'newline$'</IU>
1042             ...'newline$'</U>'newline$'
1043             if syllnext$ <> ""
1044                 printline <U Quest='u_quest$' Emph='u_emph$'
1045                     ...Susp='u_susp$'>
1046                     ...'newline$'<IU img='file$' id='file$'>
1047             endif
1048         endif
1049
1050         emp$ = " !"
1051         when = endsWith (syll$, emp$)
1052         if when = 1

```

```

1053     print </S><Break Type='4'/></F>
1054     ...'newline$'<Border Type='emp'/>
1055     ...'newline$'</IU>
1056     ...'newline$'</U>'newline$'
1057     if syllnext$ <> ""
1058         printline <U Quest='u_quest$' Emph='u_emph$'
1059             ...Susp='u_susp$'>
1060             ...'newline$'<IU img='file$' id='file$'>
1061     endif
1062 endif
1063
1064 ssp$ = " ..,"
1065 when = endsWith (syll$, ssp$)
1066 if when = 1
1067     print </S><Break Type='4'/></F>
1068     ...'newline$'<Border Type='ssp'/>
1069     ...'newline$'</IU>
1070     ...'newline$'</U>'newline$'
1071     if syllnext$ <> ""
1072         printline <U Quest='u_quest$' Emph='u_emph$'
1073             ...Susp='u_susp$'>
1074             ...'newline$'<IU img='file$' id='file$'>
1075     endif
1076 endif
1077
1078 iu$ = " /"
1079 when = endsWith (syll$, iu$)
1080 if when = 1
1081     print </S><Break Type='3'/></F>
1082     ...'newline$'<Border Type='iu'/>
1083     ...'newline$'</IU>'newline$'
1084     if i < n - 1 and syllnext$ <> "#"
1085         printline <IU img='file$' id='file$'>
1086     elseif i < n - 1 and syllnext$ = "#"
1087         if durnext < 0.50
1088             pause$ = "s"
1089         elseif durnext > 0.50 and durnext < 1.00
1090             pause$ = "l"
1091         elseif durnext > 1.00
1092             pause$ = "xl"
1093         endif
1094         if pause$ = "xl"
1095             print <Pause Sec='durnext:2' Type='pause$'> # </Pause>
1096             ...'newline$'<IU img='file$' id='file$'>'newline$'
1097         else
1098             print <Pause Sec='durnext:2' Type='pause$'>
1099             ...'newline$'<IU img='file$' id='file$'>'newline$'
1100         endif
1101         elseif i = n - 1
1102             printline </U>
1103         endif
1104     endif
1105 endif
1106
1107 rep$ = " [/"
1108 when = endsWith (syll$, rep$)
1109 if when = 1
1110     print </S><Break Type='3'/></F><Restart Type='rep'/>
1111     ...'newline$'<Border Type='rep'/>
1112     ...'newline$'</IU>'newline$'
1113     if i < n - 1 and syllnext$ <> "#"
1114         printline <IU img='file$' id='file$'>
1115     elseif i < n - 1 and syllnext$ = "#"
1116         if durnext < 0.50
1117             pause$ = "s"
1118         elseif durnext > 0.50 and durnext < 1.00
1119             pause$ = "l"
1120         elseif durnext > 1.00
1121             pause$ = "xl"
1122         endif
1123         if pause$ = "xl"
1124             print <Pause Sec='durnext:2' Type='pause$'> # </Pause>
1125             ...'newline$'<IU img='file$' id='file$'>'newline$'
1126         else
1127             print <Pause Sec='durnext:2' Type='pause$'>
1128             ...'newline$'<IU img='file$' id='file$'>'newline$'
1129         endif
1130         elseif i = n - 1
1131             printline </U>

```

```

1132     endif
1133   endif
1134
1135   prt$ = " [//]"
1136   when = endsWith (syll$, prt$)
1137   if when = 1
1138     print </S><Break Type='3'/></F><Restart Type='prt'/>
1139     ...'newline$'<Border Type='prt'/>
1140     ...'newline$'</IU>'newline$'
1141     if i < n - 1 and syllnext$ <> "#"
1142       printline <IU img='file$' id='file$'>
1143       elsif i < n - 1 and syllnext$ = "#"
1144         if durnext < 0.50
1145           pause$ = "s"
1146         elsif durnext > 0.50 and durnext < 1.00
1147           pause$ = "l"
1148         elsif durnext > 1.00
1149           pause$ = "xl"
1150         endif
1151         if pause$ = "xl"
1152           print <Pause Sec='durnext:2' Type='pause$'> # </Pause>
1153           ...'newline$'<IU img='file$' id='file$'>'newline$'
1154           else
1155             print <Pause Sec='durnext:2' Type='pause$'>
1156             ...'newline$'<IU img='file$' id='file$'>'newline$'
1157           endif
1158         elsif i = n - 1
1159           printline </U>
1160         endif
1161       endif
1162
1163   tot$ = " [//]"
1164   when = endsWith (syll$, tot$)
1165   if when = 1
1166     print </S><Break Type='3'/></F><Restart Type='tot'/>
1167     ...'newline$'<Border Type='tot'/>
1168     ...'newline$'</IU>'newline$'
1169     if i < n - 1 and syllnext$ <> "#"
1170       printline <IU img='file$' id='file$'>
1171       elsif i < n - 1 and syllnext$ = "#"
1172         if durnext < 0.50
1173           pause$ = "s"
1174         elsif durnext > 0.50 and durnext < 1.00
1175           pause$ = "l"
1176         elsif durnext > 1.00
1177           pause$ = "xl"
1178         endif
1179         if pause$ = "xl"
1180           print <Pause Sec='durnext:2' Type='pause$'> # </Pause>
1181           ...'newline$'<IU img='file$' id='file$'>'newline$'
1182           else
1183             print <Pause Sec='durnext:2' Type='pause$'>
1184             ...'newline$'<IU img='file$' id='file$'>'newline$'
1185           endif
1186         elsif i = n - 1
1187           printline </U>
1188         endif
1189       endif
1190     endif
1191
1192 # -> Select TextGrid file again to analyse the following interval
1193
1194     select TextGrid 'praatname$'
1195
1196     endif
1197
1198   endif
1199
1200 # -> End of Interval (syll) loop
1201
1202   endfor
1203 #-#-#-#(for i from 2 to n-1)
1204
1205 # -> Print end of a single Sample
1206
1207   printline </Sample>
1208
1209 # -> Select and remove all objects generated by Praat for this Sample
1210

```

```

1211  select TextGrid 'praatname$'
1212  plus Sound 'praatname$'
1213  plus IntervalTier phon
1214  plus IntervalTier f0
1215  plus IntervalTier rhythm
1216  plus TextTier tones
1217  plus IntervalTier 'name$'
1218  plus TextGrid grid
1219  plus Pitch 'praatname$'
1220  plus Intensity 'praatname$'
1221  plus Spectrogram 'praatname$'
1222  Remove
1223
1224  # -> End of file (i_file) loop
1225
1226  endfor
1227  #-#-#-#-#-#-#(for i_file to n_files)
1228
1229  select Strings list
1230  Remove
1231
1232  # -> Print end of Transcript
1233
1234    printline </Set>
1235    ...'newline$'</Transcript>
1236
1237  # -> Write contents of Info Window with a given name
1238
1239  fappendinfo 'xml_dir$'\ 'xml_file$'.xml
1240
1241  clearinfo
1242
1243  #++++++#
1244  #
1245  # -> End of program Create_XML(U+).praat written by Eduardo Velázquez
1246  #
1247  #++++++#

```


Anexo 7: Subprograma ReviewTextGrids.praat

Líneas	Secciones
16 – 19	Definición del formulario
23 – 24	Creación de la lista de archivos de sonido
28 – 62	Ciclo que procesará todos los archivos especificados
36 – 41	Lectura de los archivos TextGrid y Sound con el mismo nombre de base
45 – 48	Apertura de ventana del editor de TextGrids y aparición de un mensaje de alerta
52 – 57	Escritura del archivo TextGrid, eliminación de los archivos de la ventana Praat objects y reinicio del ciclo
67	Fin del subprograma

```

01 #+-----+
02 #
03 # This program was last updated on 24th March, 2007 by Eduardo Velázquez
04 #
05 # ReviewTextGrids.praat will display a form asking for the directory
06 # where Sound and TextGrid files are kept. All Sound files in the speci-
07 # fied directory will be listed and assorted. A loop will go through the
08 # specified files, reading and opening them. A message will prompt:
09 # "When ready, press 'Continue'". After clicking on the button, the
10 # TextGrid file will be saved.
11 #
12 #+-----+
13
14 # -> Definition of Form
15
16 form Review TextGrid files
17     comment Enter directory where Sound and TextGrid files are kept:
18     sentence dir C:\
19 endform
20
21 # -> Create list of Sound files
22
23 Create Strings as file list... list 'dir$'\*.wav
24 Sort
25
26 # -> Loop that goes through all the specified files
27
28 n_files = Get number of strings
29 for i_file to n_files
30     select Strings list
31     file_name$ = Get string... i_file
32     file$ = file_name$ - ".wav"
33
34 # -> Read in the TextGrid and Sound files with that base name
35
36     Read from file... 'dir$'\file$.wav
37     Read from file... 'dir$'\file$.TextGrid
38     praatname$ = selected$ ("TextGrid")
39
40     select Sound 'praatname$'
41     plus TextGrid 'praatname$'
42
43 # -> Open TextGrid window and prompt a message
44
45     Edit
46     editor TextGrid 'praatname$'
47     pause When ready, press 'Continue'.
48     endeditor
49
50 # -> Save TextGrid, remove files from Praat objects and continue
51
52     select TextGrid 'praatname$'

```


Anexo 8: Subprograma RecalculateF0var&F0std.praat

Líneas	Secciones
15 - 26	Asignación de variables para el porcentaje de variación de f0 y para los valores estandarizados de f0
31 - 56	Localización y extracción de la frecuencia del intervalo actual, el posterior y los dos previos
65 - 228	Extracción de valores para las variables de frecuencia
232 - 235	Despliegue de la ventana del editor de TextGrids
239	Fin del subprograma

```

01 #+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---#
02 #
03 # This program was last updated on 6th June, 2007 by Eduardo Velázquez
04 #
05 # RecalculateF0var&F0std.praat is a tool created to work on finished
06 # syllabically segmented TextGrid files containing eight tiers, where
07 # f0 values are in tier 3, f0var in tier 4 and f0std in tier 5. These
08 # two latter will be recalculated. TextGrid Editor window must be clos-
09 # ed before clicking on the button and it will be re-opened when ready.
10 #
11 #+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---#
12
13 # -> Set predefined percent f0 variation and standard f0 values
14
15 praatname$ = selected$ ("TextGrid")
16
17   select TextGrid 'praatname$'
18   n = Get number of intervals... 3
19   n_ortho = Get number of intervals... 8
20
21   f0prev = 0
22   f0std = 100
23   f0var = 0
24   f0vst = 0
25
26   for i to n
27
28   # -> Localize and extract Pitch of current, next, and two previous
29   # -> intervals (first and last intervals must always be empty)
30
31   select TextGrid 'praatname$'
32     f0 = Get label of interval... 3 i
33     ti = Get starting point... 3 i
34     f0$ = "'f0'"
35     if f0$ = ""
36       Set interval text... 3 i 0
37     endif
38     f0 = Get label of interval... 3 i
39     f0$ = "'f0'"
40
41     if i < n - 1
42       f0next = Get label of interval... 3 i+1
43     endif
44
45     if i > 2
46       f0prev = Get label of interval... 3 i-1
47       f0prev2 = Get label of interval... 3 i-2
48       f0prev$ = "'f0prev'"
49       f0prev2$ = "'f0prev2'"
50       if f0prev$ = ""
51         f0prev$ = "0"
52       endif
53       if f0prev2$ = ""
54         f0prev2$ = "0"
55       endif
56     endif

```

```

57
58 # -> Extract values for the Pitch variables
59 # -> (f0 value of annotated intervals without perceivable pitch last
60 # -> must be marked with 0, standardized f0 will be the average of
61 # -> and next intervals and will be displayed between parenthesis.)
62 # -> (In TextGrids containing more than one utterance the values for
63 # -> f0var and f0std will be reset to 0% and 100.)
64
65     if n_ortho < 4
66         tortho2_1 = Get starting point... 8 3
67         tortho2_2 = Get end point... 8 3
68
69         if i > 2 and f0 <> 0 and ti <> tortho2_1
70             f0var = 100 * (f0 - f0prev) / f0prev
71             f0vst = f0var / 100 * 12
72             f0std = (1 + (f0var / 100)) * f0stdprev
73             f0varnext = 100 * (f0next - f0) / f0
74             f0stdnext = (1 + (f0varnext / 100)) * f0std
75             if f0prev$ = "0"
76                 f0var = 100 * (f0 - f0prev2) / f0prev2
77                 f0vst = f0var / 100 * 12
78                 if i = 3
79                     f0std = 100
80                 elseif i >= 4
81                     f0std = (1 + (f0var / 100)) * f0stdprev2
82                 endif
83             endif
84         endif
85
86     elseif i > 2 and f0 = 0 and ti <> tortho2_1
87         f0var = 0
88         f0vst = 0
89         f0varnext = 100 * (f0next - f0prev) / f0prev
90         f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
91         f0std = (f0stdnext + f0stdprev) / 2
92         f0var$ = "'f0var'"
93         f0vst$ = "'f0vst'"
94         f0std$ = "'f0std'"
95
96         elseif ti = tortho2_1
97             f0var = 0
98             f0vst = 0
99             f0varnext = 100 * (f0next - f0prev) / f0prev
100            f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
101            f0std = 100
102            f0var$ = "'f0var'"
103            f0vst$ = "'f0vst'"
104            f0std$ = "'f0std'"
105            endif
106
107     elseif n_ortho = 4
108         tortho2_1 = Get starting point... 8 3
109         tortho2_2 = Get end point... 8 3
110         tortho3_1 = Get starting point... 8 4
111         tortho3_2 = Get end point... 8 4
112
113         if i > 2 and f0 <> 0 and ti <> tortho2_1 and ti <> tortho3_1
114             f0var = 100 * (f0 - f0prev) / f0prev
115             f0vst = f0var / 100 * 12
116             f0std = (1 + (f0var / 100)) * f0stdprev
117             f0varnext = 100 * (f0next - f0) / f0
118             f0stdnext = (1 + (f0varnext / 100)) * f0std
119             if f0prev$ = "0"
120                 f0var = 100 * (f0 - f0prev2) / f0prev2
121                 f0vst = f0var / 100 * 12
122                 if i = 3
123                     f0std = 100
124                 elseif i >= 4
125                     f0std = (1 + (f0var / 100)) * f0stdprev2
126                 endif
127             endif
128         endif
129
130     elseif i > 2 and f0 = 0 and ti <> tortho2_1 and ti <> tortho3_1
131         f0var = 0
132         f0vst = 0
133         f0varnext = 100 * (f0next - f0prev) / f0prev
134         f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
135         f0std = (f0stdnext + f0stdprev) / 2
136         f0var$ = "'f0var'"

```

