

Síntesis del habla con emociones en el dominio de las conversaciones virtuales

Yesika Laplaza Miras

TESIS DOCTORAL UPF / 2013

DIRECTORES DE LA TESIS

Dr. Toni Badia Cardús

Dr. Juan María Garrido Almiñana

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL
LLENGUATGE



A Robert i Júlia

Agradecimientos

Mientras redactaba esta tesis pensaba en este apartado, en el qué pondría y cómo, creyendo que sería el más fácil de todos. Ahora, cuando realmente me dispongo a escribirlo, me doy cuenta que es el más difícil, que tengo mucho que agradecer a todas las personas de mi alrededor y que mis palabras no pueden reflejar ni un ápice lo que realmente siento. A pesar de ello, quiero agradecer a cada una de ellas todo su apoyo y su comprensión, y en especial:

A Robert por su apoyo incondicional y su paciencia infinita. Por estar siempre ahí, en los buenos momentos y en los no tan buenos, asumiendo responsabilidades para hacerme la vida más sencilla. Porque tú también has sufrido y padecido esta tesis. Te quiero.

A Júlia, mi pequeña gran maestra, quien en tan poco tiempo me ha enseñado infinidad de cosas, muchas de las cuales no aparecen en los libros. Lamento que nuestro tiempo compartido de juegos y risas se haya reducido tanto en estos últimos meses.

A mis padres Eduardo y Antonia, gracias a ellos soy todo lo que soy. Gracias por apoyarme en todas mis decisiones y por la gran ayuda que habéis sido en estos últimos meses ocupándoos de mi moquito y animándome cada día.

A toda mi familia y mi familia política, demasiado extensa para nombrarlos uno a uno, por todo su apoyo, afecto, comprensión y por alentarme en los días más difíciles.

A mis amigos: Cris, Juan, Dani, Anna, Javi, Sergi, Dídac, Neus, Roger, Rubén, Fully, Laura, Gilmar, Alejandro, a mis amigos de la Universidad y de trabajo, y muchos más de una interminable lista. Muchas gracias a todos vosotros por todos los momentos compartidos, por las risas y los momentos de diversión y evasión.

A mis directores Toni Badia por acogerme en Barcelona Media y darme la oportunidad de iniciar esta investigación, y Juanma Garrido por su confianza, la gran ayuda y conocimiento aportado, y

por todos sus esfuerzos puestos para ver esta tesis finalizada. Mil gracias.

A las personas que me han cedido sus conversaciones virtuales privadas para poder realizar esta tesis: Tamara, Alex, Dani, Ari, Marc, Dora, Gemma, Robert, y Mar. Muchas gracias a todos por todo el material aportado.

Agradecer también a Barcelona Media la confianza puesta en esta tesis y su gran aportación en la grabación del locutor.

Finalmente, quiero agradecer la colaboración desinteresada de todos los evaluadores, cuya intervención ha sido indispensable para concluir esta investigación.

Resumen

Esta tesis, al centrarse en la generación de voz sintética en el dominio de las conversaciones virtuales en español, trata dos aspectos diferentes del proceso de la conversión de texto en habla:

Por un lado, parte de esta investigación gira alrededor de la normalización-corrección de los mensajes que los usuarios escriben mientras chatean con amigos o conocidos, cuyo lenguaje dista considerablemente del texto estándar y normativo que emplean estos conversores. Estos textos presentan numerosas abreviaturas, emoticonos, sustituciones de grafías o repeticiones de ellas, haciendo que el texto si es procesado por los normalizadores convencionales de los CTH resultara incomprensible por el oyente.

Por otro lado, el habla generada debe corresponderse con el dominio especificado. En las conversaciones virtuales, que se asemejan más a una conversación oral que a un discurso escrito, predomina un afán por comunicar situaciones, transmitir estados emocionales, opiniones, sentimientos, etc. Por lo tanto, la voz desarrollada en esta investigación pretende reflejar esta expresividad, concretamente se pretende generar enunciados en los que se transmitan emociones mediante la modelación de parámetros prosódicos. Para lograr este propósito se parte del conversor de texto a habla de la empresa escocesa Cereproc.

Palabras clave: Conversión de texto en habla, conversaciones virtuales, normalización, emociones, prosodia.

Abstract

This thesis, by focusing on the generation of synthetic speech in the domain of virtual Spanish conversations, addresses two different aspects of the Text-to-Speech process.

On one hand, most of this research revolves around the standardization-correction of messages that users type while chatting with friends or acquaintances, whose language varies considerably from standard text and the policy using these systems. These texts have many abbreviations, emoticons, photographic substitutions or repetitions of them, making the text incomprehensible to the listener if it is processed by the conventional normalisers of TTS.

On the other hand, the speech generated should match the specified domain. In virtual conversations, which are more like a conversation than written speech, the prevailing situations are eagerness to communicate, convey emotional states, feelings, etc. Therefore, the voice developed in this research is intended to reflect this expression, specifically aims to generate statements, which emotions are transmitted through prosodic modeling. To achieve this purpose, the text-to-speech system by the Scottish company CereProc is addressed.

Key words: text-to-speech, virtual conversations, normalization, emotion, prosody.

Prólogo

Como se verá a lo largo de esta investigación, el objetivo de los conversores de texto a habla es poder reproducir oralmente cualquier tipo de texto de una forma inteligible para el oyente y mediante una voz sintética lo más natural posible.

Acabo de decir que uno de los objetivos de los conversores de texto a habla es la capacidad de reproducir sonoramente *cualquier* tipo de texto, pero es necesario hacer una pequeña aclaración al respecto. Actualmente estos sistemas están diseñados para tratar única y exclusivamente texto normativo, por lo tanto sólo se podrá reproducir sonoramente *cualquier tipo de texto normativo*. Esto implica que textos que presenten anomalías con respecto al lenguaje estándar pueden conllevar problemas que afecten tanto a la naturalidad como a la propia inteligibilidad del mensaje.

En cuanto a la inteligibilidad, el resultado obtenido mediante síntesis por corpus, método más empleado por los conversores comerciales, es más que satisfactorio. En cambio, la naturalidad de la voz no se encuentra al mismo nivel. En los últimos años se ha despertado un creciente interés en mejorar este aspecto, intentando que las voces sintéticas se asemejen el máximo posible a la voz humana, capaz de aportar información verbal relevante para expresar la intencionalidad, la actitud o el estado emocional del emisor.

Por lo tanto, en esta tesis se pretende ampliar el tipo de textos tratables por los conversores de texto a habla, intentando que sea posible la afirmación '*cualquier* tipo de texto', e investigar la viabilidad de emplear conocimiento lingüístico para mejorar la expresividad natural de los sistemas de síntesis de habla estándar en el mundo industrial y comercial actual.

Índice

Agradecimientos	iii
Resumen	v
Abstract	vii
Prólogo	ix
Índice de figuras	xix
Índice de tablas	xxiii
Capítulo 1 INTRODUCCIÓN	1
1.1. Objetivos de la tesis	2
1.2. Contenido de la tesis	4
Capítulo 2 ESTADO DE LA CUESTIÓN	7
2.1. Conversión de texto en habla	7
2.1.1. Funcionamiento de los conversores de texto a habla basados en corpus	8
2.1.1.1. Creación del locutor sintético	9
2.1.1.2. Procesamiento del lenguaje natural	10
2.1.1.3. Generación del habla sintética de salida	15
2.2. Las conversaciones virtuales	16
2.2.1. Aspectos técnicos: Breve introducción al funcionamiento de los chats	17
2.2.1.1. Chats no vinculados a un registro previo	18
2.2.1.2. Chats vinculados a un registro previo o mensajería instantánea	20
2.2.2. Aspectos discursivos	21
2.2.2.1. El chat, entre lo oral y lo escrito	22
2.2.2.2. La conversación virtual <i>versus</i> la conversación real	23

2.2.2.3. El texto escrito oralizado _____	27
2.2.3. Características lingüísticas _____	28
2.2.3.1. Economía, rapidez y eficacia _____	29
2.2.3.2. Estrategias oralizadoras _____	32
2.2.3.3. Recursos paralingüísticos _____	33
a. Interjecciones y onomatopeyas _____	34
b. La acotación icónica _____	35
c. Los emoticonos _____	36
2.2.4. Similitudes con otros lenguajes _____	41
2.2.5. Las conversaciones virtuales y la generación de voz sintética _____	42
2.3. Síntesis de habla de conversaciones virtuales _____	44
2.3.1. La normalización de textos y los mensajes de las conversaciones virtuales _____	45
2.3.1.1. Basados en reglas _____	46
a. Corrección de textos _____	46
2.3.1.2. Basados en métodos estadísticos _____	49
a. Traducción automática _____	49
b. Técnicas de reconocimiento de voz _____	50
2.3.1.3. Valoración de las técnicas _____	52
2.3.2. La problemática de la generación de síntesis expresiva ____	54
2.3.2.1. Emociones _____	55
a. Teoría de las emociones _____	55
b. Clasificación de las emociones _____	57
2.3.2.2. Parámetros del habla relacionados con la emotividad_	67
a. Propiedades relacionadas con la intensidad _____	69
b. Propiedades relacionadas con la duración _____	69
c. Frecuencia fundamental (F0) y entonación _____	70
2.3.2.3. Modelado de las emociones para la síntesis de voz expresiva _____	72
a. Síntesis paramétrica de la prosodia _____	73
b. Síntesis por corpus _____	74

c. Síntesis basada en Modelos Ocultos de Markov (HMM)	75
2.4. Resumen	76
Capítulo 3 EL NORMALIZADOR DE TEXTOS	79
3.1. El normalizador de textos TexAFon	80
3.2. Análisis lingüístico de las conversaciones virtuales	84
3.2.1. El corpus	85
3.2.2. Metodología	90
3.2.2.1. Determinación de los tokens incorrectos	90
3.2.2.2. Clasificación de los tokens	91
3.2.2.3. Identificación/clasificación de errores	94
3.2.3. Cuantificación de errores	97
3.2.3.1. Un error por palabra	98
a. Sustitución de caracteres	98
b. Eliminación de caracteres	102
c. Inserción de caracteres	103
d. Transposición de caracteres	105
3.2.3.2. Dos errores por palabra	106
a. Sustitución y eliminación de caracteres:	107
b. Dos sustituciones	108
3.2.4. Resumen: implicaciones para la normalización de textos	109
3.3. TexAFon 2.0: Adaptación del normalizador TexAFon para conversaciones virtuales	111
3.3.1. Clasificación de tokens	115
3.3.2. Expansión de los tokens	117
3.3.2.1. Tratamiento de palabras con símbolos	117
3.3.2.2. Tratamiento de palabras con signos	118
3.3.2.3. Tratamiento de palabras con dígitos	119
3.3.2.4. Tratamiento de abreviaturas	120
3.3.2.5. Tratamiento de palabras	121
a. Identificación de palabras normalizables	121

b.	Generación de tokens normalizados _____	122
1.	Eliminación de grafías repetidas _____	122
2.	Tratamiento de los fenómenos más frecuentes _____	124
3.	Modificaciones generales _____	125
3.3.3.	Selección de la forma normalizada _____	127
3.4.	Resumen _____	130

Capítulo 4 GENERACIÓN DE LA VOZ SINTÉTICA EXPRESIVA _____ 135

4.1.	El corpus _____	136
4.1.1.	Contenido del corpus _____	136
4.1.1.1.	Corpus neutro _____	136
4.1.1.2.	Corpus emocional _____	137
a.	Selección de enunciados _____	138
b.	Contenido del corpus emocional _____	140
4.1.2.	Selección del locutor _____	142
4.1.3.	Grabación del corpus _____	144
4.1.4.	Edición y revisión de la señal de audio _____	145
4.2.	Generación de habla expresiva mediante técnicas basadas en corpus _____	146
4.2.1.	Procedimiento de generación de habla expresiva con el SDK de Cereproc _____	146
4.3.	Generación de habla expresiva mediante modelos prosódicos _____	149
4.3.1.	Generación de los modelos prosódicos _____	150
4.3.1.1.	Parámetros no melódicos _____	150
4.3.1.2.	Parámetros melódicos _____	155
4.3.2.	Resultados _____	158
4.3.2.1.	Parámetros no melódicos _____	158
c.	Datos _____	158
d.	Conclusiones _____	162
4.3.2.2.	Parámetros melódicos _____	163

a.	Parámetros globales _____	163
b.	Patrones melódicos _____	168
1.	Patrones iniciales _____	169
2.	Patrones interiores _____	175
3.	Patrones finales no final de enunciado _____	180
4.	Patrones finales de enunciado _____	186
c.	Conclusiones _____	192
4.3.3.	Generación del habla expresiva _____	193
4.3.3.1.	Parámetros no melódicos _____	194
4.3.3.2.	Parámetros melódicos _____	195
4.4.	Resumen _____	196
Capítulo 5	<i>EVALUACIÓN</i> _____	201
5.1.	El normalizador de textos _____	201
5.1.1.	Procedimiento _____	201
5.1.2.	Resultados _____	204
5.1.3.	Discusión _____	208
5.2.	Síntesis de voz expresiva _____	210
5.2.1.	Procedimiento _____	211
5.2.2.	Resultados _____	214
5.2.2.1.	Calidad de la voz sintética (CS) _____	215
5.2.2.2.	Similitud con el locutor original (SS) _____	216
5.2.2.3.	Identificación de las emociones (IE) _____	218
5.2.2.4.	Fuerza de las emociones (FE) _____	219
5.2.2.5.	Rendimiento emocional _____	220
5.2.2.6.	Comparación síntesis por corpus y síntesis paramétrica _____	222
5.2.3.	Discusión _____	223
5.3.	Resumen _____	226

Capítulo 6 CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN _____ **229**

6.1. Normalización del texto _____ **229**

6.2. Generación de síntesis expresiva _____ **233**

REFERENCIAS _____ **237**

Apéndice A CONTENIDO DEL CORPUS DE TEXTO ____ **251**

A.1. Estados emocionales y fisiológicos aislados _____ **251**

A.2. Actos de habla _____ **253**

A.3. Combinación de dos elementos _____ **253**

A.4. Combinación de tres elementos _____ **256**

Apéndice B CORPUS DE GRABACIÓN _____ **257**

B.1. Corpus emocional _____ **257**

B.1.1. Admiración _____ **257**

B.1.2. Afecto _____ **258**

B.1.3. Alegría _____ **259**

B.1.4. Alegría nivel 2 más saludo _____ **260**

B.1.5. Alteración _____ **261**

B.1.6. Burla _____ **263**

B.1.7. Burla más rechazo _____ **264**

B.1.8. Decepción _____ **265**

B.1.9. Despedida _____ **266**

B.1.10. Diversión _____ **267**

B.1.11. Duda _____ **268**

B.1.12. Enfado _____ **269**

B.1.13. Enfado más rechazo _____ **270**

B.1.14. Extrañeza _____ **271**

B.1.15. Interés _____ **272**

B.1.16. Orgullo _____	273
B.1.17. Rechazo _____	274
B.1.18. Resignación _____	276
B.1.19. Saludo _____	277
B.1.20. Sorpresa _____	278
B.1.21. Tristeza _____	279

***Apéndice C ANÁLISIS DE LOS PATRONES MELÓDICOS
DEL CORPUS DE GRABACIÓN _____ 281***

C.1. Parámetros globales _____	281
C.1.1. Grupo fónico inicial _____	282
C.1.2. Grupo fónico interior _____	283
C.1.3. Grupo fónico final _____	284
C.2. Patrones iniciales _____	285
C.2.1. Una sílaba _____	285
C.3. Patrones interiores _____	300
C.4. Patrones finales no finales de enunciado _____	341
C.4.1. Una sílaba _____	341
C.4.2. Dos sílabas _____	343
C.4.3. Tres sílabas _____	352
C.5 Patrones finales de enunciado _____	354
C.5.1. Una sílaba _____	354
C.5.2. Dos sílabas _____	358
C.5.3. Tres sílabas _____	370

Índice de figuras

<i>Figura 1: Diagrama de bloques de los procesos que forman parte de un sistema de CTH basado en concatenación de unidades (Iriondo, 2008)</i>	<i>_9</i>
<i>Figura 2 Conversación virtual pública extraída de Inforchat</i>	<i>19</i>
<i>Figura 3: Conversación privada entre Robert y Mandarina INA a través del programa informático Messenger</i>	<i>21</i>
<i>Figura 4: Ejemplo de una conversación privada entre dos usuarios en la que un turno de habla de Resurgir (en negrita) se ve interrumpido por un mensaje de Vafalungo (en verde), produciendo un solapamiento</i>	<i>27</i>
<i>Figura 5: Ejemplo de la escritura empleada en la red social Facebook donde se observan algunos de los rasgos presentes en las conversaciones virtuales</i>	<i>42</i>
<i>Figura 6: Inventario de errores empleado en Moré et al. (2005) para catalogar los fenómenos observados en los correos electrónicos en castellano y catalán</i>	<i>48</i>
<i>Figura 7: Modelo circunflejo tridimensional de Plutchik (2001)</i>	<i>58</i>
<i>Figura 8: Inventario de emociones desarrollado en el proyecto HUMAINE</i>	<i>64</i>
<i>Figura 9: Etiquetas de los estados emocionales presentados en I3Media</i>	<i>65</i>
<i>Figura 10: Etiquetas de los estados fisiológicos presentados en I3Media</i>	<i>66</i>
<i>Figura 11: Resumen de los efectos vocales que tienen cinco emociones básicas en el habla, traducido de Murray y Arnott (1993) y extraído de Iriondo (2008)</i>	<i>68</i>
<i>Figura 12: Arquitectura general del TexAFon</i>	<i>81</i>
<i>Figura 13: Distribución de los fenómenos posibles hallados en un error por palabra</i>	<i>98</i>

<i>Figura 14: Distribución de los diferentes tipos de sustituciones observadas en el corpus de análisis</i>	99
<i>Figura 15: Arquitectura general de TexAFon 2.0</i>	114
<i>Figura 16: Ejemplos extraídos del modelo de lenguaje específico de las conversaciones virtuales</i>	128
<i>Figura 17: Algunos enunciados etiquetados del corpus de las conversaciones virtuales</i>	140
<i>Figura 18: Dos ejemplos de enunciados identificados como enfado y sorpresa respectivamente presentados en el contexto en el que se desarrollaba la conversación para facilitar la interpretación al locutor durante la grabación</i>	144
<i>Figura 19: Enunciados del corpus neutro y emocional con sus respectivas etiquetas que indican el dominio al que pertenecen (letra que aparece al inicio) y el número del enunciado dentro del corpus general</i>	147
<i>Figura 20: Ejemplo del etiquetado XML empleado para realizar un enunciado seleccionando los diferentes segmentos dentro del corpus emocional representativo de las conversaciones virtuales</i>	148
<i>Figura 21: Etiquetado empleado para realizar la frecuencia fundamental, la amplitud y la duración de un enunciado determinado según los valores especificados en las etiquetas</i>	148
<i>Figura 22: Ejemplo de notación y líneas de referencia del enunciado “¿Quiere alguien explicarme de qué se trata?” extraído de Garrido (2011)</i>	157
<i>Figura 23: Ejemplo de patrón inicial de enunciado VIO_PM0 en el GA ‘ya’ dentro del enunciado neutro ‘Ya son ciento doce los soldados de Estados Unidos muertos en Irak en mayo, el peor mes de dos mil siete.’</i>	173
<i>Figura 24: Ejemplo de patrón inicial de enunciado VIO en el GA ‘luis al’ dentro del enunciado neutro ‘Luis Alberto Camacho Girones; plaza de Portugal, uno: cero seis cero cero uno; Badajoz.’</i>	174
<i>Figura 25: Ejemplo de patrón inicial de enunciado VIO en el GA ‘no’ dentro del enunciado emocional ‘No, yo me he argumentado’</i>	174
<i>Figura 26: Ejemplo de patrón inicial de enunciado VIO_PM0 en el GA ‘y tú’ dentro del enunciado emocional ‘Y tú una furcia.’</i>	175

- Figura 27* Ejemplo de patrón interior de enunciado *VI0_PM0* en el GA 'dos', dentro del enunciado neutro 'Piden cinco años para un empresario por dar tarta de marihuana a dos jóvenes.' _____ 179
- Figura 28:* Ejemplo de patrón interior de enunciado *VI0_PM0* en el GA 'todo' dentro del enunciado emocional 'Porque lo entiendes. El amor está por encima de todo eso' _____ 179
- Figura 29:* Ejemplo de patrón ascendente final no final de enunciado *VI0_PM0_PF0* en el GA 'tres' dentro del enunciado neutro 'Noventa y tres, ochocientos setenta y nueve, cuarenta y nueve, catorce' _____ 181
- Figura 30:* Ejemplo de patrón descendente final no de enunciado *PI0_VI1_VM1* en el GA 'shopping' dentro del enunciado neutro 'La larga noche de los museos y la larga noche del shopping, en que los establecimientos alargaron su horario de apertura al público' _____ 182
- Figura 31:* Ejemplo de patrón descendente final no final de enunciado *VF0* en el GA 'nir' dentro del enunciado emocional 'Luego se tirará unos días sin venir, como cuando ganó el Madrid en el Camp Nou' _____ 182
- Figura 32:* Ejemplo de patrón circunflejo final no final de enunciado *VI0_PI1_VF1* en el GA 'jodas' dentro del enunciado emocional 'Hombre, es que no me jodas, le pongo mi foto y me come, así, directamente' _ 183
- Figura 33:* Ejemplo de patrón descendente final de enunciado *VF0* en el GA 'liz' dentro del enunciado neutro 'Es triste que en el ciclismo no haya un día feliz' _____ 187
- Figura 34:* Ejemplo de patrón descendente final de enunciado más frecuente en el corpus neutro cuyo GA está formado por dos sílabas: *PI0_VI1_VM1* en el GA 'mento' dentro del enunciado neutro 'Es salvajismo de los prejuicios sin fundamento' _____ 188
- Figura 35:* Ejemplo de patrón descendente final de enunciado *VF0* en el GA 'mil' dentro del enunciado emocional 'Estoy rodeado de buenas noticias y eventos mil' _____ 188
- Figura 36:* Ejemplo de patrón circunflejo final de enunciado *VM1* en el GA 'diando' dentro del enunciado emocional 'Odio lo que estoy estudiando' _____ 189
- Figura 37:* Distribución de las palabras correctas e incorrectas presentes en el texto original y la versión corregida automáticamente mediante *TexAFon 2.0* _____ 204

<i>Figura 38: Distribución de las palabras incorrectas después del proceso de normalización</i>	205
<i>Figura 39: Comparación de la distribución de los resultados obtenidos en la corrección de palabras erróneas y falsos positivos del sistema de Telefónica I+D (Armenta et al. (2003 b) y el desarrollado en esta investigación</i>	209
<i>Figura 40: Interfaz web usada por los evaluadores para realizar la evaluación de la voz</i>	214
<i>Figura 41: Resultados obtenidos en la prueba de evaluación de la calidad de la voz</i>	215
<i>Figura 42: Resultados obtenidos en la prueba de similitud entre la voz original y la sintética</i>	217
<i>Figura 43: Resultados obtenidos en la identificación del grado de intensidad con la que se realizaba el enunciado</i>	220
<i>Figura 44: Distribución de los valores medios obtenidos en cuatro pruebas de la evaluación</i>	222

Índice de tablas

<i>Tabla 1: Aspectos orales y escritos en la comunicación mediada por ordenador</i>	23
<i>Tabla 2: Comparativa entre la conversación virtual y la conversación real propuesta por López Quero (2003)</i>	25
<i>Tabla 3: Sustituciones grafemáticas más frecuentes en el lenguaje 2.0 comentadas por Morala (2001)</i>	31
<i>Tabla 4: Ejemplo extraído del diccionario online diccionariosms en el que se muestran los posibles significados que pueden tener los emoticonos %-), :-/ y ;o)</i>	38
<i>Tabla 5: Inventarios de emociones básicas propuestos por varios autores (Cowie et al., 2003)</i>	62
<i>Tabla 6: Lista de categorías consideradas en el preproceso</i>	82
<i>Tabla 7: Información sobre las conversaciones que forman el corpus de análisis</i>	89
<i>Tabla 8: Listado de los diferentes fenómenos observados en las palabras consideradas incorrectas del primer análisis, y la frecuencia de aparición de cada uno de ellos</i>	93
<i>Tabla 9: Ejemplos de pares de palabras cuyas formas presentan algunas diferencias entre sí</i>	94
<i>Tabla 10: Algunos ejemplos de palabras incorrectas etiquetadas extraídos de la tabla resultante</i>	96
<i>Tabla 11: Recuento de los errores o desviaciones observadas en el corpus de análisis clasificados según el número de anomalías presentes en cada palabra no correcta</i>	97
<i>Tabla 12: Resumen de las grafías sustituidas de forma sistemática y su correspondiente equivalencia</i>	100
<i>Tabla 13: Combinaciones sintácticas más frecuentes en las que se ha elidido el espacio en blanco. El porcentaje se ha calculado respecto al total de elisiones</i>	103

<i>Tabla 14: Distribución de las grafías que han sido insertadas, las cuales no se corresponden ni con la repetición de grafías o espacios en blanco ni por proximidad en el teclado, según si dicha grafía pertenece a la palabra anterior o posterior</i>	105
<i>Tabla 15: Distribución de la transposición de caracteres en las palabras no monosilábicas</i>	106
<i>Tabla 16: Distribución de los errores cuando se comenten 2 en una misma palabra</i>	107
<i>Tabla 17: Distribución de las diferentes combinaciones de errores a causa de la eliminación y sustitución de diversos caracteres en una misma palabra</i>	108
<i>Tabla 18: Distribución de las diferentes combinaciones de errores a causa de la doble sustitución de diversos caracteres en una misma palabra</i>	109
<i>Tabla 19: Equivalencias de vocales y dígitos empleadas en el corpus de análisis</i>	119
<i>Tabla 20: Dos ejemplos del funcionamiento, paso a paso, de TexAFon 2.0</i>	129
<i>Tabla 21: Distribución de la información de la 'tabla_vocal_prepausal.txt'</i>	152
<i>Tabla 22: Ejemplos extraídos de la tabla 'medias_dur_voc_prepausal.summary'</i>	154
<i>Tabla 23: Ejemplos extraídos de la tabla 'ratio_neutro_dur_vocal_prepausal.summary'</i>	154
<i>Tabla 24: Resultado del análisis ANOVA realizado para cada parámetro seleccionado</i>	158
<i>Tabla 25: Medias obtenidas para cada uno de los parámetros seleccionados, clasificadas según el tipo de emoción</i>	161
<i>Tabla 26: Valores medios de los puntos iniciales y finales y pendientes de las rectas superior (P) e inferior (V) en las diferentes emociones analizadas y en habla neutra</i>	165
<i>Tabla 27: Valores medios para el rango inicial y final en las diferentes emociones analizadas y en habla neutra</i>	167

<i>Tabla 28: Clasificación de los patrones melódicos neutros y emocionales según la modalidad oracional</i>	168
<i>Tabla 29: Clasificación de los patrones melódicos de las frases enunciativas según el número de sílabas que contienen</i>	169
<i>Tabla 30: Patrones iniciales más frecuentes en los grupos acentuales de una sílaba con acento en la primera sílaba</i>	170
<i>Tabla 31: Patrones iniciales más frecuentes en los grupos acentuales de dos sílabas con acento en la primera sílaba</i>	171
<i>Tabla 32: Patrones iniciales más frecuentes en los grupos acentuales de tres sílabas con acento en la primera sílaba</i>	172
<i>Tabla 33: Patrones interiores más frecuentes en los grupos acentuales de una sílaba con acento en la primera sílaba</i>	176
<i>Tabla 34: Patrones interiores más frecuentes en los grupos acentuales de dos sílabas con acento en la primera sílaba</i>	177
<i>Tabla 35: Patrones interiores más frecuentes en los grupos acentuales de tres sílabas con acento en la primera sílaba</i>	178
<i>Tabla 36: Resumen de los patrones finales no finales de enunciado más frecuentes del corpus emocional y neutro, de los grupos acentuales formados por 1, 2 y 3 sílabas, y cuya tónica es la primera</i>	180
<i>Tabla 37: Clasificación de los patrones finales no finales de enunciado, cuyo GA está formado por una sílaba</i>	184
<i>Tabla 38: Clasificación de los patrones finales no finales de enunciado, cuyo GA está formado por dos sílabas, y la primera es la tónica</i>	184
<i>Tabla 39: Clasificación de los patrones finales no finales de enunciado, cuyo GA está formado por tres sílabas, y la primera es la tónica</i>	185
<i>Tabla 40: Resumen de los patrones finales de enunciado más frecuentes del corpus emocional y neutro, de los grupos acentuales formados por 1, 2 y 3 sílabas, y cuya tónica es la primera</i>	186
<i>Tabla 41: Clasificación de los patrones finales de enunciado, cuyo GA está formado por una sílaba</i>	190
<i>Tabla 42: Clasificación de los patrones finales de enunciado, cuyo GA está formado por dos sílabas, y la primera es la tónica</i>	191

<i>Tabla 43: Clasificación de los patrones finales de enunciado, cuyo GA está formado por tres sílabas, y la primera es la tónica</i>	191
<i>Tabla 44: Algunos ejemplos extraídos de las tres versiones diferentes (texto original, normalización manual y normalización automática) necesarias para realizar la evaluación de TexAFon 2.0</i>	202
<i>Tabla 45: En negrita, ejemplos de palabras incorrectas que se han corregido mal y cuyo proceso de corrección sólo implicaba una modificación. A la derecha se encuentra la palabra en la intervención original de la conversación virtual, en el centro la corrección realizada manualmente, y a la derecha la propuesta de normalización de TexAFon 2.0</i>	206
<i>Tabla 46: Resultados obtenidos para la calidad de la voz sintética según el tipo de emoción realizada</i>	216
<i>Tabla 47: Resultados obtenidos sobre la similitud de la voz sintética respecto a la voz original del locutor según el tipo de emoción realizada</i>	217
<i>Tabla 48: Matriz de confusión de los resultados promediados de la prueba referente a la identificación de las emociones</i>	218
<i>Tabla 49: Resultados obtenidos sobre la fuerza con la que se realizaba cada emoción</i>	219
<i>Tabla 50: Resultados obtenidos al calcular el rendimiento emocional del sistema en general y en cada una de las emociones evaluadas</i>	221
<i>Tabla 51: Resultados obtenidos en la prueba 3 al comparar los enunciados generados mediante síntesis por corpus y los modelados paramétricamente</i>	223
<i>Tabla 52: Resultados obtenidos en la identificación de las emociones por los diferentes sistemas evaluados en IberSpeech 2012 más los datos obtenidos en esta investigación</i>	224
<i>Tabla 53: Resultados sin normalizar del rendimiento emocional obtenidos por los sistemas evaluados en ALBAYZIN 2012, más la valoración alcanzada en esta investigación</i>	225

Capítulo 1

INTRODUCCIÓN

En los últimos años, se ha observado un auge en el número de textos que se escriben y leen. Personas que antes sólo accedían a la lectura y escritura de textos por motivos estrictamente laborales o en ámbitos de educación, ahora lo hacen de manera más asidua y motivados por unos aspectos más lúdicos.

Las nuevas tecnologías han revolucionado el mundo de las comunicaciones, no sólo en el medio empleado, sino también en el lenguaje y en las relaciones sociales. Hoy en día, la gran mayoría de las personas tiene cuentas de correo electrónico, perfiles registrados en varias redes sociales, lee diarios en línea, puede buscar información de cualquier tipo sin necesidad de salir de casa, opinar sobre productos determinados o servicios, etc.

Esta facilidad para acceder a medios socializadores en línea se ve truncada para aquellas personas que presentan minusvalías visuales. La velocidad con la que se desarrollan las conversaciones virtuales, dominio en el cual se centra esta tesis, hace que sea imposible acceder a ellas con las herramientas de ampliación del texto. Convertir los textos en audio sería una de las posibles soluciones para que estas personas pudieran acceder a este tipo de servicios con total libertad, utilizando un conversor de texto a habla (CTH).

Actualmente, los CTH presentan un alto grado de inteligibilidad al tratar textos neutros y normativos; en cambio, en lo que a la naturalidad de los enunciados se refiere, no se encuentran a la misma altura, ya que no son capaces de reproducir todas las cualidades de la voz humana.

Los CTH comerciales están orientados principalmente al sector servicios, donde los textos a reproducir son normativos y predomina el tono neutro, aunque en los últimos años se está investigando para

intentar dotar de expresividad a estas voces, y hacerlas más atractivas para el oyente. Esta tesis se centra en el dominio de las conversaciones virtuales, un ámbito que se aleja del uso convencional de estos sistemas en varios sentidos:

- Por un lado, el lenguaje que se emplea en las conversaciones virtuales se aleja de la normatividad y las convenciones lingüísticas de la lengua escrita. Los usuarios emplean abreviaturas, emoticonos, eliminan unas grafías, repiten otras, etc., fenómenos lingüísticos que los CTH no pueden procesar adecuadamente ya que no esperan estas anomalías en el texto de entrada, por lo que si intentan reproducir el texto acústicamente el resultado es algo incomprensible para el oyente.
- Por otra parte, las conversaciones virtuales son conversaciones escritas que mantienen diversas personas, y como toda conversación informal entre amigos o conocidos, impera la necesidad de comunicar y expresar las emociones, estados anímicos, sentimientos, o impresiones. Reproducir oralmente los enunciados como si se tratara de texto neutro sería incongruente con el dominio seleccionado. Los CTH actuales no son capaces de generar habla sintética natural; por tanto es imprescindible abordar la generación de voz sintética expresiva, tratando aquellas emociones que son representativas de este dominio.

1.1. Objetivos de la tesis

El objetivo principal de esta tesis es investigar las posibilidades de obtención de una producción sonora inteligible y natural de los mensajes que conforman las conversaciones virtuales utilizando un CTH comercial basado en síntesis por corpus. Para ello se ha partido del CTH de la empresa escocesa Cereproc. De esta manera se pretende mejorar el rendimiento al hacer síntesis en una situación comunicativa determinada.

El funcionamiento de este conversor, como el de la mayoría de los CTH comerciales, se puede dividir en dos procesos o etapas: una primera etapa que se encarga de procesar lingüísticamente el texto, es decir, se encarga de expandir ortográficamente todos los elementos no normativos como los números, las abreviaturas, las direcciones de correo electrónico, etc. que puedan aparecer en el texto, de transcribir fonéticamente los enunciados, y de predecir aspectos prosódicos como la disposición y duración de las pausas. En la segunda etapa es donde se crea la señal acústica. Para ello se seleccionan, dentro de una base de datos desarrollada para tal fin, y concatenan los segmentos más favorables teniendo en consideración los aspectos predichos en la fase anterior.

Los esfuerzos de investigación se han centrado en dos líneas diferentes, pero complementarias, de cara al desarrollo de un CTH capaz de leer conversaciones virtuales:

- Normalización del texto de entrada: a través de un análisis del lenguaje empleado en las conversaciones virtuales, lo suficientemente minucioso para permitir la extracción de conclusiones relevantes en el uso de determinadas estrategias de comunicación, se pretende crear un normalizador de textos basado en reglas y diccionarios, que permita tanto la expansión como la corrección de los mensajes de los chats. De este modo, se amplía la tipología de textos que pueden ser tratados por los conversores de texto a habla, actualmente reducida exclusivamente a los textos normativos.
- La generación de voz sintética expresiva, adecuada al dominio de las conversaciones virtuales: se pretende hallar unos parámetros prosódicos relevantes que permitan la generación de voz expresiva con éxito, es decir, que las reproducciones de las diferentes emociones sean correctamente identificables. Para lograr este objetivo se parte de un análisis automático del corpus oral.

Los objetivos de la tesis se relacionan así con algunas de las líneas de investigación en el campo de la conversión texto-habla más activas en la actualidad, como son la generación de habla en dominios determinados, la síntesis expresiva y de emociones, o el tratamiento de texto no normativo de cara a la síntesis. Para ello se van a explorar aproximaciones basadas tanto en conocimiento lingüístico como basadas en corpus.

1.2. Contenido de la tesis

La tesis comienza con el capítulo 2, en el cual se expone el estado de la cuestión de los elementos tratados en esta investigación: i) la conversión de texto-habla; ii) las conversaciones virtuales, dominio bajo el cual se desarrolla toda la investigación; iii) la síntesis de habla de conversaciones virtuales.

El capítulo 3 trata la normalización de textos en el dominio de las conversaciones virtuales. Se presenta un análisis de los fenómenos lingüísticos que se han observado en las conversaciones virtuales, proceso imprescindible para diseñar y configurar un óptimo proceso de normalización-corrección de textos para un dominio tan particular como el de las conversaciones virtuales. En este epígrafe también se comenta la organización y el funcionamiento de TexAFon, herramienta con la cual se realiza la normalización de los textos, y las diferentes adaptaciones que se han ido desarrollando para poder tratar texto no normativo.

Seguidamente, en el capítulo 4, se presenta la investigación relacionada con la voz sintética expresiva: la creación del corpus, la selección del locutor profesional, las características de la grabación, la edición y revisión del corpus oral, el análisis de los parámetros prosódicos y los resultados obtenidos, y finalmente, los diferentes procesos realizados para generar y modelar la voz sintética con el objetivo de intentar transmitir las diferentes emociones analizadas.

En el capítulo 5 se detallan las evaluaciones realizadas tanto para el normalizador de textos como para la voz expresiva. Se realiza una

minuciosa explicación de cómo se ha llevado a cabo cada una de ellas, comentando los resultados obtenidos.

Finalmente, en el capítulo 6 se exponen las principales conclusiones, y las futuras líneas de investigación que se abren a consecuencia de las aportaciones de este trabajo.

Capítulo 2

ESTADO DE LA CUESTIÓN

En este apartado se revisan los dos temas centrales de esta tesis: los conversores de texto a habla basados en corpus, y las conversaciones virtuales, dominio en el cual se desarrolla esta investigación. La combinación de ambos presenta una doble problemática, ya que tratar con este tipo de discurso implica una modificación de la normalización de los textos, y la generación de habla expresiva, concretamente en la generación de voz sintética capaz de expresar ciertas emociones.

2.1. Conversión de texto en habla

El objetivo principal de los conversores de texto en habla (CTH) es conseguir una producción sonora de alta calidad, inteligible y natural, mediante un proceso sencillo, de cualquier texto digital. Es decir, se pretende conseguir una producción sonora de cualquier texto lo más parecida posible a la lectura realizada por un ser humano.

Actualmente, las voces sintetizadas que se producen mediante sistemas de CTH están dirigidas a unas aplicaciones más comerciales que requieren una voz de gran calidad y en la gran mayoría de los casos se han de realizar con una entonación neutra, como por ejemplo la atención al cliente de una empresa o el ofrecimiento de información sobre un determinado producto. Estas voces presentan una gran inteligibilidad, pero su naturalidad no está al mismo nivel. En los últimos años, los esfuerzos se han centrado particularmente en este sentido, en dotar a este tipo de voces de la capacidad expresiva del habla humana, capaz de aportar información verbal relevante para expresar la intencionalidad, la

actitud o el estado emocional del emisor. De este modo se ampliarían las aplicaciones posibles de estas voces a ámbitos más informales donde predomina un habla informal y expresiva.

A lo largo de la historia de la síntesis de voz han existido diferentes aproximaciones que se han empleado para generar voz artificial con un objetivo comercial. Sistemas como los de InfoVox (Aida-Zade et al., 2010) o ETI-Eloquence (Hertz et al., 1999) están basados en síntesis por formantes, en cambio CTH como los de Loquendo (Quazza et al., 2001), o Telefónica I+D (Armenta *et al.*, 2003) emplean técnicas de síntesis por corpus. Esta última aproximación es la más empleada por los CTH comerciales ya que la voz sintética resultante presenta una gran calidad, en cambio muestra ciertos inconvenientes como la necesidad de disponer siempre de un corpus oral lo suficientemente amplio de cada uno de los dominios en el que se desee aplicar la voz, o que el grado de naturalidad con el que se generan los enunciados no es equiparable al de inteligibilidad o calidad.

A continuación se presenta el funcionamiento de los CTH basados en síntesis por corpus, que es la aproximación en la que se basa el CTH de la empresa Cereproc (Aylett *et al.* 2006 ; Aylett *et al.*, 2007 ; Aylett *et al.* ; 2007 b; Garrido *et al.*, 2008), empleado en esta investigación.

2.1.1. Funcionamiento de los conversores de texto a habla basados en corpus

Para que los CTH basados en la síntesis por corpus puedan llevar a cabo la producción sonora de textos es necesario que se realicen dos procesos diferentes, como se puede observar en la Figura 1. La primera fase es la creación del locutor sintético. Éste es un proceso previo a la síntesis, se realiza sólo una vez, y es imprescindible para que pueda llevarse a cabo la generación de nuevos enunciados orales. El segundo proceso es el propio de la generación de enunciados mediante síntesis, que se realiza de forma automática y en tiempo real. Está formado por dos módulos: el procesamiento del lenguaje natural y la generación del habla sintética de salida.

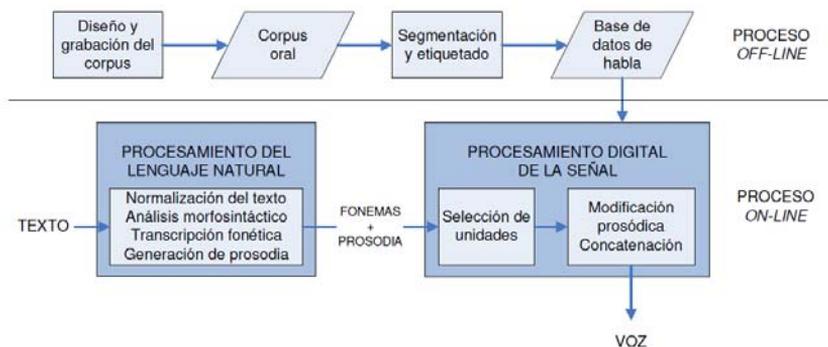


Figura 1: Diagrama de bloques de los procesos que forman parte de un sistema de CTH basado en concatenación de unidades (Iriando, 2008)

2.1.1.1. Creación del locutor sintético

El corpus oral con el que se realiza la voz final es uno de los aspectos fundamentales que hay que tener en cuenta para obtener una voz de alta calidad, inteligible y agradable. Es un proceso previo a la síntesis del habla.

Para obtener la base de datos necesaria para generar la voz sintética se han de realizar una serie de pasos previos:

- a) **Diseño y selección del corpus:** En el momento de diseñar y seleccionar los segmentos que formarán el corpus se ha de tener en cuenta dos aspectos: uno, el dominio para el cual se genera la voz, y dos, el tamaño del corpus.

El diseño del corpus ha de estar estrechamente ligado al marco de trabajo del CTH, con contenidos controlados y orientados a la aplicación deseada, ya que esto influirá notablemente en la calidad de la señal de audio (Black y Lenzo, 2000; Chu *et al.*, 2002).

El corpus debe procurar una cobertura fonética y prosódica suficiente para el dominio tratado, y para ello se suele emplear métodos automáticos como los algoritmos voraces o *greedy* (van Santen *et al.*, 1997), que aseguran que todos los fonemas de una lengua se encuentran en el corpus seleccionado al menos una vez.

- b) **Selección del locutor y grabación del corpus:** La selección del locutor es un aspecto fundamental para obtener una voz sintética de calidad. Para el óptimo desarrollo de la voz sintética, el locutor o locutora en cuestión ha de poseer ciertas cualidades dependiendo del dominio o dominios en que se enmarca la aplicación, como una buena voz y una buena dicción. En el caso de síntesis expresiva, además ha de poseer unas buenas dotes interpretativas para representar determinadas emociones de la manera más natural posible. La grabación del corpus se suele realizar en un estudio insonorizado para evitar posibles ruidos en la generación de la voz posterior.
- c) **Segmentación y etiquetado:** Una vez el corpus está grabado, es necesario segmentar y etiquetar, bien de manera automática o semiautomática, toda aquella información relevante para la posterior selección de unidades: inicio y final de los alófonos, su duración, intensidad, etc.

Estas tres fases son las que permiten crear la base de datos a la que el módulo de procesamiento de la señal recurrirá para seleccionar las unidades que necesita para formar su segmento fónico, y construir así la voz sintetizada.

2.1.1.2. Procesamiento del lenguaje natural

Mediante el procesado del texto se pretende extraer toda la información necesaria para la correcta generación del habla de

salida. Sería ideal poder obtener un análisis morfológico, sintáctico, semántico, o pragmático del texto para poder comprender en su totalidad todo el enunciado sin dar cabida a posibles errores interpretativos, y conseguir así la misma lectura del texto que realizaría un ser humano. A pesar de que esto sería lo ideal, actualmente la mayoría de CTH no realizan unos análisis tan detallados ya que el procesado de lenguaje natural no consigue un rendimiento de calidad en todos estos aspectos. En esta fase del procesamiento del lenguaje no solo se trata la expansión ortográfica de todos los elementos, sino también conocer los aspectos morfosintácticos más relevantes para la síntesis, realizar una correcta transcripción fonética del enunciado, o predecir a partir del propio texto la estructura prosódica que éste debe tener. A partir de esta información, el módulo de generación del habla sintética de salida podrá crear el enunciado deseado con el propósito de obtener la mayor calidad posible.

Son muchos los autores que han explicado en detalle las diferentes fases que se han de llevar a cabo en procesamiento del lenguaje natural referente a la síntesis del habla (Dutoit 1999, Llisterri *et al.* 2004). A continuación se presenta una breve explicación de cada uno de ellos.

- a) **Normalización:** Es el primer módulo que se aplica. En los textos pueden aparecer elementos como dígitos, siglas, abreviaturas, horas, etc., lo que se denomina palabras no estándar (NSWs, *non-standard words* Sproat, 2001) que deben ser tratadas de una manera especial para que el sistema las pueda expandir ortográficamente y posteriormente leer de forma correcta. Por ejemplo, si en el texto aparece ‘XXI cursa’ el normalizador debería identificar ‘XXI’ como un número romano y no como dos equis y una i; además, una vez que se ha identificado como un número romano, se ha de convertir a número arábigo y decidir si se va a expandir como un número ordinal (vigésimo primero) o cardinal (veintiuno), además, ha de tener en cuenta el género y número del sustantivo al que complementa, ya que no sería correcto que el módulo de normalización expandiera ‘veintiuno cursa’, ni ‘vigésimo

primero cursa’, la solución correcta a este caso concreto sería ‘vigésimo primera cursa’ aunque también se podría aceptar ‘veintiuna cursa’.

Normalmente, este módulo de normalización no corrige los posibles errores ortográficos o sintácticos que pueda presentar el texto. Sólo trata textos normativos y escritos correctamente. Cualquier texto que se aleje de esta normatividad podría presentar graves problemas de inteligibilidad al final del proceso, en la generación de la señal de audio.

- b) **Análisis morfosintáctico:** El siguiente paso es el análisis lingüístico del texto. Lo ideal sería poder analizar todos los textos morfológica, sintáctica, semántica y pragmáticamente para poder conseguir una comprensión e interpretación de lo que se está leyendo, y realizar así la misma lectura que haría un humano. A pesar de esto, la mayoría de los CTH solo realizan un análisis morfosintáctico.

El análisis morfológico consiste en etiquetar las palabras según su categoría gramatical. Estas etiquetas, aparte de la categoría, también pueden aportar tanto información sobre el género y número en el caso de los sustantivos y adjetivos, como el tiempo, modo, persona y el número en el caso de los verbos.

La utilización de estos procesadores morfológicos permite desambiguar las palabras homógrafas, y por tanto mejorar la prosodia, ya que ciertas palabras, según sea la categoría gramatical a la que pertenecen, llevan acento léxico o no¹, variando así las inflexiones tonales. Estos aspectos son muy importantes para poder comprender correctamente el texto.

Los procesos sintácticos identifican las unidades sintácticas y establecen relaciones de jerarquía entre ellas,

¹ Los determinantes, los pronombres átonos, las conjunciones, las preposiciones y los términos de tratamiento se consideran palabras átonas, y por tanto sin acento léxico.

contribuyendo así a detectar los límites prosódicos y facilitando la inserción de pausas no marcadas ortográficamente.

- c) **Transcripción fonética:** El texto es transcrito mediante un alfabeto fonético, normalmente en SAMPA (Speech Assessment Methods Phonetic Alphabet)², sin que ninguna persona intervenga en el proceso. Existen dos aproximaciones, una basada en reglas y otra que utiliza diccionarios. La elección de uno u otro método depende del grado de correspondencia entre la ortografía y la pronunciación de la lengua.

Para procesar lenguas como el español o el catalán normalmente se suelen utilizar transcritores basados en reglas, ya que existe un alto grado de correspondencia entre los grafemas y los fonemas. Estos sistemas suelen disponer de un pequeño diccionario que contemplan las excepciones a las reglas.

En el caso del inglés o el francés, en los que la pronunciación no es en muchos casos deducible de la ortografía, lo más aconsejable es utilizar diccionarios de unidades léxicas ya transcritas, y remitir a un conjunto de reglas para resolver los posibles cambios a consecuencia de la silabificación.

- d) **Asignación de aspectos prosódicos:** En la última etapa del procesamiento, a partir de unas reglas previamente definidas, se calculan los parámetros prosódicos como son la duración, la intensidad o la frecuencia fundamental (F0) para cada sonido de acuerdo con las particularidades propias de cada lengua, y se asignan etiquetas con la información

² Alfabeto fonético legible por ordenador mediante caracteres ASCII basado en el Alfabeto Fonético Internacional. Cada idioma tiene su propia tabla de representación, y ésta no es compatible con otros idiomas. Para solucionar este problema se ha creado el X-SAMPA, una tabla sin diferencias específicas entre lenguas. <http://www.phon.ucl.ac.uk/home/sampa/index.html>

referente a las modificaciones prosódicas que deben realizarse en las posiciones señaladas (Llisterri et al. 2004).

En el habla natural la duración de los segmentos depende de factores externos como el acento, la velocidad de elocución, los fonemas (o la unidad básica de la cual se parte -fonema, difonema, o sílaba, entre otros) que preceden y siguen al segmento en cuestión, etc. Existen dos tipos de modelos para predecir este factor, los modelos basados en reglas (Klatt, 1979) o los modelos estadísticos como los basados en redes neuronales (Campbell, 1990) o los basados en árboles de regresión y clasificación (van Santen, 1994).

La intensidad, al considerarse un factor de difícil control y no relevante para la síntesis neutra, no se tiene en consideración en muchos CTH. Aun así, existen algunos modelos para predecirla: Bartkova et al. (1993), Bagshaw (1998). En los últimos años se ha observado un creciente interés en incorporar y modificar este factor en la síntesis del habla gracias al interés despertado en incluir expresiones emocionales.

Finalmente, los movimientos melódicos o las curvas de entonación es el rasgo que recibe mayor atención (Llisterri et al., 2003), ya que su predicción puede evitar la monotonía y conseguir una mayor variabilidad en el habla sintetizada.

Existen varios modelos prosódicos, que dependen de la orientación teórica que se toma, el modelo de anotación empleado, de las necesidades del proyecto en cuestión, y del propio sistema de CTH. Estos métodos se pueden clasificar en dos grupos: cuantitativos en los que se encontraría el modelo de Fujisaki (Fujisaki & Nagashima, 1969) o el modelo Tilt (Taylor, 2000); y los cualitativos como los modelos de Secuencia de Tonos (ToBI, Tone and Break Indices) (Silverman, et al., 1992), el IPO ('t Hart et al., 1990) o el modelo INTSINT (Hirst et al., 2000).

2.1.1.3. Generación del habla sintética de salida

Existen diferentes técnicas para llevar a cabo la generación de la señal de voz, como por ejemplo las basadas en reglas (Klatt, 1987), en concatenación de unidades, o las que emplean Modelos Ocultos de Markov (HMM) (Aylett & Yamagashi, 2008). Como ya se ha comentado anteriormente, la mayoría de sistemas comerciales, y en concreto el CTH de Cereproc, son sistemas basados en síntesis por corpus, un subtipo de la síntesis por concatenación de unidades.

Estos sistemas crean la señal vocal gracias a la concatenación de una serie de unidades acústicas adecuadamente seleccionadas de entre las que constituyen la base de datos oral, o locutor sintético. Estas unidades elegidas no están determinadas a priori, sino que se seleccionan de una manera dinámica en función de las secuencias de fonemas que se desea generar, y de su contexto (Balestri *et al.* 1999), empleando algoritmos que tienen en cuenta tanto el coste de unidad, como el coste de concatenación. De esta manera se seleccionan las secuencias más largas que resultan idóneas para ser concatenadas con otras, y que correspondan a las partes de texto que se desea sintetizar. A través de un sencillo ejemplo será más fácil comprender este proceso. Si deseamos sintetizar el siguiente enunciado “Todos corrieron hacia la casa de madera”, y en el corpus oral disponemos de enunciados como:

- a) “Todos corrieron hacia los coches”,
- b) “La casa parecía abandonada”,
- c) “Todos corrieron sin mirar atrás” y
- d) “Siempre supo que algún día viviría en la casa de madera”

El conversor debe elegir los fragmentos que se corresponden, palabra por palabra, con el enunciado que se desea generar, y que a su vez sea el segmento más amplio posible. En este caso, debería

seleccionar el inicio de la frase a) “Todos corrieron hacia”, y el segmento de la oración d) correspondiente a “la casa de madera”, omitiendo los fragmentos “la casa” del enunciado b) y “todos” de la c), ya que las otras oraciones (a y d) disponen de unos fragmentos mayores al coincidir más palabras iguales entre el texto que se desea sintetizar y los enunciados que conforman el corpus.

Además de prestar atención al tamaño de los segmentos seleccionados, también se tiene especial cuidado, como comenta Llisterri *et al.* (2004) “en reducir al mínimo la manipulación de la señal acústica para no alterar su calidad y evitar discontinuidades en los puntos de concatenación, de modo que se obtenga una voz sintetizada con un buen nivel de naturalidad”.

Si las características prosódicas de los segmentos seleccionados son lo suficientemente parecidas a los pretendidos por la selección, la concatenación de las diferentes ondas es suficiente para obtener el enunciado oral deseado. Si por el contrario, los segmentos no se asemejan lo suficiente a la prosodia requerida, es necesario realizar modificaciones en algunos parámetros como la duración o la frecuencia fundamental. Para esta etapa de la generación del habla sintética, la gran mayoría de sistemas comerciales emplean técnicas TD-PSOLA, que mediante algoritmos puede modificar estos parámetros de forma eficiente (Charpentier *et al.* 1989, Dutoit *et al.* 1996).

2.2. Las conversaciones virtuales

Internet ha revolucionado el mundo de la comunicación y la información. Nos permite mantener el contacto con personas que se encuentran a miles de kilómetros de distancia, mantenernos informados en todo momento de lo que hacen, lo que piensan, lo que les preocupa, e incluso conocer gente nueva sin necesidad de salir de casa.

Son diversos los medios que nos permiten esto: los correos electrónicos, las videoconferencias, las conversaciones virtuales,

también denominadas chats, los blogs, y el gran número de redes sociales que han surgido en los últimos años, como por ejemplo Facebook, Twitter o Tuenti.

Este capítulo se centrará únicamente en las conversaciones virtuales, analizadas por diversos autores tanto en castellano como en catalán (Mayans 2000, 2002, 2002 b; Torres 1999; Yus 2001 y 2010; Pons 2002; López Quero 2003; Sanmartín 2007; Kallsen 2007; entre otros), proporcionando al lector una visión general de su funcionamiento y de sus peculiares características tanto discursivas como lingüísticas.

2.2.1. Aspectos técnicos: Breve introducción al funcionamiento de los chats

Las conversaciones virtuales o chats (término proveniente del inglés, que en español significa charlar) tienen su origen, como señala Torres (1999), en el año 1988, cuando el finlandés J. Oikarinen diseñó el primer sistema multichat en tiempo real. Pero no fue hasta 1991, con la guerra del Golfo, situación que requería un sistema de comunicación que transmitiera información segundo a segundo, cuando se empezó a popularizar.

El chat es una aplicación informática, un *software*, que haciendo uso de la conexión a Internet, permite la comunicación escrita de manera inmediata y simultánea entre múltiples personas aunque éstas se encuentren a kilómetros de distancia.

Existen distintos sistemas que permiten este tipo de comunicación, con unas características específicas según si están vinculados a un registro previo o no. Por un lado están los chats, como los que se pueden encontrar en www.terra.es, o www.ozu.es donde el usuario accede sin necesidad de estar registrado en la web; y por otro lado encontramos chats como Messenger, dependiente de Hotmail, o los proporcionados por servidores de correo electrónico como gmail o de redes sociales como Facebook, para los cuales es imprescindible estar registrado para poder acceder a ellos. El funcionamiento de ambos chats es muy similar, las diferencias se basan principalmente

en el modo de acceso, el diseño y las aplicaciones extras que ofrecen algunos sistemas, ya que el modo en el que se comunican los usuarios, lo que es la conversación virtual en sí, no difiere en gran medida según el tipo de chat. A pesar de esto, se ha decidido explicar muy brevemente el funcionamiento de ambos sistemas ya que en un inicio la mayoría de estudios sobre las conversaciones virtuales se centraron en los chats no vinculados a registros previos, y en los últimos años se ha observado un aumento de referencias a la mensajería instantánea, tan de moda actualmente.

2.2.1.1. Chats no vinculados a un registro previo

Para poder acceder a ellos, el usuario únicamente necesita introducir un *nickname* o apodo bajo el cual se relacionará con el resto de usuarios. Estos nicks tienen dos funciones bien diferenciadas: bien son utilizados para mantener la identidad real de la persona oculta, es decir, el resto de usuarios no puede deducir *a priori* a qué sexo pertenece, la edad que tiene, su nacionalidad, etc., o por el contrario son empleados como reclamo para otros usuarios proporcionando la información anteriormente comentada, sus inclinaciones sexuales, etc.

Normalmente, estos servidores contienen una serie de canales, salas o *chatrooms* en los que se encuentran los usuarios. Estos canales pueden ser variables en número, e incluso algunos programas permiten la creación de nuevas salas por parte de los usuarios. Un mismo usuario puede estar conectado en varias salas a la vez.

El listado de usuarios que participan en cada canal suele encontrarse a la derecha de la pantalla. Éstos pueden enviar mensajes al conjunto del grupo de personas si escriben en la parte inferior del empujador. El usuario, cada vez que escribe un enunciado y pulsa la tecla retorno, envía un mensaje a un servidor, y posteriormente, en cuestión de milésimas de segundo, es publicado en la pantalla del ordenador, permitiendo así su lectura por parte de todos los usuarios. La pantalla se renueva permanentemente con las aportaciones de los participantes (Crystal, 2002) exigiendo de este modo una rápida lectura y escritura de los mensajes.

A parte de poder conversar con un grupo de personas de manera pública, donde todos los usuarios del canal pueden leer e intervenir en ella, también se puede llevar a cabo conversaciones privadas con cada uno de los participantes, donde la conversación entablada única y exclusivamente puede ser leída y contestada por los miembros que la conforman, normalmente dos personas.

Los participantes en la conversación también tienen la posibilidad de escoger la tipografía y el color con los que escribir sus textos, pudiendo así llamar la atención sobre el resto de usuarios.

El diseño de estas aplicaciones puede variar de unas a otras, pero su funcionamiento será muy similar. La Figura 2 muestra la interfaz gráfica típica de las conversaciones virtuales públicas.



Figura 2 Conversación virtual pública extraída de Inforchat

2.2.1.2. Chats vinculados a un registro previo o mensajería instantánea

Tanto Messenger como el chat que proporciona Gmail son programas más completos ya que ofrecen una serie de posibilidades que los chats no vinculados a un registro previo, o el que proporciona la red social Facebook, no poseen, como son el uso de webcams o conversaciones de voz. Messenger también ofrece la posibilidad de compartir ficheros y de jugar a juegos online con otros usuarios.

Para acceder a estos servidores el usuario debe estar registrado en la web. Necesita una dirección de correo electrónico y una contraseña para entrar al sistema.

Cuando una persona va a entrar en un chat no vinculado a un registro previo no sabe con cuántas personas se va a encontrar en el canal seleccionado, en cambio en los chats vinculados a un registro previo el usuario siempre sabrá con cuántas personas puede encontrarse ya que es necesario ser agregado o formar parte de los contactos para poder entablar una conversación virtual. Estos contactos podrán estar en línea o no, en cambio en los chats no vinculados a un registro previo siempre estarán en línea.

Mientras que en los chats sin vinculación a registros previos el *nickname* juega un papel muy importante a la hora de atraer usuarios, en estos chats su funcionalidad ya no es trascendental porque los usuarios ya se conocen. Normalmente, estos programas no permiten utilizar un apodo, sino que éste ya viene prefijado por el nombre del usuario y no puede ser cambiado; sin embargo Messenger sí que permite el uso de *nicks*, y estos pueden tener una longitud mayor que los empleados en los chats no vinculados a registros. Los usuarios pueden escribir frases completas, refranes e incluso fragmentos de canciones.

En este tipo de chats las conversaciones son privadas, normalmente desarrolladas exclusivamente por dos usuarios, aunque en ocasiones este número puede verse incrementado al invitar a terceras personas.

La elección de la tipografía, el tamaño y el color solo es posible en Messenger, en otros chats como los de Gmail o Facebook no es posible realizar ningún cambio al respecto.

La Figura 3 recoge un ejemplo de una conversación privada desarrollada en Messenger entre dos usuarios en la que se puede apreciar los aspectos comentados.



Figura 3: Conversación privada entre Robert y Mandarinina INA a través del programa informático Messenger

2.2.2. Aspectos discursivos

El chat emplea un canal gráfico donde los usuarios continuamente escriben y leen los mensajes que van apareciendo en la pantalla del ordenador. A pesar de que el discurso escrito permite una planificación y una reelaboración del mismo, en los chats no se da este aspecto, la conversación se planifica sobre la marcha, como si

de una conversación oral cara a cara se tratara, y la rapidez y la fluidez de los mensajes son dos aspectos primordiales.

2.2.2.1. El chat, entre lo oral y lo escrito

Tradicionalmente, se ha considerado el registro escrito como un modo de expresión reflexivo, distante y formal, mientras que el registro oral es la antítesis de éste, siendo sus principales características la proximidad, la espontaneidad y la informalidad (Biber, 1986). A pesar de esta drástica distinción también se pueden encontrar registros orales más formales como por ejemplo una conferencia, y registros escritos más informales y espontáneos como las cartas personales.

La dicotomía existente entre lo oral y lo escrito queda difuminada, una vez más, en las conversaciones virtuales. Mayans (2000) afirma que “un medio como los canales de conversación de un chat no puede sino romper semejante dicotomía, creando lo que podríamos llamar, en terminología de Clifford Geertz (1993), un género confuso”.

Uno de los aspectos que más llama la atención sobre los chats es la combinación de la escritura con unos modos de verbalización propios del habla. Son muchos los autores (López Quero 2003; Pons 2002; Sanmartín 2007; Yus 2001) que han etiquetado las conversaciones virtuales como híbridos entre lo oral y lo escrito, incluso hay quienes han llamado al lenguaje de internet ‘habla escrita’ (Crystal, 2002) cuyas características son la informalidad, la espontaneidad, la ausencia de una estructura elaborada y una planificación del discurso sobre la marcha; en resumen, un registro que se asemeja bastante a la lengua hablada coloquial. Y es que a pesar de realizarse en un medio escrito, las conversaciones virtuales son precisamente eso, conversaciones llevadas a cabo entre diferentes personas. La Tabla 1, creada por Voiskounsky, y extraída de Yus (2010), resume tanto los aspectos orales como los escritos que posee la comunicación mediada por ordenador.

Oral	Escrito
<ul style="list-style-type: none"> ▪ Utilización de un estilo conversacional informal (uso de la primera persona verbal, fórmulas coloquiales de tratamiento, modismos recurrentes, etc.) ▪ Búsqueda de equivalentes textuales de los rasgos de la interacción conversacional, como las interrupciones. ▪ Mensajes cortos ▪ Tema conversacional íntimo y emocional, lo que conlleva el uso de grafías y símbolos típicos de la conversación virtual. 	<ul style="list-style-type: none"> ▪ El usuario controla la composición del texto sin que el destinatario pueda intervenir en ella. ▪ Sintaxis más complicada que en los enunciados orales. ▪ En el texto el remitente hace explícita la razón de la emisión de su mensaje. ▪ Posibilidad de revisión del texto y corrección de los errores antes de ser enviado al servidor.

Tabla 1: Aspectos orales y escritos en la comunicación mediada por ordenador

2.2.2.2. La conversación virtual *versus* la conversación real

La conversación virtual presenta muchas semejanzas con la conversación informal, definida por Tusón (1991) como “espontánea, poco planificada con anterioridad y por la existencia

de información implícita basada en el conocimiento compartido en el que la gente confía para construir el sentido a través de la cooperación entre los participantes”.

Según Torres (1999), las conversaciones virtuales comparten once de los trece puntos con los que Sacks, Schegloff i Jefferson (1974) caracterizaron a la conversación espontánea:

- i. Ambas interacciones son orales
- ii. Habla una persona a la vez
- iii. El orden de los turnos no es fijo
- iv. La duración del turno no es fija
- v. La duración de la conversación no se especifica previamente
- vi. Lo que dicen los hablantes no se especifica previamente
- vii. La distribución de los turnos es libre
- viii. Puede variar el número de hablantes
- ix. El discurso puede ser continuo o discontinuo
- x. Se utilizan diferentes unidades formales de construcción de los turnos (una palabra, una oración, etc.)
- xi. Hay mecanismos para reparar las transgresiones en la toma de turno.

Los dos puntos en los que divergen, según Torres (1999), son los referentes a:

- la transición entre turnos de habla, cuyo intervalo de tiempo no existe o es muy breve en el caso de la

conversación oral y en cambio en la conversación virtual el tiempo de espera para recibir una respuesta se puede alargar durante minutos;

- y los solapamientos entre mensajes de diferentes participantes, ya que su presencia es minoritaria en una conversación virtual con respecto a los que se producen en una conversación oral.

Salvador López Quero (2003) también propone una comparativa entre ambos tipos de conversaciones, que se expone en la Tabla 2.

CONVERSACIÓN VIRTUAL	CONVERSACIÓN REAL
Textual (escrita)	No textual (oral)
Menor información contextual no verbal (vocal y visual)	Información extralingüística
Carácter sincrónico débil	carácter sincrónico pleno
Pantalla: ancho de banda estrecho (scroll factor)	No pantalla
Más lenta	Menos lenta
No turnos de habla	Turnos de habla
Múltiples interacciones simultáneas	Menos interacciones
Más dispersión geográfica	No dispersión geográfica
Anónimo: apodos (nicks)	No anónimo: personajes reales

Tabla 2: Comparativa entre la conversación virtual y la conversación real propuesta por López Quero (2003)

De esta comparativa habría que comentar dos aspectos como el carácter sincrónico y los turnos de habla. Respecto al carácter sincrónico, Yus (2001) dice que las conversaciones virtuales “adolecen de una sucesividad temporal y espacial en la producción/recepción de los mensajes que merma sus posibilidades comunicativas”, a pesar de ello, los usuarios de los chats están presentes, están conectados al canal, mientras se desarrolla la conversación, por este motivo Crystal (2002), Patterson (1996) y Yus (2001) lo consideran sincrónico. En cambio, autores como Rintel y Pittan (1997) lo consideran asincrónico. López Quero (2003) considera las conversaciones virtuales sincrónicas débiles porque “aunque los internautas están presentes, la falta de sucesividad en la producción y recepción de los mensajes, resta, obviamente, sincronía a la interlocución virtual”.

En cuanto a los turnos de habla, López Quero (2003) considera que no se dan en las conversaciones virtuales, mientras que Torres (1999), Blanco (2002) o Yus (2001) sí creen que existen. Blanco define los turnos de habla del chat como “el espacio lineal, ocupado por la intervención de un solo hablante, que está limitado en un extremo por el nombre de la persona que está en posesión de la palabra y en el otro, por el final de párrafo”.

Muy frecuentemente, encontramos estos turnos fragmentados en diversas intervenciones, y es que en el chat no conviene escribir frases muy largas ya que como ocurre en las conversaciones orales, se puede perder el interés y la atención del receptor. Además, a los chats se les debe sumar el inconveniente de que las intervenciones se vayan moviendo en la pantalla del ordenador a consecuencia de la aparición de nuevos enunciados hasta desaparecer (factor *scroll*), cosa que favorece aún más la fragmentación de enunciados. Como consecuencia de este modo de proceder, a menudo un mensaje que un usuario ha enviado fragmentado en varias intervenciones aparece entrecortado por intervenciones de otros usuarios, como ocurre en el ejemplo de la Figura 4 en que la intervención de Vafalungo, en respuesta a un comentario realizado anteriormente, aparece entre medio de un mensaje fragmentado de Resurgir, interrumpiendo su discurso.

<Resurgir> al final que dia quieres que coja
<Resurgir> yo estoy por el 11 o 12 de agosto
[...]
<Resurgir> y me han dicho que ningun problema
<Vafalungo> ok, q efectivo
<Resurgir> y que es lo que se ve
<Resurgir> curso
<Resurgir> teorico
<Resurgir> **prueba en piscina**
<Vafalungo> **ok, pues pon el día q quieras**
<Vafalungo> **o no pongas día**
<Resurgir> **y salida al mar desde la orilla**

Figura 4: Ejemplo de una conversación privada entre dos usuarios en la que un turno de habla de Resurgir (en negrita) se ve interrumpido por un mensaje de Vafalungo (en verde), produciendo un solapamiento

2.2.2.3. El texto escrito oralizado

Como bien comenta Yus (2001), “el rasgo más esencial de la conversación virtual es la cualidad oral de los textos escritos que los usuarios de Internet intercambian entre sí. Esta oralidad responde al intento de comunicar no sólo contenidos, sino también matices esenciales de la comunicación oral como por ejemplo la actitud proposicional del emisor hacia el texto que ha tecleado en su pantalla”.

Los recursos contextuales, tanto visuales como vocales, que acompañan a un enunciado oral permiten que el propósito enunciativo que tenía el emisor sea entendido correctamente por el receptor del mensaje, aunque esto no significa que no se puedan dar malentendidos. Por ejemplo si le decimos a una persona que alguien no podrá asistir a la cita mientras le guiñamos el ojo, esta persona

ha de entender que lo que le estamos diciendo es mentira, que esa persona sí asistirá pero no quiere que otra tercera persona lo sepa, o por ejemplo si alguien nos dice que no tiene miedo con una voz entrecortada y temblorosa debemos entender que en realidad sí tiene miedo pero nos lo quiere ocultar.

Aspectos prosódicos como el tono, el énfasis, la entonación, etc. que están presentes en el habla, en una conversación oral son difíciles de representar de manera escrita. Transmitir estos parámetros no es una tarea imposible de realizar, en la literatura y en los cómics hay infinidad de ejemplos de textos deformados para comunicar oralidad, pero los usuarios del chat sí necesitan diferentes recursos tipográficos para conseguirla, como por ejemplo el uso de mayúsculas o repeticiones de grafías.

También presenta numerosos problemas el intentar plasmar el paralenguaje, tanto vocal como no vocal (kinesia). En una conversación oral cara a cara, si nuestro interlocutor está sentado y mueve una pierna continuamente y a un ritmo acelerado, podemos intuir que está nervioso por algún motivo; en cambio, en una conversación virtual esta información extralingüística es difícil de transmitir y en ocasiones se pierde. A pesar de ello, los usuarios del chat poseen diferentes recursos para resolver estos contratiempos, aunque en ocasiones pueden resultar deficientes, ya que no se puede representar la gran variedad tonal y vocal que ofrecen los órganos fonadores del ser humano ni transmitir todos los gestos que un locutor realiza.

En el epígrafe siguiente se muestra las características lingüísticas que emplean los usuarios tanto para dotar de oralidad el texto como para transmitir la información deseada.

2.2.3. Características lingüísticas

En estos últimos años el lenguaje empleado en las conversaciones virtuales ha sido juzgado por su negativa influencia en otros discursos textuales de carácter formal y por el poco respeto y degradación de la normativa ortográfica. En este capítulo no se va a

profundizar en estos juicios, sino que se va a abordar desde un punto de vista estrictamente lingüístico, aportando una detallada descripción de las diversas técnicas empleadas por parte de los usuarios.

2.2.3.1. Economía, rapidez y eficacia

Como ya se ha comentado anteriormente, las conversaciones virtuales se desarrollan con gran velocidad. Los usuarios han de leer los mensajes antes de que estos desaparezcan de la pantalla del ordenador, y contestar con la misma celeridad para mantener la atención de los lectores.

Para conseguir escribir, y decir, el máximo de cosas en el mínimo tiempo posible, los usuarios del chat utilizan diferentes recursos:

Se emplean abreviaciones y acrónimos. Algunos de ellos se convierten en convenciones y forman parte del vocabulario cotidiano del chat al ser empleados con gran asiduidad. Puede causar problemas en la correcta comprensión del mensaje a una persona que no posea una alfabetización en la codificación y descodificación de este tipo de textos.

Tb también

Tp tampoco

Qtal ¿qué tal?

LOL laughing out loud (reírse en voz alta)

Es muy frecuente observar la elisión de grafías. A parte de la elisión de la 'd' intervocálica en los participios (Llisterri, 2002), que refleja claramente un habla más informal y coloquial, también se puede

apreciar la elisión de todas, o de la mayoría de las vocales de una palabra.

Partío partido

Dnd donde

Mñn Mañana

La reducción a la mínima expresión formal de algunas palabras cuyo uso es muy frecuente es otra característica de estos textos.

Nas buenas

Bro brother

Chas muchas

Es muy común sustituir unas grafías por otras o por otros símbolos. Yus (2001) incluye este tipo de sustituciones en lo que denomina (Orto)grafía homofónica, y presenta dos tipos, las sustituciones léxicas y las sustituciones grafémicas. La primera consiste en sustituir una palabra o parte de ella por otra cuya pronunciación sea igual o muy próxima a la anterior, pero su longitud es menor. El español es menos propenso a este tipo de sustituciones que otras lenguas como la inglesa, pero a pesar de ello podemos encontrar palabras como

=mente igualmente

To2 todos

lika única

Las sustituciones grafemáticas son el resultado de sustituir unas grafías por otras. Las más frecuentes y sistemáticas se recogen en la Tabla 3.

GRAFÍA	EQUIVALENCIA
K	CA QU QUE
Q	QUE
X	CH
W	BU GU
Y	LL

Tabla 3: Sustituciones grafemáticas más frecuentes en el lenguaje 2.0³ comentadas por Morala (2001)

La velocidad con la que se escriben los mensajes repercute notablemente en el número elevado de faltas de ortografía que se cometen. Algunos de estos errores son involuntarios, producidos por fallos al teclear el mensaje o por el desconocimiento del usuario, pero otros en cambio se realizan intencionadamente, como por ejemplo la ausencia de acentuación de las palabras. A pesar de esta desviación de la normativa ortográfica, tal y como afirma Yus (2001) “los usuarios suelen aceptar estas faltas como un elemento consustancial a la interacción del chat, si bien, ocasionalmente alguno se queja en público de la reiterada laxitud ortográfica de

³ Lenguaje empleado en el mundo virtual: correos electrónicos, conversaciones virtuales, redes sociales, foros, etc.

alguno de ellos”, ya que lo importante es poder hablar, y a pesar de todas estas desviaciones, los usuarios son capaces de descifrar correctamente los mensajes.

2.2.3.2. Estrategias oralizadoras

Yus (2001), que a su vez sigue a Androutsopoulos (2000), presenta una clasificación de las estrategias que los usuarios emplean para dotar al texto de oralidad. Aquí expondremos esta misma clasificación con alguna modificación.

1. (Orto)grafía fonética: El usuario plasma textualmente el enunciado tal y como sería pronunciado oralmente. En español, esta estrategia suele usarse con palabras provenientes de otras lenguas.

Imeil (E-mail)

2. (Orto)grafía coloquial: Eliminación de parte de algunas palabras, normalmente átonas, debido a su pronunciación en la cadena hablada.

Ella acaba d’adquirir l’abilidad d’estudiar de noche (Ella acaba de adquirir la habilidad de estudiar de noche)

3. (Orto)grafía regional: Representación gráfica de la forma de pronunciar de los hablantes de una región. También se denomina dialecto visual (*eye dialect* en inglés).

Comío (comido)

4. (Orto)grafía prosódica: El usuario intenta representar el patrón melódico con el que pronunciaría el mensaje a través de la escritura. Para ello puede recurrir a diversas técnicas:
 - a. Repetición tanto de letras como de signos de puntuación

Holaaaaa!

- b. Usar mayúsculas para denotar enfado o que está gritando

He dicho q NO (He dicho que no)

- c. Empleo de signos de puntuación, normalmente ausentes

No h@gas caso a ningl xtranyo... Y sobre todo mir@ al cruz@r la calle!!! (No hagas caso a ningún extraño. ¡Y sobre todo mira al cruzar la calle!)

- 5. (Orto)grafía interlingüística: se plasman con grafía fonética los préstamos lingüísticos de otras lenguas, caracterizándolas con la ortografía de lengua de destino

Standard (Estándar)

Yus (2001) también propone una categoría más, la (orto)grafía homofónica, constituida por las sustituciones léxicas y grafemáticas explicadas en el epígrafe anterior. Consideramos que este recurso se adecúa mejor a la exigencia de rapidez de este tipo de conversaciones, más que a dotar de oralidad al texto.

2.2.3.3. Recursos paralingüísticos

Como ya se ha comentado anteriormente, los participantes de las conversaciones virtuales necesitan recurrir a una serie de procesos para dotar de expresividad al mensaje, para proporcionar la información vocal y visual que se pierde al no poder ver ni oír a la otra persona mientras se conversa.

En este apartado se abordarán los recursos paralingüísticos que emplean los usuarios del chat, como son las interjecciones, las

onomatopeyas, la acotación icónica y los emoticonos, haciendo más hincapié en este último por su gran importancia y uso.

a. Interjecciones y onomatopeyas

Se ha observado un gran uso tanto de interjecciones como de onomatopeyas muy habituales en la lengua oral. Con esto el usuario pretende potenciar la expresividad de la conversación. La influencia del inglés también es palpable en estos casos.

Wow

Oooohhhhhh

Uuuuuyyyyyy

Morala (2001) comenta que la onomatopeya que representa a la risa es la que se produce con más variedad. A continuación se muestra el amplio abanico de posibles representaciones de la risa que recoge en su artículo y que respalda esta afirmación.

Jajajaja hahahaha

Jejejeje hehehehe

Jijijijihihihihih

Jojojojo hohohoho

jujujuju

Jurjurjurjur

Juo juo juo juo juo

Wajajajajaja

Jua jua jua jua

Juas juas juas juaaaaaasssss

Tanto en el caso de las onomatopeyas como de las interjecciones, también se puede aplicar cualquier técnica empleada en la (orto)grafía prosódica.

b. La acotación icónica

La acotación icónica o emote (Mayans, 2002), consiste en usar textos que describen conductas no verbales, como si de una acotación teatral se tratara.

Existen dos variedades:

- Acotación icónica comentada: mediante una oración el usuario puede expresar conductas no verbales que informan del entorno físico no presente o reflejan estados de ánimo.

<<Prophet>> esta harto porque su ordenador se ha colgado

Kali bosteza

- Acotación icónica autónoma: la conducta no verbal se expresa mediante una o dos palabras y enmarcada entre asteriscos.

<Toro> Eso que dices es gracioso *sonrisa burlona*

c. Los emoticonos

Los emoticonos, también conocidos como *smileys*, son una serie de combinaciones de caracteres ASCII que representan, normalmente, el rostro de una persona, aunque también se pueden combinar para dibujar animales, objetos o personajes famosos. A continuación se puede observar una serie de ejemplos extraídos de Torres (1999) y del diccionario en línea www.diccionariosms.com, que muestra la gran creatividad de algunos usuarios para dibujar diferentes emoticonos:

:-(persona triste
@)>--->---	rosa
(_8(V)	Homer Simpson

Tradicionalmente, para poder observar estos dibujos se debe girar la cabeza 90° a la izquierda. Los dos puntos equivalen a los ojos, el guión suele ser la nariz y el carácter que se encuentra a la derecha representa la boca.

Actualmente, se emplean del mismo modo los *kaomoji* o *kaoni*, los emoticonos japoneses, que permiten su interpretación sin necesidad de girar la cabeza, o los emoticonos animados, que aparte de representar un rostro o un objeto, también pueden realizar ciertos movimientos como guiñar un ojo, mover una mano, etc.

:-) emoticono convencional utilizado para representar la alegría

^_^ kaomoji empleado para representar la alegría

Su uso es muy frecuente en las conversaciones virtuales, mucho más que en otras comunicaciones mediadas por ordenador como los correos electrónicos, los blogs o las diversas redes sociales. Incluso toda una intención o mensaje puede quedar resumido por el único uso de un emoticono. En algunos casos, como afirma J.M. Morala (2001), se han empezado a utilizar “de forma indiscriminada, sin más intencionalidad que la de incluir uno de esos emoticones junto a cada frase, tenga o no sentido en la conversación, con un resultado más colorido que expresivo.”

Su autoría es algo confusa, ya que hay quienes afirman que fueron inventados por Kevin Mackenzie en abril de 1979 cuando sugirió el empleo de estos símbolos para poder dotar de emotividad a los textos de correo electrónico, aunque muy a su pesar, su propuesta fue objeto de mofa al considerarse una idea descabellada. Sin embargo, Scout Fahlman, profesor en la universidad de Carnegie Mellon, en Pittsburg, se proclama como inventor de estos iconos en 1982 al enviar los primeros emoticonos, formados por los dos puntos, el guión y el paréntesis (:-) , :- ()⁴.

Independientemente de quién los inventara, en ambos casos surge con la intención de dotar al texto de la capacidad de expresar los estados de ánimo o la actitud con la que se ha de interpretar el mensaje del emisor de una manera simple, intentando transmitir los diferentes gestos faciales que tienen lugar en una conversación oral cara a cara, y ausentes en este medio comunicativo. El uso de cada uno de estos iconos es completamente voluntario por parte del usuario, los emplea con la finalidad de comunicar al emisor una intención determinada.

La lista de emoticonos que un usuario puede utilizar en una conversación es muy extensa, sobre todo gracias a los listados de emoticonos prediseñados que ofrecen algunas aplicaciones en las que el usuario puede adjuntar el emoticono con solo clicar encima de él. A pesar de ello, solo unos pocos son los que se emplean con más asiduidad (Morala 2001; Crystal 2002; Mayans 2002b).