```

136     f0vst$ = "'f0vst'"
137     f0std$ = "'f0std'"
138
139     elsif ti = tortho2_1 or ti = tortho3_1
140     f0var = 0
141     f0vst = 0
142         f0varnext = 100 * (f0next - f0prev) / f0prev
143         f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
144     f0std = 100
145     f0var$ = "'f0var'"
146     f0vst$ = "'f0vst'"
147     f0std$ = "'f0std'"
148     endif
149
150 elsif n_ortho = 5
151     tortho2_1 = Get starting point... 8 3
152     tortho2_2 = Get end point... 8 3
153     tortho3_1 = Get starting point... 8 4
154     tortho3_2 = Get end point... 8 4
155     tortho4_1 = Get starting point... 8 5
156     tortho4_2 = Get end point... 8 5
157
158     if i > 2 and f0 <> 0 and ti <> tortho2_1 and ti <> tortho3_1
159     ...and ti <> tortho4_1
160         f0var = 100 * (f0 - f0prev) / f0prev
161     f0vst = f0var / 100 * 12
162     f0std = (1 + (f0var / 100)) * f0stdprev
163     f0varnext = 100 * (f0next - f0) / f0
164     f0stdnext = (1 + (f0varnext / 100)) * f0std
165     if f0prev$ = "0"
166         f0var = 100 * (f0 - f0prev2) / f0prev2
167         f0vst = f0var / 100 * 12
168         if i = 3
169             f0std = 100
170         elsif i >= 4
171             f0std = (1 + (f0var / 100)) * f0stdprev2
172         endif
173     endif
174 endif
175
176 elsif i > 2 and f0 = 0 and ti <> tortho2_1 and ti <> tortho3_1
177 ...and ti <> tortho4_1
178     f0var = 0
179     f0vst = 0
180     f0varnext = 100 * (f0next - f0prev) / f0prev
181     f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
182     f0std = (f0stdnext + f0stdprev) / 2
183     f0var$ = "'f0var'"
184     f0vst$ = "'f0vst'"
185     f0std$ = "'f0std'"
186
187     elsif ti = tortho2_1 or ti = tortho3_1 or ti = tortho4_1
188     f0var = 0
189     f0vst = 0
190         f0varnext = 100 * (f0next - f0prev) / f0prev
191         f0stdnext = (1 + (f0varnext / 100)) * f0stdprev
192     f0std = 100
193     f0var$ = "'f0var'"
194     f0vst$ = "'f0vst'"
195     f0std$ = "'f0std'"
196     endif
197
198 endif
199
200     if f0var = undefined
201         f0var = 0
202     endif
203     if f0vst = undefined
204         f0vst = 0
205     endif
206     if f0std = undefined
207         f0std = 0
208     endif
209
210     if i >= 4
211         f0stdprev2 = f0stdprev
212     endif
213
214     f0stdprev = f0std

```


Anexo 10: Hoja de estilo `Prosody.xsl`

Líneas	Secciones
18 - 1395	Plantilla principal que se aplicará a todo el elemento <code>Transcript</code>
27 - 220	Encabezado con título, enlaces al resumen y a las convenciones, así como una tabla con los datos de los participantes
224 - 233	Botones para contraer o expandir todas las tablas de imagen y sonido
237 - 330	Estructura básica de los turnos de palabra y los enunciados, ordenados según la clave de identificación de cada unidad entonativa
243 - 248	Alternativa para evitar los turnos de palabra producidos por <code>EDU</code> , así como los enunciados solapados, comentados con <code>act</code> (risas, etc.) o ininteligibles
257 - 263	Proceso para permitir que cada una de las tablas de imagen y sonido pueda abrirse de manera independiente
268	Impresión del texto de cada enunciado en el turno de palabra
278 - 282	Asignación de variables para los atributos <code>img</code> e <code>id</code> de cada unidad entonativa
274 - 323	Tablas alterables de imagen y sonido
310 - 320	Impresión de la clave de identidad de la unidad entonativa actual
325 - 328	Fragmento que deberá sacarse de los comentarios cuando el filtro de turnos y enunciados esté activo
335 - 567	Resumen de fenómenos lingüísticamente relevantes
571 - 1392	Convenciones de los signos utilizados en la transcripción para representar los fenómenos lingüísticamente relevantes
1393 - 1395	Fin de la plantilla principal
1401 - 1406	Despliegue de la duración de la pausa, anclada al signo <code>#</code>
1411 - 1425	Asignación de una imagen diferente delante de las unidades entonativas en posición distinta a la primera
1430 - 1444	Asignación de una imagen diferente delante de los enunciados en posición distinta a la primera
1446	Fin de la hoja de estilo

```

01 <!-- Prosody.xsl - Version 5 -->
02 <!-- Declaration of XML document -->
03 <!-- Declaration of XSL document -->
04 <!-- Date: 25th November 2004 -->
05 <!-- Last Update: 18th October 2006 -->
06 <!-- Written by: Eduardo Velazquez -->
07
08 <xsl:stylesheet version="1.0"
09     xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
10     xmlns="http://www.w3.org/TR/xhtml1/strict">
11
12 <xsl:output
13     method="html"
14     encoding="utf-8" />
15

```



```

16 <!-- Primary template to be applied to the whole Transcript -->
17
18 <xsl:template match="Transcript">
19   <html>
20     <head>
21       <title>Transcript of File: <xsl:value-of select="Header/File"/></title>
22     </head>
23
24 <!-- Header with Title, links to Summary and Coventions, as well as Participants'
25     data-->
26
27     <body bgcolor="#FFFFFF">
28       <a name="top" />
29       <h2>
30         <p align="center">
31           <font face="Times New Roman, Times, serif">
32             Transcript of File: <xsl:value-of select="Header/File"/>
33           </font>
34         </p>
35       </h2>
36       <p align="center">
37         <font size="-1"><a href="#sum" title="Go to Summary"><em>Summary</em></a>
38         and
39         <a href="#conv" title="Go to Conventions"><em>conventions</em></a> at the
40         bottom)</font></p>
41       <xsl:if test="Header">
42         <table width="80%" border="1" align="center">
43           <tr>
44             <td>
45               <table width="100%" border="0" align="center">
46                 <tr>
47                   <td width="20%" valign="top">
48                     <font face="Arial, Helvetica, sans-serif" size="2">
49                       <b>Title:</b></font></td>
50                   <td width="80%">
51                     <font face="Arial, Helvetica, sans-serif" size="2">
52                       <xsl:value-of select="Header/Title"/></font></td>
53                   </tr>
54                   <tr>
55                     <td width="20%" valign="top">
56                       <font face="Arial, Helvetica, sans-serif" size="2">
57                         <b>Data:</b></font></td>
58                     <td width="80%">
59                       <font face="Arial, Helvetica, sans-serif" size="2">
60                         <xsl:value-of select="Header/Data"/></font></td>
61                     </tr>
62                   <tr>
63                     <td width="20%" valign="top">
64                       <font face="Arial, Helvetica, sans-serif" size="2">
65                         <b>Place:</b></font></td>
66                     <td width="80%">
67                       <font face="Arial, Helvetica, sans-serif" size="2">
68                         <xsl:value-of select="Header/Place"/></font></td>
69                     </tr>
70                   <tr>
71                     <td width="20%" valign="top">
72                       <font face="Arial, Helvetica, sans-serif" size="2">
73                         <b>Date:</b></font></td>
74                     <td width="80%">
75                       <font face="Arial, Helvetica, sans-serif" size="2">
76                         <xsl:value-of select="Header/Date"/></font></td>
77                     </tr>
78                   <tr>
79                     <td width="20%" valign="top">
80                       <font face="Arial, Helvetica, sans-serif" size="2">
81                         <b>Topic:</b></font></td>
82                     <td width="80%">
83                       <font face="Arial, Helvetica, sans-serif" size="2">
84                         <xsl:value-of select="Header/Topic"/></font></td>
85                     </tr>
86                   <tr>
87                     <td width="20%" valign="top">
88                       <font face="Arial, Helvetica, sans-serif" size="2">
89                         <b>Situation:</b></font></td>
90                     <td width="80%">
91                       <font face="Arial, Helvetica, sans-serif" size="2">
92                         <xsl:value-of select="Header/Situation"/></font></td>
93                     </tr>
94                   </tr>

```

```

95 <tr>
96 <td width="20%" valign="top">
97 <font face="Arial, Helvetica, sans-serif" size="2">
98 <b>Source:</b></font></td>
99 <td width="80%">
100 <font face="Arial, Helvetica, sans-serif" size="2">
101 <xsl:value-of select="Header/Source"/></font></td>
102 </tr>
103 <tr>
104 <td width="20%" valign="top">
105 <font face="Arial, Helvetica, sans-serif" size="2">
106 <b>Class:</b></font></td>
107 <td width="80%">
108 <font face="Arial, Helvetica, sans-serif" size="2">
109 <xsl:text> </xsl:text>
110 <xsl:value-of select="Header/Class/@Type1"/><xsl:text>,
111 </xsl:text><xsl:value-of select="Header/Class/@Type2"/>
112 <xsl:text>,
113 </xsl:text><xsl:value-of select="Header/Class/@Type3"/>
114 </font></td>
115 </tr>
116 <tr>
117 <td width="20%" valign="top">
118 <font face="Arial, Helvetica, sans-serif" size="2">
119 <b>Length (in seconds):</b></font></td>
120 <xsl:variable name="length">
121 <xsl:value-of select="Header/Length"/>
122 </xsl:variable>
123 <td width="80%">
124 <font face="Arial, Helvetica, sans-serif" size="2">
125 <xsl:value-of select="format-number($length, '#.##')"/>
126 </font></td>
127 </tr>
128 <tr>
129 <td width="20%" valign="top">
130 <font face="Arial, Helvetica, sans-serif" size="2">
131 <b>Words:</b></font></td>
132 <td width="80%">
133 <font face="Arial, Helvetica, sans-serif" size="2">
134 <xsl:value-of select="Header/Words"/></font></td>
135 </tr>
136 <tr>
137 <td width="20%" valign="top">
138 <font face="Arial, Helvetica, sans-serif" size="2">
139 <b>Acoustic quality:</b></font></td>
140 <td width="80%">
141 <font face="Arial, Helvetica, sans-serif" size="2">
142 <xsl:value-of select="Header/Acoustic_quality/@Type"/>
143 </font></td>
144 </tr>
145 <tr>
146 <td width="20%" valign="top">
147 <font face="Arial, Helvetica, sans-serif" size="2">
148 <b>Transcriber:</b></font></td>
149 <td width="80%">
150 <font face="Arial, Helvetica, sans-serif" size="2">
151 <xsl:value-of select="Header/Transcriber"/></font></td>
152 </tr>
153 <tr>
154 <td width="20%" valign="top">
155 <font face="Arial, Helvetica, sans-serif" size="2">
156 <b>Revisor:</b></font></td>
157 <td width="80%">
158 <font face="Arial, Helvetica, sans-serif" size="2">
159 <xsl:value-of select="Header/Revisor"/></font></td>
160 </tr>
161 <tr>
162 <td width="20%" valign="top">
163 <font face="Arial, Helvetica, sans-serif" size="2">
164 <b>Comments:</b></font></td>
165 <td width="80%">
166 <font face="Arial, Helvetica, sans-serif" size="2">
167 <xsl:value-of select="Header/Comments"/></font></td>
168 </tr>
169 </table>
170 </td>
171 </tr>
172 </table>
173 </table>

```

```

174 </xsl:if>
175 <xsl:if test="Header/Participants">
176 <h4><p align="center"><font face="Arial, Helvetica, sans-
177 serif">Participants:</font></p></h4>
178 <table border="1" width="80%" align="center">
179 <tr>
180 <th><font face="Arial, Helvetica, sans-serif" size="2">
181 <b>Name</b></font></th>
182 <th><font face="Arial, Helvetica, sans-serif" size="2">
183 <b>Sex</b></font></th>
184 <th><font face="Arial, Helvetica, sans-serif" size="2">
185 <b>Age</b></font></th>
186 <th><font face="Arial, Helvetica, sans-serif" size="2">
187 <b>Education</b></font></th>
188 <th><font face="Arial, Helvetica, sans-serif" size="2">
189 <b>Occupation</b></font></th>
190 <th><font face="Arial, Helvetica, sans-serif" size="2">
191 <b>Role</b></font></th>
192 <th><font face="Arial, Helvetica, sans-serif" size="2">
193 <b>Origin</b></font></th>
194 </tr>
195 <xsl:for-each select="Header/Participants/Speaker">
196 <tr>
197 <td><xsl:value-of select="Name"/></td>
198 <td><div align="center"><xsl:value-of select="Sex/@Type"/></div></td>
199 <td><div align="center"><xsl:value-of select="Age/@Type"/></div></td>
200 <td><div align="center">
201 <xsl:value-of select="Education/@Type"/></div></td>
202 <td><xsl:value-of select="Occupation"/></td>
203 <td><xsl:value-of select="Role"/></td>
204 <td><xsl:value-of select="Origin"/></td>
205 </tr>
206 </xsl:for-each>
207 <tr>
208 <td colspan="7"><font face="Arial, Helvetica, sans-serif" size="2">
209 x = unknown <br />
210 Age: A (18-25), B (26-40), C (40-60), D (&gt;60)<br />
211 Education: 1 (primary school or less), 2 (high school), 3
212 (graduated or university students), x (unknown)<br />
213 </font>
214 </td>
215 </tr>
216 </table>
217 <p/>
218 <hr/>
219 <p/>
220 </xsl:if>
221
222 <!-- Buttons to contract or to expand all Picture-and-Sound Tables -->
223
224 <div>
225 <center>
226 <p>
227 <font face="Arial, Helvetica, sans-serif">
228 <a href="javascript:sweeptoggle('contract')">Contract All</a> |
229 <a href="javascript:sweeptoggle('expand')">Expand All</a>
230 </font>
231 </p>
232 </center>
233 </div>
234
235 <!-- Basic structure of turns and utterances, ordered by intonation units' ID-->
236
237 <xsl:for-each select="Text/Turn">
238 <xsl:sort select="U/IU/@id" />
239
240 <!-- Alternative to last line in order to avoid turns produced by EDU, as well as
241 overlapped, act-commented (laughs, etc.) and unintelligible utterances -->
242
243 <!-- <xsl:variable name="tname"><xsl:value-of select="@Name" /></xsl:variable>
244 <xsl:variable name="edu">EDU</xsl:variable>
245 <xsl:if test="not($tname = $edu)">
246 <xsl:if test="not(U/IU/Overlap)">
247 <xsl:if test="not(U/IU/Comment/@act)">
248 <xsl:if test="not(U/IU/Unintelligible)">
249 -->
250 <xsl:variable name="n"><xsl:number /></xsl:variable>
251 <table width="100%" border="0">

```