⁴ El texto original que envió a la comunidad virtual puede visitarse en la siguiente dirección: http://research.microsoft.com/en-us/um/people/mbj/Smiley/Joke_Thread.html

Símbolo	Significado	Popularidad
%)-)	Borracho	1
	Estoy confuso	1
	Me tienes loco!	1
:-/	Estoy indeciso	2
	Medio serio, medio preocupado	1
	No me ralles	1
;o)	Contento	1
	Guiño	1
	Suerte	1

Tabla 4: Ejemplo extraído del diccionario online dictionariosms en el que se muestran los posibles significados que pueden tener los emoticonos %)-), :-/ y ;o)

Autores como Mayans (2002b) o Sanmartín (2007) reconocen que estos iconos, del mismo modo que ocurre en una conversación oral cara a cara, pueden ser polisémicos, ya que su interpretación no depende tanto del icono en sí, sino del contexto en el que se encuentra. Esta polisemia junto a las dificultades que presentan algunos emoticonos para ser interpretados correctamente han conllevado la realización de diversos inventarios, indicando el significado o significados que puede tener cada uno de ellos, e incluso se han creado diccionarios on-line como ‘diccionariosms’ (<http://www.diccionariosms.com/contenidos/>) en los que diferentes usuarios han incluido los emoticonos con el significado con el que ellos lo emplean. En la Tabla 4 se puede observar el significado que han otorgado a cada emoticono y el número de veces que diferentes usuarios le han asignado la misma equivalencia.

Varios autores han realizado una clasificación de estos símbolos según su funcionalidad (Wilson 1993; Marcoxia 2000); siendo el más completo es el que presenta Torres (1999) en el que clasifica los emoticonos en cinco puntos diferentes. Para facilitar la comprensión de esta clasificación se han extraído una serie de ejemplos del corpus de conversaciones virtuales creado para esta investigación.

1. Indicadores de los sentimientos y emociones del emisor: alegría, tristeza, sorpresa, etc.

[Pelblau] pero ahora estas de vacaciones paternas no???

[Luisiko] las cojo el lunes, así gano un día :-)

2. Indicadores de ironía y humor que ayudan a la interpretación del mensaje.

[Liadara] mmas interesante que el?

[Liadara] mmmmmmm

[Liadara] esto que

[Liadara] deja q piense ...

[Liadara] ¿nadie?

[david19bcn] as de diamantes es muy interesante

[david19bcn] **xDDD**

[princess_] dejalooooooooooooooooooooo

3. Indicadores de complicidad, manifestando el conocimiento de un código compartido y restringido a un grupo de usuarios determinado.

[Pelblau] ahora ja estamos viviendo juntos

[Luisiko]ya ya

[Luisiko] eso ya lo sabía

[Luisiko] por eso hay que dar otro pasao mas :-)

4. Preservadores de la imagen, los cuales permiten seguir las máximas de cortesía.

[La_Xula] pos io soi del real ijo

[pibe16] ais... xD si hoy os toca perder...

[La_Xula] jajaja

[La_Xula] malooooooooooooooooooooo

[pibe16] si no... a reventar el autobus del valencia

[La_Xula] espero k no

[pibe16] que se rompe los cuernos contra el barça y contra vosotros os deja paso **xD**

5. Amenazadores de la imagen, utilizados para remarcar el carácter de mofa o burla del mensaje.

[Pulg0sah] tu cara es un insulto

[Pulg0sah] para la humanidad

[AiToR_TuGa] es un atentado a los 5 sentidos

[AiToR_TuGa] el malnacido

[Pulg0sah] hahaha

[_KuuuuN_] habria k veros a vosotros

[_KuuuuN_] bixos

[_KuuuuN_] xDDDDDDDDDDDDDDDD

2.2.4. Similitudes con otros lenguajes

Lenguajes empleados en los mensajes de móvil, foros, algunas páginas web de opinión, como por ejemplo *Ciao!* (www.ciao.es), o redes sociales presentan similitudes con el empleado en las conversaciones virtuales.

Sobre todo se observan errores o desviaciones provocados por la rapidez con la que se escribe, como por ejemplo la sustitución sistemática de grafías, uso de abreviaturas o errores provocados al teclear. También se contemplan técnicas para representar la prosodia (Moré *et al.*, 2005) o el uso de emoticonos.

Esto ocurre principalmente cuando el discurso tiene un carácter más informal, cotidiano y próximo tanto al emisor como al receptor del mensaje.

Los usuarios de estos sistemas explotan al máximo las posibilidades que les proporciona el teclado, ofreciendo combinaciones de caracteres originales y creativos.

La Figura 5 muestra un ejemplo del lenguaje empleado en la red social Facebook.

prosódicos mediante la repetición de grafías y signos de puntuación, o mediante el empleo de mayúsculas, también se puede reflejar el habla coloquial eliminando algunas partes de los enunciados (Sanmartín, 2007), normalmente formas átonas como ‘me dormío’ en vez de ‘me he dormido’, escribiendo palabras del modo en el que serían pronunciadas en una zona concreta (‘de Cai’, en vez de ‘de Cádiz’), e incorporando interjecciones, cuyo uso es muy frecuente en las conversaciones orales.

Para evitar posibles equívocos en la interpretación de los mensajes a causa de la ausencia del canal visual y auditivo, los usuarios de las conversaciones virtuales emplean varias estrategias para intentar solventar este problema como por ejemplo el uso emoticonos. Estos últimos pueden representar personajes famosos u objetos, aunque son más empleados para representar los gestos faciales que una persona realizaría en una conversación oral, transmitiendo así ciertas emociones como la alegría (: -)), la tristeza (:-() o el enfado (X-(), manifestar la ironía o el humor del mensaje (: -p, :-)), la existencia de una cierta complicidad en la información que se comparte (; -)) o por el contrario, mostrar un carácter cortés o de burla (: -), XD).

Todos estos aspectos se han de tener en consideración a la hora de generar la voz sintética. Los CTH como ya se ha mencionado antes y como se verá en los siguientes epígrafes, sólo tratan texto normativo, por lo que la falta de normatividad al realizar desviaciones de la norma ortográfica es uno de los problemas que presenta este tipo de textos (Armenta *et al.* 2003b), pero no el único. Además de ser indispensable un módulo que corrija/normalice estos fenómenos para que el texto final sea comprensible, los CTH también deberían identificar la información emocional que los usuarios transmiten en sus mensajes a través de diversas estrategias oralizadoras, y dotar a la señal acústica de salida con los rasgos fónicos y prosódicos con los que fue escrito el mensaje. Es decir, ante un enunciado como ‘Holaaaaaa!!!!’ en el que se repiten tanto la grafía ‘a’ como el signo de cierre de interrogación, el módulo de procesamiento del lenguaje del CTH debería normalizar el enunciado eliminando los caracteres repetidos (¡Hola!), pero no debería obviar el resto de información que está transmitiendo el usuario, que se muestra muy alegre o muy sorprendido al saludar a alguien como indica el uso reiterado de una

misma grafía. Estos aspectos presentan dificultades para ser expresados oralmente mediante un CTH convencional, ya que estos no están diseñados para tener en consideración toda esta información, ni modelar la voz sintética en consecuencia.

En lo referente a las interjecciones, como ya se ha comentado más arriba, son unos elementos muy presentes en el habla oral coloquial, y también son muy empleados en las conversaciones virtuales, ya que aportan gran expresividad al mensaje oral. Este es el motivo por el cual algunos CTH, como el de la empresa Loquendo, han incluido el uso de interjecciones, ya que de este modo logran dotar a sus locutores sintéticos de una mayor expresividad.

2.3. Síntesis de habla de conversaciones virtuales

A partir de la descripción aportada tanto de la conversión de texto a habla basada en síntesis por corpus como de las conversaciones virtuales, se puede intuir las dificultades que presenta la generación de voz artificial en este dominio. Por un lado está la peculiar manera de escribir los mensajes, factor que hace que sea insuficiente el procesamiento del lenguaje natural que se desarrolla en estos sistemas, y por otro lado está la necesidad de generar voz expresiva acorde a este dominio, capaz de reproducir unos enunciados más expresivos, capaces de transmitir las emociones, los actos de habla, etc. emitidos por los usuarios. Mejorando estos dos aspectos se desea ampliar el ámbito de trabajo de los CTH comerciales que tenían hasta el momento, en los que sólo se trata texto normativo y se genera habla neutra. Concretamente, en esta investigación, se pretende trabajar en un dominio restringido como es el habla expresiva, delimitado únicamente al ámbito de las conversaciones virtuales, dominio que requiere una adaptación de los modelos actuales de procesado del lenguaje a consecuencia de las peculiaridades lingüísticas, y generar un habla sintética expresiva que simule los estados emocionales más representativos de este subdominio concreto.

A continuación se presentan las diferentes técnicas desarrolladas para tratar textos informales y no normativos (2.3.1.) y la problemática existente alrededor de la generación de habla expresiva mediante CTH basados en corpus (2.3.2.), que abarca desde conceptos tan básicos como es la propia definición de lo que se considera una emoción (2.3.2.1.) hasta los parámetros que permiten analizar adecuadamente y en su totalidad el habla expresiva (2.3.2.2.), o incluso la propia generación y modelado prosódico (2.3.2.3.).

2.3.1. La normalización de textos y los mensajes de las conversaciones virtuales

Como se ha podido observar por el epígrafe 2.2. Las conversaciones virtuales, del mismo modo que otros tipos de textos de la red, presentan unas características lingüísticas y gráficas (empleo de abreviaturas, uso de emoticonos, errores ortográficos provocados por la rapidez en la escritura y la falta de tiempo para releer y corregir el texto, etc.) que se alejan, en mayor o menor medida, del discurso escrito estándar y normativo. Por otro lado, un CTH, sistema con el que se pretende generar la voz sintética en esta tesis, sólo puede tratar texto correcto y normativo. Si se intentara reproducir enunciados de las conversaciones virtuales con los sistemas de normalización de texto que actualmente poseen los CTH, el resultado sería una reproducción ininteligible para el oyente, el cual necesitará recurrir al texto escrito para poder comprender el mensaje en cuestión. Es obvio, por tanto, que el procesamiento lingüístico que requieren estos tipos de textos es mayor del que presentan los normalizadores estándares que poseen los CTH actuales.

En los últimos años se ha observado un creciente interés por tratar de normalizar este tipo de textos, en concreto los mensajes de móvil (SMS), entendiendo ‘normalizar’ en un sentido mucho más amplio: reescribir un texto de las características de los SMS o las conversaciones virtuales, utilizando una ortografía más convencional para que pueda ser legible tanto para un humano

como para una máquina⁵ (Yvon, 2010). Para lograr este objetivo se han empleado varios métodos diferentes: los basados en reglas y los que emplean métodos estadísticos.

2.3.1.1. Basados en reglas

Estos sistemas actúan del mismo modo que un experto humano en un dominio de conocimiento concreto. Ante una situación o un problema determinado, gracias al conocimiento previo, realizan una serie de procesos para solucionarlo.

a. Corrección de textos

Esta es la técnica más explorada en el pasado (Kenneth 1991; Toutanova 2002). Se considera la normalización como una tarea de corrección, basándose en que el *token* de entrada presenta ‘ruido’, desviaciones, interferencias en la forma de la palabra o palabras correctas.

La corrección se realiza palabra por palabra, prestando especial atención a aquellas que no se encuentran en el diccionario de trabajo, el cual contiene todas las palabras consideradas correctas. Esta técnica se basa en la presunción de que la mayoría de las palabras del texto están correctamente escritas, y por lo tanto, si aparecen en el diccionario no se debe realizar ningún tipo de corrección sobre ellas. Las palabras que no se encuentran en el diccionario son modificadas mediante una serie de reglas. A menudo, se recurre a métodos estadísticos para seleccionar la corrección más idónea para cada una de las palabras tratadas, como

⁵ De ahora en adelante, cuando se haga referencia a la normalización de los mensajes de las conversaciones virtuales se hará en este sentido más amplio, el de corregir y expandir todos los elementos no estandarizados de un texto.

en el caso del corrector ortográfico desarrollado por Armenta *et al.* (2003b).

En este contexto, la mejor corrección o correcciones X de una palabra incorrecta Y se recupera del diccionario mediante la combinación de la probabilidad del contexto individual independiente de X con la probabilidad del modelo de error, que calcula la probabilidad de errores de escritura de Y para X . El componente clave es el modelo de error, que no sólo ha de captar los errores ortográficos producidos por grafías próximas, sino que también ha de tener presente las similitudes fonéticas (Yvon, 2010).

Este modelo probabilístico también puede basarse en los modelos ocultos de Markov (HMM) (Yvon, 2010).

Para poder diseñar un sistema de corrección, y controlar así la cobertura de correcciones es útil conocer tanto los errores que pueden aparecer en el tipo de texto con el que se trabaja como la frecuencia de aparición de los mismos (Oliva, 1997).

No existe una tipología universal de errores disponible para cualquier sistema que se desee construir, ya que dependiendo del idioma y del estudio que se realice pueden variar considerablemente tanto los tipos de errores identificados como los ejes sobre los que se define la clasificación (Díaz, 2005). La creación de una tipología de errores siempre ha de estar ligada al objetivo para el cual ha sido desarrollada.

Veronis (1988) distingue entre errores de competencia debidos a carencias cognitivas de los hablantes y errores de actuación o fortuitos. Este tipo de clasificación puede resultar útil en textos en los que haya habido una revisión previa del discurso, aunque en ocasiones es difícil identificar si un error ha sido fortuito o no en un texto revisado (Díaz, 2005).

Es importante que cada tipo de error esté perfectamente definido y no pueda confundirse con otros errores. Para ello se puede crear una jerarquía en la que en cada nivel se recojan unos casos determinados, evitando crear muchos subniveles para evitar posibles solapamientos.

-
1. Errores no intencionales
 - 1.1 Errores de actuación
 - 1.2 Errores de competencia
 - 1.2.1 Ortográficos
 - 1.2.1.1 Acentos
 - 1.2.1.2 Confusión fonema-grafema
 - 1.2.1.3 Composición y separación de símbolos
 - 1.2.1.4 Capitalización
 - 1.2.1.5 Errores en abreviaturas y acrónimos
 - 1.2.2 Léxicos
 - 1.2.2.1 Barbarismos
 - 1.2.2.2 Confusiones recurrentes
 - 1.2.2.3 Reproducción oral
 - 1.2.2.4 Adaptación de préstamos a la normativa del idioma
 - 1.2.3 Sintácticos
 - 1.2.4 Cohesión
 - 1.2.4.1 Errores de tiempo verbal
 - 1.2.4.2 Errores anafóricos
 - 1.2.4.3 Errores de puntuación
 2. Desviaciones intencionales
 - 2.1 Cambio de lengua
 - 2.1.1 Léxico
 - 2.1.1.1 Expresivo
 - 2.1.1.2 Terminológico
 - 2.1.2 Frasal
 - 2.2 Nuevas formas de expresividad textual
 - 2.2.1 Ortográficas
 - 2.2.1.1 Innovaciones ortográficas
 - 2.2.1.2 Falta de acentuación sistemática
 - 2.2.2 Léxicas
 - 2.2.2.1 Vocabulario propio de usuarios de internet
 - 2.2.2.2 Registro informal (parecido al lenguaje oral)
 - 2.2.2.3 Reproducción prosódica
 - 2.2.2.4 Vocabulario sms
 - 2.2.3 Visuales
 - 2.2.4 Pragmáticas
 - 2.2.5 Puntuación simplificada
 - 2.2.6 Sintaxis simplificada
 3. Terminología
 - 3.1 Terminología del dominio
 - 3.2 Terminología de la comunidad
-

Figura 6: Inventario de errores empleado en Moré et al. (2005) para catalogar los fenómenos observados en los correos electrónicos en castellano y catalán

Moré *et al.* (2005) presentan un inventario de errores para los correos electrónicos escritos en catalán y castellano en el que clasifican los errores en tres grandes grupos: desviaciones intencionales, errores no intencionados y terminología. Para cada uno de estos apartados se crearon diferentes subniveles para clasificar los diversos fenómenos observados en los textos. En la Figura 6 se puede ver esta clasificación.

En el caso de las conversaciones virtuales del español no hay una clasificación de errores publicada específica para este tipo de discurso.

Telefónica I+D creó un sistema para corregir mensajes cortos de móvil y correos electrónicos (Armenta *et al.* 2003b) basado en esta técnica. El resultado obtenido fue satisfactorio, ya que el porcentaje de palabras correctas del texto final era de un 87%.

2.3.1.2. Basados en métodos estadísticos

a. Traducción automática

El texto de estos mensajes es tratado como si estuviera escrito en un idioma distinto. Se considera el lenguaje empleado en los mensajes como si fuera el idioma extranjero, y el texto expandido y corregido el idioma de destino al cual se desea traducir el texto. Del mismo modo que en la traducción automática, los mismos textos escritos en lenguaje 2.0 y expandidos se almacenan, creando así un corpus paralelo con el que trabajar.

Emplear este método puede parecer exagerado (Choudhury *et al.*, 2007) ya que la relación existente entre ambas ‘lenguas’ es muy estrecha.

Este sistema puede ser fácil de implementar si se tiene una base de datos suficientemente amplia. Aún así, Yvon (2010) considera que este método presentaría ciertos problemas, ya que la creatividad léxica de los usuarios es infinita y variable, siempre están buscando

nuevas formas con las que comunicar la información deseada de la manera más original posible y en el menor tiempo, recurriendo a abreviaturas, acrónimos o combinaciones de grafías y dígitos o grafías y símbolos. El lenguaje empleado en este tipo de discursos es dinámico, lo que dificulta poder capturarlo en una base de datos más estática como la de los traductores automáticos, donde las frases escritas en un lenguaje 2.0 y las frases normalizadas se almacenan en la memoria en vez de ser modificadas.

Para el inglés, Aw et al. (2006) han realizado un sistema de corrección de SMS basándose en esta técnica. Usando el sistema de evaluación BLEU (Papineni *et al.* 2001) obtuvo un resultado de 0,81 para un corpus formado por 5.000 mensajes. En el caso del español no existen estudios de normalización de este tipo de textos empleando este método, aunque recientemente se ha observado el uso de esta técnica en la normalización estándar, la expansión de NSWs, para sistemas de CTH, concretamente para la normalización de números y abreviaturas (López-Ludeña et al., 2012).

b. Técnicas de reconocimiento de voz

A menudo el lenguaje 2.0 está más próximo a las representaciones fonéticas que a la forma ortográfica normativa. Los usuarios codifican las palabras mediante abreviaturas, sustituciones de unas grafías por otras creando nuevas formas léxicas que si son descodificadas correctamente reproducen la pronunciación fonética de aquello a lo que se quería referir la persona que escribió el mensaje. Es decir, para agilizar el proceso de escritura, reducir el tiempo y el número de caracteres empleados (muy importante en SMS o en Tweets que presentan un número limitado de caracteres), y dotar de oralidad a los textos, los usuarios de estos medios pueden emplear diferentes técnicas como:

- la sustitución de caracteres por otros cuya reproducción fónica sea igual o parecida a la que

sustituye, como por ejemplo el hecho de escribir ‘imxta’ en vez de ‘importa’;

- la sustitución de determinados caracteres por números como ‘to2’ por ‘todos’ o ‘lika’ por ‘única’;
- empleando contracciones léxicas propias del habla coloquial como ‘man’ en vez de ‘me han’ o ‘mestás’ en vez de ‘me estás’;
- escribiendo palabras extranjeras tal y como serían pronunciadas por una persona, por ejemplo ‘plis’ en vez de ‘please’ o ‘japi’ en vez de ‘happy’.

Gracias a estas características fonéticas del texto, se cree que el uso de técnicas empleadas en el reconocimiento automático del habla (ASR, *Automatic Speech Recognition*) puede ser útil para tratar correctamente estos textos.

Esta técnica consiste en crear un conjunto de secuencias fonéticas que representen todas las pronunciaciones posibles para cada una de las palabras que conforman el mensaje escrito. Posteriormente se realiza una conversión inversa, en la que las secuencias fonéticas se convierten en secuencias de palabras para acceder al diccionario de acceso, y poder seleccionar la mejor secuencia de palabra mediante el empleo de un modelo de lenguaje estadístico. En este caso, a diferencia de los sistemas de ASR empleados para reconocer el habla de una persona, la incertidumbre tanto de los fonemas como de los límites entre palabras son menores.

Kobus et al. (2008) emplean esta técnica para la normalización de SMS escritos en francés. Los resultados obtenidos aún son mejorables, ya que de un corpus formado por 2.998 mensajes de móvil (SMS), presenta un WER (*Word Error Rate*) de 19,79%, siendo las sustituciones, sobre todo las que hacen referencia a palabras cortas como ‘j <-> je’, las que recogen un mayor número de errores. Además, según Kobus *et al.* (2008), muchas de las palabras que aparecen en los mensajes son correctas, y durante el proceso de fonetización se introduce cierta ambigüedad que el

modelo de lenguaje no es capaz de compensar, transcribiendo así una palabra errónea.

2.3.1.3. Valoración de las técnicas

Las tres aproximaciones comentadas aquí son validas para tratar de normalizar un texto no normativo como es el de las conversaciones virtuales.

El método que realiza la normalización como si de una traducción automática se tratara puede obtener buenos resultados, sobre todo por la gran proximidad que hay entre las dos lenguas (lenguaje 2.0 y lenguaje estándar de un mismo idioma), pero para obtener una normalización de calidad se requiere de grandes corpus paralelos.

Además, el lenguaje 2.0 es dinámico, es decir, continuamente los usuarios de las conversaciones virtuales y de otros ámbitos 2.0 crean nuevas maneras de expresarse que denotan una gran creatividad y originalidad. Por ejemplo, la conjunción ‘aunque’ puede ser escrita de diversas formas (‘aunque’, ‘aunq’, ‘aunke’, ‘aunk’, ‘alke’, ‘alk’, o ‘alq’) y todas ellas son válidas en este tipo de discursos. Esta variabilidad en la forma de escribir esta palabra en concreto no surgió a un mismo tiempo, sino que probablemente primero se emplearan las abreviaturas ‘aunq’, ‘aunke’ y ‘aunk’, y posteriormente surgieran las otras. Este factor repercute negativamente en la correcta normalización de los textos, ya que se requerirá una continua ampliación del corpus de trabajo para recoger las nuevas formas de escritura.

La aproximación basada en el reconocimiento de voz presenta dos ventajas con respecto a la aproximación basada en traducción automática:

1. No requiere de grandes corpus para normalizar el texto.
2. El dinamismo presente en el lenguaje 2.0 no es un inconveniente, ya que las nuevas formas pueden ser tratadas a partir de las reglas creadas.

Gracias a que el lenguaje 2.0 a menudo se muestra muy próximo a la representación fonética, como por ejemplo ‘to2 <-> todos’ o ‘a3bt <-> atrévete’, mediante una serie de reglas, como las empleadas en el reconocimiento de la voz, se puede normalizar de forma satisfactoria este tipo de palabras.

El problema de esta técnica reside en las palabras correctas del texto. A pesar de que este tipo de lenguaje es anárquico respecto a las normas ortográficas, la mayoría de las palabras de un mensaje están correctamente escritas, y este sistema podría modificar estas palabras correctas generando nuevos errores en el texto final.

En el caso concreto de la normalización empleando técnicas de corrección basadas en reglas, a pesar de necesitar un corpus para analizar y observar los fenómenos lingüísticos presentes en él, no es imprescindible que el tamaño del mismo sea tan grande como el requerido en la aproximación basada en traducción automática.

Dos aspectos negativos que presenta este tipo de aproximación son la complejidad que a menudo muestran las propias reglas y los problemas de interdependencia; en ocasiones se necesita crear nuevas reglas para poder tratar correctamente las diferentes excepciones de un proceso, y también se ha de tener en cuenta el orden en el que se ejecutan las diferentes reglas, ya que un error en la cronología puede ocasionar errores importantes en la ejecución del programa e incluso afectar negativamente en la normalización del texto final. Aun así, gracias a un análisis previo del corpus, se puede crear una serie de reglas, lo suficientemente generales y flexibles para poder tratar la gran mayoría de las palabras incorrectas del texto, incluso aquellas variantes surgidas recientemente. Gracias a que se trabaja con un diccionario para saber si la palabra a tratar es correcta o no, se reduce considerablemente el número de falsos positivos y de nuevos errores en el texto final.

Esta última aproximación es la que presenta unos mejores resultados, y sus reglas son más duraderas en el tiempo, es decir, a diferencia de la aproximación basada en traducción automática, no requiere una ampliación de reglas o de palabras a incluir en los diferentes diccionarios tan frecuentemente.

2.3.2. La problemática de la generación de síntesis expresiva

En el campo de las tecnologías del habla, el término ‘emoción’ suele asociarse con la ‘expresividad’, y con todo aquello que tiene que ver con un habla menos formal y más natural. Normalmente, el habla neutra o formal, la empleada en los CTH, no suele transmitir ningún tipo de emotividad, en cambio el habla menos formal suele ser más expresiva porque implica con mayor frecuencia la transmisión de contenidos emocionales.

En un dominio como el de las conversaciones virtuales, donde imperan unas relaciones sociales más informales y relajadas, el lenguaje y el habla empleados por los usuarios son más expresivos. Las personas que entablan conversaciones suelen comentar sus estados de ánimo, sus opiniones respecto a ciertos temas alejándose de un entorno formal y neutro.

Al centrarse esta investigación en un dominio informal, la generación de voz sintética debe reflejar las características de dicho dominio, por lo que la voz generada ha de ser expresiva, con la capacidad de transmitir determinadas emociones. Resultaría extraño e incongruente que una voz pensada para reproducir mensajes de este dominio lo hiciera de una manera neutra, sin expresar ninguna emoción. Esto podría producir rechazo por parte de los usuarios, ya que el texto sonoro no sería representativo de este ámbito.

Hay que señalar que la expresividad del habla no se relaciona exclusivamente con la expresión de emociones, sino con otros aspectos, ligados por ejemplo al tipo de situación comunicativa, o aspectos pragmáticos. En esta tesis se ha abordado únicamente la realización fónica de las emociones y los actos de habla, aspectos que caracterizan más (aunque no de forma exclusiva) la situación comunicativa de las conversaciones virtuales.

En este epígrafe se presenta tanto la problemática existente entre la definición del propio término ‘emoción’, como la generación de voz sintética expresiva.

2.3.2.1. Emociones

Las emociones presentan una serie de problemas que van desde la propia definición de qué se considera una emoción, hasta la clasificación de las diferentes emociones. Los términos ‘emoción’ y ‘estado emocional’ se entremezclan con otros conceptos afines, como los estados de ánimo, las actitudes, los sentimientos, etc., haciendo difícil su definición y delimitación.

A continuación se abordan estos aspectos presentando el marco teórico existente en la actualidad.

a. Teoría de las emociones

El principal problema con el que se encuentra una persona que desea trabajar con emociones es la falta de una metodología aceptada para describirlas. Las diferentes definiciones que existen dependen de la disciplina que se emplea.

A continuación se exponen las cuatro perspectivas teóricas más influyentes de la tradición de la psicología de los últimos 125 años descritas por Cornelius (2000).

Perspectiva Darwiniana: Esta corriente se inicia con el libro de Darwin ‘The Expression of Emotion in Man and Animals’ publicado en 1872. Darwin defiende como idea principal de su teoría que las emociones son fenómenos desarrollados como funciones importantes de supervivencia, seleccionadas para resolver ciertas dificultades a las que la especie humana ha tenido que hacer frente. Por esta razón, los comportamientos son similares en todos los seres humanos e incluso en algunos animales.

Perspectiva Jamesiana: Fue inspirada por los escritos sobre la emoción de William James, y en concreto por el artículo ‘What is an emotion?’ publicado en 1884. Del mismo modo que Darwin, considera las emociones como adaptaciones al entorno relacionadas con la supervivencia, pero al contrario que éste, no se detiene en sus

manifestaciones sino que se centra en la explicación de la propia naturaleza de las emociones. Según esta perspectiva, siempre que hay una emoción, el ser humano experimenta una serie de cambios corporales como una respuesta automática y evolutiva del entorno. Para los seguidores de esta corriente es imposible la existencia de una emoción sin un cambio corporal.

Perspectiva Cognitivista: Esta perspectiva se basa en los estudios que Magda Arnold realizó sobre las emociones, aunque sus orígenes son mucho más antiguos, datan de los filósofos helenísticos. De las cuatro aproximaciones, ésta es la más dominante ya que se ha incorporado dentro de las otras tres. Para Arnold y sus seguidores, el pensamiento y las emociones son inseparables; las emociones son completamente dependientes de una evaluación que consiste en juzgar qué acontecimientos son buenos o malos. Mientras James se centraba en responder de qué modo tenía lugar la percepción de la emoción, cómo aparecían los cambios corporales, Arnold intenta responder qué tipo de percepción es. Cada emoción está relacionada con un patrón específico y diferente de evaluación.

Perspectiva Constructivista social: De las cuatro perspectivas mencionadas es la más joven, más diversa y la que provoca mayor controversia. Esta corriente ha estado muy ligada a la antropología y la sociología, aunque en los años 80 se aproximó a la psicología de la mano de James Averill y Rom Harré. Para los constructivistas sociales la cultura juega un papel primordial tanto en la organización de las emociones como en su variabilidad de niveles, ya que es la que determina el proceso de evaluación mediante las reglas sociales fijadas. Las emociones son consideradas productos culturales fijados por las reglas sociales, ya que son las reglas culturales específicas las que dictan cómo, cuándo y quién experimenta y expresa las emociones.

A partir de estas cuatro corrientes se puede definir a los estados emocionales como estados y procesos de los sujetos, en su mayoría relacionados con la afectividad, los cuales tienen una serie de manifestaciones externas a través de varios canales como son el propio mensaje lingüístico, la prosodia, la quinesia o determinadas alteraciones fisiológicas como la sudoración, temblores, etc.

Estos estados internos, a su vez se pueden diferenciar en dos tipos:

- **Estados emocionales:** relacionados principalmente con la afectividad (enfado, alegría, tristeza, etc.) y en segundo término con la racionalidad y el intelecto (ironía), ya sean intensos o no, y de mayor o menor duración en el tiempo (Cowie *et al.*, 2001).
- **Estados fisiológicos:** relacionados con reacciones fisiológicas o sensaciones, como por ejemplo el estrés o el cansancio.

Algunas de las manifestaciones internas que transmiten las personas cuando experimentan un estado emocional o fisiológico se reflejan en el mensaje lingüístico y en el habla.

b. Clasificación de las emociones

Al no haber una definición única y universal de qué se considera una emoción, tampoco existe un inventario objetivo y común. Esto, sumado a las diferencias en las connotaciones semánticas que se observan en algunos términos cuando son traducidos de una lengua a otra, dificulta todavía más la creación de una clasificación universal basada en criterios objetivos y cuantificables. Aun así, son varias las propuestas presentadas a lo largo de los años. En la Tabla 5 se recogen las más conocidas.

Si se presta atención a estas propuestas, se puede observar que en algunos casos se emplean términos diferentes para referirse a estados emocionales muy parecidos, como es el caso de *'fright'* y *'fear'*, o diferentes términos para referirse a distintos niveles de intensidad dentro de un mismo estado emocional, como por ejemplo *'irritation'* y *'hot anger'* propuesta por Banse y Scherer en 1996.

La mayoría de los listados existentes sólo hacen referencia a las emociones básicas (Ekman, 1999), también denominadas plenas

(Scherer, 1999) o primarias (Plutchik, 2001). Se consideran emociones básicas aquellas cuya forma es más intensa, y a partir de las cuales se generan todas las demás variaciones y combinaciones (emociones subyacentes o secundarias), aunque tampoco hay un criterio único para definir las ni un número determinado de emociones.

A pesar de los inventarios mostrados en la Tabla 5, la mayoría de las propuestas que se han desarrollado en la síntesis de habla recogen menos de 10 emociones, siendo el conjunto más empleado el denominado ‘The Big Six’ (Cornelius, 2000) que incluye las seis emociones básicas: la alegría, el asco, el enfado, el miedo, la sorpresa y la tristeza.

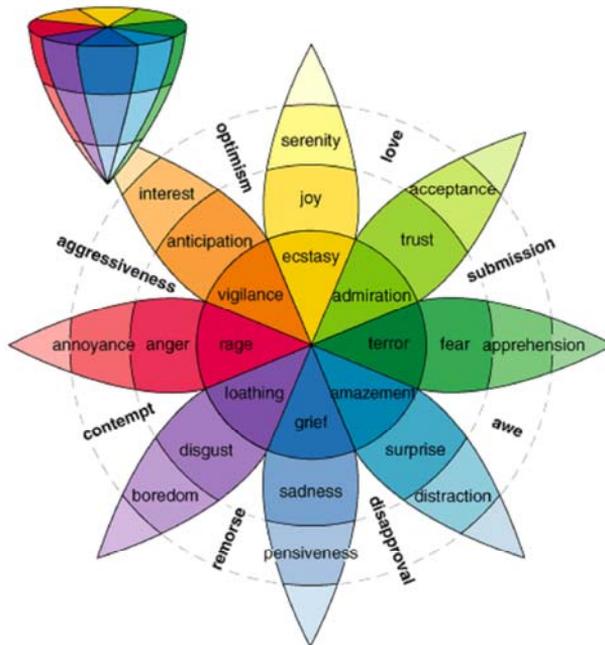


Figura 7: Modelo circunflejo tridimensional de Plutchik (2001)

<i>Lazarus (1999)</i>	<i>Ekman (1999)</i>	<i>Buck (1999)</i>	<i>Lewis-Haviland (1993)</i>	<i>Banse-Scherer (1996)</i>	<i>Cowie et. Al (1999)</i>
Anger	Anger	Anger	Anger/Hostility	Rage / Hot anger Irritation / Cold anger	Angry
Fright	Fear	Fear	Fear	Fear / Terror	Afraid
Sadness	Sadnes / Distress	Sadness	Sadness	Sadness / Dejection Grief / Desperation	Sad
Anxiety		Anxiety	Anxiety	Worry / Anxiety	Worried
Happiness	Sensory Pleasure	Happiness	Happiness	Happiness Elation (Enjoy)	Happy
	Amusement		Humor		Amused
	Satisfaction				Pleased

	Contentment	Interested			Content
		Curious			Interested
		Surprised			
	Excitement				Excited
		Bored		Boredom / Indifference	Bored
		Burn out			Relaxed
Disgust	Disgust	Disgust	Disgust		
	Contempt	Scorn		Contempt / Scorn	
Pride	Pride	Pride	Pride		
		Arrogance			

Jealousy		Jealousy			
Envy		Envy			
Shame	Shame	Shame	Shame	Shame / Guilty	
Guilt	Guilt	Guilt	Guilt		
	Embarrassment		Embarrassment		Disappointed
Relief	Relief				
Hope					Confident
Love			Love		Loving
Compassion		Pity			Affectionate

Aesthetic		Moral rapture Moral indignation			
------------------	--	------------------------------------	--	--	--

Tabla 5: Inventarios de emociones básicas propuestos por varios autores (Cowie *et al.*, 2003)

También se ha observado cierto grado de confusión en determinadas emociones, mientras que en otras, al ser completamente opuestas no existe tal equivocación. Plutchik (2001), siguiendo la tradición de investigadores precedentes como por ejemplo Schlosberg (1941), representa mediante una estructura circular su propuesta de emociones básicas. Tal y como se puede observar en la Figura 7, encontramos en el centro las 8 emociones que Plutchik considera básicas. Cada una de estas emociones tiene una escala de intensidad representada verticalmente. Las emociones con mayores similitudes se encuentran próximas entre sí, y las emociones que presentan mayores diferencias están situadas en el lado opuesto (separadas 180 grados). Las emociones situadas en los espacios en blanco son emociones que combinan dos emociones básicas.

Además de esta clasificación de las emociones partiendo de la intensidad con que se muestran, existen dos parámetros más que permiten tal categorización, por un lado teniendo en cuenta la evaluación o valencia, es decir, si la emoción implica una evaluación positiva o negativa sobre una cosa, persona o hecho, y la activación, que indica si la emoción implica una disposición a actuar por parte de la persona que la siente (Whissel, 1989).

Estas propuestas son muy útiles para dominios muy determinados, pero son insuficientes para poder describir el conjunto de emociones que pueden darse en una situación comunicativa no controlada como el habla real. En estas situaciones, los contenidos emocionales que transmite una persona pueden aparecer entremezclados con otros, y ser difíciles de identificar. Un ejemplo de esto es el fragmento que se muestra a continuación extraído de la novela 'El código Da Vinci' de Dan Brown (2003).

¡Están todos muertos! – Sor Sandrine gritó al auricular del teléfono de su residencia. Estaba dejando un mensaje en el contestador automático-. ***¡Por favor, que alguien responda!*** ***¡Están todos muertos!***

Los tres primeros números de la lista habían arrojado resultados terribles —una viuda histérica, un detective investigando en plena noche en el lugar de un crimen y un lacónico cura que consolaba a una familia

destrozada. Los tres contactos estaban muertos. Y ahora, al llamar al cuarto y último número —el que en teoría no debía marcar a menos que no pudiera contactar con los tres anteriores—, le salía un contestador.

Los enunciados marcados en negrita presentan emociones entremezcladas, se puede observar cierto grado de miedo, impaciencia ya que Sor Sandrine ha llamado con anterioridad a tres personas más sin haber podido hablar con el destinatario que deseaba, y también se observa desánimo, ya que es la cuarta llamada realizada y nuevamente no es atendida, le salta el contestador automático. La identificación de las emociones en el texto depende en gran medida de cada individuo, quizá otras personas perciban otras emociones a las comentadas aquí.

Towards a standard list of emotion words (combining LIMSI / GERG / QUB lists)

anger	amusement	courage
contempt	delight	hope
disgust	elation	pride
annoyance	excitement	satisfaction
irritation	happiness	trust
anxiety	joy	doubt
embarrassment	pleasure	envy
fear	calm	frustration
helplessness	content	guilt
powerlessness	serene	shame
worry	relaxed	tension
boredom	relieved	stress
disappointment	affection	shock
hurt	empathy	surprise
sadness	friendliness	interest
despair	love	politeness

18

Figura 8: Inventario de emociones desarrollado en el proyecto HUMAINE

En un intento más de proporcionar una clasificación más amplia de las emociones, Douglas-Cowie *et al.* (2006), en el marco del proyecto HUMAINE proponen un nuevo inventario que contiene 48 etiquetas diferentes, y que se recoge en la Figura 8. Como ya ocurría en clasificaciones anteriores, estados emocionales muy próximos aparecen bajo dos etiquetas diferentes como es el caso de ‘happiness’ y ‘joy’ o ‘calm’ y ‘serene’. Tampoco existe una distinción entre estados emocionales y estados fisiológicos.

Positive	Neutral	Negative
Affection	Surprise	Mockery
Joy	Indifference	Irony
Fun		Boredom
Trust		Doubt
Excitement		Distrust
Interest		Dejection
Complicity		Disappointment
Pride		Resignation
Relief		Worry
Compassion		Envy
Admiration		Disapproval
		Nostalgia
		Shame
		Guilt
		Sadness
		Disgust
		Impatience
		Anger
		Impotence
		Fear

Figura 9: Etiquetas de los estados emocionales presentados en I3Media

Positive	Neutral	Negative
Pleasure	Sleepiness	Fatigue
Relaxation	Agitation	Pain
		Sickness
		Cold
		Heat

Figura 10: Etiquetas de los estados fisiológicos presentados en I3Media

En el proyecto I3Media (Garrido *et al.*, 2012 b) se propone un inventario más amplio que contiene 126 etiquetas distribuidas de la siguiente manera:

- 33 estados emocionales
- 9 estados fisiológicos
- niveles de intensidad para cada etiqueta, en la que el nivel 1 equivale a un nivel bajo, el 2 a un nivel medio, y el 3 a una intensidad alta.

Cada emoción es combinada con cada uno de los diferentes niveles de intensidad, obteniendo así etiquetas como ‘anger_1’, ‘anger_2’ o ‘anger_3’. Tanto los estados emocionales como los estados fisiológicos son clasificados según su valencia en positivos, negativos y neutros. Además, también cuenta con una clasificación de elementos paralingüísticos, elementos que también pueden ayudar en la transmisión de expresividad: 15 elementos paralingüísticos comunes a las tres lenguas en las que se enmarca el estudio, como por ejemplo la risa, el lloro, besos, etc., y una serie de interjecciones que varía en número según el idioma (36 para el español, 40 para catalán y 46 para el inglés). En la Figura 9 y la Figura 10 se puede observar la clasificación de estados emocionales y fisiológicos respectivamente.

En las conversaciones virtuales puede aparecer cualquier tipo de estado emocional o fisiológico, e incluso algunos enunciados

pueden reflejar la combinación de dos o más emociones, aspecto muy frecuente en situaciones comunicativas reales. Por este motivo, el inventario de las seis emociones básicas tan habitual en la síntesis de habla expresiva es insuficiente para reflejar todos los estados emocionales y fisiológicos que tienen lugar en una conversación real. Por lo tanto, cuanto más amplio sea el inventario que se emplee, más detallado y más representativo del dominio será el corpus oral y por tanto la nueva voz sintética.

Para esta investigación se ha decidido emplear el listado del proyecto de I3Media como base para esta investigación, ya que es uno de los que contiene un mayor número de etiquetas emocionales y de elementos paralingüísticos bien definidos y delimitados, los cuales, a diferencia de otros inventarios como por ejemplo el de HUMAINE, no generan posibles confusiones entre las diferentes etiquetas.

2.3.2.2. Parámetros del habla relacionados con la emotividad

Si mantenemos una conversación telefónica con alguien podemos intuir si está contento, triste, enfadado, nervioso, etc. sólo escuchándole. Por lo tanto, está claro que un hablante puede transmitir una serie de emociones empleando únicamente su voz. Para determinar cómo la voz puede llegar a transmitir emotividad se han realizado diversos estudios que demuestran que cada emoción posee unas características acústicas determinadas que hacen que un oyente pueda llegar a reconocerlas (Martínez y Rojas 2011, entre otros).

Entre los diversos estudios sobre la correspondencia entre el habla y las emociones, destaca el realizado por Murray y Arnott (1993), que presenta un resumen de los trabajos más relevantes en este ámbito. A través de esta recopilación se puede observar que la gran mayoría de estos estudios coincidían en los efectos vocales de algunas emociones (véase Figura 11).

	Miedo	Alegría	Tristeza	Enfado	Asco
Velocidad del habla	Ligeramente más rápida	Más rápida o más lenta	Ligeramente más lenta	Mucho más rápida	Mucho más lenta
Promedio de F ₀	Mucho más alta	Más alta	Ligeramente más baja	Mucho más alta	Mucho más baja
Rango de F ₀	Más amplio	Más amplio	Ligeramente más estrecho	Más amplio	Ligeramente más amplio
Cualidad de la voz	Jadeante	Estrepitosa	Resonante	Sonoridad irregular	Ruidosa
Cambios de F ₀	Abruptos en sílabas tónicas	Suaves inflexiones ascendentes	Inflexiones descendentes	Normal	Amplios en inflexiones descendentes finales
Articulación	Tensa	Normal	Arrastrada	Precisa	Normal

Figura 11: Resumen de los efectos vocales que tienen cinco emociones básicas en el habla, traducido de Murray y Arnott (1993) y extraído de Iriondo (2008)

La cuantificación de estos parámetros expuestos por Murray y Arnott (1993) resulta confusa. Es necesaria una mayor precisión en la cuantificación de estos parámetros para poder obtener unos modelos acústicos concluyentes. Cowie *et al.* (2001) realiza un estudio más amplio al analizar 14 estados emocionales diferentes los cuales son descritos según sus características acústicas, su contorno melódico, el tono, la cualidad de la voz y una quinta categoría denominada ‘otros’.

Los parámetros más analizados en el estudio de las emociones son la frecuencia fundamental, tanto su rango como su media, la intensidad, y la duración, ya que varían según la emoción que se desea transmitir (Blondet 2006; Llisterri *et al.* 2004b)

Estudios anteriores (Montero 2003, Iriondo 2007) demuestran que estos parámetros son significativos en la realización de enunciados emotivos.

A continuación se comentan las propiedades acústicas de los parámetros prosódicos seleccionados en esta investigación para analizar y generar posteriormente la voz sintética expresiva.

a. Propiedades relacionadas con la intensidad

Es la unidad con la que se mide la energía de la señal acústica. Normalmente los CTH no suelen ajustar y normalizar este parámetro para que no produzca distorsiones en la concatenación de unidades.

A partir del interés creciente en incorporar emociones en la producción de habla sintética, este aspecto del habla ha ido tomando protagonismo, convirtiéndose en un parámetro importante ya que la diferencia de energía en frecuencias altas o bajas da lugar a diferentes emociones (Llisterri *et al.*, 2004).

b. Propiedades relacionadas con la duración

- 1) **Velocidad de elocución:** Está relacionado con la rapidez o lentitud con la que una persona lee un texto. Para calcular este factor es necesario determinar cuál va a ser la unidad de partida, y la unidad de tiempo. Tradicionalmente se ha empleado para ello el segmento (Riedi, 1998), pero éste puede variar y emplearse otras unidades como el difonema (O'Shaughnessy *et al.*, 1988) o la sílaba (Campbell, 1992). Este cálculo presenta ciertas dificultades de medición, ya que la duración de un fonema determinado está condicionada al contexto en el que se da lugar.

En el caso de la expresión de emociones se ha observado que la tristeza presenta una menor velocidad de elocución, y en cambio la alegría muestra una duración menor (Montero, 2003), es decir, las personas cuando transmiten un enunciado alegre lo realizan de una forma más rápida.

- 2) **Duración de la vocal prepausal:** Éste es un parámetro que no ha sido tenido en cuenta en estudios anteriores sobre la expresión de las emociones, pero parece ser relevante en ciertas emociones como son la sorpresa (¿Cómoooo?) o la alegría (¡Qué bieeeeeen!!!).

c. Frecuencia fundamental (F0) y entonación

La F0 es la onda simple de frecuencia más baja entre las que forman una onda sonora compleja periódica, y se corresponde con la frecuencia de vibración de las cuerdas vocales. Se suele medir en hercios (Hz).

La entonación es la sensación perceptiva que provocan las variaciones de F0, intensidad y duración que se dan en una oración, aportando con ello información sobre la modalidad oracional del enunciado (enunciativa, interrogativas, exclamativas) o la transmisión de información emocional (Montero 2003).

Tanto trabajos dedicados a la identificación de parámetros prosódicos relevantes en la generación de habla sintética emotiva (Montero 2003; Iriondo 2007) como estudios fonéticos (Garrido 2011) respaldan la idea de que hay dos tipos de parámetros relevantes: los que afectan a todo el enunciado (parámetros globales) y los que tienen lugar a un nivel más localizado, en una o varias sílabas, como son los patrones melódicos locales, constituidos por los patrones acentuales presentes al inicio e interior de un enunciado, y los patrones entonativos que tienen lugar en posición final y final de enunciado.

Respecto a los parámetros globales, diferentes estudios (Montero 2003, Garrido 2011) muestran que la mayoría de las emociones presentan diferencias tanto en el rango como en la altura global de F0 con respecto a los enunciados neutros.

En lo referente a los patrones locales, la propuesta de Garrido (2001) toma el grupo acentual (GA) como ámbito natural de estos patrones en español. El GA está formado por una sílaba tónica y todas las sílabas átonas que la preceden hasta llegar a la siguiente sílaba tónica o al final del enunciado. Esta unidad también ha sido empleada en la descripción de los contornos melódicos de otras lenguas (Thorsen 1978) y muestra similitudes con otras unidades empleadas en la fonología prosódica como es el 'pie' (Nespor y Vogel 1986) o el sintagma acentual (Jun y Fougeron 2000).

En el caso específico de los patrones entonativos (los que aparecen en posición final de un grupo entonativo), estos se pueden clasificar en tres tipos diferentes según la dirección del movimiento tonal

(Navarro 1944, Garrido 2001): ascendentes, descendentes, y ascendente-descendente o circunflejos.

Los tonemas ascendentes suelen asociarse con oraciones interrogativas absolutas y en determinadas exclamativas, mientras que los tonemas descendentes, en posición final de enunciado, se suelen relacionar con oraciones enunciativas o interrogativas parciales (Quilis 1993). Los tonemas circunflejos parecen estar asociados con los enunciados exclamativos, y de índole expresiva (Navarro 1944, Quilis 1993).

Existen dos modelos principales de análisis de contornos de F0, que a grandes rasgos se corresponden con los dos puntos de vista dentro de la lingüística: la fonología y la fonética:

1) Modelos fonológicos:

- a. **Modelo ToBI (Tone and Break Indices):** es un sistema de transcripción (Silverman *et al.*, 1992) originado a partir del modelo fonológico propuesto en 1980 por Pierrehumbert para el inglés de Estados Unidos (Pierrehumbert, 1980).

Se originó como un modelo de análisis fonológico que posteriormente fue adaptado para su uso en CTH.

Analiza cada contorno de F0 como una serie de eventos de pitch y una serie de índices de separación entre palabras.

2) Modelos acústicos o fonéticos:

- a. **Modelo del IPO:** Se trata de un modelo desarrollado por l'Institute for Perception Research y se basa en la estilización y estandarización de contornos (Montero, 2003). El modelo Garrido (Garrido, 1996), empleado en esta investigación, se basa en esta aproximación.

- b. **Modelo INTSINT:** Fue desarrollado por el Centre National de la Recherche Scientifique de la Université de Provence (Hirst *et al.*, 2000). Mediante una serie de símbolos se transcribe la actividad de la F0.

2.3.2.3. Modelado de las emociones para la síntesis de voz expresiva

En los últimos años se ha observado un aumento en el interés de introducir conocimientos prosódicos en la generación de la síntesis para poder dotarla de mayor naturalidad y expresividad. Y es que mediante la prosodia (la entonación, el ritmo o la intensidad) los seres humanos pueden expresar emociones, enfatizar ciertos elementos del discurso, etc.

El diseño e implementación de sistemas de síntesis de habla con emociones comenzó a finales de los años 80 con el Affect Editor diseñado por Janet Cahn (1989) y Hamlet de Iain Murray y Arnott (1995). Ambos sistemas parten de la síntesis por formantes y emplean el sintetizador DECtalk.

Affect Editor experimenta con los parámetros prosódicos típicos (nivel medio de F0, pendiente de la curva de F0, velocidad media de locución, etc.), con algunos parámetros de fuente (fricción) y un parámetro articulatorio que permite una mejor precisión de la articulación de los fonemas.

Hamlet dispone de un proceso preacústico que con la ayuda de un nuevo sistema de reglas específicas para la síntesis de emociones, permite incrementar y disminuir el número de parámetros prosódicos y la calidad del habla, y de un proceso acústico en el que se modifica la entonación y las duraciones por defecto.

En el caso del español hay pocos trabajos documentados de análisis y síntesis de emociones, aunque en los últimos años se ha observado un aumento. Encontramos la tesis doctoral de Montero (2003), la investigación de Iriondo *et al.* (2004) que además del castellano también se centra en el catalán, la investigación de Barra-Chicote *et*

al. (2010) sobre las emociones del español, y el proyecto I3Media (Garrido et al., 2012 b) para el castellano, catalán e inglés.

A continuación se presentan las diferentes aproximaciones empleadas para la generación de síntesis expresiva.

a. Síntesis paramétrica de la prosodia

Tradicionalmente se ha denominado síntesis paramétrica a la síntesis del habla basada en modelos de fuente y filtro. Actualmente existe una controversia en qué se considera síntesis paramétrica.

Algunos autores (Fordyce 1998) definen la síntesis paramétrica como aquella que selecciona unos parámetros concretos y los modifica para obtener la voz sintética deseada.

Otros investigadores consideran esta aproximación como un modelo estadístico (Shi *et al.* 2004, Holm *et al.* 2000), como las basadas en Modelos Ocultos de Markov (Black *et al.* 2007, King 2011).

En esta investigación se va a considerar síntesis paramétrica aquella que combina ambos aspectos, es decir la que realiza una selección de unos parámetros concretos gracias a un estudio previo del corpus, y que posteriormente modifica los parámetros seleccionados mediante una serie de reglas a partir de unos cálculos estadísticos previamente realizados.

Como contrapartida a esta flexibilidad, las voces modeladas mediante síntesis paramétrica pueden sufrir pérdida en la calidad con la que se generan los enunciados. El bajo rendimiento de esta aproximación se relaciona también con una falta de modelos prosódicos lo suficientemente detallados del habla emocional, que hacen que las aproximaciones paramétricas en la síntesis de la prosodia sean todavía mejorables.

b. Síntesis por corpus

En esta aproximación el modelado de las emociones es el resultado del proceso general de la selección de unidades, lo que provoca un menor control de los parámetros prosódicos que se desean modelar.

Para que el resultado sea satisfactorio, estos sistemas necesitan un corpus de gran tamaño, representativo del dominio deseado, que permitan seleccionar los segmentos necesarios para formar el enunciado sonoro, y además estos segmentos han de poseer unas características prosódicas determinadas. Esto hace que este tipo de síntesis sea costosa no solo en términos económicos ya que ampliar el sistema a un nuevo dominio implica la grabación de un nuevo corpus, y con ello la contratación del locutor y el estudio de grabación, sino también en términos de tiempo ya que se han de crear unos textos nuevos, y segmentar y etiquetar los distintos enunciados, marcando el inicio y final de cada uno de los alófonos, su duración, intensidad, etc. En resumen, siempre que se desee realizar una voz para un nuevo dominio específico será necesario crear un nuevo corpus oral con todas las implicaciones que tiene (véase 2.1.1.1.).

Esquerra *et al.* (2004) emplean esta técnica para generar voz sintética emocional a partir de la base de datos INTERFACE, un corpus oral grabado por dos locutores (uno masculino y otro femenino) formado por un corpus controlado simulando las seis emociones básicas: alegría, asco, enfado, miedo, sorpresa y tristeza.

A pesar de no haber desarrollado modelos de prosodia para las emociones, y por lo tanto la duración como la entonación reflejaban un estilo neutro, Esquerra *et al.* (2004) asegura que “la audición de las frases sintetizadas permite identificar en la mayoría de los casos la emoción a partir de la cual se ha generado”, y por lo tanto la emoción que se intenta simular.

c. Síntesis basada en Modelos Ocultos de Markov (HMM)

La generación de voz sintética a partir de modelos estadísticos como los HMM ha despertado un gran interés en los últimos años entre la comunidad científica (Tokuda et al. 1995) gracias a los buenos resultados obtenidos.

Esta aproximación emplea los modelos ocultos de Markov para modelar y generar información sobre el espectro, la F0 y la duración de cada uno de los segmentos del habla. A partir de unos determinados alófonos, se seleccionan aquellos cuyos valores tengan una mayor probabilidad de ocurrencia, siempre según los datos observados en el corpus de entrenamiento (Violante, 2012).

Esta aproximación presenta una serie de ventajas que la hacen atractiva:

- No necesita grandes corpus orales para generar la voz sintética, y por lo tanto el espacio de la memoria de datos es menor.
- Permite modelar las características de la voz fácilmente, permitiendo así la generación de voz expresiva sin necesidad de disponer de un corpus oral emocional.
- Mediante pequeñas modificaciones puede ser aplicado a varios idiomas diferentes.

De las tres aproximaciones comentadas, la síntesis por corpus es la que resulta más costosa en términos económicos ya que requiere grandes bases de datos para cada uno de los dominios en los que se desea trabajar, en este caso concreto, para cada una de las emociones (Black, 2003). Los sistemas basados en HMM no requieren corpus tan grandes, pero en contrapartida, la prosodia de la voz sintética no es tan rica en detalles en comparación con el habla natural.

La síntesis paramétrica, tal y como se entiende en este trabajo, permite un gran control de los parámetros a modificar ya que son seleccionados previamente. Además, esta selección se realiza aplicando conocimiento lingüístico extraído a partir de un modelo concreto, y las modificaciones se llevan a cabo previo un análisis fonético que permite conocer si el parámetro en cuestión es significativo, y cómo se ha de llevar a cabo dicha modificación. Un aspecto negativo que cabe destacar es que la manipulación de la señal puede afectar a la calidad de la voz sintética (Barra-Chicote *et al.*, 2010)

Actualmente, a pesar de las ventajas teóricas que ofrece la síntesis paramétrica, en términos generales, la síntesis basada en HMM es la que mejores resultados obtiene, ya que permite generar habla expresiva sin perder calidad a un bajo coste. Aun así, se ha decidido abordar la síntesis empleando una aproximación paramétrica ya que permite emplear conocimiento lingüístico específico.

2.4. Resumen

Esta tesis tiene como objetivo crear un locutor sintético para un dominio muy concreto como son las conversaciones virtuales. Las peculiaridades de este ámbito hacen que sea necesario abordar dos puntos fundamentales de los CTH. Por un lado, como consecuencia del tipo de lenguaje empleado en estas conversaciones, es necesario adaptar o modificar el procesado del lenguaje para poder tratar correctamente estos textos; y por otro lado, al tratarse de una situación comunicativa real, es preciso que el locutor sintético refleje el dominio en el que se centra; en este caso, el habla ha de ser más expresiva que las voces sintéticas comerciales.

Como ya se ha explicado anteriormente, los usuarios de las conversaciones virtuales emplean una serie de estrategias para comunicar y transmitir información, sentimientos, opiniones, etc. lo más rápido posible para mantener la atención del receptor del mensaje. Esto hace que el lenguaje resultante sea en ocasiones muy diferente al normativo y estándar de la lengua. Ante mensajes como los de las conversaciones virtuales en los que abundan las abreviaturas, las sustituciones de grafías, las repeticiones de

vocales, consonantes o signos de puntuación, los emoticonos, las interjecciones, etc. los normalizadores actuales de los CTH no pueden tratar correctamente todos estos fenómenos, y muy a menudo ofrecen una lectura ininteligible. Ante estos resultados es necesario modificar el módulo encargado del procesamiento del lenguaje con el fin de poder tratar estos textos no normativos con la misma eficacia con la que se tratan los textos normativos y estándar. Como se ha observado en el epígrafe existen varias aproximaciones que permiten la normalización de este tipo de textos: las basadas en reglas como las empleadas en la corrección de textos, y métodos estadísticos como los que utilizan la traducción automática o las técnicas basadas en reconocimiento de voz. En esta investigación se ha optado por la aproximación basada en reglas, ya que a partir del conocimiento lingüístico se puede crear una serie de reglas duraderas en el tiempo, lo suficientemente generales y flexibles para poder tratar correctamente el mayor porcentaje posible de palabras consideradas incorrectas, permitiendo a su vez reducir el número de falsos positivos y de nuevos errores en el texto final.

Para lograr extraer todo el conocimiento lingüístico, diseñar y crear un normalizador capaz de tratar correctamente este tipo de textos es necesario realizar un análisis lingüístico lo más exhaustivo posible.

En cuanto al locutor sintético, una etapa imprescindible para su creación es la elaboración y posterior grabación del corpus. Éste ha de reflejar el tipo de habla que tiene lugar en el dominio para el cual se ha creado la voz. En este caso, al tratarse de una situación comunicativa real, los textos escritos reflejan una mayor expresividad que otros tipos de textos como los periodísticos, o informativos. En esta tesis, la expresividad del locutor se centra exclusivamente en la transmisión de estados emocionales, fisiológicos y actos de habla. Para asegurar que el corpus y la voz resultantes sean lo más representativos posibles del dominio es necesario etiquetar todas las oraciones del corpus textual con la emoción o emociones que desean transmitir los usuarios. Para esta tarea se ha tomado como punto de partida el inventario de I3Media (Garrido *et al.* 2012) ya que es uno de los más amplios elaborados hasta el momento para la CTH.

Existen diferentes aproximaciones para lograr obtener una voz expresiva: síntesis por corpus, mediante modelos ocultos de Markov y síntesis paramétrica. En esta investigación se ha decidido realizar

el modelado de la voz expresiva mediante síntesis paramétrica, ya que permite un mayor control de los parámetros a modificar (intensidad, amplitud, velocidad de elocución, frecuencia fundamental, entonación y duración de la vocal prepausal), y aplicar un modelo prosódico concreto obtenido a partir de un análisis estadístico.

Capítulo 3

EL NORMALIZADOR DE TEXTOS

Como ya se ha comentado anteriormente, una de las fases de los CTH que se ha de realizar para generar la voz sintética es procesar el texto de entrada que se desea reproducir. El sistema ha de ser capaz de transformar las formas no estandarizadas como los dígitos, las fechas, o las abreviaturas entre otros, en transcripciones ortográficas para que pueda ser correctamente pronunciado por el sintetizador e inteligible para el oyente.

Las conversaciones virtuales, a diferencia de los textos convencionales, como las noticias o los textos informativos, presentan grandes diferencias lingüísticas, factor que hace más complejo su procesamiento lingüístico y normalización para ser tratadas del mismo modo que cualquier otro texto normativo.

Por ello, el primer objetivo de investigación fijado fue el desarrollo de un módulo de procesamiento de textos para la síntesis capaz de normalizar el texto de entrada de las conversaciones virtuales. Para ello se ha seguido una aproximación basada en reglas con el objetivo de explorar hasta qué punto los resultados obtenidos con este tipo de textos son adecuados para esta tarea. En este caso, lo novedoso no está relacionado con el modo en el que se desarrolla la normalización sino en el tipo de textos que se trata. Las conversaciones virtuales tienen una codificación lingüística propia que hace que a menudo se alejen de la escritura estándar y normativa, haciendo que su interpretación sea dificultosa o incomprensible para alguien que no está familiarizado con ellas.

Se ha partido de TexAFon, un módulo ya existente desarrollado para procesar texto normativo en español, y compatible con el motor de Cereproc. Este normalizador ha sido creado por investigadores del grupo de lingüística computacional (GLiCom) de la universidad Pompeu Fabra y el grupo de voz y lenguaje de Barcelona Media (Garrido et al., 2012a), al cual se le han realizado

una serie de modificaciones encaminadas a procesar correctamente este tipo de textos.

Estas modificaciones se han diseñado a partir de un análisis lingüístico de un corpus de conversaciones virtuales recogido para este fin. A partir de este análisis y siguiendo una aproximación típica de la corrección de textos en la que la definición de palabra correcta juega un papel importante, se ha podido establecer una jerarquía entre los diferentes procesos de corrección.

Al contrario de otros sistemas como el de Moré et al. (2005) que antes de realizar la traducción automática llevan a cabo una corrección del texto, en esta investigación la normalización y corrección se realiza de manera simultánea. De esta manera la corrección se aplica sobre aquellas palabras que lo necesiten, reduciendo el número de posibles palabras correctas modificadas por el sistema, y mejorando así el tiempo de respuesta.

A continuación se presenta la herramienta TexAFon, y seguidamente se abordan todos los pasos que se han llevado a cabo tanto en el análisis lingüístico, el cual permite identificar los fenómenos que se desarrollan en este tipo de textos, como las modificaciones que se han realizado en el normalizador, para normalizar/corregir de manera óptima estos textos.

3.1. El normalizador de textos TexAFon

Para la normalización de textos se ha partido de la herramienta TexAFon, que permite la normalización de textos, la transcripción fonética, la silabificación, y la segmentación prosódica de un texto digital en español o catalán de forma automática. Está concebida para realizar todos los procesos lingüísticos aplicados en los textos de entrada de los conversores de texto a voz, aunque también se puede emplear con otros propósitos como la transcripción fonética de textos o la creación automática de diccionarios fonéticos.

Esta herramienta está desarrollada en Python⁶. Presenta una arquitectura modular (Figura 12) que facilita el desarrollo de nuevas aplicaciones, nuevas lenguas y la conexión con otros módulos externos y aplicaciones. Tal y como se puede observar, unos módulos son independientes del idioma, mientras que otros, en cambio, dependen del idioma en el que esté escrito el texto.

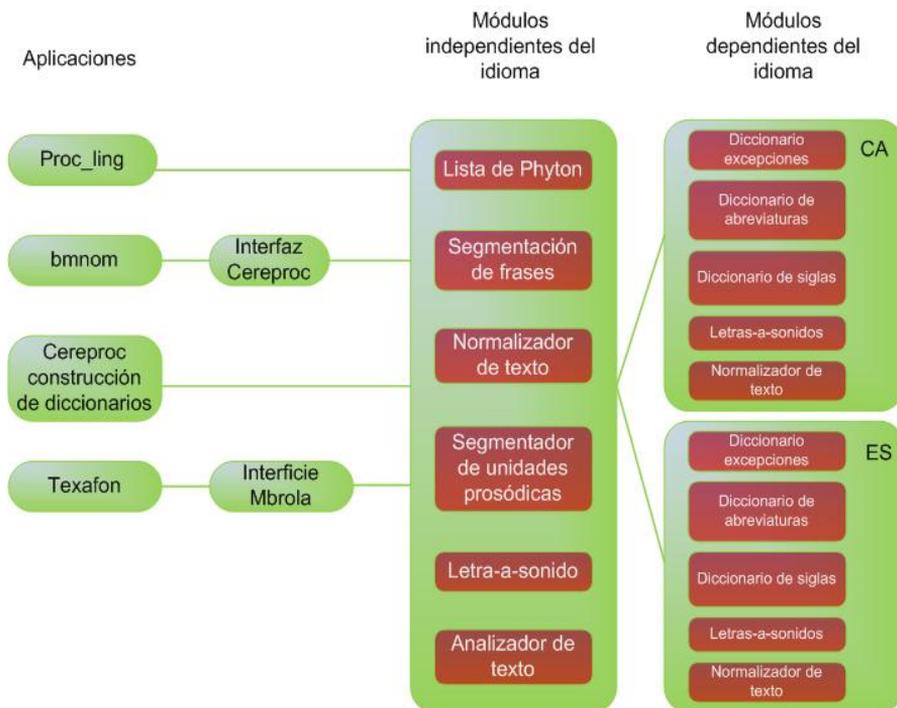


Figura 12: Arquitectura general del TexAFon

‘Segmentación de frases’ es el módulo encargado de segmentar en frases el texto de entrada basándose en los signos de puntuación (puntos, cierre de interrogativas o exclamativas) que indican el final de un enunciado.

‘Normalizador de texto’ es el módulo que se encarga propiamente de la normalización del texto. Abreviaturas, fechas, direcciones de

⁶ Lenguaje de programación diseñado por Guido van Rossum en 1991. Página oficial: <http://www.python.org>

correo electrónico, dígitos, etc. se expanden en su forma ortográfica. Se realiza en dos etapas:

1. **Clasificación:** Cada palabra del texto es identificada como una de las categorías incluidas en la Tabla 6. Esta identificación se realiza mediante reglas, y es independiente al idioma. En algunos casos, recurre a diccionarios específicos de cada idioma para poder identificar algunos tokens, como por ejemplo las abreviaturas.
2. **Expansión:** Cada palabra es procesada de una manera determinada, dependiendo de su etiqueta, hasta obtener su correspondiente versión ortográfica. Este proceso depende completamente del idioma seleccionado.

palabra	porcentaje	Número entero	número decimal
abreviatura ordinal	fecha	hora	número romano
número teléfono	cifras	Dirección electrónica	sigla
abreviatura	Cadena con símbolos	Cadena con signos	Cifras letras y signos
Cifras y letras	Cifras y signos	Letras y signos	signo
símbolo	temperatura	Letra aislada	expresión monetaria

Tabla 6: Lista de categorías consideradas en el preproceso

El módulo ‘segmentador de unidades prosódicas’ se encarga de segmentar el texto de entrada en unidades prosódicas. Este es un aspecto relevante para los conversores de texto a habla, ya que permiten la predicción de parámetros prosódicos como la posición y duración de las pausas, la duración de fonemas y los contornos de F0.

1. **Grupos entonativos:** Se realiza antes de la transcripción fonética. Para detectarlos se basa en los signos de puntuación (punto, coma, paréntesis, etc.) que aparecen en el texto, en los que el locutor debería haber realizado una pausa. Cada grupo entonativo está formado por el enunciado comprendido entre pausa y pausa.
2. **Grupos acentuales:** Se realiza después de la transcripción fonética ya que necesita información respecto a los límites de sílabas que ésta proporciona. Cada grupo acentual está formado por una sílaba tónica y el resto de átonas hasta encontrar la siguiente sílaba tónica del enunciado.

La transcripción fonética y la predicción del acento se desarrollan dentro del módulo ‘letra-a-sonido’ (lts). Tanto en español como en catalán, existe una correspondencia entre la ortografía y su pronunciación, motivo por el cual la transcripción fonética se realiza mediante reglas. La transcripción se realiza en dos etapas:

1. **Proceso ‘palabra por palabra’:** cada palabra se comprueba si pertenece al listado de excepciones y se transcribe fonéticamente, bien usando el diccionario con las excepciones en el que se indica cuál ha de ser su transcripción fonética, o bien empleando reglas.
2. **Proceso grupo entonativo:** La transcripción fonética de cada palabra también incluye marcas

de límite de sílaba o palabra. En esta parte del proceso se presta especial atención a los fonemas que se encuentran entre los límites de palabras contiguas, modificando la transcripción original si el contexto así lo requiere.

La transcripción fonética resultante se realiza en dos formatos diferentes: en SAMPA (Wells *et al.*, 1989), y el empleado por el conversor texto voz de Cereproc (Garrido *et al.*, 2008).

Finalmente, el módulo ‘analizador de texto’ indica el tipo de modalidad oracional a la que pertenece cada oración: afirmativa, interrogativa o exclamativa. En el caso de las interrogativas también indica si es total o parcial.

Respecto a los módulos dependientes al idioma se encuentran el ‘letra-a-sonido, encargado de la silabificación y la predicción del acento, y el ‘normalizador de texto’ que se encarga de la expansión del texto. Además, también existen tres diccionarios, uno de excepciones, el cual contiene las palabras cuya transcripción fonética es irregular, otro de abreviaturas, y un tercero de siglas.

Después de todo el proceso, el normalizador devuelve 7 ficheros que contienen diferente tipo de información, permitiendo ser aplicados según las necesidades del momento. Para profundizar más sobre las características de esta herramienta el lector puede consultar Garrido *et al.* (2012a).

3.2. Análisis lingüístico de las conversaciones virtuales

En este apartado se presenta el corpus de análisis y los pasos que se han seguido para desarrollar el análisis lingüístico sobre este corpus específico.

En esta investigación se ha abordado la normalización desde una perspectiva ‘de corpus’, es decir, se parte de un corpus recogido con el objetivo de ser analizado lingüísticamente. Además, y a diferencia de otros estudios (Mayans, 1999, 2000, 2002; Torres,

1999; Yus, 2001 y 2010; Pons, 2002; López Quero, 2003; Sanmartín, 2007; Kalsson 2007), el corpus es tratado de forma computacional, hecho que permite obtener datos de frecuencia de cada uno de los fenómenos observados. De este modo se podrá organizar y estructurar las modificaciones que se deberán llevar a cabo para optimizar la nueva normalización. El objetivo es, en primer lugar, determinar qué elementos no normativos ('incorrectos') se pueden encontrar en este tipo de textos, y una vez identificados, establecer una tipología. Esta tipología servirá para determinar el tratamiento posterior que deberá hacerse en el módulo de proceso lingüístico para normalizar estos elementos.

3.2.1. El corpus

El corpus empleado para el análisis lingüístico está formado por 45 conversaciones, de las cuales 38 son privadas y extraídas de chats vinculados a un registro previo, en concreto de Messenger y Gmail, y las 7 restantes fueron extraídas de chats no vinculados a un registro, de www.inforchat.com, y son públicas. Todas ellas se obtuvieron durante los años 2008 y 2009. La edad declarada de los usuarios está comprendida entre los 18 y los 35 años aproximadamente.

El corpus está formado por 8.780 intervenciones diferentes, con una media de 4,68 palabras por enunciado, cifra similar a la obtenida en otros estudios (Crystal, 2002). Contiene un total de 40.676 tokens⁷. En la Tabla 7 puede verse su distribución.

⁷ Se entiende por token toda aquella cadena alfanumérica comprendida entre dos espacios en blanco.

Tipo de conversación	Conversación	nº intervenciones	nº usuarios activos	nº tokens	media palabras por intervención
Privadas	1	4	2	11	2,75
	2	33	2	148	4,484848485
	3	50	2	178	3,56
	4	42	2	236	5,619047619
	5	130	2	579	4,453846154
	6	27	2	184	6,814814815
	7	30	2	174	5,8
	8	13	2	51	3,923076923
	9	12	2	107	3,923076923
	10	1	1	15	15
	11	1	1	2	2

	12	76	2	314	4,131578947
	13	1	1	2	2
	14	34	2	106	3,117647059
	15	117	2	389	3,324786325
	16	40	2	131	3,275
	17	35	2	143	4,085714286
	18	4	2	6	1,5
	19	82	2	346	4,219512195
	20	179	2	663	3,703910615
	21	22	2	66	3
	22	119	2	515	4,327731092
	23	24	2	89	3,708333333

	24	5	1	10	2
	25	1	1	1	1
	26	7	2	44	6,285714286
	27	76	2	264	3,473684211
	28	133	2	658	4,947368421
	29	13	2	77	5,923076923
	30	140	2	540	3,857142857
	31	432	2	2291	5,303240741
	32	183	2	922	5,038251366
	33	88	2	602	6,840909091
	34	75	2	389	5,186666667
	35	115	2	684	5,947826087

	36	93	2	654	7,032258065
	37	163	2	819	5,024539877
	38	131	2	695	5,305343511
Públicas	1	376	32	2043	5,433510638
	2	924	28	4745	5,135281385
	3	1267	46	5955	4,700078927
	4	2029	57	8727	4,301133563
	5	294	22	1315	4,472789116
	6	969	61	3635	3,75128999
	7	190	12	1151	6,057894737
TOTAL	45	8780		40676	4,682989222

Tabla 7: Información sobre las conversaciones que forman el corpus de análisis

3.2.2. Metodología

3.2.2.1. Determinación de los tokens incorrectos

Primeramente, antes de realizar cualquier tipo de comparación, fue imprescindible definir qué se iba a considerar correcto y qué incorrecto. Se resolvió que para decidir si una palabra era correcta, ésta debía ser cotejada en un diccionario y aparecer en él. Para ello fue necesario disponer de un diccionario general en el que aparecieran las formas de las palabras del español. La Fundación Barcelona Media cedió un diccionario que contiene más de 900 mil palabras⁸. Este diccionario fue creado a partir de un corpus no revisado, ya que a pesar de recoger nombres y adjetivos flexionados, verbos conjugados e incluso palabras derivadas y compuestas del español y algunas palabras de idiomas extranjeros como el inglés o el francés, no contiene todas las palabras de la lengua española, y además se ha observado la presencia de palabras inexistentes en español como ‘aafe’, ‘akpkamni’, o ‘canones’. Por lo tanto, no se puede considerar ni completo ni normativo.

En este primer análisis cada token del corpus original que contenía todas las conversaciones virtuales fue cotejado automáticamente con las palabras de este diccionario. Si la palabra se encontraba en el diccionario se consideraba correcta, en cambio, si no estaba en él, se extraía para ser analizada. A consecuencia de las características del diccionario de trabajo, se observaron ciertos errores en la identificación de las palabras correctas e incorrectas del texto. Palabras que no son correctas en la lengua española fueron consideradas correctas al encontrarse en el diccionario, y palabras que sí son correctas en español fueron etiquetadas como incorrectas al no aparecer en el diccionario. En el caso de las palabras correctas etiquetadas como incorrectas, se añadieron al diccionario de trabajo y no se tuvieron en consideración en el análisis lingüístico. En cambio, las palabras incorrectas que habían sido consideradas como

⁸ Agradezco a la Fundación Barcelona Media tanto la cesión de este diccionario como su colaboración en la grabación del corpus oral, indispensable para la creación del locutor sintético Edu.

correctas, al no poder localizarlas todas ellas de un modo automático o semiautomático, tampoco se tuvieron en consideración en el primer análisis, pero sí en el segundo que se realizó de un modo más preciso.

En esta tesis, a diferencia de otros estudios como por ejemplo el realizado por Moré et al. (2005), no se ha tenido en cuenta si el error se ha producido por desconocer la norma ortográfica, si ha sido a causa de un descuido del usuario, o si bien ha sido intencionado, ya que esta información no es relevante en un texto que no ha sido revisado previamente antes de ser enviado, y como bien afirma Díaz (2005) en muchas ocasiones realizar tal distinción es dificultoso.

3.2.2.2. Clasificación de los tokens

Como ya se ha comentado en el estado de la cuestión (2.3.1.1. Basados en reglas) existen varios inventarios de errores, pero ninguno lo suficientemente universal como para ser empleado en el diseño de cualquier sistema.

Este primer análisis, y su consecuente clasificación de los tokens, se realizó de un modo *naïf*, partiendo de los diferentes fenómenos observados por otros autores y recogidos en el epígrafe 2.2.3. *Características lingüísticas*, del estado de la cuestión de esta tesis. Se indicó el tipo de anomalía o anomalías que se observaban en cada una de las palabras etiquetadas como incorrectas: es decir, se marcó si se trataba de una abreviatura, un emoticono, si la palabra presentaba un carácter de más, o uno o varios de menos, si parte de la palabra había sido sustituida por otra, etc.

Las palabras que presentaban un mismo tipo de anomalía fueron agrupadas y contabilizadas de forma completamente automática gracias a una serie de scripts en Python diseñados para esta tarea. El resultado obtenido puede verse en la Tabla 8 que se muestra a continuación.

Fenómeno	Nº de casos
Acentuación	652
Repetición	629
Sustitución de 1 grafía	507
Eliminación de un espacio en blanco	404
Eliminación de una grafía	309
Elemento paralingüístico	306
Palabra correcta	221
Nombre propio	165
Inserción de una grafía	154
Sustitución de 1 vocal por un número	150
Extranjerismo	128
Otros	82
Transposición de dos grafías	75
Dígito	65
Sustitución de 2 grafías	54
Fonetización	48
Abreviatura	45
Emoticono	40
Símbolo	29
Sustitución de 2 vocales por números	29
Palabra con símbolo	25
Eliminación de dos grafías	23
Eliminación de dos espacios en blanco	20
Esqueleto consonántico	19
Dirección de correo electrónica	18
Signo	13
Inserción de dos grafías	7

Sigla	7
Eliminación de tres espacios en blanco	4
Eliminación de tres grafías	3
Hora	3
Inserción de un espacio en blanco	3
Sustitución de 3 vocales por números	2
Eliminación de cuatro grafías	1
Número romano	1

Tabla 8: Listado de los diferentes fenómenos observados en las palabras consideradas incorrectas del primer análisis, y la frecuencia de aparición de cada uno de ellos

Algunas de estas categorías, como son las abreviaturas, los dígitos, las direcciones de correo electrónico, las horas, los números romanos, las palabras con símbolos, las siglas, los signos, o los símbolos, ya eran correctamente normalizadas por el normalizador TexAFon. Por ello, se decidió realizar un segundo análisis, teniendo en cuenta ahora aquellos fenómenos que el normalizador no trataba, simplificando así el inventario de desviaciones.⁹

En esta segunda clasificación, las conversaciones virtuales originales fueron cotejadas de manera automática, mediante un script de Python, con una versión del texto normalizado manualmente. En esta normalización manual únicamente se tuvo presente los errores/desviaciones que se habían realizado en cada una de las palabras que formaban el texto original; no se tuvo en cuenta aspectos gramaticales como el empleo de signos de puntuación ni si su uso era correcto o no. Los pares de palabras que presentaban alguna diferencia, ya fuera en el número o en la forma de los caracteres, fueron extraídos a un fichero donde estos pares estaban organizados en columnas; en la primera se presentaban las

⁹ Nótese que se emplea el término ‘incorrecto’ y ‘desviación’ para designar el mismo concepto: todo token, cadena alfanumérica que se encuentra entre dos espacios en blanco, que presenta errores ortográficos o desviaciones, intencionadas o no.

palabras originales del texto, y la segunda columna contenía las formas normalizadas. La Tabla 9 muestra algunos ejemplos de pares de palabras presentes en este fichero. Posteriormente, cada uno de los pares del fichero ha sido etiquetado según las modificaciones que habían sufrido las palabras originales de la conversación virtual respecto a la versión normalizada formando así una tabla. Para el etiquetado de las anomalías se empleó el inventario de errores que se explica a continuación.

Texto original	Texto normalizado manualmente
Cayate	Cállate
Everano	verano
Prverso	Perverso
Xk	porque
Yaves	Ya ves

Tabla 9: Ejemplos de pares de palabras cuyas formas presentan algunas diferencias entre sí

3.2.2.3. Identificación/clasificación de errores

Para decidir cómo iba a ser el inventario de desviaciones, se partió de los resultados obtenidos en el primer análisis lingüístico. Al observar que ciertos fenómenos ya eran tratados por TexAFon, este segundo análisis se centró en aquellas anomalías que afectaban exclusivamente a los tokens que serían etiquetados como ‘palabra’ por el normalizador. A partir del primer análisis se comprobó que estas alteraciones podían afectar tanto a las grafías de un token como al número de palabras que formaban el token original, y que

estas alteraciones siempre giraban alrededor de unos mismos aspectos (eliminación de una grafía, eliminación de un espacio en blanco, inserción de una grafía o de un espacio en blanco, sustitución de grafías, etc.). Por ello, se decidió realizar el etiquetado en tres niveles diferentes, los cuales proporcionan toda la información necesaria sobre cada fenómeno o fenómenos producidos en cada una de las palabras, ya que cada uno de ellos indica:

1. Si la modificación afecta únicamente a caracteres (C) o al número de tokens (T)
2. El tipo de modificación general:
 - a. Eliminación (E)
 - b. Inserción (I)
 - c. Sustitución (S)
 - d. Transposición (T)
3. El caso concreto de dicha modificación:
 - a. Acentuación (A)
 - b. Grafías sistemáticas (G)
 - c. Léxico (L)
 - d. Signo de puntuación (P)
 - e. Otros (O)

La Tabla 10 recoge algunos ejemplos de palabras incorrectas extraídas del corpus de análisis con las etiquetas que indican el tipo de desviación o desviaciones que presenta cada una de ellas. La primera columna corresponde a la palabra original que el usuario escribió, la palabra de la segunda columna es la normalización-corrección propuesta en la normalización manual, y las tres siguientes columnas corresponden a las etiquetas de las desviaciones observadas en cada una de las palabras. Para que su

comprensión sea más sencilla, a continuación se comentarán más en detalle dos de estos ejemplos.

Palabra original	Palabra corregida manualmente	Nivel de modificación	Modif. general	Modif. concreta
Algregada	Agregada	C	S	O
Q	Que	C	S	G
Ahoar	Ahora	C	T	O
Osea	O sea	T	E	P
Corto...jamás	Corto... jamás	T C	E S	P A

Tabla 10: Algunos ejemplos de palabras incorrectas etiquetadas extraídos de la tabla resultante

Las palabras ‘algregada’ y ‘agregada’ presentan el mismo número de tokens, uno, por lo tanto en la tercera columna se indica con una ‘c’ que la modificación afecta únicamente a los caracteres. En este caso, se ha añadido una letra extra en el momento de escribir la palabra por lo que se indica que la palabra ha sufrido una inserción mediante la etiqueta ‘i’. Finalmente, como dicha inserción no se corresponde con una grafía sistemática, ni léxica, ni es un signo de puntuación, ni tiene relación con la acentuación se indica que pertenece a otra categoría (‘o’).