```

252     <tr>
253         <td width="7%" align="left" valign="top">
254
255 <!-- Every Turn's Picture-and-Sound tables may be opened independently -->
256
257         <h4 onClick="expandcontent(this, 'sc{$n}')"
258             style="cursor:hand; cursor:pointer">
259             <span class="showstate"></span>
260             <font face="Courier New, Courier, mono">
261                 <strong>&#160;*<xsl:value-of select='@Name' />: </strong>
262             </font>
263         </h4>
264
265 <!-- Print text of every Utterance in a Turn -->
266
267             <xsl:apply-templates /><br/>
268
269 <!-- Set values of every intonation unit's attribute 'img' and 'id' -->
270
271 <!-- Switchable tables -->
272
273         <div id="sc{$n}" class="switchcontent">
274
275             <xsl:for-each select="U/IU">
276                 <xsl:if test="@img">
277                     <xsl:variable name="img">
278                         <xsl:value-of select="@img" /></xsl:variable>
279                     <xsl:variable name="id">
280                         <xsl:value-of select="@id" /></xsl:variable>
281                     <xsl:variable name="i"><xsl:number /></xsl:variable>
282
283 <!-- Picture-and-Sound Table -->
284
285                 <table border="1">
286                     <tr>
287                         <td colspan="22">
288                             <center>
289                                 
290                                 <br/><br/>
291                                 <form name="button" method="get"
292                                     action="HTML/Sounds/{$img}.wav"
293                                     accept="audio/wav" target="_self">
294                                     <a href="javascript:audio({$img})()">
295                                         <input name="Play" type="submit"
296                                             id="Play" value="    Play audio fragment    "/></a>
297                                 </form>
298                             </center>
299                         </td>
300                     </tr>
301                 </table>
302             </xsl:if>
303         </xsl:for-each>
304     </div>
305
306 </td>
307
308 <!-- Print current intonation unit's ID -->
309
310         <td width="20%" align="right" valign="top">
311             <xsl:for-each select=".">
312                 <xsl:for-each select="U/IU">
313                     <xsl:if test="@id">
314                         <h5>
315                             <font face="Courier New, Courier, mono" size="-2">
316                                 <xsl:value-of select="@id" /></font>
317                         </h5>
318                     </xsl:if>
319                 </xsl:for-each>
320             </xsl:for-each>
321
322 </td>
323 </tr>
324 </table>
325 <!-- Uncomment when turn and utterance filter is turned on! -->
326 <!-- </xsl:if>
327 </xsl:if>
328 </xsl:if>
329 -->
330 </xsl:for-each>

```

```

331
332
333 <!-- Summary of linguistic relevant phenomena -->
334
335     <hr/>
336     <p>
337     <h3 align="center">
338         <a name="sum">Summary</a>
339     </h3>
340     </p>
341     <table width="80%" border="0" align="center">
342     <tr>
343         <td width="40%">
344             <font size="-1" face="Arial, Helvetica, sans-serif">
345                 <strong>Turns:</strong>
346             </font>
347         </td>
348         <td width="12%">
349             <div align="right">
350                 <xsl:value-of select="count(..//Turn)"/>
351             </div>
352         </td>
353         <td width="16%"/>
354         <td width="40%">
355             <font size="-1" face="Arial, Helvetica, sans-serif">
356                 <strong>Intonation units:</strong>
357             </font>
358         </td>
359         <td width="12%">
360             <div align="right">
361                 <xsl:value-of select="count(..//IU)"/>
362             </div>
363         </td>
364     </tr>
365     <tr>
366     <td>
367         <font size="-1" face="Arial, Helvetica, sans-serif">
368             <strong>Utterances:</strong>
369         </font>
370     </td>
371     <td>
372         <div align="right">
373             <xsl:value-of select="count(..//U)"/>
374         </div>
375     </td>
376     <td/>
377     <td>
378         <font size="-1" face="Arial, Helvetica, sans-serif">
379             <strong>Syllables:</strong>
380         </font>
381     </td>
382     <td>
383         <div align="right">
384             <xsl:value-of select="count(..//S)"/>
385         </div>
386     </td>
387 </tr>
388 <tr>
389 <td>
390     <font size="-1" face="Arial, Helvetica, sans-serif">
391         <strong>Assertive utterances:</strong>
392     </font>
393 </td>
394 <td>
395     <div align="right">
396         <xsl:value-of select="count(..//Border[@Type='nt'])"/>
397     </div>
398 </td>
399 <td/>
400 <td>
401     <font size="-1" face="Arial, Helvetica, sans-serif">
402         <strong>Elisions/Liaisons:</strong>
403     </font>
404 </td>
405 <td>
406     <div align="right">
407         <xsl:value-of select="count(..//S/Break[@Type='0'])"/>
408     </div>
409 </td>

```

```

410     </td>
411 </tr>
412 <tr>
413     <td>
414         <font size="-1" face="Arial, Helvetica, sans-serif">
415             <strong>Interrogative utterances:</strong>
416         </font>
417     </td>
418     <td>
419         <div align="right">
420             <xsl:value-of select="count(..//Border[@Type='qst'])"/>
421         </div>
422     </td>
423 </tr>
424 <tr>
425     <td>
426         <font size="-1" face="Arial, Helvetica, sans-serif">
427             <strong>Segmental elongations:</strong>
428         </font>
429     </td>
430     <td>
431         <div align="right">
432             <xsl:value-of select="count(..//Elongation)"/>
433         </div>
434 </tr>
435 <tr>
436     <td>
437         <font size="-1" face="Arial, Helvetica, sans-serif">
438             <strong>Suspended utterances:</strong>
439         </font>
440     </td>
441     <td>
442         <div align="right">
443             <xsl:value-of select="count(..//Border[@Type='ssp'])"/>
444         </div>
445     </td>
446 </tr>
447 <tr>
448     <td>
449         <font size="-1" face="Arial, Helvetica, sans-serif">
450             <strong>Unintelligible productions:</strong>
451         </font>
452     </td>
453     <td>
454         <div align="right">
455             <xsl:value-of select="count(..//Unintelligible)"/>
456         </div>
457 </tr>
458 <tr>
459     <td>
460         <font size="-1" face="Arial, Helvetica, sans-serif">
461             <strong>Self-interrupted utterances:</strong>
462         </font>
463     </td>
464     <td>
465         <div align="right">
466             <xsl:value-of select="count(..//Border[@Type='slf'])"/>
467         </div>
468     </td>
469 </tr>
470 <tr>
471     <td>
472         <font size="-1" face="Arial, Helvetica, sans-serif">
473             <strong>Repetitions:</strong>
474         </font>
475     </td>
476     <td>
477         <div align="right">
478             <xsl:value-of select="count(..//Restart[@Type='rep'])"/>
479         </div>
480     </td>
481 </tr>
482 <tr>
483     <td>
484         <font size="-1" face="Arial, Helvetica, sans-serif">
485             <strong>Externally interrupted utterances:</strong>
486         </font>
487     </td>
488     <td>

```

```

489         <div align="right">
490             <xsl:value-of select="count(..//Border[@Type='ext'])"/>
491         </div>
492     </td>
493 </td>
494 <td>
495     <font size="-1" face="Arial, Helvetica, sans-serif">
496         <strong>Partial reformulations:</strong>
497     </font>
498 </td>
499 <td>
500     <div align="right">
501         <xsl:value-of select="count(..//Restart[@Type='prt'])"/>
502     </div>
503 </td>
504 </tr>
505 <tr>
506     <td>
507         <font size="-1" face="Arial, Helvetica, sans-serif">
508             <strong>Overlaps:</strong>
509         </font>
510     </td>
511     <td>
512         <div align="right">
513             <xsl:value-of select="count(..//Overlap[@Type='pas'])"/>
514         </div>
515     </td>
516 </td>
517 <td>
518     <font size="-1" face="Arial, Helvetica, sans-serif">
519         <strong>Total reformulations:</strong>
520     </font>
521 </td>
522 <td>
523     <div align="right">
524         <xsl:value-of select="count(..//Restart[@Type='tot'])"/>
525     </div>
526 </td>
527 </tr>
528 <tr>
529     <td>
530         <font size="-1" face="Arial, Helvetica, sans-serif">
531             <strong>Pauses (longer than 1 second):</strong>
532         </font>
533     </td>
534     <td>
535         <div align="right"><xsl:value-of
536         select="count(..//Pause[@Type='xl'])"/></div>
537     </td>
538 </td>
539 <td>
540     <font size="-1" face="Arial, Helvetica, sans-serif">
541         <strong>Total time (in seconds):</strong>
542     </font>
543 </td>
544 <td>
545     <div align="right">
546         <xsl:variable name="length">
547             <xsl:value-of
548             select="..//preceding-sibling::Header/Length"/>
549         </xsl:variable>
550         <xsl:value-of
551         select="format-number($length, '#.##')"/>
552     </div>
553 </td>
554 </tr>
555 </table>
556 <br/>
557 <div align="right">
558 <a href="#top">
559     <strong>
560         <font face="Arial, Helvetica, sans-serif">Top</font>
561     </strong>
562     
563 </a>
564 </div>
565 <hr/>

```

```

568
569 <!-- Transcript's sign conventions to represent linguistic relevant phenomena -->
570
571 <h3 align="center">
572 <a name="conv">Conventions</a>
573 </h3>
574 <table width="100%" border="0" align="center">
575 <tr>
576 <td colspan="3">
577 <div align="center">
578 <font size="-1">
579 <strong>
580 <font face="Arial, Helvetica,
581 sans-serif">Various</font>
582 </strong>
583 </font>
584 </div>
585 </td>
586 <td colspan="3">
587 <div align="center">
588 <font size="-1">
589 <strong>
590 <font face="Arial, Helvetica, sans-serif">
591 Utterances and Intonation Units</font>
592 </strong>
593 </font>
594 </div>
595 </td>
596 <td colspan="3">
597 <div align="center">
598 <font size="-1">
599 <strong>
600 <font face="Arial, Helvetica,
601 sans-serif">Acoustic Analysis</font>
602 </strong>
603 </font>
604 </div>
605 </td>
606 </tr>
607 <tr>
608 <td width="8%">
609 <div align="right">
610 <font size="-1">
611 <font face="Courier New, Courier, mono">
612 <strong>*ABC:</strong>
613 </font>
614 </div>
615 </td>
616 <td width="1%"/>
617 <td width="24%">
618 <font size="-1">
619 number of turn and speaker identification</font>
620 </td>
621 <td width="7%">
622 <div align="right">
623 <font size="-1"/></font>
624 </div>
625 </td>
626 <td width="1%"/>
627 <td width="24%">
628 <font size="-1">
629 border of intonation unit (line break)</font>
630 </td>
631 <td width="8%">
632 <div align="right">
633 <font size="-1">ABC</font>
634 </div>
635 </td>
636 <td width="1%"/>
637 <font size="-1"/>
638 </td>
639 <td width="26%">
640 <font size="-1">
641 normated orthographical transcription</font>
642 </td>
643 </tr>
644 </tr>
645 </tr>

```



```

646         <td>
647             <div align="right">
648                 <font size="-1">&lt; .. &gt;</font>
649             </div>
650         </td>
651     </td>
652     <td>
653         <font size="-1">
654             simultaneously-produced text (overlap)</font>
655     </td>
656     <td>
657         <div align="right">
658             <font size="-1">/ ..</font>
659         </div>
660     </td>
661 </td>
662 <td>
663     <font size="-1">
664         continuation of previous information unit</font>
665 </td>
666 <td>
667     <div align="right">
668         <font size="-1">Phon</font>
669     </div>
670 </td>
671 <td>
672 <td>
673     <font size="-1"/>
674 </td>
675 <td>
676     <font size="-1">
677         phonetic transcription (Praat or UTF-8)</font>
678 </td>
679 </tr>
680 <tr>
681     <td>
682         <div align="right">
683             <font size="-1">[&lt;] &lt; .. &gt;</font>
684         </div>
685     </td>
686 </td>
687 <td>
688     <font size="-1">overlapping text</font>
689 </td>
690 <td>
691     <div align="right">
692         <font size="-1">//</font>
693     </div>
694 </td>
695 </td>
696 <td>
697     <font size="-1">
698         border of assertive or exclamative utterance</font>
699 </td>
700 <td>
701     <div align="right">
702         <font size="-1">Beg</font>
703     </div>
704 </td>
705 <td>
706     <font size="-1"/>
707 </td>
708 <td>
709     <font size="-1">
710         beginning time of syllable in seconds</font>
711 </td>
712 </tr>
713 <tr>
714     <td>
715         <div align="right">
716             <font size="-1">
717                 
718                 ,
719                 
720             </font>
721         </div>
722     </td>
723 </td>
724 <td>

```

```

725         <font size="-1">
726         link to analysis menu (on mouse over!)</font>
727     </td>
728     <td>
729         <div align="right">
730             <font size="-1">?</font>
731         </div>
732     </td>
733 </td>
734 <td>
735     <font size="-1">
736     border of interrogative utterance</font>
737 </td>
738 <td width="8%">
739     <div align="right">
740         <font size="-1">End</font>
741     </div>
742 </td>
743 <td width="1%">
744     <font size="-1"/>
745 </td>
746 <td width="34%">
747     <font size="-1">
748     ending time of syllable in seconds</font>
749 </td>
750 </tr>
751 <tr>
752     <td>
753         <div align="right">
754             <font size="-1">
755                 
756             </font>
757         </div>
758     </td>
759 </td>
760 <td>
761     <font size="-1">new utterance in a turn</font>
762 </td>
763 <td>
764     <div align="right">
765         <font size="-1">...</font>
766     </div>
767 </td>
768 </td>
769 <td>
770 <td>
771     <font size="-1">utterance suspension</font>
772 </td>
773 <td>
774     <div align="right">
775         <font size="-1">Dur</font>
776     </div>
777 </td>
778 <td>
779     <div align="right">
780         <font size="-1">
781             <font size="-1"/>
782         </font>
783     </div>
784 </td>
785 <td>
786     <div align="left">
787         <font size="-1">
788             <font size="-1">
789             duration of syllable in seconds</font>
790         </font>
791     </div>
792 </td>
793 </tr>
794 <tr>
795     <td>
796         <div align="right">
797             <font size="-1">
798                 
799             </font>
800         </div>
801 </td>
802 </td>
803 <td>

```

```

804         <font size="-1">
805             new intonation unit in an utterance</font>
806     </td>
807     <td>
808         <div align="right">
809             <font size="-1">+</font>
810         </div>
811     </td>
812 </td>
813 <td>
814     <font size="-1">self-interruption</font>
815 </td>
816 <td>
817     <div align="right">
818         <font size="-1">Durvar</font>
819     </div>
820 </td>
821 <td>
822     <font size="-1"/>
823 </td>
824 <td>
825     <font size="-1">
826         variation of syllable duration (previous Dur = 100%)</font>
827 </td>
828 </tr>
829 <tr>
830     <td>
831         <div align="right">
832             <font size="-1">
833                 
835                 ,
836                 
838             </font>
839         </div>
840     </td>
841 </td>
842 <td>
843     <font size="-1">
844         expand or contract detailed analysis table</font>
845 </td>
846 <td>
847     <div align="right">
848         <font size="-1">+</font>
849     </div>
850 </td>
851 </td>
852 <td>
853     <font size="-1">
854         external interruption</font>
855 </td>
856 <td>
857     <div align="right">
858         <font size="-1">F0min</font>
859     </div>
860 </td>
861 <td>
862     <font size="-1"/>
863 </td>
864 <td>
865     <font size="-1">
866         fundamental frequency minimum in hertz</font>
867 </td>
868 </tr>
869 <tr>
870 <td>
871     <td>
872         <div align="right">
873             <font size="-1">#</font>
874         </div>
875     </td>
876 </td>
877 <td>
878     <font size="-1">pause longer than 1 second</font>
879 </td>
880 <td>
881     <div align="right">
882         <font size="-1">[</font>

```

```

883         </div>
884     </td>
885 </td>
886 <td>
887     <font size="-1">repetition</font>
888 </td>
889 <td>
890     <div align="right">
891         <font size="-1">F0max</font>
892     </div>
893 </td>
894 <td>
895     <font size="-1"/>
896 </td>
897 <td>
898     <font size="-1">
899     fundamental frequency maximum in hertz</font>
900 </td>
901 </tr>
902 <tr>
903 <td>
904     <div align="right">
905         <font size="-1">xxx</font>
906     </div>
907 </td>
908 <td/>
909 <td>
910     <font size="-1">unintelligible production</font>
911 </td>
912 <td>
913     <div align="right">
914         <font size="-1">[//]</font>
915     </div>
916 </td>
917 <td/>
918 <td>
919     <font size="-1">partial reformulation</font>
920 </td>
921 <td>
922     <div align="right">
923         <font size="-1">F0dBmax</font>
924     </div>
925 </td>
926 <td>
927     <font size="-1"/>
928 </td>
929 <td>
930     <font size="-1">
931     fundamental frequency at intensity maximum in hertz</font>
932 </td>
933 </tr>
934 <tr>
935 <td colspan="3">
936     <div align="center">
937         <font size="-1">
938             <strong>
939                 <font face="Arial, Helvetica,
940                 sans-serif">Comments</font>
941             </strong>
942         </font>
943     </div>
944 </td>
945 <td>
946     <div align="right">
947         <font size="-1">[///]</font>
948     </div>
949 </td>
950 <td/>
951 <td>
952     <font size="-1">total reformulation</font>
953 </td>
954 <td>
955     <div align="right">
956         <font size="-1">F0</font>
957     </div>
958 </td>
959 <td>
960     <font size="-1"/>

```

```

961         </td>
962         <td>
963             <font size="-1">
964                 mean fundamental frequency in hertz</font>
965         </td>
966     </tr>
967     <tr>
968         <td>
969             <div align="right">
970                 <font size="-1">[%act: ..]</font>
971             </div>
972         </td>
973     </tr>
974     <td/>
975     <td>
976         <font size="-1">speaker's action or speech act</font>
977     </td>
978     <td colspan="3">
979         <div align="center">
980             <font size="-1">
981                 <strong>
982                     <font face="Arial, Helvetica,
983                         sans-serif">Syllable Breaks</font>
984                 </strong>
985             </font>
986         </div>
987     </td>
988     <td>
989         <div align="right">
990             <font size="-1">F0var</font>
991         </div>
992     </td>
993     <td>
994         <font size="-1"/>
995     </td>
996     <td>
997         <font size="-1">
998             F0 variation in percent (previous F0 = 100%)</font>
999     </td>
1000 </tr>
1001 <tr>
1002     <td height="17">
1003         <div align="right">
1004             <font size="-1">[%sit: ..]</font>
1005         </div>
1006     </td>
1007     <td/>
1008     <td>
1009         <font size="-1">situation</font>
1010     </td>
1011     <td>
1012         <div align="right">
1013             <font size="-1">..&#183;..</font>
1014         </div>
1015     </td>
1016     <td/>
1017     <td>
1018         <font size="-1">
1019             syllable break (only polysyllabic words!)</font>
1020     </td>
1021     <td>
1022         <div align="right">
1023             <font size="-1">F0vst</font>
1024         </div>
1025     </td>
1026     <td>
1027         <font size="-1"/>
1028     </td>
1029     <td>
1030         <font size="-1">
1031             F0 variation in semitones (12 st = 100%)</font>
1032     </td>
1033 </tr>
1034 <tr>
1035     <td>
1036         <div align="right">
1037             <font size="-1">[%add: ..]</font>
1038         </div>
1039     </td>

```

```

1040 <td/>
1041 <td>
1042 <font size="-1">addressed person</font>
1043 </td>
1044 <td>
1045 <div align="right">
1046 <font size="-1">
1047 ..
1048 <strong>
1049 <font size="3">.</font>
1050 </strong>
1051 ..
1052 </font>
1053 </div>
1054 </td>
1055 <td/>
1056 <td>
1057 <font size="-1">elision, liaison</font>
1058 </td>
1059 <td>
1060 <div align="right">
1061 <font size="-1">F0std</font>
1062 </div>
1063 </td>
1064 <td>
1065 <font size="-1"/>
1066 </td>
1067 <td>
1068 <font size="-1">
1069 standardized fundamental frequency (1st F0 = 100Hz)</font>
1070 </td>
1071 </tr>
1072 <tr>
1073 <td>
1074 <div align="right">
1075 <font size="-1">[%par: ..]</font>
1076 </div>
1077 </td>
1078 <td/>
1079 <td>
1080 <font size="-1">paralinguistic elements</font>
1081 </td>
1082 <td>
1083 <div align="right">
1084 <font size="-1">
1085 ..
1086 <strong>
1087 <font size="3">.</font>
1088 </strong>
1089 </font>
1090 </div>
1091 </td>
1092 <td/>
1093 <td>
1094 <font size="-1">lexical word boundary</font>
1095 </td>
1096 <td>
1097 <div align="right">
1098 <font size="-1">dBmin</font>
1099 </div>
1100 </td>
1101 <td>
1102 <font size="-1"/>
1103 </td>
1104 <td>
1105 <font size="-1">intensity minimum</font>
1106 </td>
1107 </tr>
1108 <tr>
1109 <td>
1110 <div align="right">
1111 <font size="-1">[%alt: ..]</font>
1112 </div>
1113 </td>
1114 <td/>
1115 <td>
1116 <font size="-1">normative variant or proposal</font>
1117 </td>
1118 </tr>

```

```

1119         <td>
1120             <div align="right">
1121                 <font size="-1">..</font>
1122             </div>
1123         </td>
1124     </td>
1125     <td>
1126         <font size="-1">fragment</font>
1127     </td>
1128     <td>
1129         <div align="right">
1130             <font size="-1">dBmax</font>
1131         </div>
1132     </td>
1133     <td>
1134         <font size="-1"/>
1135     </td>
1136     <td>
1137         <font size="-1">intensity maximum</font>
1138     </td>
1139 </tr>
1140 <tr>
1141     <td>
1142         <div align="right">
1143             <font size="-1">[%com: ..]</font>
1144         </div>
1145     </td>
1146     <td/>
1147     <td>
1148         <font size="-1">transcripotor's comments</font>
1149     </td>
1150     <td colspan="3">
1151         <div align="center">
1152             <font size="-1">
1153                 <strong>
1154                     <font face="Arial, Helvetica,
1155                         sans-serif">Rhythm</font>
1156                 </strong>
1157             </font>
1158         </div>
1159     </td>
1160     <td>
1161         <div align="right">
1162             <font size="-1">tdBmax</font>
1163         </div>
1164     </td>
1165     <td>
1166         <font size="-1"/>
1167     </td>
1168     <td>
1169         <font size="-1">
1170         time of intensity maximum in seconds</font>
1171     </td>
1172 </tr>
1173 <tr>
1174     <td>
1175         <div align="right">
1176             <font size="-1">[%lan: ..]</font>
1177         </div>
1178     </td>
1179     <td/>
1180     <td>
1181         <font size="-1">language</font>
1182     </td>
1183     <td>
1184         <div align="right">
1185             <font size="-1">( 0 )</font>
1186         </div>
1187     </td>
1188     <td>
1189         <font size="-1"/>
1190     </td>
1191     <td>
1192         <font size="-1">free foot</font>
1193     </td>
1194     <td>
1195         <div align="right">
1196             <font size="-1">dB</font>
1197         </div>

```

```

1198         </div>
1199     </td>
1200     <td>
1201         <font size="-1"/>
1202     </td>
1203     <td>
1204         <font size="-1">
1205             mean intensity of syllable in decibels</font>
1206     </td>
1207 </tr>
1208 <tr>
1209     <td>
1210         <div align="right">
1211             <font size="-1">[&pho: ..]</font>
1212         </div>
1213     </td>
1214     <td/>
1215     <td>
1216         <font size="-1">phonetic phenomenon</font>
1217     </td>
1218     <td>
1219         <div align="right">
1220             <font size="-1">S</font>
1221         </div>
1222     </td>
1223     <td>
1224         <font size="-1"/>
1225     </td>
1226     <td>
1227         <font size="-1">focused stressed syllable</font>
1228     </td>
1229     <td>
1230         <div align="right">
1231             <font size="-1">dBvar</font>
1232         </div>
1233     </td>
1234     <td>
1235         <font size="-1"/>
1236     </td>
1237     <td>
1238         <font size="-1">
1239             intensity variation in percent (previous dB = 100%)</font>
1240     </td>
1241 </tr>
1242 <tr>
1243     <td>
1244         <div align="right">
1245             <font size="-1">[-..-]</font>
1246         </div>
1247     </td>
1248     <td/>
1249     <td>
1250         <font size="-1">segmental reconstruction</font>
1251     </td>
1252     <td>
1253         <div align="right">
1254             <font size="-1">W</font>
1255         </div>
1256     </td>
1257     <td>
1258         <font size="-1"/>
1259     </td>
1260     <td>
1261         <font size="-1">focused unstressed syllable</font>
1262     </td>
1263     <td>
1264         <div align="right">
1265             <font size="-1">Stress</font>
1266         </div>
1267     </td>
1268     <td>
1269     <td>
1270         <font size="-1"/>
1271     </td>
1272     <td>
1273         <font size="-1">
1274             presence or absence of stress on syllable</font>
1275     </td>
1276 </tr>

```



```

1277 <tr>
1278 <td>
1279 <div align="right">
1280 <font size="-1">[%...: laugh]/>
1281 </div>
1282 </td>
1283 <td/>
1284 <td>
1285 <font size="-1">comment in English</font>
1286 </td>
1287 <td>
1288 <div align="right">
1289 <font size="-1">s</font>
1290 </div>
1291 </td>
1292 <td>
1293 <font size="-1"/>
1294 </td>
1295 <td>
1296 <font size="-1">unfocused stressed syllable</font>
1297 </td>
1298 <td>
1299 <div align="right">
1300 <font size="-1">Focus</font>
1301 </div>
1302 </td>
1303 <td>
1304 <font size="-1"/>
1305 </td>
1306 <td>
1307 <font size="-1">
1308 presence or absence of focus on syllable</font>
1309 </td>
1310 </tr>
1311 <tr>
1312 <td>
1313 <div align="right">
1314 <font size="-1">[%qlt: ..] ( .. )</font>
1315 </div>
1316 </td>
1317 <td/>
1318 <td>
1319 <font size="-1">affected area</font>
1320 </td>
1321 <td>
1322 <div align="right">
1323 <font size="-1">w</font>
1324 </div>
1325 </td>
1326 <td>
1327 <font size="-1"/>
1328 </td>
1329 <td>
1330 <font size="-1">unfocused unstressed syllable</font>
1331 </td>
1332 <td>
1333 <div align="right">
1334 <font size="-1">Tone</font>
1335 </div>
1336 </td>
1337 <td>
1338 <font size="-1"/>
1339 </td>
1340 <td>
1341 <font size="-1">
1342 boundary, prenuclear or nuclear tone (SpToBI, etc.)</font>
1343 </td>
1344 </tr>
1345 <tr>
1346 <td>
1347 <div align="right">
1348 <font size="-1">[..]</font>
1349 </div>
1350 </td>
1351 <td>
1352 <font size="-1"/>
1353 </td>
1354 <td>

```

```

1355         <font size="-1">phonetic value of syllable</font>
1356     </td>
1357     <td>
1358         <div align="right">
1359             <font size="-1">( . . . . )</font>
1360         </div>
1361     </td>
1362     <td>
1363         <font size="-1"/>
1364     </td>
1365     <td>
1366         <font size="-1">limits of a single foot</font>
1367     </td>
1368     <td>
1369         <div align="right">
1370             <font size="-1">Rhythm</font>
1371         </div>
1372     </td>
1373     <td>
1374         <font size="-1"/>
1375     </td>
1376     <td>
1377         <font size="-1">
1378             syllable's role on rhythmical structure (v. Rhythm, left)
1379         </font>
1380     </td>
1381 </tr>
1382 </table>
1383 <br/>
1384 <div align="right">
1385     <a href="#top">
1386         <strong>
1387             <font face="Arial, Helvetica, sans-serif">Top</font>
1388         </strong>
1389         
1390     </a>
1391 </div>
1392 </body>
1393 </html>
1394 </xsl:template>
1395 <!-- End of primary template -->
1396 <!-- Display pause duration anchored to # -->
1400 <xsl:template match="Pause[@Type='xl']">
1401     <xsl:variable name="Sec"><xsl:value-of select="@Sec" /></xsl:variable>
1402     <xsl:variable name="Type"><xsl:value-of select="@Type" /></xsl:variable>
1403     <a href='javascript:alert("Sec={\$Sec}, Type={\$Type}")' title="Sec={\$Sec},
1404     Type={\$Type}">
1405     <br /><xsl:apply-templates/></a></xsl:template>
1406 <!-- Set a different image in front of intonation units other than in first
1407 position -->
1410 <xsl:template match="IU[position()>1]"><br />
1411     <xsl:choose>
1412     <xsl:when test="IU/F/S">
1413         <a href="#" onMouseOver="showmenu(event,linkset[0])"
1414         onMouseOut="delayhidemenu()">
1415         
1416         </a>
1417         <xsl:apply-templates />
1418     </xsl:when>
1419     <xsl:otherwise>
1420         
1421         <xsl:apply-templates />
1422     </xsl:otherwise>
1423     </xsl:choose>
1424 </xsl:template>
1425 <!-- Set a different image in front of utterances other than in first position in
1426 Turn -->
1429 <xsl:template match="U[position()>1]"><br />
1430     <xsl:choose>
1431     <xsl:when test="IU/F/S">
1432         <a href="#" onMouseOver="showmenu(event,linkset[0])"
1433         </xsl:when>

```

```
1434     onMouseOut="delayhidemenu()">
1435     
1436     </a>
1437     <xsl:apply-templates />
1438 </xsl:when>
1439 <xsl:otherwise>
1440     
1441     <xsl:apply-templates />
1442 </xsl:otherwise>
1443 </xsl:choose>
1444 </xsl:template>
1445
1446 </xsl:stylesheet>
1447
1448 <!-- End of XSLT stylesheet Prosody.xsl, created by Eduardo Velazquez -->
```

Anexo 11: Hoja de estilo `Prosody-set.xsl`

Líneas	Secciones
18 - 668	Plantilla principal que se aplicará a todo el elemento <code>Transcript</code>
25 - 71	Predefinición del estilo y color de fuente para los enlaces y el menú emergente
76 - 77	Código externo de JavaScript <code>dynamic.js</code> para el menú emergente y las tablas expandibles de imagen y sonido
82 - 92	Título y enlaces al resumen y a las convenciones
96 - 105	Botones para contraer o expandir todas las tablas de imagen y sonido
109 - 330	Estructura básica de los turnos de palabra y los enunciados, ordenados según la clave de identificación de cada unidad entonativa
118 - 124	Proceso para permitir que cada una de las tablas de imagen y sonido pueda abrirse de manera independiente
129 - 229	Asignación de variables a los valores de <code>F0</code> , <code>F0std</code> , <code>F0vst</code> , <code>F0var</code> , <code>dBvar</code> , <code>Durvar</code> , <code>rhythm</code> y <code>tone</code> de cada sílaba
233 - 269	Valor de cada variable según la manera de ordenarse y desplegarse en el menú emergente
280 - 284	Asignación de variables para los atributos <code>img</code> e <code>id</code> de cada unidad entonativa
274 - 425	Tablas alterables de imagen y sonido
288 - 385	Tabla de visión de conjunto con todos los atributos fonéticos de la sílaba
389 - 408	Tabla de imagen y sonido
389 - 424	Impresión de la clave de identidad de la unidad entonativa actual
431 - 660	Resumen de fenómenos lingüísticamente relevantes
663	Convenciones de los signos utilizados en la transcripción para representar los fenómenos lingüísticamente relevantes (Omitido por ser igual que en <code>Prosody.xsl</code>)
665 - 668	Fin de la plantilla principal
674 - 719	Despliegue de los datos fonéticos de cada sílaba cuando el apuntador pasa sobre ellas y generación de ventanas de alerta cuando sean apretadas
723 - 730	Despliegue de la duración de la pausa, anclada al signo #
735 - 740	Interpretación y formato de las disyunciones entre sílabas (sinalefas -0- y disyunciones intersilábicas entre palabras -1-)
745 - 759	Asignación de una imagen diferente delante de las unidades entonativas en posición distinta a la primera
764 - 779	Asignación de una imagen diferente delante de los enunciados en posición distinta a la primera
781	Fin de la hoja de estilo