En la palabra ‘osea’, la modificación afecta al número de tokens (‘t’) ya que de uno pasa a dos ‘o sea’. En este caso el usuario ha eliminado un carácter (‘e’), en concreto un signo de puntuación (‘p’), ya que los espacios en blanco se han incluido dentro de los signos de puntuación.

3.2.3. Cuantificación de errores

A parte de la identificación de los diferentes errores, también se ha llevado a cabo una cuantificación de los mismos, ya que conocer la frecuencia de aparición de cada uno de ellos resulta imprescindible para controlar la cobertura del normalizador (Oliva, 1997).

Se ha detectado un total de 8.768 errores/desviaciones, que representa el 21,5% del total del corpus, distribuido de la siguiente manera:

NÚMERO ERRORES POR PALABRA	NÚMERO DE CASOS
1 ERROR	7476 (85,27%)
2 ERRORES	1083 (12,35%)
3 ERRORES	180 (2,05%)
4 O MÁS ERRORES	29 (0,33%)

Tabla 11: Recuento de los errores o desviaciones observadas en el corpus de análisis clasificados según el número de anomalías presentes en cada palabra no correcta

A continuación se explicará con más detalle los errores más frecuentes tanto de un error por palabra como de dos errores, ya que entre ambos representan más del 90 por ciento de los errores cometidos en las conversaciones virtuales.

3.2.3.1. Un error por palabra

Como se puede observar en la Figura 13, la sustitución de caracteres es el fenómeno más frecuente: representa el 71%, bastante por delante de la eliminación (17%), la inserción (10%) y la transposición (2%). A continuación se comentará más en detalle cada uno de estos fenómenos.

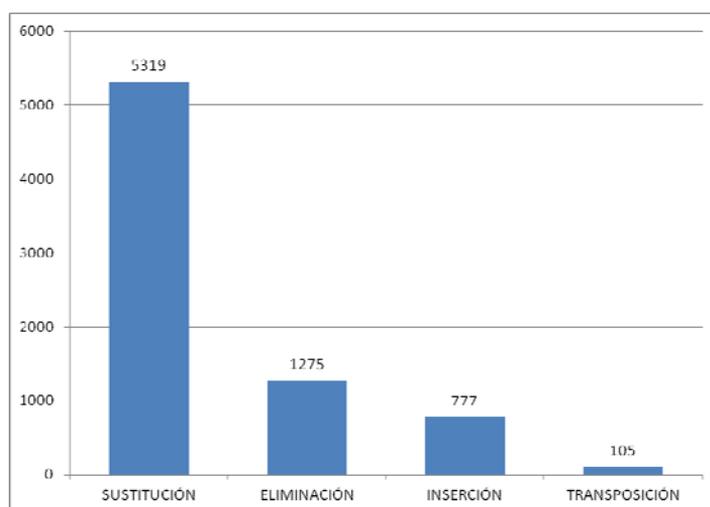


Figura 13: Distribución de los fenómenos posibles hallados en un error por palabra

a. *Sustitución de caracteres*

En la sustitución de caracteres se encuentra la sustitución sistemática de grafías, la acentuación, sustituciones léxicas y otras sustituciones. En la Figura 14 se muestra su distribución.

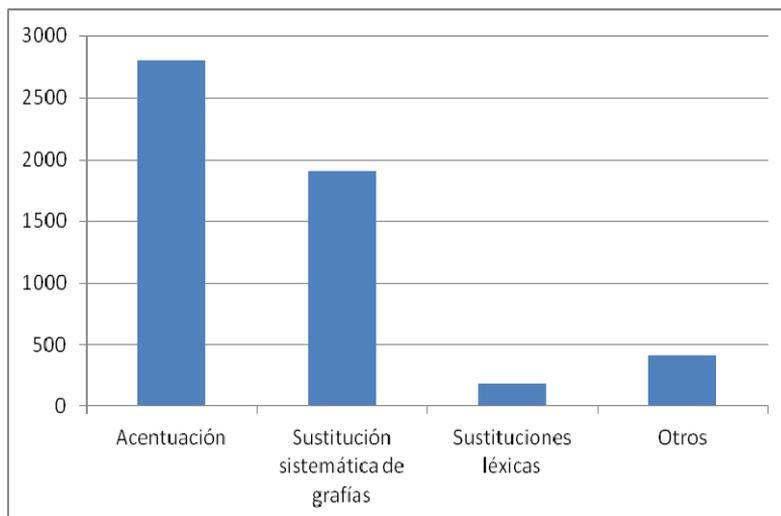


Figura 14: Distribución de los diferentes tipos de sustituciones observadas en el corpus de análisis

- ✓ ***Acentuación:*** Es el fenómeno que más tiene lugar en las conversaciones virtuales. En el 99,9% de los casos, las palabras se han escrito sin acento gráfico, sólo en un 0,1% las palabras aparecían acentuadas cuando no deberían estarlo.
 - Déis (deis)
 - Pais (país)

- ✓ ***Sustitución sistemática de grafías:*** también denominadas sustituciones grafemáticas por Yus (2001), (véase el epígrafe 2.2.3.1. Economía, rapidez y eficacia). Representa el 35,83% respecto a todas las sustituciones. En la Tabla 12 se puede observar por un lado las grafías que se han sustituido de forma sistemática, los segmentos por los cuales han sido sustituidas y el porcentaje de distribución en relación a todas las sustituciones sistemáticas de grafías.
 - Trankilo (tranquilo)

- Peqeña (pequeña)
- Muxo (mucho)
- Wenas (buenas)

Q	QUE (28,6%) QUÉ (10%) QU (3,5%)
K	QU (17,15%) QUE (12% C (5%) QUÉ (4%)
X	CH (6,24%) POR (3%) PER (0,5%) PAR (0,2%)
W	BU (3,3%) GU (0,78%) U (0,31%) HU (0,15%) GUA (0,05%)
Y	LL (2,25%) I (0,31%)
I	Y (2,46%)

Tabla 12: Resumen de las grafías sustituidas de forma sistemática y su correspondiente equivalencia

- ✓ **Sustituciones léxicas:** Bajo esta etiqueta se recogen las abreviaturas y acrónimos que los usuarios del chat suelen utilizar, algunas de ellas de modo muy habitual. Suponen el 3,62% de todas las sustituciones:

- Tb (también)
- msn (Messenger)
- Pq (porque)
- Xk (porque)

- ✓ **Otras sustituciones:** En este apartado se incluyen todas aquellas sustituciones que no se han podido enmarcar en los tres tipos anteriores. Representan el 7,76%.

Algunas de ellas son claramente intencionadas, como el empleo de dígitos en vez de vocales (47%), y otras son involuntarias, causadas por la no revisión del mensaje y su rápida escritura, como el hecho de escribir una grafía por otra sin relación aparente (41,5%), o porque se encuentren próximas en el teclado (38,3%). Otras situaciones, en cambio, pueden ser intencionadas o deberse a un desconocimiento de la norma escrita por parte del usuario, como es el empleo indiscriminado de grafías cuya fonética es similar (37,4%). En algunas situaciones, un mismo error puede converger entre varios de los aquí mencionados, como el caso de la 'b' y la 'v', que son contiguas en el teclado y su reproducción fónica es igual.

- Atenci0n (atención)
- Tajadn (tajada)
- Saves (sabes)
- Empiezes (empieces)

b. Eliminación de caracteres

La eliminación de caracteres se da en un 17% de los casos.

Las conversaciones virtuales se caracterizan por no utilizar los signos de puntuación (comas, puntos, interrogantes o exclamaciones tanto de apertura como de cierre, etc.) en sus mensajes. A pesar de esto, en este análisis del texto no se han tenido en cuenta estas supresiones; se ha centrado toda la atención en los espacios en blanco y las grafías.

Las vocales son las grafías que más se han eliminado; representan el 40% de los casos, y pueden eliminarse todas las vocales de una palabra o algunas de ellas. Este es un aspecto muy frecuente en las conversaciones virtuales, ya que la ausencia de vocales permite la comprensión del mensaje sin grandes dificultades, y su rápida escritura.

- Mñn (mañana)
- Llgar (llegar)

En segundo lugar se encuentra la eliminación del espacio en blanco (26,27%). En la mayoría de los casos esto tiene lugar cuando le precede un signo de puntuación (54,8%), normalmente puntos suspensivos. Existe una serie de combinaciones con cierta tendencia a aparecer escritas juntas (38%), como una única palabra, como por ejemplo ‘apartir’ (preposición + verbo), ‘delas’ (preposición + determinante), ‘esque’ (verbo + pronombre), etc. En la Tabla 13 se muestran las combinaciones más frecuentes.

La grafía ‘h’ también es una de las más eliminadas por los usuarios del chat (22.35%), sobre todo cuando se encuentra al inicio de palabra, ya que en español no se pronuncia.

- Ola (hola)
- Aces (haces)

El 11.3% restante responde a eliminaciones de otras consonantes, especialmente a inicio o final de palabra.

COMBINACIONES	FRECUENCIA APARICIÓN
PREPOSICIÓN + DET.	16 (5,57%)
VERBO + PRONOMBRE	9 (3,13%)
PREPOSICIÓN + PRONOMBRE	7 (2,43%)
PREPOSICIÓN + VERBO	7 (2,43%)
PRONOMBRE + PRONOMBRE	7 (2,43%)
PRONOMBRE + VERBO	7 (2,43%)
ADVERBIO + PRONOMBRE	6 (2,09%)

Tabla 13: Combinaciones sintácticas más frecuentes en las que se ha elidido el espacio en blanco. El porcentaje se ha calculado respecto al total de elisiones

c. Inserción de caracteres

La inserción de caracteres ocupa el tercer puesto del ranking con un 10% de los casos.

No existe un grupo de grafías que sistemáticamente se inserten, aunque el factor más extendido es la repetición de caracteres (66%), tanto de grafías como de signos de puntuación (cuando estos son transcritos). Esta repetición puede ser variable en número.

- Besotee (besote)
- Amigaaaaaaaaa (amiga)

También es frecuente (7,46%), aunque en menor medida, la inserción de grafías que se encuentran próximas a otras en el teclado del ordenador. Es fácil que, con la rapidez a la que se ha de escribir los mensajes, el usuario pulse alguna tecla extra cercana.

- Vuelasn (vuela)
- Chuilena (chilena)

La inserción también puede afectar a los espacios en blanco. Un usuario puede introducir un espacio en blanco cuando no corresponde, normalmente entre grafías de una misma palabra, convirtiendo lo que debería ser una palabra en dos. Este fenómeno es minoritario ya que su presencia se reduce al 3% de todos los casos de inserción.

- No sotros (nosotros)
- Al go (algo)

Hay otras grafías que se han insertado y no se pueden clasificar en ninguno de los casos anteriores. Representan el 23,6%. En este caso, las inserciones de caracteres suele realizarse con mayor frecuencia al final de la palabra (75%). En un inicio se pensó que estas inserciones al final de la cadena, y también al inicio de ella, podía deberse a restos de la palabra siguiente o anterior, ya que a causa de la velocidad que prima en los chats, el usuario puede presionar a deshora la tecla correspondiente al espacio en blanco,

pero sólo el 19% corresponde con este hecho. En la Tabla 14 muestra las situaciones en las que ha sido insertada cada una de las grafías analizadas y la frecuencia con la que se ha realizado.

	INICIO DE PALABRA	INTERIOR DE PALABRA	FINAL DE PALABRA
PALABRA ANTERIOR O POSTERIOR	16 (9,2%)	-	17 (9,77%)
OTROS	11 (6,32%)	16 (9,2%)	114 (65,52%)

Tabla 14: Distribución de las grafías que han sido insertadas, las cuales no se corresponden ni con la repetición de grafías o espacios en blanco ni por proximidad en el teclado, según si dicha grafía pertenece a la palabra anterior o posterior

d. Transposición de caracteres

La transposición de caracteres es la alteración del orden de dos grafías consecutivas. Este fenómeno es el que presenta una menor frecuencia de aparición respecto a los otros tres observados anteriormente. Sólo se han hallado 105 casos de transposición de caracteres contiguos, que equivalen al 2%.

El 20% de los casos tiene lugar en palabras monosílabas, y el 80% restante en palabras con más de una sílaba.

- Qeu (que)
- Productitvo (productivo)
- Hya (hay)
- Evrdad (verdad)

En el caso de la transposición, al contrario que en la sustitución, no se puede decir que existan unas grafías que suelen transponerse sistemáticamente. A pesar de que el mayor número de casos tenga lugar al final de la palabra, tampoco se puede afirmar que este fenómeno depende de la posición en la que se encuentra dentro de la misma palabra, ya que la distribución es bastante similar en todos los casos, como se puede observar en la Tabla 15.

PRIMERAS DOS GRAFÍAS	INTERIOR PALABRA	ÚLTIMAS DOS GRAFÍAS
23 (27,38%)	26 (30,95%)	35 (41,66%)

Tabla 15: Distribución de la transposición de caracteres en las palabras no monosilábicas

3.2.3.2. Dos errores por palabra

Existen diez posibles combinaciones de errores cuando coexisten dos errores en una misma palabra. Tal y como se puede observar en la Tabla 16, la distribución de la combinación de dos errores en una misma palabra es dispar: ciertas combinaciones son muy propensas a darse, mientras que otras en cambio son muy poco frecuentes.

A continuación se expondrá en más detalle las dos combinaciones más frecuentes: sustitución y eliminación, y dos sustituciones en una misma palabra.

COMBINACIÓN ERRORES	Nº CASOS
sustitución -sustitución	321
sustitución - inserción	183
sustitución - eliminación	391
sustitución - transposición	16
inserción – inserción	23
inserción - eliminación	53
inserción - transposición	5
eliminación - eliminación	74
eliminación - transposición	5
transposición - transposición	12

Tabla 16: Distribución de los errores cuando se comenten 2 en una misma palabra

a. Sustitución y eliminación de caracteres:

Dentro de esta categoría más genérica podemos encontrar diferentes tipos de combinaciones: la sustitución de una vocal acentuada por otra no acentuada y la eliminación de un espacio en blanco, la sustitución de unas grafías sistemáticas y la elisión de otro carácter, etc. En la Tabla 17 se puede observar la distribución de las posibles combinaciones. Algunas de ellas, como la sustitución de vocales acentuadas y la eliminación de signos de puntuación, y la sustitución de grafías sistemáticas y la eliminación de signos de puntuación, presentan un gran número de casos, mientras que otras combinaciones tienen lugar en muy pocas ocasiones, como la sustitución léxica y la eliminación de signos de puntuación, o no se han producido en ningún caso en el corpus de estudio, como la

sustitución de otros caracteres y la eliminación de grafías sistemáticas.

Al contrario de lo que ocurría en la eliminación de los signos de puntuación cuando era el único error en la palabra, la eliminación del espacio en blanco se produce con mayor frecuencia entre dos palabras (58%) que en vez de precedida por un signo de puntuación (42%). Aún así, cuando tiene lugar después de un signo de puntuación, éste suele ser puntos suspensivos.

Nuevamente, del mismo modo que en un error por palabra, la gran mayoría de las sustituciones están relacionadas con la acentuación y la sustitución de grafías sistemáticas.

		Eliminación		
		Signo de puntuación	Grafías sistemáticas	Otros
Sustitución	Acentuación	100	43	56
	Grafías sistemáticas	83	29	34
	Sustituciones léxicas	3	0	0
	Otros	7	0	36

Tabla 17: Distribución de las diferentes combinaciones de errores a causa de la eliminación y sustitución de diversos caracteres en una misma palabra

b. Dos sustituciones

De las diez posibles combinaciones que pueden darse dentro de la doble sustitución de caracteres, sólo se han observado seis, y algunas de ellas con un número muy reducido de casos (Tabla 18).

Nuevamente, los dos fenómenos que han ocurrido con gran frecuencia en un error por palabra, la sustitución de vocales acentuadas y la sustitución sistemática de grafías, vuelven a ser los

más frecuentes cuando se combinan y coexisten en una misma palabra. Esta combinación es la más frecuente, con gran diferencia, respecto al resto de las posibles combinaciones.

El caso menos frecuente es la combinación de dos errores de acentuación. Como ya se ha comentado con anterioridad, los usuarios del chat suelen prescindir de escribir el acento de las palabras, pero en ocasiones, y sobre todo según el usuario, sí los escriben. Estos tres casos han ocurrido por escribir con acento la vocal errónea de una palabra que sí debería ir acentuada.

acentuación - grafía sistemática	165
grafía sistemática - otros	52
otro -otro	44
grafía sistemática - grafía sistemática	29
acentuación - otro	28
acentuación - acentuación	3

Tabla 18: Distribución de las diferentes combinaciones de errores a causa de la doble sustitución de diversos caracteres en una misma palabra

3.2.4. Resumen: implicaciones para la normalización de textos

La gran mayoría de estudios relacionados con las conversaciones virtuales exponen un inventario de características lingüísticas observadas en este tipo de discurso, como por ejemplo la repetición de grafías o signos de puntuación para intentar expresar la prosodia con la que se realizaría un enunciado en una conversación oral, el empleo de abreviaturas y acrónimos para ahorrar tiempo a la hora de escribir los mensajes, el uso de emoticonos para transmitir

ciertas emociones, etc. Son muchos y variados los fenómenos que se emplean en las conversaciones virtuales, pero ningún estudio ha cuantificado con qué frecuencia se suceden, ni cuántas desviaciones puede llegar a presentar un mismo token. En esta investigación sí se ha tenido en consideración este aspecto, realizando un cálculo automático no sólo de la frecuencia de aparición de cada uno de los diferentes fenómenos contemplados en el corpus sino también del número de alteraciones que puede presentar una misma palabra.

Se ha observado que cada token del texto original puede llegar a presentar hasta más de cuatro desviaciones, siendo más frecuente que sólo presenten una o dos.

En el caso de un error/desviación por palabra, los tres fenómenos más frecuentes son la ausencia de acentuación, la sustitución sistemática de grafías y la eliminación de grafías, en especial de la supresión de vocales.

Respecto a la combinación de los dos fenómenos más frecuentes en un mismo token, tal y como se ha comentado más arriba, se trata de una sustitución más una eliminación de caracteres, o de dos caracteres sustituidos por otras grafías.

Moré *et al.* (2005) presentan un estudio de los errores lingüísticos observados en los correos electrónicos en castellano y catalán que afectan a la traducción automática de los mismos, realizando un cálculo de aparición de cada uno de ellos. La tipología de errores recogida en este estudio es más amplia que la presentada en esta tesis; aún así, a pesar de que la escritura de correos electrónicos permite una revisión y una mayor elaboración del mismo, se observa que la ausencia de acentuación, del mismo modo que ocurre en las conversaciones virtuales, es el fenómeno más frecuente.

Analizar y cuantificar las anomalías observadas en este tipo de textos permite conocer más a fondo qué fenómenos tienen lugar en estos textos, y esto, a su vez, posibilita diseñar y organizar un módulo de normalización-corrección que permita tratar los casos más reiterativos y en un orden concreto para priorizar así los casos más frecuentes frente a los más minoritarios. Como se ha comentado, los tokens que presentan una o dos desviaciones representan el 97,62% de los casos: por lo tanto, en esta investigación únicamente se van a tratar estos casos, descartando la

normalización de tokens con tres o más desviaciones ya que la probabilidad de que aparezcan en una conversación virtual escrita es prácticamente mínima e intentar normalizarlos podría provocar que el porcentaje de normalizaciones correctas disminuyera considerablemente. En el caso de las palabras con una o dos desviaciones, aparte de tratar los fenómenos comentados anteriormente en los epígrafes 2.3.1 y 2.3.2, también se van a normalizar aquellos tokens que contienen símbolos o signos como por ejemplo ‘=mente’ o ‘tod@s’.

El orden en el que se desarrollan los diferentes procesos del normalizador es un aspecto importante que se ha tenido en cuenta gracias al análisis lingüístico realizado previamente. Sabiendo con qué asiduidad se producen las diferentes desviaciones, se ha podido organizar el módulo de tal manera que se prioriza, nuevamente, los casos más frecuentes frente a los que han presentado un menor número de apariciones. La sustitución de vocales no acentuadas por otras acentuadas y la sustitución de grafías sistemáticas son las dos operaciones que se realizan en primer lugar, mientras que la transposición de caracteres se realiza siempre en último lugar ya que su presencia en el análisis del corpus fue inferior a las otras.

En el epígrafe siguiente se muestra en qué ha consistido la adaptación del módulo de normalización TexAFon.

3.3. TexAFon 2.0: Adaptación del normalizador TexAFon para conversaciones virtuales

Como ya se ha ido mencionando reiteradamente a lo largo de esta tesis, las conversaciones virtuales presentan unas llamativas características lingüísticas que aleja estos textos de los textos normativos y convencionales, haciendo que los tratamientos en procesado del lenguaje natural realizados hasta ahora sean insuficientes para tratarlos correctamente.

En el caso de TexAFon, que fue diseñado para tratar texto normativo, ocurre exactamente lo mismo: su funcionamiento es deficiente cuando intenta aplicar su proceso lingüístico en textos no normativos como los estudiados en esta investigación. Hay tokens que no reciben la etiqueta de la categoría adecuada, o palabras que

no se expanden correctamente, como por ejemplo los emoticonos que son identificados como signos de puntuación y expandidos como tal, o las palabras cuyas vocales han sido substituidas por números, en las que estos números son expandidos y no convertidos a vocales como por ejemplo ‘n0’ que es normalizado como ‘ene cero’ en vez de ‘no’. Esto hace que sea necesario plantearse una modificación del sistema para operar correctamente con textos no normativos.

La adaptación del normalizador está orientada a tratar los fenómenos observados en el análisis lingüístico (3.2. *Análisis lingüístico de las conversaciones virtuales*), integrándolos dentro del proceso de normalización ya existente. En este caso, no se ha desarrollado un módulo totalmente independiente, previo al normalizador, como en el caso de Armenta *et al.* (2003 b) o Moré *et al.* (2005), sino que se ha considerado más útil integrar los diferentes procesos dentro del flujo del normalizador. De este modo se aprovecha la identificación que realiza el normalizador de cada uno de los tokens para corregirlo/normalizarlo según nuestros intereses. Por ejemplo, palabras como ‘tb’ o ‘pk’ deberían ser identificadas como abreviaturas, y su correspondiente expansión sería extraída de un diccionario de abreviaturas específico del léxico de las conversaciones virtuales. Por el contrario, las direcciones URL (*Uniform Resource Locator*) no se deben intentar corregir, sino que el sistema debería leerlas tal y como las ha escrito el usuario, deletreando aquellos fragmentos que no puedan ser pronunciados correctamente. A partir de estos dos breves ejemplos, queda más clara la relevancia de aprovechar el funcionamiento del sistema de TexAFon, el cual identifica ante qué tipo de palabra estamos, y permite decidir si se va a aplicar un proceso de normalización previo a la expansión del mismo, y en caso afirmativo la normalización se puede desarrollar de distintas formas, siempre dependiendo del tipo de palabra a tratar.

Esta aproximación ha conllevado realizar cambios que afectan a varios puntos del proceso de normalización del sistema (véase Figura 15), como ha sido la creación de una serie de nuevos elementos:

1. **Un diccionario general:** El diccionario empleado para el primer análisis lingüístico del texto ha sido incrementado

con topónimos, antropónimos e hipocorísticos extraídos del Instituto Nacional de Estadística (www.ine.es), y palabras de ámbito más coloquial halladas en el corpus de análisis. Actualmente contiene cerca de un millón de entradas.

2. **Un diccionario de abreviaturas específicas del lenguaje de Internet:** cuenta con más de 1100 entradas distintas, extraídas en su mayoría de [diccionarioSMS \(www.diccionariosms.com\)](http://www.diccionariosms.com) y del propio corpus de análisis. Recoge abreviaturas empleadas en este tipo de discursos, que en el lenguaje escrito normativo no se emplean, como por ejemplo al+ (almas), alke (aunque), tb (también), etc.
3. **Un diccionario de emoticonos:** contiene 805 emoticonos, obtenidos del corpus de análisis, de la página web [diccionarioSMS, o de canal 3000 \(http://www.canal3000.com/moviles/em_oticonos01.htm\)](http://www.canal3000.com/moviles/em_oticonos01.htm), y listados publicados en diversos artículos como por ejemplo Torres (1999).
4. **Un modelo de lenguaje específico para las conversaciones virtuales,** el cual se ha desarrollado contabilizando los pares de bigramas que aparecen en el corpus de análisis.
5. Un módulo **‘modifica_palabra’** que realiza una serie de modificaciones en los tokens etiquetados como ‘palabra’ para obtener una o varias normalizaciones para cada palabra incorrecta.

Además, ha sido necesario realizar ciertas modificaciones en algunos procesos de ‘segmentación de frases’ y del ‘normalizador de texto’ del español para tratar correctamente otros fenómenos como los observados en ‘palabra con signo’, ‘palabra con símbolo’, ‘palabras con dígitos’, ‘abreviaturas’, ‘emoticonos’ y ‘letra aislada’.

La corrección de textos no normativos es optativa, por tanto, si el usuario escoge la normalización de textos normativos, el texto es expandido siguiendo el mismo funcionamiento que TexAFon 1.0. Si, por el contrario, el usuario activa la normalización de textos no normativos, aparte de realizar la expansión de las palabras, también se realiza una corrección completamente automática del texto, es decir, sin la intervención del usuario.

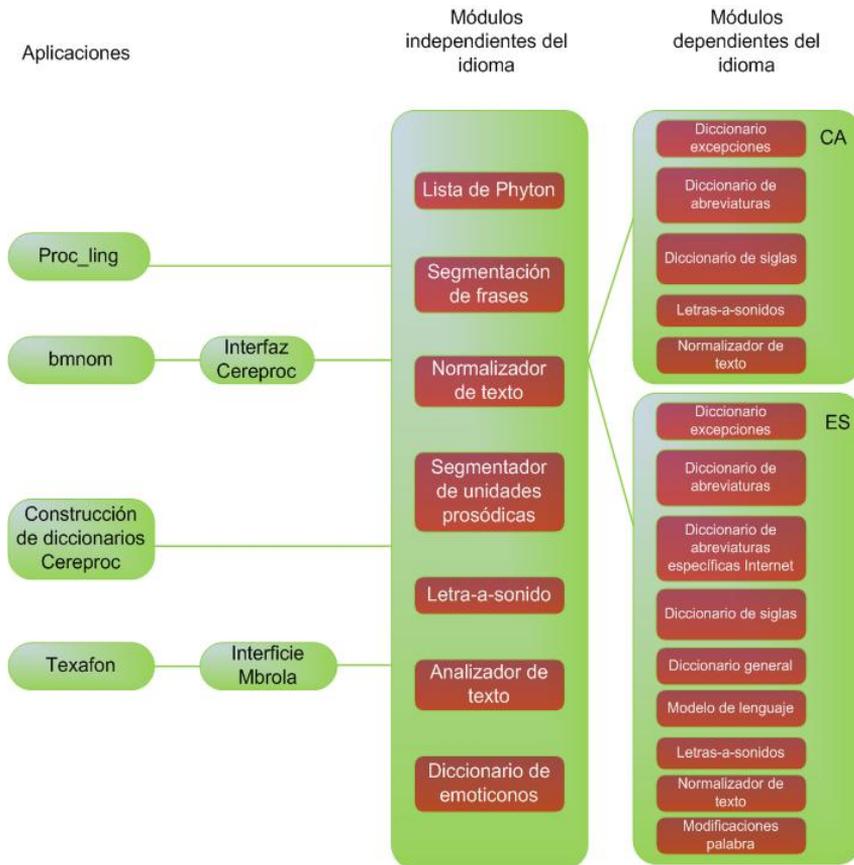


Figura 15: Arquitectura general de TexAFon 2.0

Finalmente, para poder gestionar adecuadamente los casos de ambigüedad en el proceso de normalización, se ha desarrollado e integrado un modelo de lenguaje basado en bigramas, específico para las conversaciones virtuales.

El proceso de normalización se divide en tres etapas diferentes:

1. **Clasificación de tokens:** Cada uno de los tokens de entrada, según las características que presentan, son identificados como pertenecientes a una de las 26 posibles etiquetas mencionadas anteriormente en la Tabla 6.
2. **Expansión de los tokens:** Dependiendo de cómo hayan sido etiquetados los tokens anteriormente se llevará a cabo

diferentes procesos para obtener la normalización y expansión de cada uno de ellos.

3. **Selección de la forma normativa:** Si durante el proceso de normalización se ha localizado más de una posibilidad de palabra normalizada, se lleva a cabo un proceso de desambiguación mediante el empleo de un modelo de lenguaje específico de las conversaciones virtuales para seleccionar la palabra más apropiada según el contexto en el que se encuentra.

Cada una de estas fases se comenta más ampliamente a continuación.

3.3.1. Clasificación de tokens

Manteniendo el mismo funcionamiento que en TexAFon 1.0, los tokens del texto de entrada, antes de ser normalizados y expandidos, son clasificados según la categoría a la que pertenecen. Así, por ejemplo, los tokens que combinan grafías y dígitos se identifican como ‘cifras_y_letras’, y los que presentan grafías y símbolos son etiquetados como ‘cadena_con_símbolo’. Los tokens que no pertenecen a ninguna categoría del inventario se identifican como ‘palabra’, haciendo que esta categoría sea la más frecuente.

A pesar de que este listado es amplio y suficiente para tratar texto normativo, en el caso de las conversaciones virtuales ha sido necesario ampliarlo añadiendo las etiquetas ‘emoticono’ y ‘risa’ para que las cadenas que pertenezcan a estas dos categorías no sean confundidas con otras como ‘cadena_con_signos’ o ‘palabra’, y a consecuencia de ello ser mal expandidas.

En ambos casos, el sistema es capaz de identificar un emoticono o una expresión de risa que contenga un número variable de repeticiones tanto de uno como de varios de los caracteres que lo conforman.

Para identificar correctamente los emoticonos fue imprescindible la adaptación del módulo ‘segmenta_frase’. Este módulo, como se ha comentado en el apartado 3.1. *El normalizador de textos TexAFon*,

se basa en los signos de puntuación para separar las frases de un texto. Los emoticonos están formados también por signos de puntuación combinados con otros caracteres ASCII, siendo este aspecto precisamente el que causa problemas en el correcto funcionamiento del módulo. Ante el enunciado “yo quiero ser feliz :_(“, TexAFon indicaría que está formado por dos frases “yo quiero ser feliz : “ y “_(“, desvirtuando el mensaje original que el usuario quiso transmitir cuando lo escribió. Actualmente, en TexAFon 2.0, se comprueba si el signo de puntuación forma parte de un emoticono, y si es así, el emoticono no se fragmenta, se mantiene y se continúa la búsqueda de otro signo de puntuación que sí corresponda con un final de oración. De este modo no se producen truncamientos entre fragmentos de una misma oración, ni se expanden elementos incoherentes como sería ‘guión bajo abre paréntesis’ (_() del ejemplo anterior.

Cuando este módulo identifica un emoticono, realizando una consulta en el diccionario de emoticonos, éste es marcado con la etiqueta ‘[emoticono]’ y de este modo no es leído por el sistema. Actualmente no se intenta interpretar qué tipo de información desea transmitir el usuario al emplearlo ya que esto requiere un análisis textual y pragmático más complejo que el realizado en esta tesis. Sería necesario un módulo que clasificara los emoticonos según la funcionalidad con la que se han transcrito (indicadores de sentimientos y emociones, indicadores de ironía y humor, indicadores de complicidad, preservadores de la imagen o amenazadores de la imagen), y en el caso de indicar una emoción sería necesario identificar de cuál se trata. Aun así, sería interesante tratarlos en un futuro ya que a través de su uso se podría identificar la emoción o emociones que desea transmitir el usuario al incluirlo en su mensaje y realizar la lectura del enunciado considerando esta información aportada automáticamente.

En la etapa de clasificación, si el usuario escoge la normalización con corrección, pensada para textos de Internet, no normativos, el número de grafías que se consideran ‘letra aislada’ se ve reducido. Se han excluido de la lista las grafías ‘d’, ‘k’, ‘l’, ‘m’, ‘n’, ‘q’, ‘s’, ‘t’, y ‘x’, que son precisamente las abreviaturas formadas por una única grafía que los usuarios emplean con mayor frecuencia para economizar tiempo. Ahora, estas grafías están incluidas en el diccionario de abreviaturas específico de las conversaciones virtuales junto a sus correspondientes expansiones.

3.3.2. Expansión de los tokens

En la aproximación seguida aquí, el proceso de ‘corrección’ o normalización se ve como una ampliación del concepto de normalización aplicado a estos textos ya correctos: los diferentes elementos ‘anómalos’ del texto de las conversaciones virtuales se tratan como casos específicos. Ello ha obligado a modificar una serie de procedimientos de expansión de tokens ya existentes en TexAFon: cadena_con_símbolos, cadena_con_signos, cifras_y_letras, abreviaturas, y sobre todo, palabra.

El motivo por el cual se ha decidido realizar una adaptación de los diferentes procesos de expansión es que la actual organización de TexAFon permite normalizar mejor las diferentes anomalías observadas para cada categoría, ya que facilita la aplicación de diferentes estrategias según las características de cada token. Al limitar las modificaciones a realizar según la categoría a la que pertenece el token de entrada, no hay una interferencia entre los diferentes procesos, haciendo que el número de normalizaciones posibles quede más acotado, y reduciendo así el número de posibles correcciones erróneas. Por ejemplo, ante el token ‘to2’, si no se realizase una primera clasificación y una expansión acorde con el tipo de categoría con la que ha sido etiquetada, se iniciaría un proceso de normalización que implicaría la sustitución de cada uno de los caracteres por otros, eliminación de grafías, inserción de grafías, y transposición de los diferentes caracteres ralentizando el proceso de normalización y posiblemente finalizando sin obtener la normalización correcta. En cambio con el planteamiento desarrollado en esta investigación, se reduce el tiempo del procesado, y se obtiene la normalización deseada sin grandes contratiempos.

3.3.2.1. Tratamiento de palabras con símbolos

En el corpus se puede encontrar cadenas como `amig@s`, `ignorad@` o `Bar$a`. En la versión anterior, estos tokens serían expandidos

como dos o más palabras: una cadena más la expansión del símbolo: ‘amig arroba ese’, ‘ignorad arroba’ o ‘bar dólares a’. Estos tokens son identificados y etiquetados como ‘Cadena_con_símbolos’.

Antes de expandir los símbolos presentes en la palabra, se realiza una consulta en el diccionario de abreviaturas específicas de las conversaciones virtuales (diccionario_abreviaturas_chat), y si la palabra se encuentra en dicho diccionario se toma la expansión propuesta en este diccionario. Si por el contrario la palabra no está en el diccionario consultado, la expansión se realiza como en TexAFon 1.0 expandiendo el símbolo.

3.3.2.2. Tratamiento de palabras con signos

Tal y como se ha explicado en el subapartado 2. *Eliminación de caracteres*, la unión de dos palabras al no introducir un espacio en blanco después del signo de puntuación y convirtiéndola en una sola, es un fenómeno muy recurrente en los mensajes de las conversaciones virtuales. TexAFon los procesa como un token del tipo ‘Cadena_con_signos’. En la primera versión del normalizador, estos signos de puntuación también eran expandidos, como si se tratara de una parte más de la lectura del texto. El token ‘bueno...mañana’ en la versión 1.0, la original, se expandiría de la siguiente manera: ‘bueno punto punto punto mañana’. Esta forma de procesar el texto no es adecuada para el texto de las conversaciones virtuales, ya que de este modo la voz sintética también leería los puntos. En la versión 2.0, los puntos suspensivos se identifican y se marcan como una etiqueta de signo de puntuación ‘bueno [etiqueta_puntos_suspensivos] mañana’, de este modo, el locutor sintético no la pronuncia, sólo realiza una pausa entre ambas palabras.

3.3.2.3. Tratamiento de palabras con dígitos

En las conversaciones virtuales analizadas se ha observado el empleo de números en sustitución de las vocales de una palabra. Este fenómeno se da en el 2,2% de las veces en relación al número total de errores/desviaciones del corpus. No todas las vocales pueden ser sustituidas por números, ni todos los números pueden ser interpretados como vocales. En el corpus se ha encontrado la distribución recogida en la Tabla 19.

VOCAL DÍGITO	
A	4
E	3
I	1
O	0

Tabla 19: Equivalencias de vocales y dígitos empleadas en el corpus de análisis

Este tipo de cadenas se etiquetan en TexAFon como ‘Cifras_y_letras’, por lo que su expansión se ha tratado dentro del procedimiento que expande estos tokens. En esta nueva versión del normalizador, si la cadena formada por letras y números contiene cualquiera de los cuatro dígitos de la tabla, se sustituyen los dígitos por la vocal correspondiente, acentuada y sin acentuar, y se coteja en el diccionario general. Si la palabra se encuentra en el diccionario se considera que la normalización ha sido correcta y se devuelve la nueva palabra; si por el contrario la palabra no se encuentra en el diccionario después de las sustituciones pertinentes, la expansión se realiza como en la versión 1.0, expandiendo las grafías y los dígitos con la diferencia de que esta expansión se

realiza como si la palabra resultante fuera una y no varias. Por ejemplo, ante ‘to2’, como el número 2 no equivale a ninguna vocal posible, se realiza la expansión de los caracteres como si se tratara de una única palabra ‘todos’ y se comprueba en el diccionario para decidir si la expansión ha sido correcta o no. En el supuesto caso de que ‘todos’ no estuviera incluida en el diccionario general se procedería a realizar la expansión como si de dos palabras diferentes se tratara (‘to dos’).

3.3.2.4. Tratamiento de abreviaturas

En el corpus de análisis se ha observado un uso elevado de abreviaturas específicas de las conversaciones virtuales, como por ejemplo ‘xk’ (porque), ‘anq’ (aunque) o ‘aptc’ (apetece).

El tratamiento para estos casos es el mismo que se ha empleado en el caso de las abreviaturas de TexAFon 1.0. con dos diferencias. Como en la versión anterior, se comprueba si la palabra en cuestión se encuentra en el diccionario específico y se toma la expansión propuesta para dicha palabra. En la normalización de las conversaciones virtuales, los diccionarios consultados son dos en vez de uno. A parte del diccionario de abreviaturas estandarizadas, se creó un diccionario específico para estos casos: el ‘diccionario de abreviaturas específicas del lenguaje de Internet’. Este diccionario, tal y como se ha comentado con anterioridad, recoge abreviaturas específicas del lenguaje 2.0.

La segunda novedad reside en el número de posibles normalizaciones propuestas para cada palabra. Mientras que en el diccionario de abreviaturas sólo hay una posible normalización para cada una de las palabras que conforman el diccionario, en el específico para el lenguaje de Internet puede presentar más de una propuesta, como es el caso de ‘n’ que puede expandirse por ‘en’, ‘ni’, o ‘no’. En estos casos es necesario recurrir a una tercera fase, en la que se seleccione el mejor candidato (véase el epígrafe 3.3.3. *Selección de la forma normalizada*).

Si la oración de entrada fuera ‘porke no t cambias el nombre poseidon’, ‘t’ sería etiquetada como una abreviatura ya que se encuentra en el diccionario de abreviaturas específico del lenguaje

de Internet. Esta abreviatura en concreto presenta tres posibles normalizaciones: ‘te’, ‘tu’ y ‘tú’. En esta etapa del proceso, se toman las tres propuestas que aparecen en el diccionario, y posteriormente en la tercera fase de la normalización y con la ayuda de un modelo de lenguaje se selecciona la expansión más adecuada teniendo en cuenta el contexto en el que se encuentra (en este caso concreto debería ser ‘te’).

3.3.2.5. Tratamiento de palabras

La mayoría de los fenómenos considerados se dan en tokens que TexAFon etiqueta como ‘palabra’, como es el caso de ‘cierto’, ‘besito’, ‘olaaa’, o ‘miembro’. El proceso de expansión de estos tokens es más complejo que los anteriores, e incluye varias fases:

1. Determinación de si se trata de una palabra normalizable: no todos los tokens etiquetados así debían ser sometidos a los procesos de normalización implementados, por lo que el primer paso debía ser determinar si se trata de una palabra que debía normalizarse o no.
2. Generación de posibilidades de expansión, haciendo llamadas a diferentes procedimientos de expansión.

a. Identificación de palabras normalizables

En este caso, una vez más, ha sido necesario definir qué se va a considerar correcto y que no. Para ello se ha seguido el mismo procedimiento que en el primer análisis lingüístico: se considera incorrecta o errónea toda aquella cadena que no se encuentra ni en el diccionario general ni en el diccionario de abreviaturas de chat.

A diferencia de Borrell *et al.* (2008), no se ha podido excluir todas aquellas palabras escritas con mayúsculas para evitar así la corrección de palabras correctas como los nombres propios, ya que en las conversaciones virtuales los usuarios no suelen utilizar

mayúsculas cuando escriben, a excepción de querer connotar enfado o que el usuario está gritando, y en este caso todo el enunciado puede aparecer en mayúsculas, y los nombres propios aparecen muy frecuentemente en minúsculas.