```
01 <!-- Prosody-set.xsl - Version 2 -->
02 <!-- Declaration of XML document -->
03 <!-- Declaration of XSL document -->
04 <!-- Date: 25th November 2004 -->
05 <!-- Last Update: 21st June 2007 -->
06 <!-- Written by: Eduardo Velazquez -->
07
08 <xsl:stylesheet version="1.0"
09     xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
10     xmlns="http://www.w3.org/TR/xhtml1/strict">
11
12 <xsl:output
13     method="html"
14     encoding="utf-8" />
15
16 <!-- Primary template to be applied to the whole Transcript -->
17
18 <xsl:template match="Transcript">
19     <html>
20     <head>
21         <title>Transcript of Fragments:</title>
22
23     <!-- Font style and color settings for links and Pop-it Menu -->
24
25         <style type="text/css">
26             <![CDATA[
27             #popitmenu{
28                 position: absolute;
29                 background-color: white;
30                 border:1px solid black;
31                 font: normal 12px Verdana;
32                 line-height: 18px;
33                 z-index: 100;
34                 visibility: hidden;
35             }
36
37             #popitmenu a{
38                 text-decoration: none;
39                 padding-left: 6px;
40                 color: black;
41                 display: block;
42             }
43
44             #popitmenu a:hover{ /*hover background color*/
45                 background-color: #CCFF9D;
46             }
47
48             a:hover {
49                 font-weight: bold;
50                 color: #FF0000;
51                 text-decoration: none;
52             }
53
54             a:link {
55                 font-weight: normal;
56                 color: #000000;
57                 text-decoration: none;
58             }
59
60             a:visited {
61                 color: #000000;
62                 text-decoration: none;
63             }
64
65             a:active {
66                 color: #000000;
67                 text-decoration: none;
68             }
69         ]]>
70     </style>
71
72 <!-- External JavaScript code "dynamic.js" for Pop-it Menu and Expandable
73     Picture-and-Sound Tables -->
74
75     <script src="dynamic.js" type="text/javascript">
76     </script>
77 </head>
78
79
```

```

80 <!-- Title and links to Summary and Coventions -->
81
82 <body bgcolor="#FFFFFF">
83   <a name="top" />
84   <h2>
85   <p align="center">
86   <font face="Times New Roman, Times, serif">Transcript of Fragments:</font>
87   </p>
88   </h2>
89   <p align="center">
90   <font size="-1"><a href="#sum" title="Go to Summary"><em>Summary</em>
91   </a> and <a href="#conv" title="Go to Conventions"><em>conventions</em>
92   </a> at the bottom)</font></p>
93
94 <!-- Buttons to contract or to expand all Picture-and-Sound Tables -->
95
96   <div>
97     <center>
98     <p>
99     <font face="Arial, Helvetica, sans-serif">
100    <a href="javascript:sweeptoggle('contract')">Contract All</a> |
101    <a href="javascript:sweeptoggle('expand')">Expand All</a>
102    </font>
103    </p>
104    </center>
105  </div>
106
107 <!-- Basic structure of turns and utterances, ordered by intonation units' ID-->
108
109   <xsl:for-each select="Set/Sample">
110   <xsl:sort select="U/IU/@id" />
111   <xsl:variable name="n"><xsl:number /></xsl:variable>
112   <table width="100%" border="0">
113     <tr>
114       <td width="7%" align="left" valign="top">
115
116 <!-- Every Turn's Picture-and-Sound tables may be opened independently -->
117
118       <h4 onClick="expandcontent(this, 'sc{$n}')"
119       style="cursor:hand; cursor:pointer">
120         <span class="showstate"></span>
121         <font face="Courier New, Courier, mono">
122         <strong>&#160;*<xsl:value-of select='@Name' />: </strong>
123         </font>
124       </h4>
125
126 <!-- Set values for each syllable F0, F0std, F0vst,
127 F0var, dBvar, Durvar, rhythm and tone -->
128
129       <xsl:for-each select="U">
130         <xsl:variable name="F0s">
131           <xsl:for-each select="IU/F/S">
132             <xsl:choose>
133               <xsl:when test="position() != last()">
134                 <xsl:value-of select="@F0" /> -
135               </xsl:when>
136               <xsl:otherwise>
137                 <xsl:value-of select="@F0" />
138               </xsl:otherwise>
139             </xsl:choose>
140           </xsl:for-each>
141         </xsl:variable>
142         <xsl:variable name="F0stds">
143           <xsl:for-each select="IU/F/S">
144             <xsl:choose>
145               <xsl:when test="position() != last()">
146                 <xsl:value-of select="@F0std" /> -
147               </xsl:when>
148               <xsl:otherwise>
149                 <xsl:value-of select="@F0std" />
150               </xsl:otherwise>
151             </xsl:choose>
152           </xsl:for-each>
153         </xsl:variable>
154         <xsl:variable name="F0vars">
155           <xsl:for-each select="IU/F/S">
156             <xsl:choose>
157               <xsl:when test="position() != last()">

```

```

158         <xsl:value-of select="@F0var" /> -
159     </xsl:when>
160     <xsl:otherwise>
161         <xsl:value-of select="@F0var" />
162     </xsl:otherwise>
163 </xsl:choose>
164 </xsl:for-each>
165 </xsl:variable>
166 <xsl:variable name="F0vsts">
167     <xsl:for-each select="IU/F/S">
168         <xsl:choose>
169             <xsl:when test="position() != last()">
170                 <xsl:value-of select="@F0vst" /> -
171             </xsl:when>
172             <xsl:otherwise>
173                 <xsl:value-of select="@F0vst" />
174             </xsl:otherwise>
175         </xsl:choose>
176     </xsl:for-each>
177 </xsl:variable>
178 <xsl:variable name="dBvars">
179     <xsl:for-each select="IU/F/S">
180         <xsl:choose>
181             <xsl:when test="position() != last()">
182                 <xsl:value-of select="@dBvar" /> -
183             </xsl:when>
184             <xsl:otherwise>
185                 <xsl:value-of select="@dBvar" />
186             </xsl:otherwise>
187         </xsl:choose>
188     </xsl:for-each>
189 </xsl:variable>
190 <xsl:variable name="Durvars">
191     <xsl:for-each select="IU/F/S">
192         <xsl:choose>
193             <xsl:when test="position() != last()">
194                 <xsl:value-of select="@Durvar" /> -
195             </xsl:when>
196             <xsl:otherwise>
197                 <xsl:value-of select="@Durvar" />
198             </xsl:otherwise>
199         </xsl:choose>
200     </xsl:for-each>
201 </xsl:variable>
202 <xsl:variable name="Rhythm">
203     <xsl:for-each select="IU/F/S">
204         <xsl:choose>
205             <xsl:when test="position() != last()">
206                 <xsl:value-of select="parent::F/@Wt" /> -
207             </xsl:when>
208             <xsl:otherwise>
209                 <xsl:value-of select="parent::F/@Wt" />
210             </xsl:otherwise>
211         </xsl:choose>
212     </xsl:for-each>
213 </xsl:variable>
214 <xsl:variable name="Tones">
215     <xsl:for-each select="IU/F/S">
216         <xsl:choose>
217             <xsl:when test="position() != last()">
218                 <xsl:value-of select="@Tone" /> -
219             </xsl:when>
220             <xsl:otherwise>
221                 <xsl:value-of select="@Tone" />
222             </xsl:otherwise>
223         </xsl:choose>
224     </xsl:for-each>
225 </xsl:variable>
226 <xsl:variable name="ID">
227     <xsl:value-of select="IU/@img" />
228 </xsl:variable>
229
230 <!-- Variable value as ordered and displayed on a Pop-It Menu -->
231
232
233     <script type="text/javascript">
234     linkset[<xsl:value-of select='\$n' />]=
235     '<a href="javascript:alert(&quot;{'$F0s}&quot;)">
236     F0 sequence in Hz</a>'

```

```

237 linkset[<xsl:value-of select='$n' />]+=
238 ' <a href="javascript:alert(&quot;{$F0stds}&quot;)">
239 Standardized F0s</a>'
240 linkset[<xsl:value-of select='$n' />]+=
241 ' <a href="javascript:alert(&quot;{$F0vars}&quot;)">
242 F0 variation in %</a>'
243 linkset[<xsl:value-of select='$n' />]+=
244 ' <a href="javascript:alert(&quot;{$F0vsts}&quot;)">
245 F0 variation in st</a>'
246 linkset[<xsl:value-of select='$n' />]+=
247 ' <a href="javascript:alert(&quot;{$dBvars}&quot;)">
248 dB variation in %</a>'
249 linkset[<xsl:value-of select='$n' />]+=
250 ' <a href="javascript:alert(&quot;{$Durvars}&quot;)">
251 Syll duration in %</a>'
252 linkset[<xsl:value-of select='$n' />]+=
253 ' <hr />' //Optional Separator
254 linkset[<xsl:value-of select='$n' />]+=
255 ' <a href="javascript:alert(&quot;{$Rhythm}&quot;)">
256 Rhythm structure</a>'
257 linkset[<xsl:value-of select='$n' />]+=
258 ' <a href="javascript:alert(&quot;{$Tones}&quot;)">
259 Tone structure</a>'
260 </script>
261 </xsl:for-each>
262 </td>
263 <td width="73%" valign="top">
264 <a href="" onMouseOver="showmenu(event,linkset[{$n}])"
265 onMouseOut="delayhidemenu()">
266 </a>
267 <xsl:apply-templates />
268 <br/>
269
270
271
272 <!-- Set values of every intonation unit's attribute 'img' and 'id' -->
273
274 <!-- Switchable tables -->
275
276 <div id="sc{$n}" class="switchcontent">
277
278 <xsl:for-each select="U">
279 <xsl:if test="IU/@img">
280 <xsl:variable name="img"><xsl:value-of select='IU/@img' />
281 </xsl:variable>
282 <xsl:variable name="id"><xsl:value-of select='IU/@id' />
283 </xsl:variable>
284 <xsl:variable name="i"><xsl:number /></xsl:variable>
285
286 <!-- Overview table with all syllable's phonetic attributes -->
287
288 <table border="1">
289 <xsl:if test="IU/F/S">
290 <tr>
291 <th><font size="1">Orth</font></th>
292 <th><font size="1">Phon</font></th>
293 <th><font size="1">Beg</font></th>
294 <th><font size="1">End</font></th>
295 <th><font size="1">Dur</font></th>
296 <th><font size="1">Durvar</font></th>
297 <th><font size="1">F0min</font></th>
298 <th><font size="1">F0max</font></th>
299 <th><font size="1">F0dBmax</font></th>
300 <th><font size="1">F0</font></th>
301 <th><font size="1">F0var</font></th>
302 <th><font size="1">F0vst</font></th>
303 <th><font size="1">F0std</font></th>
304 <th><font size="1">dBmin</font></th>
305 <th><font size="1">dBmax</font></th>
306 <th><font size="1">tdBmax</font></th>
307 <th><font size="1">dB</font></th>
308 <th><font size="1">dBvar</font></th>
309 <th><font size="1">Stress</font></th>
310 <th><font size="1">Focus</font></th>
311 <th><font size="1">Tone</font></th>
312 <th><font size="1">Rhythm</font></th>
313 </tr>
314 <xsl:for-each select="IU/F/S">
315 <tr>

```



```

316 <td><font size="1">
317 <xsl:value-of select='@Orth' />
318 </font></td>
319 <td><font size="1">
320 <xsl:value-of select='@Phon' />
321 </font></td>
322 <td><font size="1">
323 <em><xsl:value-of select='@Beg' /> s</em>
324 </font></td>
325 <td><font size="1">
326 <em><xsl:value-of select='@End' /> s</em>
327 </font></td>
328 <td><font size="1">
329 <xsl:value-of select='@Dur' /> s
330 </font></td>
331 <td><font size="1">
332 <xsl:value-of select='@Durvar' />%
333 </font></td>
334 <td><font size="1">
335 <em><xsl:value-of select='@F0min' /> Hz</em>
336 </font></td>
337 <td><font size="1">
338 <em><xsl:value-of select='@F0max' /> Hz</em>
339 </font></td>
340 <td><font size="1">
341 <em><xsl:value-of select='@F0dBmax' /> Hz</em>
342 </font></td>
343 <td><font size="1">
344 <xsl:value-of select='@F0' /> Hz
345 </font></td>
346 <td><font size="1">
347 <xsl:value-of select='@F0var' />%
348 </font></td>
349 <td><font size="1">
350 <xsl:value-of select='@F0vst' /> st
351 </font></td>
352 <td><font size="1">
353 <xsl:value-of select='@F0std' /> Hz
354 </font></td>
355 <td><font size="1">
356 <em><xsl:value-of select='@dBmin' /> dB</em>
357 </font></td>
358 <td><font size="1">
359 <em><xsl:value-of select='@dBmax' /> dB</em>
360 </font></td>
361 <td><font size="1">
362 <em><xsl:value-of select='@tdBmax' /> s</em>
363 </font></td>
364 <td><font size="1">
365 <xsl:value-of select='@dB' /> dB
366 </font></td>
367 <td><font size="1">
368 <xsl:value-of select='@dBvar' />%
369 </font></td>
370 <td><font size="1">
371 <xsl:value-of select='@Stress' />
372 </font></td>
373 <td><font size="1">
374 <xsl:value-of select='@Focus' />
375 </font></td>
376 <td><font size="1">
377 <xsl:value-of select='@Tone' />
378 </font></td>
379 <td><font size="1">
380 <xsl:value-of select='../@Wt' />
381 </font></td>
382 </tr>
383 </xsl:for-each>
384 </xsl:if>
385
386 <!-- Picture-and-Sound Table -->
387
388 <tr>
389 <td colspan="22">
390 <center>
391 
392 <br/><br/>
393 <form name="button" method="get"
394

```

```

395         action="HTML/wav/{$img}.wav" accept="audio/wav"
396         target="_self">
397         <a href="javascript:audio{$img}()">
398         <input name="Play" type="submit" id="Play"
399         value="    Play audio fragment    "/></a>
400     </form>
401     </center>
402 </td>
403 </tr>
404 </table>
405 </xsl:if>
406 </xsl:for-each>
407 </div>
408 </td>
409
410 <!-- Print current intonation unit's ID -->
411
412     <td width="20%" align="right" valign="top">
413         <xsl:for-each select=".">
414         <xsl:for-each select="U">
415         <xsl:if test="IU/@id">
416         <h5>
417         <font face="Courier New, Courier, mono" size="-2">
418             <xsl:value-of select='IU/@id' /></font>
419         </h5>
420         </xsl:if>
421         </xsl:for-each>
422         </xsl:for-each>
423     </td>
424 </tr>
425 </table>
426
427 </xsl:for-each>
428
429 <!-- Summary of linguistic relevant phenomena -->
430
431 <hr/>
432 <p>
433     <h3 align="center">
434         <a name="sum">Summary</a>
435     </h3>
436 </p>
437 <table width="80%" border="0" align="center">
438     <tr>
439         <td width="40%">
440             <font size="-1" face="Arial, Helvetica, sans-serif">
441                 <strong>Samples:</strong>
442             </font>
443         </td>
444         <td width="12%">
445             <div align="right">
446                 <xsl:value-of select="count(..//Sample)"/>
447             </div>
448         </td>
449         <td width="16%"/>
450         <td width="40%">
451             <font size="-1" face="Arial, Helvetica, sans-serif">
452                 <strong>Intonation units:</strong>
453             </font>
454         </td>
455         <td width="12%">
456             <div align="right">
457                 <xsl:value-of select="count(..//IU)"/>
458             </div>
459         </td>
460     </tr>
461 </table>
462
463     <td>
464         <font size="-1" face="Arial, Helvetica, sans-serif">
465             <strong>Utterances:</strong>
466         </font>
467     </td>
468     <td>
469         <div align="right">
470             <xsl:value-of select="count(..//U)"/>
471         </div>
472 </td>
473 </td>

```