Esta decisión presenta un claro inconveniente, habrá palabras que se considerarán correctas ya que existen y se encuentran en el diccionario, pero por el contexto en el que se dan posiblemente sean incorrectas, como es el caso de ‘mas’ en la oración ‘q dientes mas bonicos tienes venus y mas blancos’, o el de ‘esta’ en ‘esta claro que el primer maestro no tuvo maestro’. Ambas palabras son correctas en español, pero en el contexto en el que se encuentran no están bien escritas, ya que ambas deben estar acentuadas. A pesar de ser incorrectas, el hecho de existir en la lengua española y aparecer en el diccionario de trabajo hace que el módulo no las identifique como incorrectas, y por lo tanto no se normalicen. Se acepta esta pequeña desventaja a cambio de no intentar corregir todas las palabras, incluso las correctas, y generar numerosas incorrecciones en el texto.

b. Generación de tokens normalizados

A partir del análisis lingüístico realizado al corpus de conversaciones virtuales, se ha diseñado y creado el módulo ‘modificaciones_palabra’, similar al propuesto por Borrell *et al.* (2008) y Armenta *et al.* (2003 b), que realiza las correcciones oportunas a cada palabra: eliminación de caracteres repetidos, sustitución de caracteres, inserción, etc.

El orden en el que se realizan las diferentes modificaciones es el siguiente:

1. Eliminación de grafías repetidas

Como se ha comentado anteriormente en el capítulo 2, los usuarios de los chats emplean la repetición de grafías para intentar transmitir los rasgos prosódicos propios de la conversación oral. Esta

repetición puede afectar a una grafía o a varias de una misma palabra, y el número de repeticiones puede ser variable. Este fenómeno, aun siendo relevante para la posterior oralización del texto, ya que estas repeticiones corresponden a alargamientos vocálicos que deberían estar presentes en la señal acústica de salida, para la correcta normalización del texto se ha optado por su eliminación. Esta información prosódica podría guardarse en una variable para el posterior modelado de la señal acústica, pero actualmente es un aspecto que no se trata, pero que sí podría llevarse a cabo en un futuro sin grandes complicaciones.

Dado que cualquier palabra puede verse afectada por este fenómeno, y ante la gran variabilidad en el número de grafías afectadas, y el número de reiteraciones que puede mostrar cada una de ellas (hooollllaaa, hoooolaaa, hoolaaaaa, holaaaaaaaaa) es imposible realizar un diccionario con todas las posibles variantes de cada palabra. Se ha optado por la creación de un script que comprueba si la palabra tiene caracteres repetidos, poniendo especial atención en combinaciones correctas de dos grafías iguales como por ejemplo 'rr' (carro), 'll' (calle), 'aa' (Aarón), 'ee' (leen), 'cc' (acción), 'oo' (coopera), 'ii' (miniinvasivo), 'mm' (gamma), y los elimina si se da el caso.

A pesar de que este fenómeno no ha sido el que mayor número de casos ha mostrado en el análisis lingüístico, se ha decidido tratar en primer lugar para evitar modificaciones innecesarias y posibles normalizaciones incorrectas. En muchos casos, las palabras que presentan repetición en alguna de sus grafías no muestran otro tipo de desviación añadida, y por lo tanto, eliminando estas reiteraciones se obtiene una palabra correcta sin necesidad de realizar más modificaciones.

Una vez eliminadas las grafías repetidas, se comprueba la palabra resultante en el diccionario. Si la palabra se encuentra en él, se considera correcta y finaliza el proceso de corrección. Si por el contrario, la palabra no se encuentra en el diccionario, se debe continuar con la corrección.

2. Tratamiento de los fenómenos más frecuentes

Gracias al análisis del corpus se ha observado que más de mitad de los tokens, exactamente el 56%, muestran algún problema relacionado con la acentuación o la sustitución sistemática de grafías. Por este motivo se ha decidido tratar estos dos fenómenos justo después de la eliminación de caracteres y antes de las modificaciones más generales. Con esta manera de proceder se logra normalizar correctamente un elevado número de casos, sin aumentar las posibles normalizaciones ni creando mayor ambigüedad al sistema, y por lo tanto, también agiliza el proceso ya que si se encuentra una posible normalización no se realizan más búsquedas.

El flujo de funcionamiento del procedimiento es el siguiente:

1. Recorriendo el token de izquierda a derecha, se va comprobando si cada una de las grafías se corresponde con una vocal, acentuada o no, o con una de las grafías cuya sustitución se realiza sistemáticamente (q, k, x, w, y, i).
2. Una vez localizadas se realizan las sustituciones correspondientes, dependiendo en cada caso de la grafía que se modifica, realizando una única modificación cada vez. En el caso de que el normalizador se encontrara con la palabra 'koxe' se generarían las siguientes posibles normalizaciones: 'coxe', 'kóxe', 'koche', 'koxé', guardando cada una de ellas en una lista.
3. Una vez finalizadas todas las sustituciones posibles se comprueba si alguna de las normalizaciones propuestas por el sistema está en el diccionario general o en el de abreviaturas específicas del lenguaje de Internet. Si una o varias palabras aparecen en los diccionarios, se finaliza el proceso de normalización-expansión; si, por el contrario, ninguna de las palabras se encuentra en el diccionario, se vuelve a realizar una sustitución de las vocales (no acentuadas por acentuadas y viceversa) y de las grafías sistemáticas, pero esta vez las modificaciones no se realizan

en la palabra original del texto sino en cada una de las palabras que han resultado de la primera modificación. Continuando con el ejemplo anterior de 'koxe', se tomaría la primera palabra modificada 'coxe' y se le realizarían las modificaciones oportunas: 'cóxe', 'coche', 'coxé'. De este mismo modo se actuaría sobre cada una de las palabras restantes. Al final del proceso se obtendría la siguiente lista: ['cóxe', 'coche', 'coxé', 'cóxe', 'koxe', 'kóche', 'kóxé', 'coche', 'kóche', 'koxe', 'koché', 'coxé', 'kóxé', 'koché', 'koxe']. Se eliminan las posibles palabras repetidas y se comprueba cada una de ellas en los diccionarios. Si alguna de ellas se encuentra en el diccionario se finaliza el proceso de normalización; si por el contrario no se ha hallado ninguna normalización correcta se pasa al tercer y último paso.

3. Modificaciones generales

El siguiente paso toma la palabra original y todas las palabras generadas a partir de la sustitución de acentuación y grafías sistemáticas. A cada uno de estos tokens se le realiza los siguientes procesos:

1. Sustitución: Se recorre toda la palabra de derecha a izquierda, sustituyendo cada una de las grafías por las del abecedario. La cadena resultante se comprueba en el diccionario, y si se encuentra en él se guarda en una lista.
 - Sen: aen, ben, cen, den, een, (...) san, sbn, scn, sdn, etc.
2. Eliminación: Recorriendo la palabra de izquierda a derecha, se elimina cada vez una grafía de la palabra. La nueva palabra es cotejada en el diccionario, realizando el mismo procedimiento que en el caso de la sustitución.
 - Colpa: olpa, clpa, copa, cola, colp
3. Inserción: De derecha a izquierda, se añade cada una de las grafías del abecedario y el espacio en blanco en cada una de

las posiciones de la palabra. Cuando se inserta el espacio en blanco, se comprueba si las dos palabras en las que se ha dividido la palabra inicial están en el diccionario. Solamente se considerará correcta si ambas están en el diccionario. Nuevamente, del mismo modo que se ha realizado en la sustitución y la eliminación, las palabras correctas son guardadas en la lista de modificaciones correctas.

- Poblema: apoblema, bpoblema, (...) p oblema, paoblema, pboblema, etc.
4. Transposición: También de izquierda a derecha, se cambia el orden de dos grafías consecutivas. Se sigue el mismo procedimiento que en los casos anteriores para comprobar y guardar las palabras correctas.
- Adisó: daisó, aidsó, adsió, adió.

Una vez que la última palabra de la lista de palabras que modificar ha sido tratada, se toma la lista de modificaciones correctas que se ha generado en este último paso. Dependiendo del contenido de esta lista se actuará de un modo o de otro:

1. La lista está vacía. A pesar de todas las modificaciones realizadas, no se ha logrado una normalización que se encuentre en los diccionarios de trabajo. Se dejará la palabra original del texto, sin ninguna modificación.
2. La lista contiene una única palabra. El proceso de modificaciones generales sólo ha obtenido una palabra que ha considerado correcta. Esta es la palabra que se tomará como correcta y aparecerá en la versión del texto normalizado.
3. La lista contiene varias palabras. Durante el proceso de expansión se puede generar más de una posible normalización para un único token. Para decidir qué palabra será la correcta es necesario recurrir al tercer nivel del proceso de normalización, el modelo de lenguaje, para que realice la selección.

3.3.3. Selección de la forma normalizada

Existe la posibilidad de que durante el proceso de expansión se haya encontrado más de una normalización para un token del texto de entrada. Para poder resolver las ambigüedades producidas por este incremento de posibles correcciones se utiliza un modelo de lenguaje creado específicamente para las conversaciones virtuales.

El hecho de que el texto original de estas conversaciones pueda presentar desviaciones respecto al lenguaje normativo hace que el empleo del modelo de lenguaje sea algo más complejo que un modelo de lenguaje convencional, ya que se ha de decidir con qué tipo de texto se realizará el modelo de lenguaje (el corpus original o el corpus normalizado) y en qué momento del proceso se recurrirá a él y cómo.

En este caso, como el modelo de lenguaje va a ser empleado al final del proceso de normalización, y solo en el caso de que haya más de una posible normalización, el modelo de lenguaje se ha creado a partir del texto normalizado manualmente, de forma automática mediante una serie de scripts de Python. Cada uno de los tokens que conforma el corpus de análisis fue emparejado con la palabra siguiente que aparecía en los mensajes. También se tuvo en cuenta la posición de inicio y final de cada mensaje marcándola con las etiquetas [inicio] y [final] respectivamente. Posteriormente, y también de forma automática gracias a un script de Python creado para este fin, se realizó un recuento del número de veces que había aparecido una misma tupla en el corpus. Finalmente, para cada una de las palabras del texto se creó una lista con todas las palabras, ordenadas alfabéticamente, con las que había formado tupla, y el número de veces que ocurría dicha combinación. La Figura 16 muestra una serie de ejemplos extraídos del modelo de lenguaje. Como se puede observar, al inicio de cada línea aparece una palabra, como por ejemplo 'guapo', y a continuación todas las palabras a las que ha precedido en el corpus, en este caso concreto 'guapo' se ha encontrado a final de enunciado ('[final]') en cuatro ocasiones, ha aparecido delante de un signo de cierre de exclamación una vez, de puntos suspensivos también una vez, y del pronombre 'tú' otra vez.

guapete qué 1
 guapo [final] 4 [signo_cierra_exclamacion]1 [signo_puntos_suspensivos] 1 tú 1
 guapos [signo_puntos_suspensivos] 1
 guapísima [signo_cierra_interrogacion] 1
 guapísimo [final] 1 [signo_cierra_exclamacion]2 [signo_cierra_interrogacion] 1
 guardado porque 1
 guardao [final] 1
 guardó [signo_puntos_suspensivos] 1
 guarrete [final] 1
 guau [final] 1
 guay [final] 9 [signo_cierra_exclamacion]2 [signo_cierra_interrogacion] 1 [signo_coma]1 el 2 espera 1
 guerra[emoticono] 1 [final] 1 [signo_cierra_interrogacion] 1 ahora 1 entre 1 mundial 2 y 1
 guerrera de 1 eres1 jajaja 1 nada 1 nas 1 qué 3 ya 1
 guggenheim también 1
 guillotina y 1
 guindoo 1

Figura 16: Ejemplos extraídos del modelo de lenguaje específico de las conversaciones virtuales

Para realizar la búsqueda de la normalización con mayor probabilidad de ser la correcta, se busca en el modelo de lenguaje la palabra normalizada anterior a la tratada del texto de entrada. Para ello, previamente, en la etapa de expansión, se ha guardado dicha palabra en una variable denominada ‘palabra_anterior’. Es importante el hecho de guardar la palabra en una variable, ya que el modelo de lenguaje ha sido creado a partir del corpus normalizado, y la lista de palabras candidatas normalizadas también están normalizadas. Si se intentara buscar la palabra original del texto en el modelo de lenguaje en muchas ocasiones el resultado sería infructuoso, provocando que la selección de candidatos no fuera satisfactoria, y por lo tanto también repercutiría negativamente en el resultado final de la normalización.

Una vez localizada la palabra anterior en el modelo de lenguaje, se comprueba si las palabras candidatas propuestas por la normalización están incluidas en la lista de palabras con las que dicha palabra había formado una tupla en el corpus de análisis, y se selecciona la palabra cuya frecuencia de aparición es mayor. Si por el contrario, ninguna de las posibles normalizaciones se encuentra en la lista, se selecciona la primera palabra modificada. Lo mismo ocurre cuando la palabra anterior no se encuentra en el modelo de lenguaje.

En la Tabla 20 se muestra un ejemplo del proceso de normalización realizado en esta investigación.

Enunciado original	Eso k t llevas	Cuando tngas 40
Identificación de los tokens	Eso [palabra] K [abreviatura] T [abreviatura] Llevas [palabra]	Cuando [palabra] Tngas [palabra] 40 [cifra]
Expansión de los tokens	Eso: [eso] K: [que, qué] T: [te, tu, tú] Llevas: [llevas]	Cuando [cuando] Tngas [angas, ingas, ingás, tajás, tangas, tanjas, tejas, tengas, tengo as, tijas, tingas, togas, tojas, tongas, tugas, tujas, tungas, unjas, vengas] 40 [cuarenta]
Selección de la forma normalizada	Eso: [eso] K: [que , qué] T: [te , tu, tú] Llevas: [llevas]	Cuando [cuando] Tngas [angas, ingas, ingás, tajás, tangas, tanjas, tejas, tengas , tengo as, tijas, tingas, togas, tojas, tongas, tugas, tujas, tungas, unjas, vengas] 40 [cuarenta]
Enunciado final	Eso que te llevas	Cuando tengas cuarenta

Tabla 20: Dos ejemplos del funcionamiento, paso a paso, de TexAFon 2.0

3.4. Resumen

Las peculiares características lingüísticas que presentan las conversaciones virtuales, en las que parece reinar un caos ortográfico absoluto entre sustituciones de grafías, eliminación de caracteres, etc., y en los que la única regla que prevalece es la de poder hablar, hacen que sea necesario plantearse una serie de cambios en los sistemas de procesado de lenguaje natural empleados hasta el momento para poder tratar estos textos de forma correcta.

El módulo de procesado del texto de un conversor texto a habla, cuya función es expandir todos los elementos no ortográficos del texto, transcribirlo fonéticamente y extraer del texto el máximo de información posible (morfológica y sintáctica principalmente, aunque también podría ser posible un análisis semántico o pragmático) con el objetivo de que el resultado obtenido mediante la lectura de la voz sintética sea el más parecido a la realizada por un ser humano, también presenta severos problemas al intentar tratar los mensajes de las conversaciones virtuales.

Uno de los objetivos de esta investigación se centra en crear un normalizador para el módulo de procesamiento del texto, compatible con el conversor de texto a habla de Cereproc, capaz de tratar correctamente texto no normativo. Para ello, se ha tomado como base TexAFon (Garrido *et al.*, 2012), que, como la gran mayoría de los procesadores existentes, está orientado a tratar texto normativo, y presenta ciertos problemas al enfrentarse a los mensajes de las conversaciones virtuales, llegando a producir severos problemas que afectan a la inteligibilidad y comprensión de los enunciados por parte de los oyentes. Esto hace que sea indispensable realizar una serie de modificaciones en la herramienta TexAFon que permita la correcta normalización de todo tipo de textos, tanto normativos como no normativos.

Para lograr este objetivo se ha escogido abordar el problema desde una perspectiva de corrección de textos basada en reglas por tres motivos:

1. Al basarse en la presunción de que la mayoría de las palabras del texto son correctas, únicamente se modifican aquellas palabras que no se encuentran en el diccionario, reduciendo considerablemente el número de falsos positivos.
2. No es necesario disponer de grandes corpus para llevar a cabo la corrección. A partir de un corpus como el de esta tesis se puede extraer reglas lo suficientemente generales para llevar a cabo el proceso de una forma óptima. Las reglas pueden perdurar en el tiempo sin necesidad de modificaciones continuas como consecuencia de la rápida evolución del lenguaje 2.0.
3. Esta aproximación permite aprovechar parte del conocimiento lingüístico y del proceso de normalización del sistema TexAFon. Identificando el tipo de token al que corresponde cada cadena del fichero de entrada, y tratándola según sus características se puede obtener mejores resultados finales.

Para poder diseñar las modificaciones necesarias para tratar correctamente los mensajes de las conversaciones virtuales es imprescindible conocer con exactitud todos y cada uno de los fenómenos lingüísticos que tienen lugar en ellas. Son muchos los autores que han analizado el lenguaje empleado en las conversaciones virtuales (Mayans, 2000, 2002, 2002b; Torres, 1999; Yus, 2001 y 2010; Pons, 2002; López Quero, 2003; Sanmartín, 2007; Kalsson 2007), pero ninguno de ellos ha proporcionado datos cuantitativos al respecto. Por este motivo se han realizado dos análisis lingüísticos sobre un corpus de conversaciones virtuales recogido con este fin.

El primer análisis se realizó sin conocer a priori los fenómenos que podían darse en estos textos, detallando las anomalías observadas en las palabras identificadas como no correctas. Al observar que determinados tokens podían ser tratados por TexAFon, como por ejemplo las direcciones de correo electrónico, o los dígitos, se decidió analizar más en detalle los tokens que no podían ser normalizados y expandidos de forma adecuada.

Este segundo análisis ha permitido observar que algunas palabras de las conversaciones virtuales pueden llegar a presentar más de cuatro desviaciones, siendo más frecuente que sólo muestren una o dos.

Los fenómenos más frecuentes observados han sido en primer lugar los relacionados con la sustitución de grafías, en los que predominan los problemas de acentuación y la sustitución sistemática de una serie de grafías como ‘k<-> c / qu/ que / qué’, o ‘x <-> por / ch / per /par’, entre otras grafías, y en segundo lugar los que atañen a la eliminación de caracteres, seguidos de la inserción de grafías, y finalmente la transposición de caracteres.

Tras la obtención de estos datos se decidió que el proceso de corrección de los tokens del texto de entrada se realizaría de manera simultánea a la expansión, y no como una etapa previa a éste como en otras investigaciones similares se ha realizado (Armenta et al., 2003; Moré et al., 2005). Aprovechando la organización de TexAFon, se realizó una adaptación del mismo para tratar las conversaciones virtuales: TexAFon 2.0. Se mantienen las dos etapas de identificación y expansión de tokens, y las modificaciones recaen sobre varios puntos del proceso como son:

1. La creación de una serie de diccionarios, en concreto tres: diccionario general, diccionario de abreviaturas específicas de las conversaciones virtuales, y un diccionario de emoticonos.
2. La adaptación de una serie de módulos como ‘emoticono’, ‘letra aislada’, ‘palabra con signo’, ‘palabra con símbolo’, ‘palabras con dígitos’, y ‘abreviaturas’.
3. El desarrollo de un nuevo módulo ‘modifica_palabra’ que realiza las modificaciones pertinentes sobre los tokens etiquetados como ‘palabra’, y que no se encuentren en el diccionario general, hasta obtener una normalización posible.
4. La creación de una tercera etapa que permite la selección de la mejor normalización en el caso de que el normalizador halle más de un posible candidato.

La realización de un modelo de lenguaje específico para este tipo de textos necesario para llevar a cabo la tercera etapa del proceso

Capítulo 4

GENERACIÓN DE LA VOZ SINTÉTICA EXPRESIVA

La generación de habla expresiva se ha abordado desde una doble perspectiva: por un lado se pretende evaluar el potencial de las técnicas de síntesis por corpus para la generación de habla emotiva a partir de un corpus oral cuyo tamaño es reducido, y por otro lado, evaluar el uso de modelos prosódicos específicos para las emociones, que ante unas mismas condiciones puede aportar mejoras en la generación de este tipo de voces. Para ello, se ha tomado como base el motor de síntesis de Cereproc.

En ambos casos, el punto de partida escogido fue el mismo: un corpus grabado por un locutor profesional con un número reducido de enunciados representativos de las emociones que se pretenden modelar. El objetivo es investigar hasta qué punto las técnicas basadas en conocimiento lingüístico permiten obtener habla expresiva natural a partir de este tipo de material en un sistema comercial basado en corpus.

Como ya se ha comentado en el estado de la cuestión, la selección del locutor es un aspecto fundamental para obtener una voz sintética de calidad cuando se usan técnicas basadas en corpus. En este caso, como el corpus creado es representativo del dominio, y contiene enunciados emocionales, el locutor profesional que ha de realizar la grabación, aparte de tener buena voz y una buena dicción, también ha de poseer unas dotes interpretativas para representar las emociones del corpus de la manera más natural posible. Por ello, para asegurarnos que se escogía al candidato más idóneo se realizó un casting, que se describe en el apartado 2.

A continuación se comenta el proceso previo llevado a cabo para obtener la voz sintética del locutor masculino para el español, Edu. Se explican los diferentes procesos que se han realizado para crear

el corpus oral con el que generar la voz sintética, y los parámetros prosódicos que se han analizado para posteriormente modelar la voz.

4.1. El corpus

El corpus de habla recogido para esta tesis tenía una doble finalidad:

1. Servir de material de base para la creación de un locutor sintético apto para la generación de habla sintética expresiva utilizando técnicas basadas en corpus.
2. Servir de material para el desarrollo de modelos prosódicos paramétricos especialmente orientados a la generación de habla expresiva.

4.1.1. Contenido del corpus

4.1.1.1. Corpus neutro

La creación de un locutor sintético para un CTH que utiliza técnicas basadas en corpus implica disponer de un corpus amplio, con una cobertura fonética de la lengua suficiente que permita la generación de cualquier tipo de enunciado, ya sea neutro o emocional. Por ello, además del corpus emocional también es necesario disponer de un corpus neutro, que permitirá tanto la generación de enunciados neutros como la comparación de diversos parámetros prosódicos para la posterior modelación de los enunciados emocionales.

El corpus neutro empleado en este trabajo está formado por 3.060 enunciados, y contiene:

- a) direcciones, tanto postales como electrónicas,

- b) números,
- c) letras,
- d) frases interrogativas y exclamativas,
- e) noticias,
- f) elementos paralingüísticos como risas, lloros e interjecciones.

Fue realizado de manera conjunta por la empresa Cereproc y Barcelona Media para la construcción de otros locutores en español compatibles con el motor de síntesis de Cereproc. Para la selección de los enunciados se empleó un algoritmo *greedy* que asegura la cobertura fonética en todos los contextos, es decir, es posible generar cualquier enunciado posible en español ya que el corpus contiene todos los segmentos necesarios para ello.

4.1.1.2. Corpus emocional

Del mismo modo que las conversaciones virtuales presentan continuamente enunciados expresivos en los que se transmiten emociones u opiniones, el corpus oral que se ha de crear para la generación de la voz sintética también ha de contener este tipo de enunciados. Debe ser representativo de las emociones que coexisten en las conversaciones virtuales. Sería ideal que el corpus emocional contenga un gran número de enunciados de todas y cada una de las emociones que aparecen o puedan aparecer en una conversación virtual, pero esto es una tarea inviable. Aún así, sí debe recoger aquellos estados emocionales más frecuentes, ya sean básicos, subyacentes, secundarios, o aparezcan entremezclados junto con otros.

El número de enunciados para cada emoción también es un aspecto importante. Cuanto mayor sea el número de enunciados representativo de cada emoción mejores resultados se pueden obtener tanto en el análisis de los parámetros prosódicos como en la síntesis, ya que, recordemos, los CTH basados en síntesis por

corpus necesitan grandes bases de datos acústicos para garantizar buenos resultados.

Además de la cantidad de enunciados, también se debe tener en consideración la cobertura fonética y prosódica, se debe garantizar una frecuencia mínima de aparición de todos los fonemas del idioma, y la presencia de oraciones enunciativas, exclamativas e interrogativas.

En esta investigación, la cobertura fonética de este subcorpus no ha sido un requisito que se ha tenido en cuenta, ya que nuevamente no es viable disponer de una cobertura fonética para cada una de las diferentes emociones seleccionadas, ya que a menudo el número de enunciados etiquetados para cada emoción es reducido y variable, oscilando entre los 949 y 1 enunciados. Aun sabiendo que es probable que esto pueda afectar a la voz sintética final, se ha decidido solventar esta posible ausencia de segmentos recurriendo al corpus neutro, el cual sí dispone de todos los segmentos de la lengua española.

En cuanto a los aspectos relacionados con la prosodia, y desde la perspectiva del análisis de los parámetros prosódicos, y no desde la síntesis, sería muy útil que los enunciados presentaran unas cualidades comunes, como por ejemplo disponer, para cada una de las emociones, de un amplio número de oraciones que presenten un mismo número de sílabas por grupo fónico, y cuya sílaba tónica estuviera en la misma posición como los corpus INTERFACE (Hozjan, 2002) o VAESS (Montero, 2003). Esto facilitaría la obtención de unos resultados más equiparables y equitativos durante el análisis, pero al trabajar con texto extraído de una situación comunicativa real, resulta extremadamente complejo cumplir también con este aspecto. Por este motivo, finalmente se ha decidido no tenerlo en cuenta a la hora de realizar la selección de los enunciados del corpus emocional.

a. Selección de enunciados

El corpus de conversaciones virtuales recogido, y utilizado para analizar el lenguaje empleado en ellas, también ha servido para crear el corpus emocional. La recopilación de conversaciones

virtuales ha sido etiquetada manualmente y posteriormente revisada parcialmente por un grupo de expertos. Este proceso ha servido para extraer un listado con las emociones más representativas de este dominio.

Para realizar el etiquetaje se ha tomado como referencia el inventario establecido en el proyecto I3Media (Garrido *et al.*, 2012 b), formado por 33 estados emocionales y 9 estados fisiológicos. Se han etiquetado las intervenciones de manera manual, según los estados emocionales o fisiológicos que iban expresando los usuarios.

Para representar lo más fielmente posible las emociones expresadas en las conversaciones virtuales, dominio en el que se utilizará la voz sintética, se han etiquetado los siguientes aspectos:

1. Estados emocionales y fisiológicos: se han marcado tanto las emociones que aparecían solas, como las que se mostraban combinadas con otros estados emocionales o actos de habla. Para ello se ha partido del inventario propuesto en el proyecto I3Media, el cual ha sufrido una serie de adaptaciones para adecuarse al contenido del corpus como es la ampliación del número de etiquetas.
2. En este caso también se ha tenido en consideración la intensidad de las emociones, marcando dos niveles diferentes: ‘_1’ para el nivel medio y ‘_2’ para un nivel alto.
3. Elementos paralingüísticos, como la risa, los besos, o el llanto, e interjecciones, como ‘¡ah!’, ‘¡oh!’, o ‘¡he!’.
4. Actos de habla: saludo, despedida, advertencia, etc.
5. Emoticonos

A continuación, en la Figura 17 se puede observar unos ejemplos extraídos del corpus con su correspondiente etiqueta

[ALEGRÍA] he sido padre
[ENFADO] [ORGULLO] sabes cuánta gente me roba el nick para parecer guay?
[ENFADO] [RECHAZO] [IMPACIENCIA] SOY_TODO_AMOR ya tú te crees que yo
me caído de un guindo o algo ...
[SALUDO] [ALEGRÍA_2] andalúuuuuuuuuuuuuuuuuuuuuu hola guapoo

Figura 17: Algunos enunciados etiquetados del corpus de las conversaciones virtuales

Como el etiquetaje de emociones a menudo puede ser una tarea subjetiva, ya que depende en gran medida de la propia experiencia, parte del corpus fue revisado por dos expertos.

Finalmente, el corpus incluye 37 etiquetas de emociones, 5 de estados fisiológicos, 55 elementos paralingüísticos, 7 actos de habla y 15 emoticonos, que aparecen aislados o combinados.

A partir de este corpus etiquetado se llevó a cabo la selección de los enunciados que debía ser grabado después por el locutor. Se decidió que contuviera ejemplos de enunciados representativos de aquellas emociones, estados fisiológicos y actos de habla cuya frecuencia de aparición fuera igual o superior al uno por ciento (Apéndice A), considerando que este porcentaje proporciona las emociones más frecuentes del corpus, y asegura la existencia de un número suficiente de enunciados. Esto implicó un análisis automático de la frecuencia de aparición de las diferentes etiquetas.

b. Contenido del corpus emocional

El corpus emocional seleccionado para su posterior grabación está formado por enunciados representativos de:

a) 15 estados emocionales aislados:

1. Admiración
2. Afecto
3. Alegría
4. Burla

5. Decepción
6. Diversión
7. Duda
8. Enfado
9. Extrañeza
10. Interés
11. Orgullo
12. Rechazo
13. Resignación
14. Sorpresa
15. Tristeza

b) 1 estado fisiológico aislado:

1. Alteración

c) 3 combinaciones de dos elementos (2 estados emocionales, y estado emocional y acto de habla):

2. Burla y rechazo
3. Enfado y rechazo
4. Alegría nivel 2 y saludo

d) 2 actos de habla sin emoción:

5. Saludo
6. Despedida

Para cada emoción se seleccionaron 25 enunciados, lo que hace que el corpus emocional esté formado por 525 enunciados (21x25, véase Apéndice B).

También se ha grabado una serie de interjecciones y elementos paralingüísticos, pero no se han tenido en cuenta en la creación de la voz sintética.

4.1.2. Selección del locutor

La elección de un locutor u otro depende del tipo de corpus emocional con el que deseemos trabajar. Si nos interesa más un habla natural que permita un mejor análisis de las emociones intentaremos grabar a una serie de hablantes durante una conversación espontánea. Esta técnica presenta varios inconvenientes como es el no poder controlar los enunciados que se van a producir, y en el caso de las emociones tampoco se puede asegurar la presencia de unas emociones determinadas y un número mínimo de realizaciones para cada una de ellas, ni tampoco se puede asegurar que la calidad de la grabación sea satisfactoria, ya que puede presentar ruidos e interferencias.

Para salvar estos problemas que presenta la recopilación del habla natural, la grabación se puede realizar en un laboratorio induciendo las emociones. En caso, las emociones más extremas quedarían fuera de lugar al no resultar demasiado ético provocar emociones como el enfado o el miedo.

La tercera y última forma de conseguir un corpus emocional es mediante la lectura e interpretación de los textos por parte de un locutor. De esta manera se puede controlar el contenido del corpus y asegurar una calidad de la grabación óptima. El inconveniente que presenta este método es que no asegura que la interpretación realizada por el locutor represente plenamente las características del habla emocional típica del habla natural.

En el caso concreto de esta investigación, al ser nuestro objetivo crear voz sintética expresiva, es necesario tanto poder controlar el corpus como la calidad de la grabación ya que será la base de datos empleada para el análisis acústico y la creación de la voz sintética. Por lo tanto, la estrategia utilizada para grabar el corpus emocional ha sido la de emplear un actor que simule las diferentes emociones.

Como ya se ha mencionado anteriormente, la selección del locutor para la grabación del corpus de voz necesario para la construcción

de un locutor sintético es un aspecto fundamental para obtener buenos resultados en sistemas basados en corpus. Por ello, para seleccionar el locutor más idóneo para la grabación del corpus se llevó a cabo un estricto proceso de selección, que permitiera determinar si el candidato seleccionado reunía las características mencionadas al inicio del capítulo.

Se partió de 13 candidatos iniciales, seleccionados para otro proyecto anterior, que *a priori* reunían dichas características. Todos ellos eran hablantes nativos de español. Se les hizo una entrevista para conocer su nivel de español y su experiencia laboral como actor o locutor. También se llevó a cabo una pequeña grabación de la lectura de una noticia para saber si existía algún tipo de interferencia con otras lenguas, como el catalán, ya que todos ellos eran bilingües castellano-catalán, y el grado de ésta.

De estos candidatos, se preseleccionaron seis, los que se consideraron que mejor cumplían con los requisitos deseados. Cada uno de los aspirantes grabó un pequeño corpus de prueba con las 6 emociones básicas más el neutro. Cada emoción tenía tres niveles de intensidad (baja-media-alta); y por cada emoción e intensidad había dos textos; en el caso del neutro se registraron seis enunciados. En total se grabaron 42 enunciados.

A partir de estos corpus orales, se desarrolló una voz sintética de cada locutor y se realizó una pequeña evaluación para escoger a la persona que se convertiría en el locutor de la nueva voz sintética. Se tomó una frase neutra, sin ninguna connotación emocional, y se generó para cada una de las emociones básicas y el neutro. Posteriormente, un grupo de investigadores con experiencia en generación de voz sintética seleccionó al locutor partiendo de esta pequeña prueba y de toda la información recopilada hasta el momento sobre cada locutor. Se escogió al locutor que mejor representaba las emociones y cuya voz artificial era más inteligible y natural.

4.1.3. Grabación del corpus

La grabación se llevó a cabo en un estudio de grabación de la Universidad Pompeu Fabra, un estudio completamente insonorizado y equipado para realizar la grabación de una manera correcta.

Para evitar posibles interferencias y ruidos, como el de pasar las hojas, los enunciados a leer por el locutor se presentaban a través de una pantalla. En un intento de facilitar al locutor la interpretación de los enunciados emocionales, para cada una de las frases a leer se le presentó el contexto en el cual había sido realizada (véase la Figura 18)

[ALBERTO38MADRID] Que todos los tíos no somos unos putos salidos. ¡Coño!
[despeinada] Alberto, tranquilo leñe
[despeinada] Que vas a asustar al personal
[ALBERTO38MADRID] Es que siempre es igual
[ALBERTO38MADRID] **Por unos pocos mierdas, pagamos el resto.**

[Liadara] El grito... ¡Dios! ¡Me remueve las entrañas!
[Nolocho] A mí en si... ¡Hombre! El tío es feo.
[Raulzgz] Lo sustituyeron por una copia , pero creo que sigue robao.
[Nolocho] Pero el colorido no está mal
[Nolocho] Sí, sí está robado.
[Liadara] ¿Qué dices? ¡Estáis flipándolo mucho!
[Nolocho] **¡Y estaba valorado en 200 kilos**
[Nolocho] **de euros!**

Figura 18: Dos ejemplos de enunciados identificados como enfado y sorpresa respectivamente presentados en el contexto en el que se desarrollaba la conversación para facilitar la interpretación al locutor durante la grabación

La grabación se realizó con un micrófono inalámbrico de diadema, hecho que permite más movilidad al locutor a la hora de interpretar algunas emociones, asegurando así que la distancia entre el locutor y el micrófono siempre fuera la misma.

La digitalización de la señal se ha llevado a cabo a una frecuencia de muestreo de 44 KHz y 24 bits en ficheros de tipo WAV.

Para grabar todo el corpus fueron necesarias 40 horas de estudio, que equivalen aproximadamente a cuatro horas de grabación real; tres horas y media de corpus neutro, y media hora de corpus emocional. Se prestó especial atención a que el locutor leyera correctamente cada enunciado, si se equivocaba en la lectura o, en el caso del corpus emocional, se consideraba que no había representado de una manera clara la emoción en cuestión, debía repetir el enunciado hasta conseguir la lectura e interpretación deseada.

4.1.4. Edición y revisión de la señal de audio

Para facilitar la tarea al locutor, la grabación se realizó en bloques con un número variable de enunciados por bloque. Posteriormente, y de manera manual, se escuchó cada uno de los bloques y se guardó cada enunciado con un nombre determinado. En el caso de que la lectura de un mismo enunciado se hubiera repetido varias veces, se seleccionó aquel que se consideró que era mejor.

Se comprobó que el audio de cada fichero correspondía exactamente con el texto que debía leer el locutor. Si hubo alguna diferencia se modificó el fichero de texto, ya que cualquier posible discrepancia entre el texto y el audio podría comportar errores en fases posteriores como la generación de la voz sintética.

Para cada señal acústica se creó, de manera automática mediante una serie de scripts, un TextGrid de Praat (Boersma y Weenink, 2009) que contenía:

- 1) la segmentación del texto en palabras,
- 2) su transcripción ortográfica,
- 3) la segmentación en alófonos,
- 4) su transcripción fonética en SAMPA,
- 5) la segmentación en sílabas,

- 6) segmentación en grupos acentuales y,
- 7) la segmentación en grupos entonativos.

4.2. Generación de habla expresiva mediante técnicas basadas en corpus

Como ya se ha comentado en el capítulo 1, uno de los objetivos marcados al inicio de esta tesis es investigar las posibilidades de obtención de una producción sonora inteligible y natural de los mensajes que conforman las conversaciones virtuales utilizando un CTH comercial basado en síntesis por corpus. Para ello se ha partido del CTH de la empresa escocesa Cereproc.

A continuación se comenta el proceso que realiza este CTH para generar la señal acústica.

4.2.1. Procedimiento de generación de habla expresiva con el SDK de Cereproc

El módulo de síntesis (*Synthesis Software Development Kit*, SDK) utiliza técnicas de síntesis por corpus para seleccionar los fragmentos más adecuados de la base de datos para concatenarlos y crear un nuevo enunciado. Este sintetizador, gracias al módulo Voice Creation Kit (VCK) permite crear voces para dominios determinados mediante la creación y selección de corpus específico para cada uno de los dominios.

La selección y concatenación de los diferentes segmentos que conformarán el enunciado oral de salida se puede realizar de dos modos diferentes: teniendo en consideración todo el corpus en general, o sólo uno de los subcorpus que conforman el mismo.

Para la primera opción, el usuario únicamente ha de escribir el enunciado que desea reproducir oralmente, y el módulo de síntesis realiza la selección y concatenación más favorable entre todos los fragmentos del corpus.

La segunda opción está orientada a la síntesis en dominios, ya que mediante un etiquetado en XML (*Extensible Markup Language*) se obliga a que el SDK seleccione los diversos segmentos dentro de un subcorpus específico, recurriendo a otros subcorpus únicamente si el segmento o segmentos necesarios para crear la señal acústica no se encuentra en el corpus especificado mediante la etiqueta. Esto es posible gracias a que los diferentes ficheros que forman el corpus están etiquetados con una letra al inicio, la cual hace referencia al dominio específico al que pertenece el enunciado. Por ejemplo, los enunciados orientados a direcciones de correo están etiquetados con la letra 'a', y los orientados a noticias poseen la letra 'n'. La Figura 19 muestra un pequeño ejemplo de varias oraciones del corpus con su respectiva etiqueta y numeración.

a1001_041 David Menéndez González; Hernando Colón, siete; cuatro uno uno cero cero, Cáceres.
a1002_060 w w w punto sellomania punto es.
n1003_020 Los tripulantes de cabina de British Airways suspenden su huelga.
n1003_129 Los muertos en accidente laboral hasta marzo fueron trescientos diecisiete, dieciocho menos que en dos mil seis.
n1004_075 El delicado momento del sector del motor en muchos mercados, en especial los europeos, ha hecho mella en las ventas de Renault.

Figura 19: Enunciados del corpus neutro y emocional con sus respectivas etiquetas que indican el dominio al que pertenecen (letra que aparece al inicio) y el número del enunciado dentro del corpus general

En el caso del corpus de las conversaciones virtuales también fue necesario otorgarle una letra identificativa. Se emplea la letra 'x', ya que ésta todavía no se había utilizado para designar ningún otro dominio, para etiquetar todos los enunciados que conforman el corpus de las conversaciones virtuales. Esto puede influir negativamente en la generación de enunciados acústicos expresivos, ya que para crear un enunciado de un determinado estado emocional el sistema puede seleccionar algunos o todos los segmentos pertenecientes a diferentes emociones. Si se desea realizar un enunciado que transmita alegría, como todos los enunciados están etiquetados con la letra 'x', es posible que se seleccionen segmentos de enunciados que transmiten enfado, sorpresa, tristeza, etc. ya que no existe una marca que diferencie las diferentes emociones que

conforman el corpus emocional. A pesar de esta posible repercusión negativa, se ha decantado por este método, ya que al etiquetar cada una de las diferentes emociones individualmente los resultados no eran óptimos a consecuencia del reducido número de casos que recogía cada etiqueta, resultando una señal acústica de baja calidad y en ocasiones ininteligible.

Esta misma letra que sirve para identificar cada uno de los dominios es la que debe aparecer en la etiqueta XML, después del atributo 'genre', tal y como aparece en la Figura 20, para indicar el habla de qué dominio es la que ha de intentar reproducir.

```
<parent>
<usel genre =-'x'>Yo llevo al niño en el coche </usel>
</parent>
```

Figura 20: Ejemplo del etiquetado XML empleado para realizar un enunciado seleccionando los diferentes segmentos dentro del corpus emocional representativo de las conversaciones virtuales

Además de poder seleccionar el subcorpus en el que realizar la búsqueda y concatenación de unidades, este CTH también permite, mediante la misma técnica de etiquetas XML, modificar la frecuencia fundamental, la amplitud y la duración de la señal acústica empleando el argumento 'sig' como en la Figura 21 (Aylett et al., 2007b). Lamentablemente, el rendimiento obtenido mediante este modo de proceder no es satisfactorio, lo que ha implicado crear un sistema alternativo para modificar de forma automática la señal acústica según las características de la emoción deseada, el cual se comentará más extensamente en el epígrafe 4.3.3. Generación del habla expresiva.

```
<usel genre =-'x'><sig f0='+5' amplitude='0.5' rate =-'0.95'>
Dejaron la deuda al cero </sig></usel>
```

Figura 21: Etiquetado empleado para realizar la frecuencia fundamental, la amplitud y la duración de un enunciado determinado según los valores especificados en las etiquetas

4.3. Generación de habla expresiva mediante modelos prosódicos

Una persona puede saber cuándo otra está enfada, contenta, triste, alterada o tiene miedo sólo con escuchar su voz, y es que las emociones alteran en general algunos de los rasgos de una producción fónica. Una persona que está contenta probablemente hablará con un valor de intensidad superior al de una persona que está triste, y una persona alterada hablará a una mayor velocidad que otra que esté sorprendida. La síntesis de voz expresiva pretende tomar estos rasgos y modelarlos para conseguir una reproducción de enunciados emotivos similares a los realizados por el ser humano.