```

474         <td>
475             <font size="-1" face="Arial, Helvetica, sans-serif">
476                 <strong>Syllables:</strong>
477             </font>
478         </td>
479         <td>
480             <div align="right">
481                 <xsl:value-of select="count(..//S)"/>
482             </div>
483         </td>
484     </tr>
485     <tr>
486         <td>
487             <font size="-1" face="Arial, Helvetica, sans-serif">
488                 <strong>Assertive utterances:</strong>
489             </font>
490         </td>
491         <td>
492             <div align="right">
493                 <xsl:value-of select="count(..//Border[@Type='nt'])"/>
494             </div>
495         </td>
496     </td/>
497     <td>
498         <font size="-1" face="Arial, Helvetica, sans-serif">
499             <strong>Elisions/Liaisons:</strong>
500         </font>
501     </td>
502     <td>
503         <div align="right">
504             <xsl:value-of select="count(..//S/Break[@Type='0'])"/>
505         </div>
506     </td>
507 </tr>
508 <tr>
509     <td>
510         <font size="-1" face="Arial, Helvetica, sans-serif">
511             <strong>Interrogative utterances:</strong>
512         </font>
513     </td>
514     <td>
515         <div align="right">
516             <xsl:value-of select="count(..//Border[@Type='qst'])"/>
517         </div>
518     </td>
519 </td/>
520 <td>
521     <font size="-1" face="Arial, Helvetica, sans-serif">
522         <strong>Syllables with more than one Mora:</strong>
523     </font>
524 </td>
525 <td>
526     <div align="right">
527         <xsl:value-of select="count(..//Mora)"/>
528     </div>
529 </td>
530 </tr>
531 <tr>
532     <td>
533         <font size="-1" face="Arial, Helvetica, sans-serif">
534             <strong>Suspended utterances:</strong>
535         </font>
536     </td>
537     <td>
538         <div align="right">
539             <xsl:value-of select="count(..//Border[@Type='ssp'])"/>
540         </div>
541     </td>
542 </td/>
543 <td>
544     <font size="-1" face="Arial, Helvetica, sans-serif">
545         <strong>Unintelligible productions:</strong>
546     </font>
547 </td>
548 <td>
549     <div align="right">
550         <xsl:value-of select="count(..//Unintelligible)"/>
551     </div>

```

```

552         </td>
553     </tr>
554     <tr>
555         <td>
556             <font size="-1" face="Arial, Helvetica, sans-serif">
557                 <strong>Self-interrupted utterances:</strong>
558             </font>
559         </td>
560         <td>
561             <div align="right">
562                 <xsl:value-of select="count(..//Border[@Type='slf'])"/>
563             </div>
564         </td>
565     </tr>
566     <tr>
567         <td>
568             <font size="-1" face="Arial, Helvetica, sans-serif">
569                 <strong>Repetitions:</strong>
570             </font>
571         </td>
572         <td>
573             <div align="right">
574                 <xsl:value-of select="count(..//Restart[@Type='rep'])"/>
575             </div>
576         </td>
577     </tr>
578     <tr>
579         <td>
580             <font size="-1" face="Arial, Helvetica, sans-serif">
581                 <strong>Externally interrupted utterances:</strong>
582             </font>
583         </td>
584         <td>
585             <div align="right">
586                 <xsl:value-of select="count(..//Border[@Type='ext'])"/>
587             </div>
588         </td>
589     </tr>
590     <tr>
591         <td>
592             <font size="-1" face="Arial, Helvetica, sans-serif">
593                 <strong>Partial reformulations:</strong>
594             </font>
595         </td>
596         <td>
597             <div align="right">
598                 <xsl:value-of select="count(..//Restart[@Type='prt'])"/>
599             </div>
600     </tr>
601     <tr>
602         <td>
603             <font size="-1" face="Arial, Helvetica, sans-serif">
604                 <strong>Overlaps:</strong>
605             </font>
606         </td>
607         <td>
608             <div align="right">
609                 <xsl:value-of select="count(..//Overlap[@Type='pas'])"/>
610             </div>
611         </td>
612     </tr>
613     <tr>
614         <td>
615             <font size="-1" face="Arial, Helvetica, sans-serif">
616                 <strong>Total reformulations:</strong>
617             </font>
618         </td>
619         <td>
620             <div align="right">
621                 <xsl:value-of select="count(..//Restart[@Type='tot'])"/>
622             </div>
623     </tr>
624     <tr>
625         <td>
626             <font size="-1" face="Arial, Helvetica, sans-serif">
627                 <strong>Pauses (longer than 1 second):</strong>
628             </font>
629         </td>
630         <td>

```

```

631         <div align="right">
632             <xsl:value-of select="count(..//Pause[@Type='x1'])"/></div>
633         </td>
634     </td>
635     <td>
636         <font size="-1" face="Arial, Helvetica, sans-serif">
637             <strong>Total time (in seconds):</strong>
638         </font>
639     </td>
640     <td>
641         <div align="right">
642             <xsl:variable name="length">
643                 <xsl:value-of select="sum(..//S/@Dur)"/>
644             </xsl:variable>
645             <xsl:value-of select="format-number($length, '#.##')"/>
646         </div>
647     </td>
648 </tr>
649 </table>
650 <br/>
651 <div align="right">
652     <a href="#top">
653         <strong>
654             <font face="Arial, Helvetica, sans-serif">Top</font>
655         </strong>
656
657         
658     </a>
659 </div>
660 <hr/>
661
662 <!-- Transcript's sign conventions for linguistic relevant phenomena -->
663 <!-- Same as in Prosody.xsl -->
664
665 </body>
666 </html>
667 </xsl:template>
668
669 <!-- End of primary template -->
670
671 <!-- Display each syllable's phonetic data, on mouse over, and alert windows -->
672
673 <xsl:template match="S">
674     <xsl:variable name="Phon"><xsl:value-of select="@Phon" />
675     </xsl:variable>
676     <xsl:variable name="Beg"><xsl:value-of select="@Beg" />
677     </xsl:variable>
678     <xsl:variable name="End"><xsl:value-of select="@End" />
679     </xsl:variable>
680     <xsl:variable name="Dur"><xsl:value-of select="@Dur" />
681     </xsl:variable>
682     <xsl:variable name="Durvar"><xsl:value-of select="@Durvar" />
683     </xsl:variable>
684     <xsl:variable name="F0"><xsl:value-of select="@F0" />
685     </xsl:variable>
686     <xsl:variable name="F0dBmax"><xsl:value-of select="@F0dBmax" />
687     </xsl:variable>
688     <xsl:variable name="F0var"><xsl:value-of select="@F0var" />
689     </xsl:variable>
690     <xsl:variable name="F0vst"><xsl:value-of select="@F0vst" />
691     </xsl:variable>
692     <xsl:variable name="F0std"><xsl:value-of select="@F0std" />
693     </xsl:variable>
694     <xsl:variable name="dB"><xsl:value-of select="@dB" />
695     </xsl:variable>
696     <xsl:variable name="dBvar"><xsl:value-of select="@dBvar" />
697     </xsl:variable>
698     <xsl:variable name="Stress"><xsl:value-of select="@Stress" />
699     </xsl:variable>
700     <xsl:variable name="Focus"><xsl:value-of select="@Focus" />
701     </xsl:variable>
702     <xsl:variable name="Tone"><xsl:value-of select="@Tone" />
703     </xsl:variable>
704     <xsl:variable name="Rhythm"><xsl:value-of select="..//Wt" />
705     </xsl:variable>
706     <xsl:variable name="img"><xsl:value-of select="@img" />
707     </xsl:variable>
708     <xsl:variable name="id"><xsl:value-of select="@id" />
709

```

```

710 </xsl:variable>
711 <a href='javascript:alert("[{$Phon}], Beg={$Beg}s,
712 End={$End}s, Dur={$Dur}s, Durvar={$Durvar}%"
713 + "\n" + "F0={$F0}Hz, F0dBmax={$F0dBmax}Hz,
714 F0var={$F0var}%, F0vst={$F0vst}st, F0std={$F0std}Hz"
715 + "\n" + "dB={$dB}dB, dBvar={$dBvar}%, Stress={$Stress},
716 Focus={$Focus}, Tone={$Tone}")'
717 title="F0={$F0}Hz, dB={$dB}dB, Dur={$Dur}s,
718 Tone={$Tone}, Rhythm={$Rhythm}"><xsl:apply-templates/></a>
719 </xsl:template>
720
721 <!-- Display pause duration anchored to # -->
722
723 <xsl:template match="Pause[@Type='xl']">
724 <xsl:variable name="Sec"><xsl:value-of select='@Sec' />
725 </xsl:variable>
726 <xsl:variable name="Type"><xsl:value-of select='@Type' />
727 </xsl:variable>
728 <a href='javascript:alert("Sec={$Sec}, Type={$Type}")'
729 title="Sec={$Sec}, Type={$Type}"><br /><xsl:apply-templates/></a>
730 </xsl:template>
731
732 <!-- Interpret and manage Breaks (liaisons -0-,
733 and intersyllabic word breaks -1-) -->
734
735 <xsl:template match="F/Break[@Type='0']">
736 <font size="-3">&#183;</font><xsl:apply-templates />
737 </xsl:template>
738
739 <xsl:template match="F/Break[@Type='1']">
740 <xsl:apply-templates />&#160;</xsl:template>
741
742 <!-- Set a different image in front of intonation units
743 other than in first position -->
744
745 <xsl:template match="IU[position()>1]"><br />
746 <xsl:choose>
747 <xsl:when test="IU/F/S">
748 <a href="#" onMouseOver="showmenu(event,linkset[0])"
749 onMouseOut="delayhidemenu()">
750 
751 </a>
752 <xsl:apply-templates />
753 </xsl:when>
754 <xsl:otherwise>
755 
756 <xsl:apply-templates />
757 </xsl:otherwise>
758 </xsl:choose>
759 </xsl:template>
760
761 <!-- Set a different image in front of utterances
762 other than in first position in Turn -->
763
764 <xsl:template match="U[position()>1]"><br />
765 <xsl:choose>
766 <xsl:when test="IU/F/S">
767 <a href="#" onMouseOver="showmenu(event,linkset[0])"
768 onMouseOut="delayhidemenu()">
769 
770 </a>
771 <xsl:apply-templates />
772 </xsl:when>
773 <xsl:otherwise>
774 
775 <xsl:apply-templates />
776 </xsl:otherwise>
777 </xsl:choose>
778 </xsl:template>
779 </xsl:template>
780
781 </xsl:stylesheet>
782
783 <!-- End of XSLT stylesheet Prosody-set.xsl, created by Eduardo Velazquez -->

```

Anexo 12: Hoja de estilo Prosody-set-table.xsl

Líneas	Secciones
18 - 126	Plantilla principal que se aplicará a todo el elemento <code>Transcript</code>
27 - 122	Estructura básica de las tablas que contienen todos los datos relevantes
128	Fin de la hoja de estilo

```

01 <!-- Prosody-set-table.xsl - Version 2 -->
02 <!-- Declaration of XML document -->
03 <!-- Declaration of XSL document -->
04 <!-- Date: 14th June 2007 -->
05 <!-- Last Update: 19th June 2007 -->
06 <!-- Written by: Eduardo Velazquez -->
07
08 <xsl:stylesheet version="1.0"
09     xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
10     xmlns="http://www.w3.org/TR/xhtml1/strict">
11
12 <xsl:output
13     method="html"
14     encoding="utf-8" />
15
16 <!-- Primary template to be applied to the whole Transcript -->
17
18 <xsl:template match="Transcript">
19     <html>
20     <head>
21         <title>Analysis tables</title>
22     </head>
23     <body bgcolor="#FFFFFF">
24
25 <!-- Basic structure of tables containing all relevant data -->
26
27 <xsl:for-each select="Set/Sample">
28     <xsl:sort select="U/IU/@id" />
29     <xsl:variable name="n"><xsl:number /></xsl:variable>
30
31 <xsl:for-each select="U">
32 <table width="100%" border="1">
33     <tr valign="top">
34         <td bordercolor="#FFFFFF"> <div align="center">
35             <table width="100%" border="0">
36                 <tr>
37                     <th><div align="center"><font size="1">ID</font>
38                         </div></th>
39                     <th><div align="center"><font size="1">Name</font>
40                         </div></th>
41                 </tr>
42                 <tr>
43                     <td><div align="center">
44                         <font size="1"><xsl:value-of select='IU/@id' /></font>
45                         </div></td>
46                     <td><div align="center">
47                         <font size="1">
48                             <xsl:value-of select='parent::Sample/@Name' /></font>
49                         </div></td>
50                 </tr>
51             </table>
52         </div></td>
53
54         <td bordercolor="#FFFFFF"> <div align="center">
55             <table width="100%" border="0">
56                 <tr>
57                     <th><div align="center"><font size="1">Orth</font>
58                         </div></th>
59                     <th><div align="center"><font size="1">F0 (Hz)</font>
60                         </div></th>
61                     <th><div align="center"><font size="1">F0var</font>
62                         </div></th>
63                     <th><div align="center"><font size="1">F0std</font>

```

```

64         </div></th>
65         <th><div align="center"><font size="1">Tones</font>
66         </div></th>
67         <th><div align="center"><font size="1">Rhythm</font>
68         </div></th>
69     </tr>
70 </xsl:for-each select="IU/F/S">
71     <tr>
72         <td><div align="center">
73             <font size="1"><xsl:value-of select="@Orth" /></font>
74         </div></td>
75         <td><div align="center">
76             <font size="1"><xsl:value-of select="@F0" /></font>
77         </div></td>
78         <td><div align="center">
79             <font size="1"><xsl:value-of select="@F0var" />%</font>
80         </div></td>
81         <td><div align="center">
82             <font size="1"><xsl:value-of select="@F0std" /></font>
83         </div></td>
84         <td><div align="center">
85             <font size="1"><xsl:value-of select="@Tone" /></font>
86         </div></td>
87         <td><div align="center">
88             <font size="1"><xsl:value-of select="../@Wt" /></font>
89         </div></td>
90     </tr>
91 </xsl:for-each>
92 </table>
93 </div></td>
94 <td bordercolor="#FFFFFF" <div align="center">
95     <table width="100%" border="0">
96         <tr>
97             <th><div align="center"><font size="1">Quest</font>
98             </div></th>
99             <th><div align="center"><font size="1">Emph</font>
100            </div></th>
101            <th><div align="center"><font size="1">Susp</font>
102            </div></th>
103        </tr>
104        <tr>
105            <td><div align="center">
106                <font size="1"><xsl:value-of select="@Quest" /></font>
107            </div></td>
108            <td><div align="center">
109                <font size="1"><xsl:value-of select="@Emph" /></font>
110            </div></td>
111            <td><div align="center">
112                <font size="1"><xsl:value-of select="@Susp" /></font>
113            </div></td>
114        </tr>
115    </table>
116 </div></td>
117 </tr>
118 </table>
119 </xsl:for-each>
120 </xsl:for-each>
121 </body>
122 </html>
123 </xsl:template>
124 </xsl:stylesheet>
125 <!-- End of XSLT stylesheet Prosody.xsl, created by Eduardo Velazquez -->

```


Anexo 13: Subprograma en JavaScript dynamic.js

```

01 //DYNAMIC.JS - Version 1.0 - Eduardo Velázquez
02
03 /*****
04 * Pop-it menu- © Dynamic Drive (www.dynamicdrive.com)
05 * This notice MUST stay intact for legal use
06 * Visit http://www.dynamicdrive.com/ for full source code
07 *****/
08 * Adapted by Eduardo Velázquez, 19th May, 2006
09 *****/
10
11 var defaultMenuWidth="150px" //set default menu width.
12
13 var linkset=new Array()
14
15 var ie5=document.all && !window.opera
16 var ns6=document.getElementById
17
18 if (ie5||ns6)
19 document.write('<div id="popitmenu" onmouseover="clearhidemenu()"
20               onmouseout="dynamichide(event)"></div>')
21
22 function iecompattest(){
23 return (document.compatMode && document.compatMode.indexOf("CSS")!=-1)?
24 document.documentElement : document.body
25 }
26
27 function showmenu(e, which, optWidth){
28 if (!document.all&&!document.getElementById)
29 return
30 clearhidemenu()
31 menuobj=ie5? document.all.popitmenu : document.getElementById("popitmenu")
32 menuobj.innerHTML=which
33 menuobj.style.width=(typeof optWidth!="undefined"? optWidth : defaultMenuWidth
34 menuobj.contentwidth=menuobj.offsetWidth
35 menuobj.contentheight=menuobj.offsetHeight
36 eventX=ie5? event.clientX : e.clientX
37 eventY=ie5? event.clientY : e.clientY
38 //Find out how close the mouse is to the corner of the window
39 var righedge=ie5? iecompattest().clientWidth-eventX : window.innerWidth-eventX
40 var bottomedge=ie5? iecompattest().clientHeight-eventY : window.innerHeight-
41 eventY
42 //if the horizontal distance isn't enough to accomodate the width of the context
43 menu
44 if (righedge<menuobj.contentwidth)
45 //move the horizontal position of the menu to the left by it's width
46 menuobj.style.left=ie5? iecompattest().scrollLeft+eventX-
47 menuobj.contentwidth+"px" :
48 window.pageXOffset+eventX-menuobj.contentwidth+"px"
49 else
50 //position the horizontal position of the menu where the mouse was clicked
51 menuobj.style.left=ie5? iecompattest().scrollLeft+eventX+"px" :
52 window.pageXOffset+eventX+"px"
53 //same concept with the vertical position
54 if (bottomedge<menuobj.contentheight)
55 menuobj.style.top=ie5? iecompattest().scrollTop+eventY-menuobj.contentheight+"px"
56 :
57 window.pageYOffset+eventY-menuobj.contentheight+"px"
58 else
59 menuobj.style.top=ie5? iecompattest().scrollTop+event.clientY+"px" :
60 window.pageYOffset+eventY+"px"
61 menuobj.style.visibility="visible"
62 return false
63 }
64
65 function contains_ns6(a, b) {
66 //Determines if 1 element in contained in another- by Brainjar.com
67 while (b.parentNode)
68 if ((b = b.parentNode) == a)
69 return true;
70 return false;
71 }
72
73
74 function hidemenu(){

```

```

75 if (window.menuobj)
76 menuobj.style.visibility="hidden"
77 }
78
79 function dynamichide(e){
80 if (ie5&&!menuobj.contains(e.toElement))
81 hidemenu()
82 else if (ns6&&e.currentTarget!= e.relatedTarget&& !contains_ns6(e.currentTarget,
83 e.relatedTarget))
84 hidemenu()
85 }
86
87 function delayhidemenu(){
88 delayhide=setTimeout("hidemenu()",500)
89 }
90
91 function clearhidemenu(){
92 if (window.delayhide)
93 clearTimeout(delayhide)
94 }
95
96 if (ie5|ns6)
97 document.onclick=hidemenu
98
99 /*****
100 * Switch Content script- © Dynamic Drive (www.dynamicdrive.com)
101 * This notice must stay intact for legal use. Last updated April 2nd, 2005.
102 * Visit http://www.dynamicdrive.com/ for full source code
103 *****/
104 * Adapted by Eduardo Velázquez, 19th May, 2006
105 *****/
106
107 var enablepersist="on"
108 //Enable saving state of content structure using session cookies? (on/off)
109 var collapseprevious="no"
110 //Collapse previously open content when opening present? (yes/no)
111
112 var contractsymbol=''
113 //HTML for contract symbol. For image, use: 
114 var expandsymbol=''
115 //HTML for expand symbol.
116
117 if (document.getElementById){
118 document.write('<style type="text/css">')
119 document.write('.switchcontent{display:none;}')
120 document.write('</style>')
121 }
122
123 function getElementbyClass(rootobj, classname){
124 var temparray=new Array()
125 var inc=0
126 var rootlength=rootobj.length
127 for (i=0; i<rootlength; i++){
128 if (rootobj[i].className==classname)
129 temparray[inc++]=rootobj[i]
130 }
131 return temparray
132 }
133
134 function sweeptoggle(ec){
135 var thestate=(ec=="expand"? "block" : "none"
136 var inc=0
137 while (ccollect[inc]){
138 ccollect[inc].style.display=thestate
139 inc++
140 }
141 revivestatus()
142 }
143
144 function contractcontent(omit){
145 var inc=0
146 while (ccollect[inc]){
147 if (ccollect[inc].id!=omit)
148 ccollect[inc].style.display="none"
149 inc++
150 }
151 }
152

```

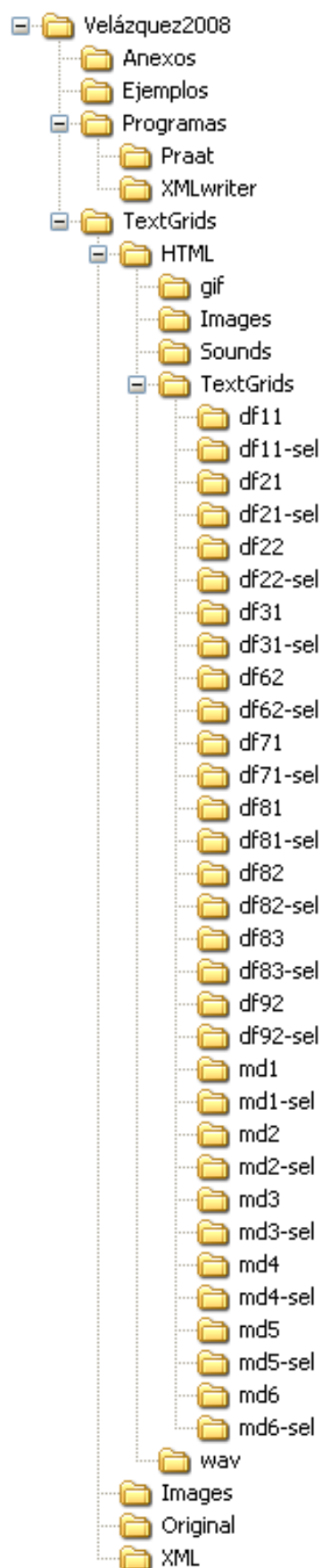
```

153 function expandcontent(curobj, cid){
154 var spantags=curobj.getElementsByTagName("SPAN")
155 var showstateobj=getElementbyClass(spantags, "showstate")
156 if (ccollect.length>0){
157 if (collapseprevious=="yes")
158 contractcontent(cid)
159 document.getElementById(cid).style.display=
160 (document.getElementById(cid).style.display!="block"?
161 "block" : "none")
162 if (showstateobj.length>0){ //if "showstate" span exists in header
163 if (collapseprevious=="no")
164 showstateobj[0].innerHTML=(document.getElementById(cid).style.display=="block"?
165 contractsymbol : expandsymbol
166 else
167 revivestatus()
168 }
169 }
170 }
171 }
172
173 function revivecontent(){
174 contractcontent("omitnothing")
175 selectedItem=getSelectedItem()
176 selectedComponents=selectedItem.split("|")
177 for (i=0; i<selectedComponents.length-1; i++)
178 document.getElementById(selectedComponents[i]).style.display="block"
179 }
180
181 function revivestatus(){
182 var inc=0
183 while (statecollect[inc]){
184 if (ccollect[inc].style.display=="block")
185 statecollect[inc].innerHTML=contractsymbol
186 else
187 statecollect[inc].innerHTML=expandsymbol
188 inc++
189 }
190 }
191
192 function get_cookie(Name) {
193 var search = Name + "="
194 var returnvalue = "";
195 if (document.cookie.length > 0) {
196 offset = document.cookie.indexOf(search)
197 if (offset != -1) {
198 offset += search.length
199 end = document.cookie.indexOf(";", offset);
200 if (end == -1) end = document.cookie.length;
201 returnvalue=unescape(document.cookie.substring(offset, end))
202 }
203 }
204 return returnvalue;
205 }
206
207 function getSelectedItem(){
208 if (get_cookie(window.location.pathname) != ""){
209 selectedItem=get_cookie(window.location.pathname)
210 return selectedItem
211 }
212 else
213 return ""
214 }
215
216 function saveswitchstate(){
217 var inc=0, selectedItem=""
218 while (ccollect[inc]){
219 if (ccollect[inc].style.display=="block")
220 selectedItem+=ccollect[inc].id+"|"
221 inc++
222 }
223
224 document.cookie=window.location.pathname+"="+selectedItem
225 }
226
227 function do_onload(){
228 uniqueidn=window.location.pathname+"firsttimeload"
229 var alltags=document.all? document.all : document.getElementsByTagName("*")
230 ccollect=getElementbyClass(alltags, "switchcontent")
231 statecollect=getElementbyClass(alltags, "showstate")

```