Para poder modelar paramétricamente la voz sintética de modo que transmita las emociones seleccionadas y grabadas en el corpus oral es necesario desarrollar unos modelos prosódicos. Del mismo modo que con el texto de las conversaciones virtuales, ha sido necesario analizar el corpus oral para poder extraer conclusiones y datos que permitan este modelado. Además de construir unos modelos, el análisis también ha permitido determinar si los parámetros seleccionados en esta investigación son significativos en la generación de habla expresiva.

En este epígrafe se desarrolla el análisis fonético de diversos parámetros prosódicos que conforman el corpus de grabación, tanto de los enunciados neutros como de los emocionales, comentando el procedimiento llevado a cabo y los resultados extraídos para posteriormente poder generar la voz sintética expresiva. La aproximación empleada para ello es el modelado paramétrico que permite modelar los parámetros seleccionados, basados en un conocimiento lingüístico, y con mayor flexibilidad que permite obtener una calidad aceptable de la señal acústica.

4.3.1. Generación de los modelos prosódicos

En la generación de enunciados emotivos confluyen diferentes aspectos rítmicos y entonativos que hacen que los oyentes puedan percibir y diferenciar diferentes estados emocionales. Desgraciadamente, aún no se ha logrado obtener un conjunto de unidades prosódicas que permita describir en su totalidad una elocución (Montero, 2003). En esta tesis se han analizado una serie de parámetros, ya analizados en otros estudios (Montero 2003, Iriondo 2008) y que se sabe que influyen en la generación de habla expresiva. Estos parámetros se han clasificado en dos tipos diferentes, los no melódicos que incluyen la velocidad de elocución, la duración de la vocal prepausal, la intensidad y la amplitud, y los parámetros melódicos formados por el rango y la altura tonal de F0, y los patrones melódicos.

4.3.1.1. Parámetros no melódicos

Del mismo modo que no todas las personas leemos un texto del mismo modo, ni hablamos igual, tampoco expresamos del mismo modo las diferentes emociones. Para conocer cómo ha transmitido las diferentes emociones el locutor, es necesario realizar un análisis acústico del corpus oral, que indicará en qué medida varían los parámetros no melódicos según la emoción realizada. Este análisis también permitirá modelar la señal acústica conforme los resultados obtenidos, y realizar así un enunciado sintético emotivo.

La obtención de estos valores se ha realizado de forma automática mediante scripts de bash, que a su vez hacían llamadas a otros scripts desarrollados en diferentes lenguajes de programación: Praat y R¹⁰ (Stowell, 2012).

Primeramente, es necesario obtener los valores de los parámetros no melódicos de cada uno de los ficheros que conforman el corpus oral. Mediante un script de bash que realiza diferentes llamadas a

¹⁰ <http://www.r-project.org/>

otros scripts de Praat se pudo obtener los diferentes cálculos. El script creó unos ficheros de salida, uno para cada uno de los parámetros analizados, con toda la información dispuesta en forma de tabla. Estos ficheros presentan una información común independientemente del parámetro analizado, como son:

- el nombre del fichero,
- el número de grupos fónicos del enunciado,
- la posición del grupo fónico en el que se está realizando el cálculo,
- el tipo de pausa,
- la modalidad oracional,
- la emoción,
- el nivel de emoción y
- el acto de habla que se desarrolla.

La Tabla 21 muestra un ejemplo de cómo aparece distribuida esta información.

Una vez obtenidos estos valores, tanto para el corpus neutro como para el emocional, se realizaron los cálculos estadísticos. Nuevamente el proceso se llevó a cabo de forma automática mediante un script en bash, que a su vez realiza diferentes llamadas a una serie de scripts desarrollados en R. En este caso, los datos obtenidos para el corpus emocional son comparados con los del corpus neutro.

Fichero	Nº GF	GF	Vocal	Duración	Tipo pausa	Modalidad oracional	Emoción	Nivel emo.	Acto de habla
Edu_x1001_000	2	1	o	0,180	FIN_FRASE	Exclamativa	Admiración	1	Indefinido
Edu_x1001_000	2	2	o_"	0,109	FIN_FRASE	enunciativa	Admiración	1	Indefinido
Edu_x1001_001	2	1	o	0,150	indefinido	Exclamativa	Admiración	1	Indefinido
Edu_x1001_001	2	2	e	0,260	FIN_FRASE	Exclamativa	Admiración	1	Indefinido
Edu_x1001_002	2	1	e	0,120	indefinido	Enunciativa	Admiración	1	Indefinido
Edu_x1001_002	2	2	e	0,209	FIN_FRASE	Enunciativa	Admiración	1	Indefinido
Edu_x1001_003	1	1	o	0,27	FIN_FRASE	Exclamativa	Admiración	1	Indefinido

Tabla 21: Distribución de la información de la 'tabla_vocal_prepausal.txt'

El script devuelve tres tipos de ficheros para cada parámetro analizado:

1. Uno, en el que se han realizado las medias de cada uno de los parámetros. La información que anteriormente aparecía para cada uno de los ficheros ahora aparece agrupada según la modalidad oracional, la emoción, el nivel de emoción y acto de habla al que pertenece. Además, también se informa del número de casos que comparten las mismas características. A partir de esta agrupación se ha realizado una media aritmética para obtener un valor medio para cada conjunto.
2. Un segundo fichero que contiene las ratios. Manteniendo la clasificación por grupos realizada para calcular las medias, se ha comparado los valores obtenidos en el corpus emocional con el neutro, y se ha calculado las diferencias observadas en cada uno de los parámetros y para cada una de las emociones. Los valores obtenidos son empleados para modelar la voz sintética dependiendo de la emoción que se desee realizar.
3. Finalmente, se crea un fichero en el que se ha calculado la ANOVA (análisis de varianza). Este cálculo permite conocer si el parámetro en cuestión es significativo o no.

En la Tabla 22 y la Tabla 23 se muestra una serie de ejemplos extraídos de los ficheros de medias y ratios de la duración de la vocal prepausal respectivamente.

Modalidad	Acto de habla	Emoción	Nivel emoción	Núm. Casos	Media duración vocal prepausal
ENUNCIATIVA	DESPEDIDA	Indefinido	Indefinido	17	0,186
ENUNCIATIVA	Indefinido	ADMIRACION	1	7	0,151
ENUNCIATIVA	Indefinido	AFECTO	1	31	0,191

Tabla 22: Ejemplos extraídos de la tabla 'medias_dur_voc_prepausal.summary'

Modalidad	Acto de habla	Emoción	Nivel emoción	Ratio duración vocal prepausal
ENUNCIATIVA	DESPEDIDA	Indefinido	Indefinido	0,798
ENUNCIATIVA	Indefinido	ADMIRACION	1	0,984
ENUNCIATIVA	Indefinido	AFECTO	1	0,777

Tabla 23: Ejemplos extraídos de la tabla 'ratio_neutro_dur_vocal_prepausal.summary'

4.3.1.2. Parámetros melódicos

Para el análisis de los parámetros relacionados con la F0 se ha aplicado el modelo descriptivo de la curvas melódicas de Garrido (1996, 2001). Dicho modelo asume que los contornos de F0 son el resultado de la superposición de una cadena de patrones de F0, que definen la evolución del contorno en el ámbito del grupo acentual (GA), sobre las líneas de referencia, que definen la altura relativa de los patrones a lo largo de un grupo fónico (GF).

Esta tesis ha abordado tanto el análisis de los patrones de F0 como las líneas de referencia. La obtención de los datos se ha realizado de forma automática, mediante la herramienta MelAn, descrita en Garrido (2010). Esta herramienta toma como punto de partida la segmentación del enunciado en unidades fonéticas. Su uso implica los siguientes procesos:

1. Estilización de los contornos de F0.
2. Anotación de los puntos de inflexión del contorno estilizado con las etiquetas de pico (P, P+) y valle (V-, V).
3. Extracción de los parámetros melódicos (GA).
4. Cálculo de las rectas de regresión para los puntos etiquetados como P y V.

La salida de la herramienta es una serie de tablas con las líneas de referencia tipo para los niveles P y V, y los patrones de F0 más frecuentes, teniendo en cuenta las diferentes emociones y actos de habla con los que se ha etiquetado cada enunciado.

En el caso de las líneas de referencia, éstas indican la altura de los niveles tonales P y V a lo largo de cada GF. De forma automática se calcula mediante regresiones de todos los puntos etiquetados como P y V de un mismo GF. De esta manera se obtiene un valor de F0 inicial para la recta y un valor para la pendiente tanto de P como de V. A partir de los valores individuales de cada GF, se calculan una serie de 'líneas-tipo' para los diferentes tipos de GF: inicial (si se encuentra al inicio de oración), interior (si se encuentra en el

interior de la oración), final (si está al final de la oración) e inicial-final (si el GF forma una oración por sí mismo). Estas líneas tipo se obtienen calculando la media de los valores individuales de cada grupo en función de las variables de análisis consideradas (emoción, nivel de emoción, acto de habla, modalidad oracional y posición del GF en el enunciado). Los valores obtenidos son el valor medio en hercios (Hz) del punto inicial de las rectas, el valor medio final y el valor medio de la pendiente.

Respecto a los patrones de F0, según el modelo propuesto por Garrido (1996, 2001, 2010) y seguido en esta tesis, se pueden definir como series recurrentes de puntos de inflexión que aparecen sistemáticamente asociados a determinadas posiciones dentro del GA que los contiene. Los GA están compuestos por una sílaba tónica y todas las sílabas átonas que la siguen hasta la próxima sílaba tónica o final de enunciado. Los puntos de inflexión se anotan mediante cuatro etiquetas diferentes: V-, V, P y P+, que comprenden desde el nivel más bajo al más alto de la altura tonal dentro del GF. Además, también aporta información sobre la sílaba del grupo acentual en la que aparece, siendo 0 la sílaba tónica, e indicando su posición con respecto al núcleo silábico de la misma: Inicio (I), cerca del centro (M) y final (F). Así, un patrón como VI0_PM0_PF1 está formado por tres puntos de inflexión, los dos primeros asociados a la sílaba tónica (0) y el último a la sílaba posterior a la tónica (1). El primer punto tiene un nivel V (valle) y está asociado al inicio (I) de la sílaba tónica, los dos últimos tienen un nivel P (pico) y están situados a la mitad (M) de la sílaba tónica y al final (F) de la sílaba posterior a la tónica respectivamente.

Todos los patrones se clasifican en función de su posición dentro del GF que lo contiene:

1. Inicial, si está al comienzo del GF.
2. Interior, si está entre el inicio y el final del GF.
3. Final no final de enunciado, si está al final del GF, pero no es final de oración.
4. Final de enunciado, si se encuentra al final del GF, y a su vez es final de oración.

Las dos primeras categorías se corresponden con los patrones acentuales, relacionados con la expresión del acento en interior de grupo entonativo. Las dos últimas categorías se corresponden con los tonemas tradicionales.

La Figura 22 muestra un ejemplo extraído de Garrido (2011) en el que se puede observar tanto el etiquetado de picos y valles que realiza la herramienta, como las líneas de referencias.

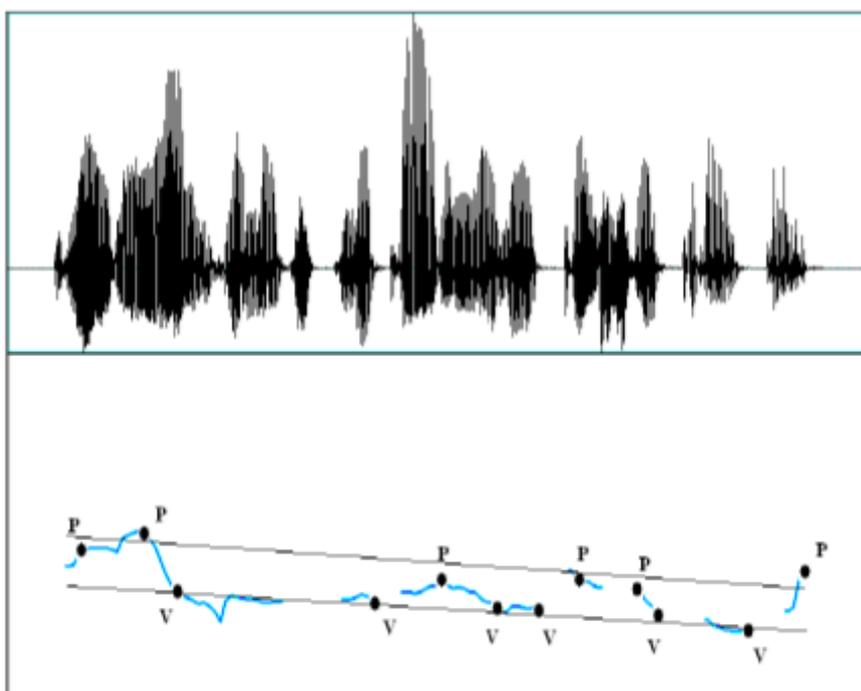


Figura 22: Ejemplo de notación y líneas de referencia del enunciado “¿Quiere alguien explicarme de qué se trata?” extraído de Garrido (2011)

Emplear este método para analizar los contornos melódicos de los diferentes enunciados permite que:

1. El análisis se realice de forma automática.

2. El análisis se realice en dos niveles diferentes: global y localmente (los dos niveles que se consideran relevantes en los contornos de F0 para la expresión de las emociones).
3. Toda la información extraída del análisis sea integrada de forma automática para modelar los contornos melódicos.

4.3.2. Resultados

4.3.2.1. Parámetros no melódicos

c. *Datos*

Se han seleccionado cinco parámetros para analizar la voz expresiva: la velocidad de elocución, la duración de la vocal prepausal, la amplitud y la intensidad.

El análisis ANOVA (Tabla 24) que se ha realizado de cada uno de los parámetros indica que los resultados extraídos son significativos en relación al tipo de emoción.

Parámetros	P<0,05
Velocidad de elocución	2,64E-03
Duración vocal prepausal	3,05E-16
Amplitud	2,89E-20
Intensidad	2,89E-20

Tabla 24: Resultado del análisis ANOVA realizado para cada parámetro seleccionado

La Tabla 25 muestra las medias obtenidas para cada uno de los parámetros seleccionados en función de la emoción expresada por

el locutor. No se observan comportamientos similares en las diferentes emociones según la polaridad (positivo y negativo) o la activación (activo/pasivo) de cada una de ellas, sino que cada emoción tiene un comportamiento particular para cada uno de los parámetros.

Las emociones afecto, burla-rechazo, diversión, duda, enfado-rechazo, extrañeza, interés, orgullo, resignación y sorpresa presentan unos valores de velocidad de elocución inferiores al neutro. El resto de las emociones (admiración, alegría, alteración, burla, decepción, enfado y tristeza) se articulan con mayor rapidez.

Todas las emociones excepto la alegría presentan una duración de la vocal prepausal mayor a la del neutro.

Prácticamente todas las emociones presentan una intensidad superior a la del neutro, excepto la admiración, el afecto, la alteración y la sorpresa.

La mayoría de las emociones presenta una amplitud inferior a la del neutro. Las emociones que no comparten este comportamiento son la alegría, la alteración, la burla y el orgullo. La burla-rechazo y el rechazo muestran la misma amplitud que el neutro.

Emoción	Nº. Casos	Media velocidad elocución¹¹	Media dur. vocal prepausal	Media intensidad (dB)	Media amplitud (Pascales)
Admiración	7	13,432	0,151	68,640	0,022
Afecto	31	12,824	0,191	69,083	0,020
Alegría	21	14,499	0,144	71,347	0,041
Alteración	12	13,672	0,185	68,499	0,032
Burla	30	13,484	0,166	69,707	0,029
Burla – Rechazo	21	12,526	0,184	69,644	0,027
Decepción	27	13,513	0,182	70,709	0,024
Diversión	18	11,859	0,185	69,539	0,025

¹¹ Se ha calculado teniendo en cuenta el número de palabras que se pronuncian en un minuto.

Duda	25	11,167	0,225	70,865	0,023
Enfado	22	13,404	0,173	69,401	0,021
Enfado – rechazo	20	12,266	0,194	67,932	0,023
Extrañeza	5	12,929	0,182	69,576	0,021
Interés	3	12,949	0,187	71,234	0,019
Orgullo	28	12,233	0,186	69,347	0,028
Rechazo	24	13,121	0,189	70,002	0,027
Resignación	41	12,688	0,186	71,683	0,021
Sorpresa	2	12,933	0,22	69,195	0,024
Tristeza	26	14,056	0,160	70,465	0,025
Neutro	4.694	13,147	0,149	69.250	0.027

Tabla 25: Medias obtenidas para cada uno de los parámetros seleccionados, clasificadas según el tipo de emoción

d. Conclusiones

En cuanto a los parámetros acústicos seleccionados para el análisis y posterior generación de la voz sintética expresiva, a pesar de que el análisis ANOVA indica que los resultados obtenidos son significativos, cada emoción tiene un comportamiento particular para cada uno de ellos.

El comportamiento de las emociones básicas observadas en el corpus presenta ciertas similitudes respecto a los resultados obtenidos en otros estudios comparables:

- **Alegría:** Presenta una velocidad de elocución, una intensidad y una amplitud superior a la del neutro. En cambio, la duración de la vocal prepausal es ligeramente inferior.
- **Enfado:** Tanto la velocidad de elocución, como la duración de la vocal prepausal y la intensidad presentan unos valores superiores a los del neutro. El valor obtenido para la amplitud es inferior con respecto al neutro.
- **Sorpresa:** Su comportamiento es completamente opuesto al de la alegría. Los valores de velocidad de elocución, de intensidad y amplitud son inferiores al neutro, mientras que la media obtenida en referencia a la duración de la vocal prepausal es considerablemente superior.
- **Tristeza:** Su modo de proceder es idéntico al observado en el enfado. La velocidad de elocución, la duración de la vocal prepausal y la intensidad presentan unos valores superiores al neutro. Por el contrario, el valor de la amplitud es inferior.

Sorprenden los resultados obtenidos en la tristeza, contradictorios a Montero (2003) donde esta emoción muestra tanto unos valores de velocidad de elocución como de duración de la vocal prepausal inferiores al neutro, o a Iriondo (2008), cuya investigación también presenta un velocidad de locución más lenta, y una intensidad menor a la del neutro. Esto parece indicar que la forma de expresar las diferentes emociones depende en mayor medida del locutor, es decir, dos locutores pueden expresar de modo distinto una misma emoción, y en ambos casos el oyente identificaría correctamente dicho estado emocional a pesar de haber sido interpretadas de distinto modo.

4.3.2.2. Parámetros melódicos

Siguiendo el modelo descrito en Garrido (2011), existen dos tipos de parámetros melódicos que intervienen en la expresión de los estados emocionales del español: los parámetros globales, que afectan a todo un enunciado, y los locales, que tienen lugar en una o varias sílabas (los patrones melódicos).

a. *Parámetros globales*

En la Tabla 26 se presentan los valores de F0 medios obtenidos para los puntos iniciales y finales de las 'líneas-tipo' superior (P) e inferior (V) y los valores de la pendiente media. Los datos de ambas tablas se presentan en función de las diferentes etiquetas de

Emoción	Núm. Casos	Valor inicial P media	Valor final P media	Pendiente P media	Valor inicial V media	Valor final V media	Pendiente V media
Admiración	1	160,93	92,08	-48,82	128,58	77,34	-36,33
Afecto	13	135,28	101,03	-23,39	108,11	75,29	-22,19
Alegría	17	174,76	134,64	-19,12	132,32	96,13	-18,17
Alteración	10	203,39	135,75	-42,55	146,78	95,25	-29,84
Burla	13	161,64	121,05	-17,82	131,59	92,50	-17,21
Burla – Rechazo	16	178,85	119,98	-36,70	135,42	96,37	-23,25
Decepción	19	126,41	97,12	-20,04	105,19	77,66	-18,75
Diversión	16	193,28	133,44	-43,13	149,86	98,27	-36,20
Duda	16	121,73	134,45	15,36	103,10	93,96	-4,81
Enfado	13	182,37	116,26	-40,60	131,40	85,79	-23,21
Enfado – rechazo	6	174,01	128,30	-21,44	134,03	87,38	-20,62
Interés	1	122,80	168,60	31,58	100,95	-7,97	21,85
Orgullo	19	171,38	82,87	-52,39	126,50	72,39	-33,49
Rechazo	20	150,39	100,13	-32,05	119,34	75,93	-26,38

Resignación	11	135,53	95,14	-25,74	110,21	83,06	-17,03
Sorpresa	2	208,64	131,35	-61,59	119,82	120,41	-1,92
Tristeza	22	123,77	85,56	-24,86	99,28	72,34	-17,38
Neutro	363	158,91	90,57	-45,30	120,27	69,18	-33,74

Tabla 26: Valores medios de los puntos iniciales y finales y pendientes de las rectas superior (P) e inferior (V) en las diferentes emociones analizadas y en habla neutra

emoción consideradas en el análisis. Sólo se han incluido los datos correspondientes a los grupos fónicos inicial-final de las oraciones enunciativas, los que constituyen un enunciado por sí mismos, y son mayoritarios en el corpus. El resto de grupos fónicos (inicial, interior y final) muestran un comportamiento muy similar al inicial-final. Se puede consultar los resultados en el Apéndice C.

Observando las tablas se puede comprobar cómo prácticamente todas las emociones presentan 'líneas-tipo' con valores finales más bajos que los iniciales, tanto en lo que se refiere a la línea inferior como a la superior. Tres son las emociones que muestran un comportamiento diferente, el valor final es superior al inicial: duda, interés, probablemente debido a que los enunciados que lo constituyen son mayoritariamente preguntas, y resignación. El análisis ANOVA de la pendiente mostró que estas diferencias son significativas en función de la variable emoción ($p < 0,05$).

Todas las emociones muestran unos niveles superiores al neutro tanto en P como en V, y tanto al inicio como al final, excepto los enunciados etiquetados como afecto, decepción, duda, interés, rechazo, resignación y tristeza.

Comparando los resultados de las diferentes emociones entre sí, se observa que emociones como la sorpresa, el enfado, la diversión, burla-rechazo, alteración y alegría presentan una altura tonal más alta que el resto de emociones.

La Tabla 27 muestra los valores medios correspondientes al rango inicial y final en los grupos inicial-final en función de las etiquetas de emoción consideradas.

No todas las emociones analizadas presentan unos niveles superiores al neutro. Afecto, burla, duda e interés muestran unos niveles inferiores en el rango inicial, mientras que orgullo y sorpresa presentan unos niveles inferiores en el rango final. Emociones como la admiración, la decepción, la resignación y la tristeza muestran un rango inferior al neutro tanto al inicio como al final. Nuevamente, el análisis ANOVA muestra que las diferencias son significativamente relevantes en función de la variable emoción.

Emoción	Nº Casos	Rango inicial P V	Rango final P V
Admiración	1	32,34	14,74
Afecto	13	27,16	25,73
Alegría	17	42,44	38,50
Alteración	10	56,61	40,49
Burla	13	30,05	28,54
Burla – Rechazo	16	43,42	25,96
Decepción	19	21,21	19,46
Diversión	16	43,42	35,16
Duda	16	18,62	40,49
Enfado	13	50,96	30,46
Enfado – Rechazo	6	39,98	40,91
Interés	1	21,85	79,20
Orgullo	19	41,79	12,17
Rechazo	20	31,04	24,20
Resignación	11	25,32	12,07
Sorpresa	2	88,81	10,93
Tristeza	22	24,48	13,21
Neutro	363	38,66	21,71

Tabla 27: Valores medios para el rango inicial y final en las diferentes emociones analizadas y en habla neutra

En algunas emociones, el rango medio final es superior al inicial, en concreto en las etiquetadas como duda, enfado-rechazo, e interés. Según Garrido (2011) esto podría relacionarse con un mayor rango local en el patrón melódico final, el tonema.

b. Patrones melódicos

Los resultados obtenidos para los patrones melódicos, acentuales y entonativos son muy extensos. Por ello, se expondrán los más frecuentes, los pertenecientes a las oraciones enunciativas (Tabla 28) y cuyo GA está formado por 1, 2 y 3 sílabas (Tabla 29). En todos los casos, se han seleccionado los grupos acentuales cuya sílaba tónica es la primera.

Modalidad oracional	Número de casos
Enunciativas	5.523
Exclamativas	616
Interrogativa total	614
Interrogativa parcial	519
Interrogativa relativa	31
Interrogativa confirmativa	18
Interrogativa alternativa	37

Tabla 28: Clasificación de los patrones melódicos neutros y emocionales según la modalidad oracional

NÚM. SÍLABAS	NEUTRO	EMOCIONAL	TOTAL
1	1.789	396	2.185
2	2.002	680	2.682
3	845	339	1.184
4	385	224	609
5	224	107	331
6	124	39	163
7	69	15	84
8	53	6	59
9	35	2	37
10	16	1	17
11	5	0	5
12	2	0	2

Tabla 29: Clasificación de los patrones melódicos de las frases enunciativas según el número de sílabas que contienen

1. Patrones iniciales

La Tabla 30, la Tabla 31 y la Tabla 32 muestran los diez patrones más frecuentes en los corpus analizados (neutro y emocional) para los grupos acentuales iniciales de 1, 2 y 3 sílabas con acento en la primera sílaba. El número de casos que presentaba cada emoción era muy reducido, por este motivo se ha decidido unificar todos los patrones y enmarcarlos dentro de una categoría más genérica, el corpus emocional.

CORPUS	PATRÓN	FREC.
NEUTRO	VI0_PM0	75
	VI0	35
	PI0	30
	VM0	24
	PM0	20
	PI0_PM0	12
	PI0_VF0	9
	VI0_VF0	8
	VI0_PFO	6
	VI0_PM0_VF0	6
	EMOCIONAL	VI0
VI0_PM0		7
VI0_VF0		5
PI0		3
PI0_PM0		3
VM0		3
0		1
PI0_VF0		1
PI0_VM0		1
VFO		1

Tabla 30: Patrones iniciales más frecuentes en los grupos acentuales de una sílaba con acento en la primera sílaba

<i>CORPUS</i>	<i>PATRÓN</i>	<i>FREC.</i>
NEUTRO	VI0_PM0	32
	PI0	30
	PI0_P11	29
	VI0_P11	26
	VI0	23
	PM0_P11	19
	PM0	15
	VI0_P10	13
	PI0_V11	12
	VI0_PM0_P11	12
	EMOCIONAL	VI0_P11
VI0_PM0		5
PI0		3
VI0		3
VI0_VF1		3
VM0_P11		3
PI0_VF0		2
VI0_PM0_PM1		2
VI0_V11_PM1		2
VM0		2

Tabla 31: Patrones iniciales más frecuentes en los grupos acentuales de dos sílabas con acento en la primera sílaba

<i>CORPUS</i>	<i>PATRÓN</i>	<i>FREC.</i>
NEUTRO	VI0_P11	17
	VM0_P11	14
	VI0_VF2	12
	PI0_PM1	10
	VI0_PM0	10
	PI0_P11_VF2	9
	PM0_P11_VI2	9
	PI0_P11_VI2	6
	PI0_VI1	6
	VI0_P11_VM2	6
EMOCIONAL	PI0_PM0_VF0_P12	1
	PI0_VF0	1
	PI0_VM0	1
	PM0	1
	VI0	1
	VI0_P10_VF2	1
	VI0_P11_VI2_VF2	1
	VI0_PM0_PM1	1
	VI0_PM0_VI1	1
	VI0_PM0_VI1_VF2	1

Tabla 32: Patrones iniciales más frecuentes en los grupos acentuales de tres sílabas con acento en la primera sílaba

Los dos patrones más frecuentes en los grupos acentuales formados por una sílaba son los mismos tanto para el corpus neutro como para el emocional (VI0_PM0 y VI0), con una frecuencia inversa.

En el corpus neutro se observa una mayor presencia de patrones con un solo punto de inflexión, pudiendo ser este alto (P) o bajo (V).

Cuando el GA está formado por dos sílabas, se observa una mayor frecuencia de patrones con dos puntos de inflexión.

En el caso de los grupos acentuales formados por 3 sílabas, los patrones correspondientes al corpus emocional no son recurrentes, cada patrón sólo aparece una sola vez. En el caso del corpus neutro, no se observa ningún patrón con una sola inflexión.

El patrón VIO_PM0 es el más frecuente tanto en el corpus emocional como neutro tanto si el GA está formado por una como por dos sílabas. En el caso de que el patrón contenga tres sílabas continúa siendo frecuente aunque en menor medida.

La Figura 23, la Figura 24, la Figura 25 y la Figura 26 muestran los patrones más frecuentes observados en el corpus de análisis. El patrón en cuestión aparece siempre sombreado.

Las dos primeras figuras corresponden a ejemplos del corpus neutro mientras que las dos últimas han sido extraídas del corpus emocional.

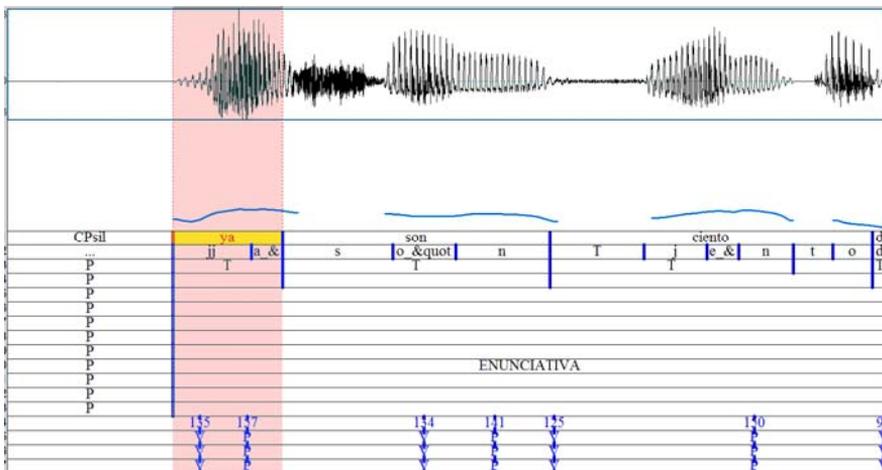


Figura 23: Ejemplo de patrón inicial de enunciado VIO_PM0 en el GA 'ya' dentro del enunciado neutro 'Ya son ciento doce los soldados de Estados Unidos muertos en Irak en mayo, el peor mes de dos mil siete.'

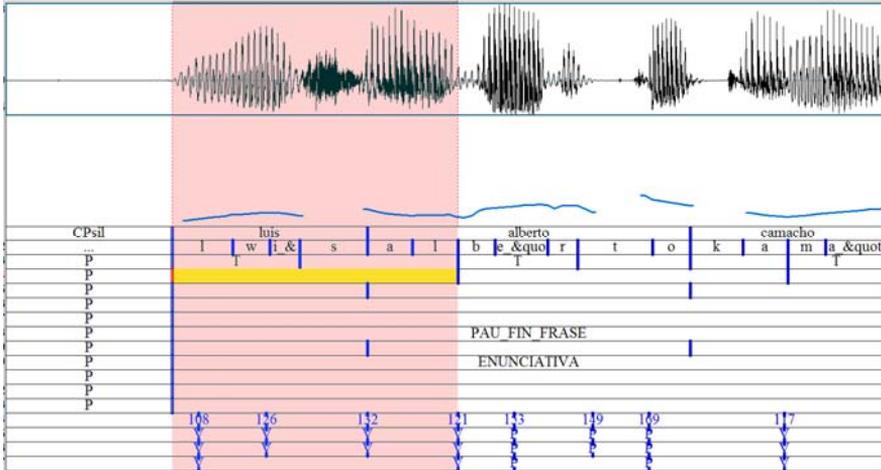


Figura 24: Ejemplo de patrón inicial de enunciado VI0 en el GA 'luis al' dentro del enunciado neutro 'Luis Alberto Camacho Girones; plaza de Portugal, uno: cero seis cero cero uno; Badajoz.'

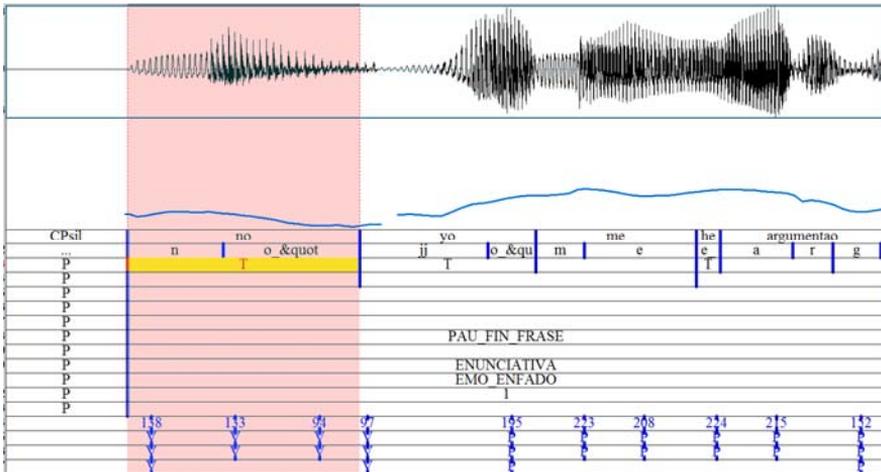


Figura 25: Ejemplo de patrón inicial de enunciado VI0 en el GA 'no' dentro del enunciado emocional 'No, yo me he argumentado'

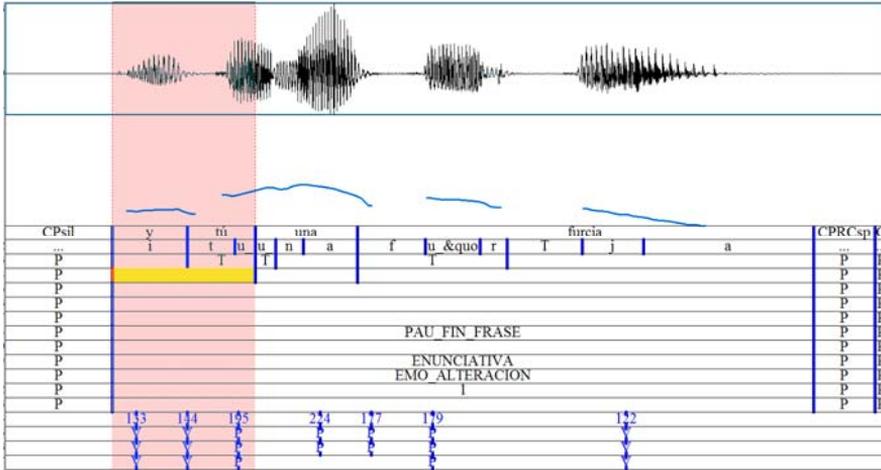


Figura 26: Ejemplo de patrón inicial de enunciado VI0_PM0 en el GA 'y tú' dentro del enunciado emocional 'Y tú una furcia.'

2. Patrones interiores

La Tabla 33, la Tabla 34 y la Tabla 35 muestran los diez patrones más frecuentes en los corpus analizados (neutro y emocional) para los grupos acentuales interiores de 1, 2 y 3 sílabas con acento en la primera sílaba. Del mismo modo que ocurría en los grupos acentuales iniciales, el número de casos que presentaba cada emoción era muy reducido, por este motivo se ha decidido unificar todos los patrones y enmarcarlos dentro de una categoría más genérica, el corpus emocional.

<i>CORPUS</i>	<i>PATRÓN</i>	<i>FREC.</i>
NEUTRO	VI0_PM0	169
	0	149
	PM0	107
	PI0	95
	VF0	47
	VI0	28

	PF0	27
	VI0_VF0	22
	VI0_PM0_VF0	20
	PI0_PF0	17
EMOCIONAL	VI0	9
	VI0_PM0	7
	VI0_VF0	4
	PI0	3
	PI0_PM0	3
	VM0	3
	0	1
	PI0_VF0	1
	PI0_VM0	1
	VF0	1

Tabla 33: Patrones interiores más frecuentes en los grupos acentuales de una sílaba con acento en la primera sílaba

<i>CORPUS</i>	<i>PATRÓN</i>	<i>FREC.</i>
NEUTRO	VI0_PM0	226
	0	188
	PI1	135
	PM0	126
	PI0	114
	VI0_PM0_VF1	59
	PF0	57
	VI0_PM0_VM1	49
	VI1	49

	VI0_P11	45
EMOCIONAL	VI0_P11	6
	VI0_PM0	5
	PI0	3
	VI0	3
	VI0_VF1	3
	VM0_P11	3
	PI0_VF0	2
	VI0_PM0_PM1	2
	VI0_V11_PM1	2
	VM0	2

Tabla 34: Patrones interiores más frecuentes en los grupos acentuales de dos sílabas con acento en la primera sílaba

<i>CORPUS</i>	<i>PATRÓN</i>	<i>FREC.</i>
NEUTRO	VI0_PM0	103
	0	96
	PI1	73
	PM0	62
	PI0	61
	PI1_VF2	35
	VI0_PM0_PI2	33
	VI0_PM0_PI1_VF2	32
	VI0_PM0_PI1_VM2	32
	VI0_P11	31
EMOCIONAL	PI0_PM0_VF0_PF2	1
	PI0_VF0	1

PI0_VM0	1
PM0	1
VI0	1
VI0_PF0_VF2	1
VI0_PI1_VI2_VF2	1
VI0_PM0_PM1	1
VI0_PM0_VI1	1
VI0_PM0_VI1_VF2	1

Tabla 35: Patrones interiores más frecuentes en los grupos acentuales de tres sílabas con acento en la primera sílaba

Los resultados obtenidos para los patrones interiores son muy parecidos a los de los patrones iniciales.

En los patrones en los que los grupos acentuales están formados por una o dos sílabas, se observa una mayor tendencia a presentar un único punto de inflexión, normalmente alto (P), en el caso del corpus neutro.

En el corpus emocional, la tendencia es realizar dos puntos de inflexión tanto en los grupos acentuales de una sílaba como de dos.

Del mismo modo que ocurría en los grupos acentuales iniciales, cuando éste está formado por 3 sílabas, los patrones no se repiten en el corpus emocional: cada patrón se realiza una única vez.

Nuevamente, el patrón VI0_PM0, que presenta una inflexión inicial V y otra P a mitad de la sílaba tónica, es de los más frecuentes. En el caso del corpus neutro es el patrón que registra un mayor número de casos en los tres grupos, y en el corpus emocional aparece como la segunda realización más frecuente en los grupos acentuales de una y dos sílabas, mientras que en el de tres sílabas no está presente.

La Figura 27 y la Figura 28 muestran un ejemplo de este patrón (VI0_PM0) en un enunciado del corpus neutro y emocional respectivamente.

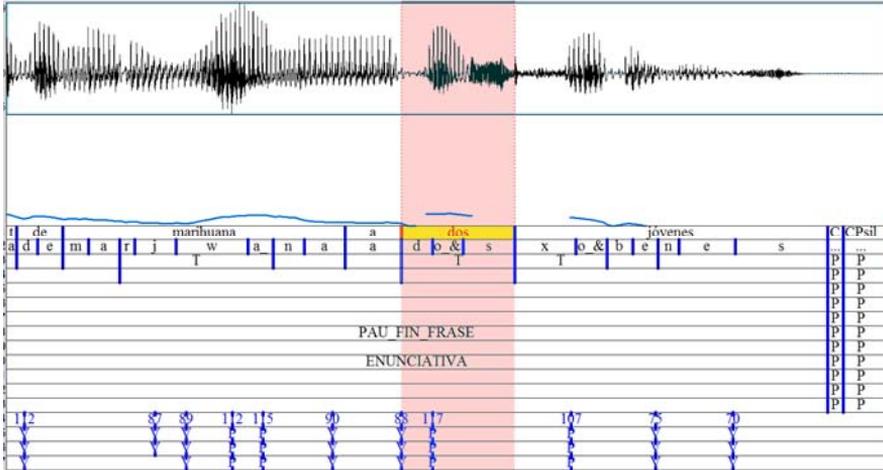


Figura 27 Ejemplo de patrón interior de enunciado VIO_PM0 en el GA 'dos', dentro del enunciado neutro 'Piden cinco años para un empresario por dar tarta de marihuana a dos jóvenes.'

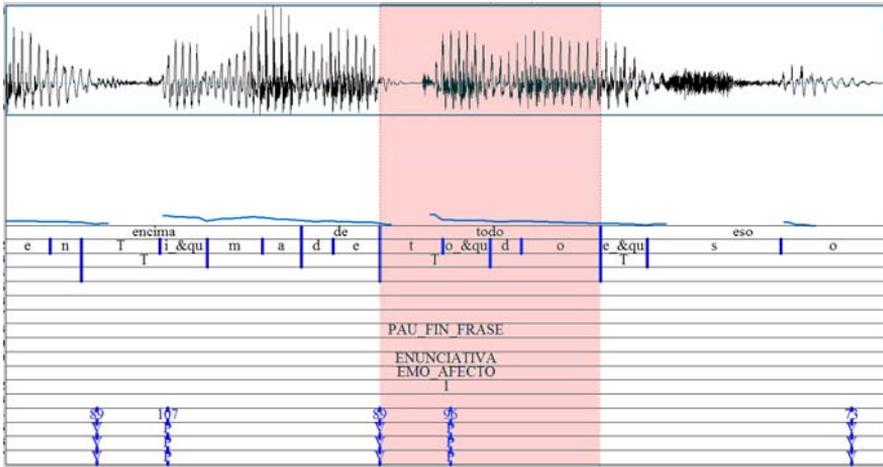


Figura 28: Ejemplo de patrón interior de enunciado VIO_PM0 en el GA 'todo' dentro del enunciado emocional 'Porque lo entiendes. El amor está por encima de todo eso'

3. Patrones finales no final de enunciado

La Tabla 36 recoge los cinco patrones más frecuentes del corpus emocional y neutro, distribuidos según el número de sílabas de cada GA. El Apéndice C muestra todos los patrones finales no finales de enunciado clasificados según el tipo de emoción, y la frecuencia con la que se han manifestado.