```
232 if (enablepersist=="on" && ccollect.length>0){
233 document.cookie=(get_cookie(uniqueidn)=="")? uniqueidn+"=1" : uniqueidn+"=0"
234 firsttimeload=(get_cookie(uniqueidn)==1)? 1 : 0 //check if this is 1st page load
235 if (!firsttimeload)
236 revivecontent()
237 }
238 if (ccollect.length>0 && statecollect.length>0)
239 revivestatus()
240 }
241
242 if (window.addEventListener)
243 window.addEventListener("load", do_onload, false)
244 else if (window.attachEvent)
245 window.attachEvent("onload", do_onload)
246 else if (document.getElementById)
247 window.onload=do_onload
248
249 if (enablepersist=="on" && document.getElementById)
250 window.onunload=saveswitchstate
251
252 /*****
```

Anexo 14: Contenido del DVD



→ Carpetas principales, hoja de cálculo Prosodia.xls

→ Documentos incluidos en los Anexos

→ Textos, TextGrids y archivos de audios de los Ejemplos

→ Archivos necesarios para instalar Praat o una versión de prueba de XMLwriter

→ Documentos de HTML de las transcripciones

→ Imágenes de las transcripciones detalladas

→ Imágenes de las transcripciones básicas

→ Archivos de audio de las transcripciones básicas

→ Archivos de TextGrid y sus correspondientes archivos de audio por cada grabación. Las carpetas con terminación '-sel' contienen los pares de archivos seleccionados para su transcripción y análisis de manera detallada en Praat.

→ Archivos de audio de las transcripciones detalladas

→ Imágenes para el formato y navegación en página *web*

→ Grabaciones originales y TextGrids primarios

→ Documentos de XML, DTD y XSLT

Anexo 15: Deutsche Zusammenfassung

Der Ausgangspunkt für die vorliegende Arbeit ist die Tatsache, dass Gesprächsteilnehmer in der Regel nicht nur den Inhalt der Sätze ihres Gesprächspartners interpretieren, sondern - aufbauend auf ihren persönlichen Erfahrungen - auch Einstellungen und Gefühle in der Sprache der anderen Teilnehmer intuitiv erkennen. Außerdem wird ein Zuhörer immer versuchen die Unterschiede, welche er auf lautlicher (oder prosodischer), morphologischer, syntaktischer oder semantischer Ebene im Vergleich zu seiner eigenen Sprechweise der jeweiligen Sprache feststellt, zu identifizieren, in Kategorien zu klassifizieren und mit seiner persönlichen Sprache und anderen bekannten Systemen zu vergleichen. So wird schließlich jeder Gesprächspartner durch die zahlreichen Hinweise, die er im Laufe des Gesprächs erhält, seine Hypothese über die soziale und geographische Herkunft der anderen Teilnehmer bestätigen. Dieser - meist unbewußte - Identifikationsprozess des Soziolekts oder Regiolekts des Gesprächspartners braucht im allgemeinen nur einige Augenblicke und hängt von der Vertrautheit des Zuhörers mit dem Sprachsystem des Sprechers ab.

Wenn Sprecher jedoch die typische Sprechweise einer bestimmten Region oder sozialen Gruppe (d.h. eine dialektale oder soziolektale Varietät, die sich von der Ihrigen unterscheidet) nachahmen, dann passen sie die charakteristischen Merkmale, welche sie mit dieser Varietät verbinden, in ihrer eigenen Sprechweise an. Das ist jedoch das beste Beispiel dafür, dass neben den Unterschieden jeder Varietät auf segmentaler Ebene auch prosodische Merkmale existieren, die erkannt, abstrahiert und sogar reproduziert werden. Folglich müssen diese charakteristischen Merkmale der einzelnen Varietäten akustische Entsprechungen haben, die leicht zu beobachten und zu quantifizieren sind.

Die Fortschritte in der Technik ermöglichen heute jedem Forscher den Zugang zu einer Vielzahl von kostenlosen Programmen zur akustischen Analyse, die hinsichtlich ihrer Qualität mit teuren Programmen gut mithalten. Ein Beispiel für solch ein kostenloses Programm mit hoher Qualität ist Praat, was von einer großen Zahl an Forschern auf der ganzen Welt benutzt wird. Dadurch werden die meisten Funktionen des Programms umgehend angewendet und so überprüft, was garantiert, dass das Programm ständig verbessert und aktualisiert wird.

Eine wichtige Eigenschaft der vorliegenden Arbeit ist, dass die Analysen anhand von spontaner Sprache durchgeführt werden. Das bedeutet zunächst, dass auf viele Vorteile der Laborsprache verzichtet werden muss. So ist beispielsweise die Qualität der Aufnahmen nicht immer wie gewünscht; die Sprecher sagen nur selten genau die Art von Sätzen und mit genau der Intensität, Geschwindigkeit und Intonation oder dem Rhythmus, wie man es erhofft, und zudem sind die Situationen, der Kontext und der Lärm in der Umgebung selten für die

Forschung und seine nachfolgende Veröffentlichung geeignet. Dies stellt jedoch auch eine besondere Herausforderung dar. Denn wenn es möglich ist, eine Methode zu entwickeln, mit der sogar solche spontanen Aufnahmen zu befriedigenden Ergebnissen in der Analyse führen, dann ist das der beste Beweis dafür, dass die Methode funktioniert, und dass sie mit jeder Art von Aufnahmen verwendet werden kann.

Konkreter gesprochen, wird die vorliegende Arbeit entlang der folgenden fünf Hypothesen entwickelt:

- H₁: Die prosodischen Faktoren, welche das Gespräch beeinflussen, sind die Intonation, der Rhythmus und die Geschwindigkeit sowie die Intensität und die Stimmqualität.
- H₂: Die theoretischen Modelle, die sich für die elektronische Untersuchung und Darstellung der Intonation des Spanischen am besten eignen, sind das metrisch autosegmentale Modell, Sp-ToBI, und das Modell der melodischen Konturen mit standardisierter Grundfrequenz. Das geeignetste Modell für den Rhythmus ist das der rhythmischen Gruppen. Die beste Art, die Geschwindigkeit und die Intensität zu untersuchen, ist die Analyse der prozentuellen Änderungen von einer Silbe zur nächsten. Die Stimmqualität ist in spontanen Aufnahmen schwer zu untersuchen.
- H₃: Die genannten Modelle reichen aus, um funktionelle Ähnlichkeiten und regionale Differenzen der Prosodie des gesprochenen Spanisch von Mexikostadt und Madrid aufzuzeigen.
- H₄: Unter Benutzung der Sprache XML ist es möglich, die verschiedenen prosodischen Phänomene hierarchisch darzustellen und durch Variablen ihre akustischen Eigenschaften zu identifizieren.
- H₅: Das Programm Praat eignet sich am besten dazu, die akustischen Analysen durchzuführen und automatisch XML Dokumente zu erstellen, welche die für eine detaillierte Analyse der Intonation, des Rhythmus, der Geschwindigkeit und der Intensität der linguistischen Einheiten der spanischen und mexikanischen Sprecher notwendigen Informationen enthalten.

Im Folgenden sollen diese Hypothesen sowie deren Ergebnisse kurz beschrieben werden.

Was die Hypothesen H₁, H₂ und H₃ angeht, so werden im ersten Teil der Arbeit die grundlegenden Konzepte zur Posodie dargestellt, da die Arbeit nicht nur die Untersuchung der Intonation, des Rhythmus und der Intensität für sich alleine genommen beabsichtigt, sondern auch einen Überblick über den Stand der Forschung geben möchte, sowie zur Behandlung der suprasegmentalen Phänomene als einer vielseitigen Einheit beitragen soll. Prosodie umfasst eben nicht nur die Analyse der Intonation und des Rhythmus sowie der

akustischen Phänomene für sich genommen, sondern alle Phänomene sollten gemeinsam untersucht werden, damit ihre Interaktionen und Abhängigkeiten untereinander berücksichtigt werden können.

Für das Spanische gibt es nur wenige Studien, welche mehr als einen Aspekt der Prosodie untersucht haben - in der Regel wird die Intonation analysiert. Der Rhythmus hat sich für die Gegenüberstellung von semantischen oder pragmatischen Unterschieden als weniger hilfreich herausgestellt, aber es ist offensichtlich, dass ein gleicher Satz, wenn er in einem speziellen Kontext melodisch und rhythmisch gesprochen wird, eine andere pragmatische Funktion übernimmt.

In dieser Arbeit wird versucht, alle für diese Untersuchung relevanten theoretischen Modelle, die bereits durch einen Autor auf das Spanische angewendet wurden und wichtige Grundgedanken vermitteln, aufzugreifen. Aus diesem Grund wurden am Ende auch Theorien gemeinsam berücksichtigt, die durch ihre eigenen Autoren als inkompatibel angesehen werden. Dies gilt insbesondere im Hinblick auf die Theorie der melodischen Konturen des Spanischen in der spontanen Sprache von Cantero (2002 und 2005) auf der einen Seite und die metrisch autosegmentale Theorie und hier insbesondere die *Spanish Tones and Break Indices*, Sp-ToBI, (Sosa 1999 und 2003, Beckman *et al.* 2002, Hualde 2003 und Prieto 2003) auf der anderen Seite. Für die Durchführung der Analyse dieser Arbeit waren jedoch beide Theorien hilfreich. Die Aspekte der Theorie der melodischen Konturen vereinfachten die Analyse der intonativen Konturen der Sätze als unabhängige und doch voneinander abhängige Einheiten, wohingegen die autosegmentale Metrik eine bessere Analyse der Töne als Komponenten dieser intonativen Konturen zuließ. Man kann sogar sagen, dass die vorliegende Studie ohne die Verbindung dieser beiden Theorien nicht in der Tiefe und unter Berücksichtigung der Interdependenzen hätte durchgeführt werden können.

Zudem hat die Theorie der melodischen Konturen auch neue Ideen geschaffen, die so durch andere Forscher noch nicht entwickelt worden waren. Im Bereich der Rhythmusanalyse wurde beispielsweise die Beschreibung der rhythmischen Gruppen von Cantero (2002) in weiterem Sinne angewendet und so ein Transkriptionssystem geschaffen, das dem der Darstellung des Gewichts der Silben in der Mehrzahl der aktuellen fonologischen Forschungsrichtungen entspricht. Beispielsweise steht die Sequenz (*s w*) für den Trochäus, wobei *s* die betonte Silbe und *w* die unbetonte Silbe darstellt und die Länge und Grenzen des Fußes oder der rhythmischen Gruppe durch Klammern angezeigt werden. Diese Darstellung war nicht nur für den Forscher als Mensch nützlich und unerlässlich für die Analyse des Rhythmus, sondern sie erlaubte auch eine leichte Erkennung und Verarbeitung

durch den Computer. So konnte beispielsweise sehr leicht gezeigt werden, welche Sequenzen unmöglich waren oder in der Umgangssprache kaum auftauchten, nämlich alle die Gruppen mit drei oder mehr unbetonten Silben nach der betonten Silbe, was die Anzahl der möglichen Kombinationen von 21 auf nur 15 oder 16 reduzierte.

Auch die Methode der Standardisierung der Grundfrequenz hat sich als sehr nützlich erwiesen und zwar nicht nur für die Analyse der Variation der Frequenz sondern vor allen Dingen auch für die Analyse der Variation der Intensität. Auch wenn es sehr wahrscheinlich ist, dass andere Forscher schon ähnliche Ideen hatten, kenne ich doch keine Arbeit, welche standardisierte Werte verwendet hat, das heißt in der ein beliebiger Wert für die erste Silbe jeder Intonationseinheit des Korpus gewählt wurde und die Werte der nachfolgenden Silben durch Addition bzw. Subtraktion der Variation zwischen der jeweiligen neuen Silbe und der ersten Silbe zum gewählten Startwert berechnet wurden.

Außerdem stellte sich heraus, dass das Inventar der Töne des Sp-ToBI (das ich nicht nur aus Gründen der Systematik und Symmetrie sondern auch aus theoretischer Kongruenz um den Ton $\uparrow L^*+H$ erweitert habe) leichter vergleichbar war, wenn man die individuellen Töne als Allotöne oder phonetische Varianten in komplementärer Distribution von nur vier Archintonemen ansieht: \acute{S} (Töne, dessen höchster Frequenzwert mit der betonten Silbe zusammenfällt: $L+H^*$, $L+\!H^*$, $\uparrow L+H^*$ y H^*), S' (Töne, dessen höchster Frequenzwert nach der betonten Silbe liegt: L^*+H , $L^*+\!H$ y $\uparrow L^*+H$), \acute{S} (Töne, dessen höchster Frequenzwert vor der betonten Silbe liegt: $H+L^*$), und S (Töne ohne oder mit kaum wahrnehmbaren Frequenzmaximum: *).

Was die Dauer der Silben angeht, so wurde diese Information und die Variation in der relativen Dauer der Silben dazu benutzt, das Tempo oder die mittlere Geschwindigkeit, mit der die Sprecher normalerweise kommunizieren, ungefähr zu ermitteln.

So wird schliesslich im Abschnitt 3.5 – in dem die vergleichende Analyse zusammengefasst wird – dargelegt, dass die beschriebene Wahl der akustischen Parameter und theoretischen Modellen es tatsächlich erlaubt, die wichtigsten wahrnehmbaren Unterschiede der Prosodie des Spanischen zwischen Madrid und Mexikostadt aufzuzeigen.

Was die Hypothesen H_4 und H_5 angeht, so wurde im Rahmen der vorliegenden Arbeit ein auf der Sprache XML basierendes Transkriptionssystem entwickelt, das die Einführung von Text und Etiketten mit struktureller Information erlaubt. Die verschiedenen Etiketten und ihre grundlegende Grammatik wurden mit dem Ziel entworfen, die hierarchischen und funktionellen Beziehungen der verschiedenen Komponenten der Prosodie darstellen zu können.

Parallel zur Entwicklung dieses Transkriptionssystems wurden die Aufnahmen ausgewählt, die für die akustische Analyse der sprachlichen Äußerungen der Sprecher aus Madrid und Mexikostadt geeignete Fragmente enthielten. Außerdem wurde eine Serie von Transkriptionsregeln festgelegt, um die Konsistenz zwischen den Texten der Transkriptionen sicherzustellen.

Zur weiteren Anpassung wurden *Scripts* in Praat entworfen, welche die akustische Analyse der in Silben unterteilten Transkriptionen automatisieren sollten. Zudem sollten dadurch die gewonnenen Informationen gleich in ein Textdokument geschrieben werden, wobei die darin zusätzlich enthaltenen Zeichen für Praat zwar nur weiterer Text waren im letztlichen Textdokument jedoch die syntaktische Struktur der Etiketten in XML darstellten.

Nachdem die entsprechenden XML Dokumente für jedes Fragment der ausgewählten Aufnahmen erstellt wurden, wurden die Information und der Text innerhalb und ausserhalb der Etiketten der Elemente und Attribute an verschiedene Formate angepasst, beispielsweise Internetseiten. Diese Seiten, welche durch die Umwandlung der Daten mit Hilfe der Sprache XSLT entstanden sind, dienten im Rahmen der vorliegenden Arbeit verschiedenen Zielen. Zum einen konnte so die gewonnene Information in einer interaktiven Internetseite dargestellt werden, welche sowohl Bilder bezüglich der Analyse der Grundfrequenz, die von Praat automatisch erzeugt wurden, enthielt als auch Verknüpfungen zu den entsprechenden Aufnahmen. (All das trug zu einer besseren Verwaltung und manuellen Auswahl der geeigneten Fragmente bei.) Zum anderen erlaubten diese Seiten eine selektive Darstellung, d.h. eine gefilterte Auswahl von Daten, welche nur die Fragmente zeigten, die von bestimmten Personen gesprochen wurden oder die bestimmte prosodische Phänomene aufwiesen, welche die akustische Analyse nicht verkomplizierten. Schliesslich unterstützte diese Darstellung auch die statistische Auswertung, da die Information auf diese Art einfach in Excel kopiert werden konnte, um dort mit der statistischen Analyse der Daten fortzufahren. Die so durch Excel gewonnene Information wurde in den Datentabellen der Abschnitte 3.2 und 3.3 zusammengefasst. Es ist allerdings wichtig hervorzuheben, dass die ausgewählten Ergebnisse nur einen Bruchteil der vielen verschiedenen Möglichkeiten bezüglich der Auswahl, der Verarbeitung, der Filter und der Unterstichproben darstellen. In den genannten zwei Abschnitten konnten zwar die Möglichkeiten der vergleichenden Analyse schon vorhergesehen werden (da die Daten der verschiedenen Sprecher und ihre Mittelwerte für Frauen und Männer schon Unterschiede und Gemeinsamkeiten erahnen ließen), dennoch wurde versucht, zunächst ausschließlich die Daten darzustellen, wobei die gemittelten Extremwerte und Mittelwerte der einzelnen Sätze aufgezeigt wurden. Insbesondere sollte

auch die Verwendung der standardisierten Werte für die Intensität und die Grundfrequenz begründet werden, durch die später die Variation dieser Parameter verglichen werden konnten, ohne dass die Ergebnisse durch nicht linguistische Faktoren beeinflusst wurden.

Nach der Darstellung der individuellen Komponenten der Analyse folgte die vergleichende Analyse im Abschnitt 3.4. Hier wurden die Mittelwerte der Sprecher aus Madrid und Mexikostadt in Grafiken dargestellt, um die wichtigsten Unterschiede und Gemeinsamkeiten in den sprachlichen Äußerungen jeder einzelnen Sprechergruppe zu visualisieren.

Die Ergebnisse waren sehr zufriedenstellend, denn sie lieferten konkrete und akustisch identifizierbare Beweise der Unterschiede, die von den Sprechern auch wahrgenommen werden. So konnte zum Beispiel gezeigt werden, dass vom Standpunkt der Sprecher aus Madrid 1) die Mexikaner die Silben langsamer und gleichmäßiger aussprechen, was dazu führt, dass ihre Sprechweise allgemein als langsamer wahrgenommen wird; 2) die Mexikaner mit geringerer Intensität oder Lautstärke sprechen und auch gleichmäßiger, das heißt sie neigen weniger dazu die Stimme in einem einzelnen Satz zu erheben oder zurückzunehmen; 3) die Mexikaner ein höheres Stimmregister benutzen, was sich auch in einer größeren Variation der Frequenz niederschlägt - das heißt sie sprechen weniger monoton; 4) die Mexikaner die Sätze häufiger mit einer hohen Intonation anfangen lassen, sie gewöhnlich die betonten Silben mit einem gleichzeitigen Frequenzmaximum aussprechen, wobei dieses höher ist als für die Sprecher von Madrid, und schliesslich die Sätze mit einer steigenden Intonation beenden, wobei sie eine höhere Frequenz verwenden - genauso wie bei fallenden Konturen die Frequenz prozentuell niedriger ist; 5) die Mexikaner im Allgemeinen in Sätzen mit gleichbleibender Kontur häufiger Töne des Typs \acute{S} verwenden, das heißt Töne deren Grundfrequenzmaximum mit der betonten Silbe zusammenfällt ($L+H^*$, $¡L+H^*$, $L+!H^*$ oder H^*) (auch wenn man beachten muss, dass die Verteilung der Nukleartöne und Endtöne in den verschiedenen Arten von Sätzen sehr stark variiert); und 6) die Mexikaner dazu neigen, kürzere rhythmische Strukturen zu formen mit einer Vorliebe für die Betonung auf der vorletzten Silbe.

Was jedoch im Vergleich von nur zwei dialektalen Varietäten der gleichen Sprache gut zu funktionieren scheint, wird sich in der Zukunft mit breiteren Datensätzen und mehreren Sprachen testen lassen müssen. Nur so werden neue Probleme identifiziert und gelöst werden können, was neue methodologische Implikationen mit sich bringen wird.

Was von der in dieser Arbeit verwendeten Methode zudem hervorgehoben werden muss, ist die Tatsache, dass sie neben der möglichen Anwendung auf die Gesamtheit der linguistischen Korpora mit Daten der gesprochenen Sprache und mit verschiedenen akustischen Qualitäten

in der Aufnahme eine Möglichkeit bietet, auf systematische Art und Weise gleichzeitig verschiedene akustische Parameter der Prosodie als einen einzigartigen und vielfältigen Forschungsbereich zu analysieren. So kann die Prosodie als eine Seite der Phonologie angesehen werden, die als ein Ganzes untersucht werden kann, obwohl sie sich aus verschiedenen Phänomenen zusammensetzt, welche die physischen Eigenschaften der Lautmaterie darstellen.

Was die Anwendungsmöglichkeiten der quantifizierten dialektalen Unterschiede, die durch die beschriebene Methode gewonnen wurden, angeht, so mögen diese für Spracherkennungssysteme hilfreich sein. Sie könnten aber auch in der Umwandlung von Text in Sprache oder der Sprachsynthese verwendet werden oder aber einfach auch zur Dokumentation von verschiedenen Sprachen und ihren dialektalen Varietäten. Die Zeit wird zeigen, inwiefern das tatsächlich geschieht.