Tipo de corpus	1 sílaba	2 sílabas	3 sílabas
Emocional	0	PI0_VM1	-
	VF0	VI0_PI1_VF1	
	VI0_PM0_VF0	PM0_VM1_PF1	
	PF0	PM1_VF1	
	PI0_VF0	VI0_PFO_VM1_VF1	
Neutro	VI0_PM0_PFO	PI0_VI1_VM1	VF0_PI2_PFO
	VI0_PM0_VF0	0	0
	VI0_PFO	PM0_VI1_VM1	PI0_VM1_VM2
	PI0_VF0	PI0_VI1_VF1	PM0_VF1_VF2
	PF0	VI1_PM1	PM0_VF1_VM2

Tabla 36: Resumen de los patrones finales no finales de enunciado más frecuentes del corpus emocional y neutro, de los grupos acentuales formados por 1, 2 y 3 sílabas, y cuya tónica es la primera

Se observa una mayor presencia de patrones en los que no hay ninguna inflexión (0), concretamente en el corpus emocional cuyos grupos acentuales están formados por una sílaba, y en el corpus neutro con dos y tres sílabas.

Tres de los cinco patrones empleados en el corpus emocional de una sílaba también aparecen en el corpus neutro (VI0_PM0_VF0, PI0_VF0 y PF0). En el caso de los grupos acentuales formados por dos sílabas no se observa este comportamiento, esto puede ser

debido al reducido número de casos que presenta el corpus emocional, cuyo número máximo de repeticiones de un mismo patrón no supera el tres.

La Figura 29 y la Figura 30 muestran los patrones más frecuentes observados en el corpus neutro para una sílaba y dos sílabas por GA respectivamente. La Figura 31 muestra el segundo patrón más frecuente de los observados en el corpus emocional formado por una sola sílaba, ya que el patrón con mayor frecuencia es '0', es decir, ausencia de inflexión. Finalmente, en la Figura 32 se muestra un ejemplo del patrón más frecuente del corpus emocional cuyo, GA está formado por 2 sílabas.

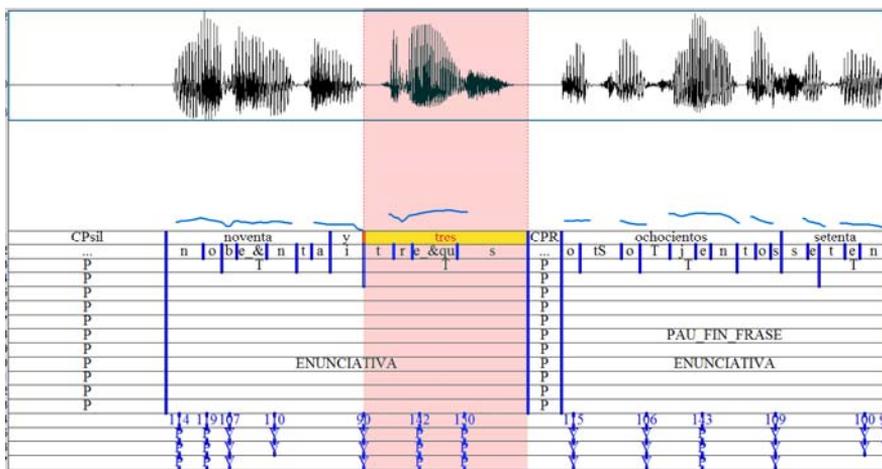


Figura 29: Ejemplo de patrón ascendente final no final de enunciado VI0_PM0_PF0 en el GA 'tres' dentro del enunciado neutro 'Noventa y tres, ochocientos setenta y nueve, cuarenta y nueve, catorce'

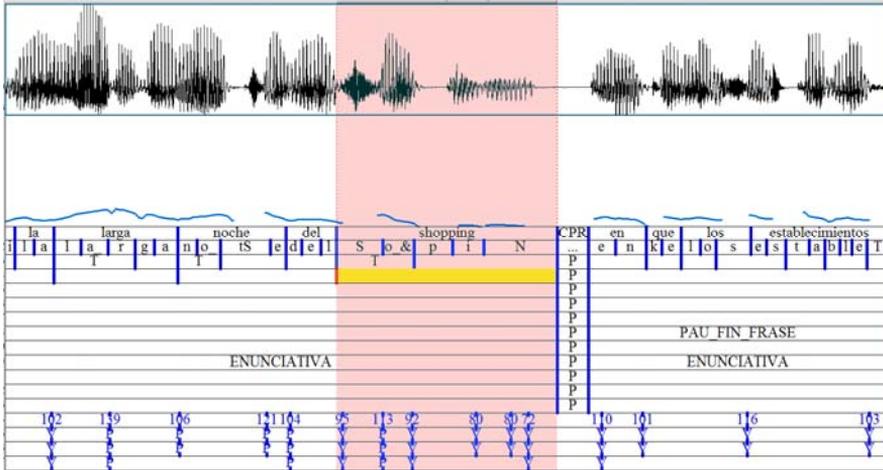


Figura 30: Ejemplo de patrón descendente final no de enunciado PI0_VI1_VM1 en el GA 'shopping' dentro del enunciado neutro 'La larga noche de los museos y la larga noche del shopping, en que los establecimientos alargaron su horario de apertura al público'

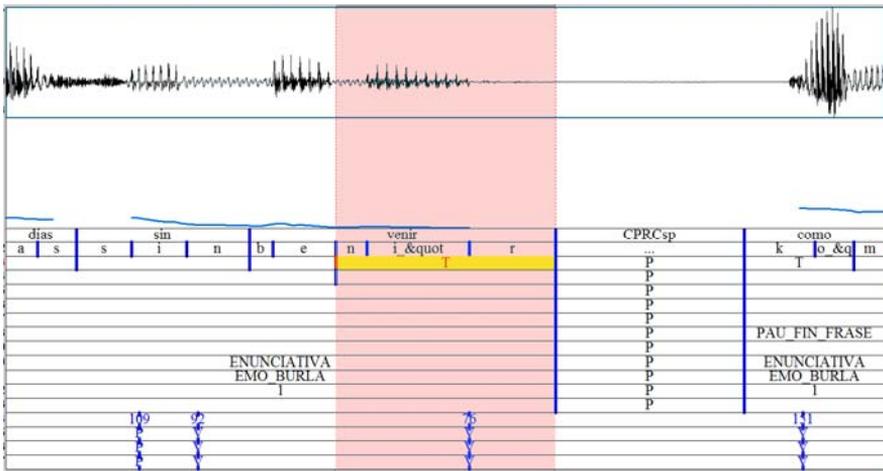


Figura 31: Ejemplo de patrón descendente final no final de enunciado VF0 en el GA 'nir' dentro del enunciado emocional 'Luego se tirará unos días sin venir, como cuando ganó el Madrid en el Camp Nou'

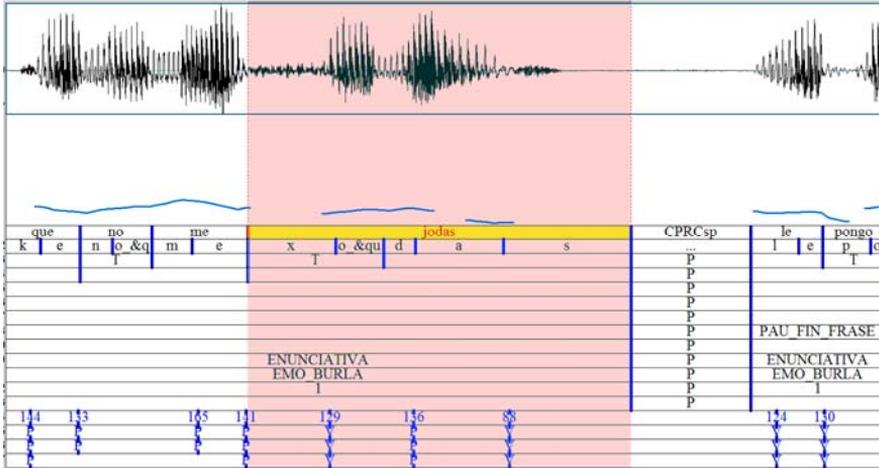


Figura 32: Ejemplo de patrón circunflejo final no final de enunciado VI0_P11_VF1 en el GA ‘jodas’ dentro del enunciado emocional ‘Hombre, es que no me jodas, le pongo mi foto y me come, así, directamente’

En el caso de los patrones finales no finales de enunciado y los patrones finales también se ha realizado una clasificación de los mismos según si son ascendentes (A), descendentes (D) o circunflejos (C). Para ello, en vez de agrupar todos los patrones emocionales bajo una sola etiqueta, se ha decidido tener en consideración las diferentes emociones para observar si existe una cierta tendencia a emplear un determinado tipo de patrón dependiendo de cada emoción. La Tabla 37, la Tabla 38 y la Tabla 39 recogen esta clasificación mostrando los patrones según el número de sílabas que conforman cada grupo entonativo; 1, 2 y 3 respectivamente.

EMOCIÓN	A	D	C
BURLA	1	2	0
DECEPCIÓN	1	0	1
DUDA	0	1	0
ENFADO	0	1	1
ENFADO-RECHAZO	0	1	0

EXTRAÑEZA	1	0	0
ORGULLO	0	2	0
RECHAZO	0	1	0
RESIGNACIÓN	3	0	0
NEUTRO	324	174	64

Tabla 37: Clasificación de los patrones finales no finales de enunciado, cuyo GA está formado por una sílaba

EMOCIÓN	A	D	C
ADMIRACIÓN	0	0	1
AFECTO	3	3	1
ALEGRÍA	1	0	0
ALTERACIÓN	0	1	0
BURLA	1	2	0
BURLA-RECHAZO	1	0	0
DECEPCIÓN	0	1	0
DUDA	1	2	0
ENFADO	0	0	1
ENFADO-RECHAZO	2	1	1
INTERÉS	1	0	1
RESIGNACIÓN	3	0	7
TRISTEZA	1	1	0
DESPEDIDA	0	0	1
NEUTRO	484	320	132

Tabla 38: Clasificación de los patrones finales no finales de enunciado, cuyo GA está formado por dos sílabas, y la primera es la tónica

TIPO CORPUS	A	D	C
EMOCIONAL	-	-	-
NEUTRO	39	31	6

Tabla 39: Clasificación de los patrones finales no finales de enunciado, cuyo GA está formado por tres sílabas, y la primera es la tónica

La observación de las diferentes tablas permite comprobar que los patrones ascendentes son los más frecuentes en el corpus neutro, independientemente del número de sílabas por el que esté formado el GA. Su uso oscila entre el 51% (en el caso de los grupos acentuales formados por dos sílabas) y el 57% de los casos (grupos acentuales constituidos por una sílaba). El segundo tipo de patrón más frecuente es el descendente. Los patrones circunflejos son minoritarios, su uso no supera el 15% en ninguno de los casos analizados.

En el caso del corpus emocional, se ha observado una mayor frecuencia de patrones descendentes, muy seguidos de los ascendentes en el caso de los GA formados por una sílaba. En cambio, en los GA formados por dos sílabas, la distribución entre los tres tipos de patrones es muy pareja.

En el caso de los GA finales formados por una sílaba, en el corpus neutro hay una tendencia a realizar prácticamente todos los patrones ascendentes (96%). En el corpus emocional el mayor número de patrones son descendentes, muy seguidos de los ascendentes. En el caso de los grupos acentuales formados por dos sílabas, los patrones circunflejos son el segundo grupo más frecuente, con una diferencia mínima de un caso respecto a los patrones ascendentes. Este aumento en la frecuencia no se debe a una mayor distribución entre las diferentes emociones, sino que el mayor número de realizaciones de patrones circunflejos tiene lugar en la emoción resignación. En el resto de emociones, o no se da este patrón o sólo se observa una vez.

Respecto a los grupos acentuales finales formados por 3 sílabas, no se ha observado ningún caso en el corpus emocional.

4. Patrones finales de enunciado

La Tabla 40 recoge los cinco patrones más frecuentes del corpus emocional (sin hacer diferencias entre las diferentes emociones que lo conforman), y neutro, clasificados según el número de sílabas del GA al que pertenecen. En el Apéndice C.5 Patrones finales de enunciado se muestran todos los patrones finales de enunciado clasificados según el tipo de emoción, y la frecuencia con la que se han manifestado.

Tipo de corpus	1 sílaba	2 sílabas	3 sílabas
Emocional	VF0	VM1	PI0_VF1_VF2
	VI0_PFO	VI0_VM1	PM0_VI1_VM2_PFO
	VI0_PM0_VF0	PI0_VM1	PM0_VM1_VM2
	VI0_VF0	VF1	VF2
	VI0_PM0_PFO	PI1_VM1	VI0_PM0_PFO_VM2
Neutro	VF0	PI0_VI1_VM1	VI1_VM2
	VI0_VF0	VM1	PI0_VI1_VM2
	PI0_VM0	VI1_VM1	VF0_VM2
	VM0	PI0_VF0_VM1	PI0_VI1_VF2
	VI0_VM0	PI0_VM1	VF1_VM2

Tabla 40: Resumen de los patrones finales de enunciado más frecuentes del corpus emocional y neutro, de los grupos acentuales formados por 1, 2 y 3 sílabas, y cuya tónica es la primera

En los grupos acentuales de una sílaba, el patrón VF0 es el más frecuente tanto en el corpus emocional como en el neutro. En los

grupos acentuales formados por dos sílabas, el patrón más frecuente en el corpus emocional (VM1) es el segundo más frecuente en el corpus neutro. En el caso de los grupos fónicos constituidos por tres sílabas, no se observa ninguna repetición entre los cinco patrones más frecuentes de ambos corpus. Esto puede deberse al reducido número de patrones que presenta el corpus emocional, unido a la poca redundancia de los mismos, ya que todos ellos sólo presentan un único caso.

La Figura 33, la Figura 34, la Figura 35 y la Figura 36 muestran unos ejemplos de los patrones más frecuentes de los grupos acentuales formados por una o dos sílabas de los corpus neutros y emocionales analizados en esta investigación.

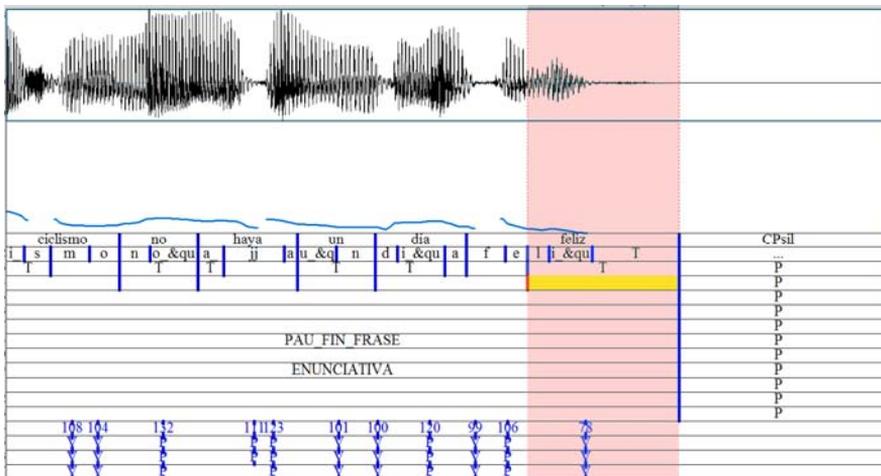
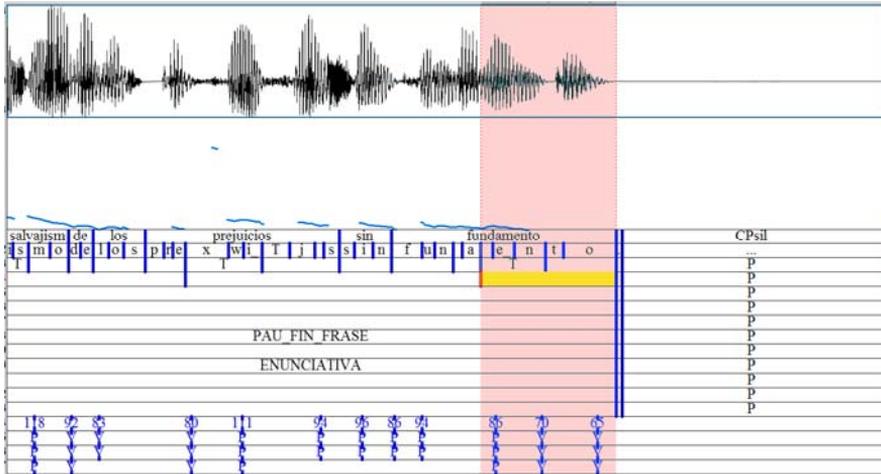


Figura 33: Ejemplo de patrón descendente final de enunciado VF0 en el GA 'liz' dentro del enunciado neutro 'Es triste que en el ciclismo no haya un día feliz'



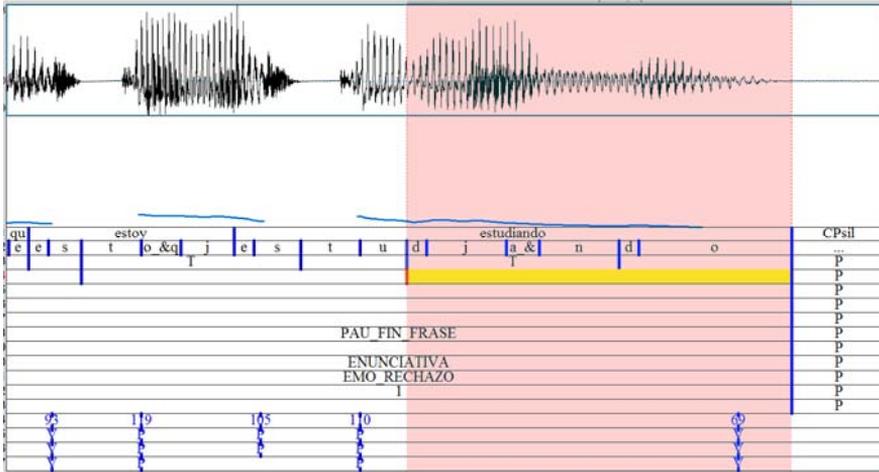


Figura 36: Ejemplo de patrón circunflejo final de enunciado VM1 en el GA 'diando' dentro del enunciado emocional 'Odio lo que estoy estudiando'

La Tabla 41, la Tabla 42 y la Tabla 43 muestran una clasificación de los patrones del corpus para los grupos acentuales finales de oración de 1, 2 y 3 sílabas respectivamente. Los patrones se han clasificado en tres posibles tipos: ascendentes (A), descendente (D) y circunflejo (C).

EMOCIÓN	A	D	C
ADMIRACIÓN	1	0	0
AFECTO	0	1	0
ALEGRÍA	1	6	0
ALTERACIÓN	0	3	0
BURLA	1	3	1
BURLA-RECHAZO	2	1	1
DECEPCIÓN	5	1	1
DIVERSIÓN	2	0	0
DUDA	5	1	0
ENFADO	0	3	3

ENFADO-RECHAZO	0	1	0
ORGULLO	1	4	0
RECHAZO	2	1	1
RESIGNACIÓN	1	6	1
SORPRESA	1	0	0
TRISTEZA	3	3	1
DESPEDIDA	0	2	1
SALUDO	2	0	0
NEUTRO	67	632	31

Tabla 41: Clasificación de los patrones finales de enunciado, cuyo GA está formado por una sílaba

EMOCIÓN	A	D	C
ADMIRACIÓN	0	0	3
AFECTO	2	11	6
ALEGRÍA	1	9	1
ALTERACIÓN	1	7	0
BURLA	2	12	1
BURLA-RECHAZO	4	6	1
DECEPCIÓN	2	8	6
DIVERSIÓN	1	5	5
DUDA	2	6	4
ENFADO	1	8	2
ENFADO-RECHAZO	1	8	1
EXTRAÑEZA	1	1	0
ORGULLO	4	11	4
RECHAZO	1	10	5
RESIGNACIÓN	3	9	3

SORPRESA	0	1	0
TRISTEZA	3	8	5
DESPEDIDA	0	5	1
SALUDO	3	12	2
NEUTRO	192	1.258	118

Tabla 42: Clasificación de los patrones finales de enunciado, cuyo GA está formado por dos sílabas, y la primera es la tónica

EMOCIÓN	A	D	C
AFECTO	0	1	0
BURLA	0	1	0
BURLA-RECHAZO	0	0	1
DIVERSIÓN	0	1	0
ENFADO-RECHAZO	0	1	0
RECHAZO	0	1	0
RESIGNACIÓN	2	0	0
NEUTRO	21	109	11

Tabla 43: Clasificación de los patrones finales de enunciado, cuyo GA está formado por tres sílabas, y la primera es la tónica

Estas tablas muestran una predominancia de los patrones descendentes en posición final de enunciado, tanto en el corpus emocional como en el neutro, e independientemente del número de sílabas por el que esté formado.

Esta predominancia es prácticamente absoluta en el corpus neutro, siempre con una frecuencia mayor del 75%. En cambio, en el corpus emocional, a pesar de que el patrón descendente es el más frecuente, se observa una mayor alternancia entre los tres tipos de patrones.

El aumento de los patrones circunflejos en el corpus emocional es un dato destacable. Mientras que en el corpus neutro este tipo de patrón era muy reducido, en el corpus emocional ha experimentado un considerable ascenso en su uso, sobre todo en el caso de los grupos fónicos formados por dos sílabas, que ha pasado del 7,5% al 22,8%. En el caso de los grupos acentuales formados por tres sílabas continúa siendo minoritario, ya que sólo se recoge una realización de este tipo, correspondiente a la combinación de emociones ‘burla-rechazo’.

Se observa un mayor empleo del patrón circunflejo en emociones como el enfado, en los grupos fónicos formados por una sílaba, y el afecto, la decepción, la diversión, la duda, el orgullo, el rechazo y la tristeza en los grupos fónicos formados por dos sílabas.

c. Conclusiones

Los resultados obtenidos tanto para la altura como para el rango tonal presentan diferencias significativas entre el corpus neutro y el corpus emocional. En cuanto a las cuatro emociones básicas, todas ellas presentan una altura tonal superior al neutro, excepto la tristeza, que es inferior. Estos resultados son equiparables a los presentados tanto por Rodríguez *et al.* (1999) como por Iriondo (2008), y parcialmente similares a los obtenidos por Garrido (2011). En el caso del rango, no todas las emociones presentan valores superiores al neutro. Los datos obtenidos se asemejan a los presentados por Montero (2003), y difieren en cierta medida de los de Garrido *et al.* (2011) en los que la tristeza presenta unos valores superiores al neutro. Esto hace pensar que probablemente estas diferencias se deban a que hay cierta variación en el modo de expresar esta emoción concreta por parte de los diferentes locutores.

Respecto a los patrones melódicos, los iniciales e interiores no presentan grandes diferencias entre el habla neutra y el habla emocional. En cambio, sí se observan diferencias entre los patrones finales empleados en ambos corpus.

Mientras que en posición inicial e interior los patrones predominantes son VI0_PM0 y VI0, en los patrones finales se observa un mayor uso de los patrones VM1, VI0_VM1 y VF0. En

esta posición también hay una mayor alternancia entre los tres tipos, siendo los circunflejos los que muestran un mayor aumento. Como bien comenta Garrido (2011), los patrones entonativos “circunflejos parecen ser una marca genérica de emotividad, más que una marca específica de una determinada emoción”, ya que se emplean los mismos patrones con independencia de la emoción a expresar.

En este análisis el uso de patrones circunflejos es disperso y dispar entre las emociones analizadas, siendo la resignación y el enfado las que más números de casos recogen, y la admiración, el afecto, el enfado-rechazo y el interés los que menos. Emociones como la alegría, la alteración, la diversión, la duda, el orgullo, el rechazo y la sorpresa no presentan ningún patrón circunflejo. Este comportamiento no se adecúa con los datos presentados por Garrido (2011) en los que emociones con un mayor nivel de activación presentan un rango y altura tonal mayor, y también una mayor frecuencia de uso de los patrones circunflejos. El reducido número de casos quizá sea la causa de estos resultados, posiblemente con una ampliación del corpus de análisis, se podrían equilibrar los resultados y asemejarse más a estudios anteriores. Estos resultados pueden afectar al modelado paramétrico de la síntesis de habla repercutiendo negativamente en la correcta identificación de las distintas emociones.

4.3.3. Generación del habla expresiva

Al no poder realizar todas las modificaciones paramétricas deseadas en la señal acústica directamente en el proceso de generación de la misma mediante el CTH de Cereproc, ya que el resultado obtenido no es satisfactorio, la generación y modelado del habla expresiva se lleva a cabo mediante una simulación del proceso. Los enunciados se generan primero con el CTH, utilizando las etiquetas XML correspondientes para seleccionar las unidades de los distintos subcorpus (neutro y emocional) según el tipo de enunciado que se desee realizar. Posteriormente, se modifica la señal, de forma automática, mediante un modelado paramétrico de la prosodia para intentar transmitir así el estado emocional o fisiológico deseado.

En el caso de querer realizar un enunciado neutro, únicamente será necesario escribir la frase que se desea reproducir. Si por el contrario, se pretende realizar un enunciado emocional es necesario emplear las etiquetas XML comentadas en el epígrafe 2.2 para que el resultado sea mejor.

El modelado paramétrico se realiza automáticamente mediante un script de Praat que modifica al mismo tiempo los parámetros melódicos y los no melódicos. Para ello emplea las tablas de las ratios generadas durante el análisis prosódico.

A continuación, a pesar de que el proceso real se lleva a cabo de forma secuencial, se exponen los procesos desarrollados de manera separada.

4.3.3.1. Parámetros no melódicos

El análisis de los parámetros no melódicos estaba formado por la velocidad de locución, la intensidad, la amplitud y la duración de la vocal prepausal. En el proceso de modelado, a pesar de resultar un parámetro significativo estadísticamente, no se va a tener en cuenta el parámetro de la duración de la vocal prepausal ya que el procedimiento de modelado de la señal no permite conocer cuándo empieza dicho alófono en cada reproducción, y por lo tanto, la modificación no podría realizarse de manera automática.

Para modelar la señal acústica según nuestros intereses, es necesario indicar en el script el tipo de modalidad oracional a la que pertenece el enunciado (enunciativa, exclamativa, interrogativa total, interrogativa parcial, interrogativa relativa, interrogativa confirmativa o interrogativa alternativa), y en función de qué aspectos se va a llevar a cabo dicho modelado, es decir, se ha de indicar el estado emocional que se desea reproducir, el nivel emocional de éste y/o el acto de habla.

El orden con el que se realizan las modificaciones es el siguiente:

1. Primero se modifica el F0, cuyo proceso se explica en el epígrafe siguiente.

2. Esta señal ya modificada es la que se toma para modelar la velocidad de locución.
3. La señal resultante de estas dos modificaciones previas es la que se emplea para modelar la intensidad y la amplitud, obteniendo así la señal acústica final.

Todas estas modificaciones se realizan según los aspectos que se indica en el momento de lanzar el script. Cada script específico para modelar cada uno de los parámetros analizados recurre a las tablas de las ratios que recogen las diferencias observadas entre el corpus neutro y el emocional. El script realiza una búsqueda de los aspectos indicados y realiza la modificación según el valor marcado. Por ejemplo, si tenemos una señal acústica generada a partir del corpus emocional del CTH de Cereproc, y cuya modalidad oracional es enunciativa y deseamos modificarla para que exprese el estado emocional de ‘sorpresa’, con un nivel de intensidad 1 (equivalente al nivel medio), el script realiza una búsqueda en la tabla con la información de las ratios hasta localizar la línea que reúne todas estas características (véase Tabla 23 para observar la distribución de la información), y así poder tomar el valor que aparece al final de la línea referente al parámetro en cuestión: por ejemplo, enunciativa, emoción sorpresa, nivel emoción 1, acto de habla indefinido. Una vez obtenido el valor que relaciona el enunciado neutro con el emocional, se emplea para modificar la señal acústica.

Este mismo proceso se realiza con cada uno de los parámetros no melódicos.

4.3.3.2. Parámetros melódicos

En el modelado de los parámetros melódicos, del mismo modo que en el análisis, se ha tenido en consideración tanto los aspectos globales como los locales.

Antes de llevar a cabo la modificación es imprescindible realizar un paso previo: disponer de un fichero ‘.contour’ con la predicción del contorno de F0 para cada GA del enunciado. Para la generación del

fichero ‘contour’ a partir del texto del enunciado se ha empleado la herramienta GenProso (Garrido, 2012d), que predice el contorno más idóneo para cada grupo entonativo a partir de las de las tablas de patrones frecuentes obtenidos en la fase de análisis, teniendo en consideración no sólo la emoción y el nivel de intensidad o el acto de habla que queremos transmitir, sino también la modalidad oracional, la posición del GA dentro del enunciado (inicial, interior, final no de enunciado, final de enunciado), el número de sílabas que contiene dicho GA y en qué lugar se encuentra la sílaba tónica. El proceso finaliza con la creación de un fichero ‘.contour’ que contiene la cadena de patrones seleccionados para cada GA que forma el texto de entrada.

Una vez obtenida toda esta información, se da paso al modelado. La reproducción de los parámetros globales y locales se realiza de forma sucesiva mediante una serie de scripts de Praat.

En el caso de los parámetros locales, la información sobre el contorno almacenada en el fichero ‘.contour’ es copiada en un nuevo tier del TextGrid, y alineada con las sílabas del enunciado. Para los parámetros globales, el proceso para modificarlos es similar al comentado en el epígrafe anterior. Un script calcula las líneas de referencia a partir de la información del nuevo tier con la información del contorno a realizar y del tipo de enunciado que se desea generar.

4.4. Resumen

En este capítulo se ha abordado la creación de la voz sintética expresiva. Para ello ha sido necesario realizar una serie de procedimientos que conllevan desde la realización del corpus hasta la generación del habla expresiva.

La creación de cualquier tipo de voz sintética requiere un proceso previo a la síntesis cuyo objetivo es crear la base de datos acústica con la que se realizará la síntesis, y una segunda etapa que consiste en la propia generación del enunciado escrito en oral.

La primera etapa se realiza una única vez por cada nueva voz sintética que se desee crear, y engloba la creación del corpus, la selección del locutor y la grabación del corpus oral.

Uno de los objetivos marcados para esta investigación es explorar la generación de habla emocional utilizando una voz sintética expresiva, capaz de transmitir unas determinadas emociones. Este hecho hace que el corpus oral esté formado por un subcorpus de contenido neutro, que en el caso de esta tesis está constituido por noticias, direcciones, números, letras, etc., y un subcorpus emocional.

En este caso concreto, al centrarse la tesis en el dominio de las conversaciones virtuales, es imprescindible que el corpus emocional sea representativo de dicho dominio. Para asegurar que esto ocurre se han seleccionado las emociones y los actos de habla más frecuentes de un corpus de conversaciones virtuales etiquetado y revisado. El corpus de grabación contiene enunciados representativos de 19 emociones, simples o combinadas, y 2 actos de habla. Para cada uno de ellos se han seleccionado 21 enunciados, lo que hace que el corpus esté formado por 525 enunciados en su totalidad.

En cuanto al tipo de locutor, al tratarse de una grabación de un corpus neutro y emocional, aparte de disponer de una buena dicción, también era necesario que tuviera unas buenas dotes interpretativas para que la lectura del corpus emocional fuera lo más natural posible, y asegurar así la correcta interpretación de las diferentes emociones. La selección del locutor se realizó a partir de un casting que valoraba ambos aspectos.

La grabación se realizó en un estudio insonorizado de la Universidad Pompeu Fabra. Posteriormente se llevó a cabo la edición y revisión del corpus oral.

Normalmente, en un sistema de síntesis por corpus comercial como el de Cereproc, que es el empleado en esta investigación, este sería todo el proceso previo a la generación de la síntesis. A partir de esta base de datos el conversor podría convertir en un enunciado oral cualquier texto.

En cambio, como el objetivo de esta tesis es realizar la síntesis del habla empleando una aproximación paramétrica, aportando conocimiento lingüístico al modelado del habla, es necesario

analizar acústicamente el corpus oral para poder extraer una serie de valores específicos del locutor escogido, que permitan modelar los enunciados con el objetivo de transmitir unas determinadas emociones.

Para el análisis se ha seleccionado una serie de parámetros melódicos y no melódicos: velocidad de elocución, duración de la vocal prepausal, intensidad, amplitud, rango de la F0, la altura tonal, y los patrones melódicos. Estos parámetros ya han sido analizados en estudios anteriores (Montero 2003, Iriondo 2008, Garrido 2012).

Durante el análisis de ambos corpus (neutro y emocional) se han creado una serie de tablas que además de permitir observar los valores de cada parámetro según la emoción realizada, también se emplean para la posterior modificación del enunciado oral.

Todos los parámetros no melódicos son significativos para la generación de habla expresiva, y cada emoción presenta un comportamiento particular, que no permite crear correspondencias entre el modo de proceder y las características de valencia y activación de las diferentes emociones. Aun así, las cuatro emociones básicas (alegría, enfado, sorpresa y tristeza) que se encuentran en el corpus muestran similitudes con los resultados obtenidos en otros estudios anteriores (Montero 2003, Iriondo 2008), a excepción de la tristeza.

Con respecto a los parámetros melódicos, los resultados obtenidos en los parámetros globales, que hacen referencia al rango de F0 y la altura tonal de las emociones básicas, son similares a otros estudios (Montero 2003) donde todas las emociones presentan un rango y una altura tonal superior al neutro a excepción de la tristeza, que muestra una altura tonal inferior.

En el caso de los patrones melódicos, el análisis se ha realizado teniendo en cuenta la posición del GA dentro de cada enunciado: inicial, interior, final y final de enunciado.

Los patrones iniciales e interiores no muestran grandes diferencias entre el habla neutra y el habla emocional, donde predominan los patrones VI0_PM0 y VI0. En cambio, en los patrones finales y finales de enunciado aparte de observar una mayor diferencia en el uso de los patrones dependiendo del tipo de habla, también se observa un mayor empleo de patrones circunflejos en el corpus

emocional, tal y como se observa en otros estudios previos (Garrido 2011, 2012c).

Finalmente, toda esta información ha sido empleada para la preparación de un método para la generación paramétrica de habla expresiva. Al no poder realizar las modificaciones deseadas en la señal acústica directamente en el proceso de generación de la misma mediante el CTH de Cereproc, se lleva a cabo una simulación del proceso realizando el modelado prosódico *a posteriori*. El modelado se realiza de forma automática mediante una serie de scripts en Praat, modificando simultáneamente los parámetros melódicos y no melódicos empleando las tablas obtenidas durante el análisis.

Capítulo 5

EVALUACIÓN

En este capítulo se presentan las evaluaciones realizadas tanto para la normalización de textos como para la voz sintética expresiva.

5.1. El normalizador de textos

El objetivo principal de la evaluación del normalizador de textos es comprobar si el grado de corrección-normalización que ofrece es adecuado.

A continuación se expone el procedimiento desarrollado para la evaluación del normalizador de textos TexAFon 2.0, y los resultados obtenidos.

5.1.1. Procedimiento

Para la evaluación se ha empleado una conversación pública extraída del servidor web www.inforchat.com durante el año 2008. Está formada por 1.222 intervenciones, y 4.069 tokens, lo que equivale al 10% del corpus de entrenamiento, y no fue empleada para el análisis lingüístico de las conversaciones virtuales.

Para poder llevar a cabo la evaluación tal y como fue planteada era necesario disponer de tres versiones diferentes del mismo corpus:

1. La versión original de los mensajes, tal y como fueron escritos por los diversos usuarios.

2. Una versión corregida de todos los errores y desviaciones que se encuentran en las conversaciones. Para asegurar que se corrigen todos los errores, la corrección se ha realizado manualmente.
3. La normalización realizada automáticamente mediante la herramienta TexAFon 2.0.

La Tabla 44 recoge una serie de enunciados tal y como aparecen en cada una de las versiones anteriormente citadas.

Texto original	Texto normalizado manualmente	Texto normalizado automáticamente
Eso no queria poner	Eso no quería poner	Eso no quería poner
Si simpaticabcn	Sí simpática Barcelona	Si simpática Barcelona
:) ami el almaaaaaa	[emoticono] a mí el alma	[signo_dos_puntos] [signo_cierra_paréntesis] a mí el alma
jajajaj es k to no pue ser	Jajaja es que todo no puede ser	Jajaja es que to no púe ser

Tabla 44: Algunos ejemplos extraídos de las tres versiones diferentes (texto original, normalización manual y normalización automática) necesarias para realizar la evaluación de TexAFon 2.0

Una vez obtenidas las tres versiones, y para poder valorar el funcionamiento del normalizador de textos TexAFon 2.0, se ha realizado una comparación automática mediante scripts de Python de cada uno de los tokens en cada una de las versiones. Cada una de las palabras del texto original ha sido comparada con la normalización propuesta automáticamente por el normalizador

TexAFon 2.0 y a su vez con la propuesta en la normalización manual, teniendo en cuenta que algunos tokens de la versión original podían expandirse en dos o más palabras.

Cuando los tokens de cada una de las versiones no presentaban ningún cambio entre sí, se consideraba que la palabra escrita en la versión original era correcta y no debía realizarse ninguna modificación. En cambio cuando se observaba alguna diferencia esta podía deberse a que:

- La palabra original es diferente a las propuestas realizadas tanto por la normalización automática como la manual: la palabra original es incorrecta y la normalización propuesta por TexAFon 2.0 es correcta.
- La palabra original y la propuesta en la normalización automática son iguales, pero la indicada en la normalización manual es diferente: la palabra original es incorrecta, TexAFon no ha detectado que era errónea y no la ha modificado, o bien sí que ha detectado que no era correcta pero no ha encontrado ninguna posible normalización, por lo que ha escrito la palabra tal cual aparecía en la versión original.
- La palabra original y la propuesta en la normalización manual son iguales, pero la normalización seleccionada en la versión automática es diferente: la palabra original era correcta y no debía ser modificada, pero TexAFon la ha considerado incorrecta al no encontrarse en ninguno de los diccionarios de trabajo y la ha modificado ofreciendo una propuesta diferente.
- Los tokens de las tres versiones son diferentes: la palabra original es incorrecta, TexAFon ha detectado que era errónea pero la propuesta que ha seleccionado no es correcta.

Para cada una de las posibilidades mencionadas se han extraído porcentajes que permiten evaluar el funcionamiento de TexAFon 2.0. A continuación se muestran los resultados obtenidos.

5.1.2. Resultados

El corpus original empleado para la evaluación presenta un total de 1.579 tokens incorrectos, lo que equivale a un 38,8% del total del corpus. Este porcentaje es considerablemente superior al observado en el corpus de análisis, que contaba con un 21,5%.

Del número total de palabras incorrectas, TexAFon 2.0 ha corregido correctamente el 70,3% de ellas, consiguiendo un porcentaje total de palabras correctas del texto de salida del 85,9%, lo que supone una mejora de 24,7%. En la Figura 37 se puede observar este incremento de palabras correctas después de la normalización automática.

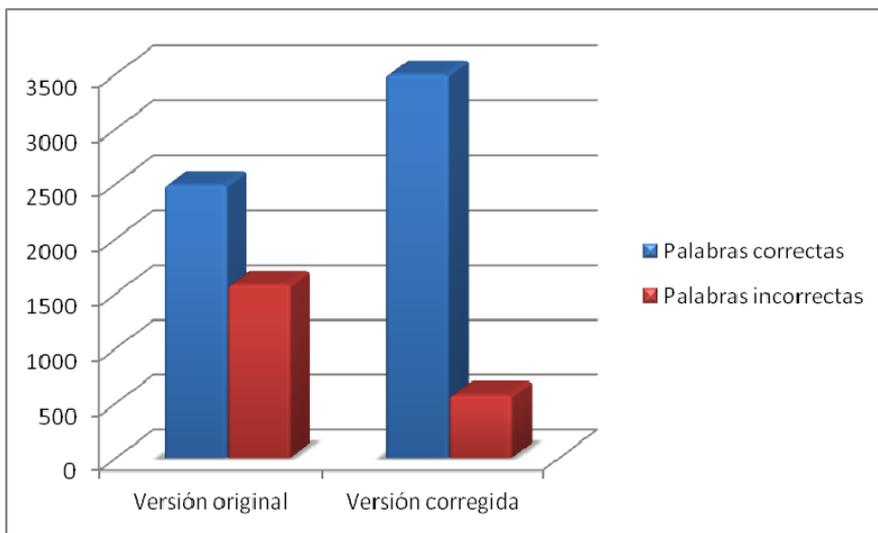


Figura 37: Distribución de las palabras correctas e incorrectas presentes en el texto original y la versión corregida automáticamente mediante TexAFon 2.0

El 14,1% restante corresponde a palabras incorrectas que no se han corregido adecuadamente, o que simplemente no se han corregido, y a palabras que el normalizador ha considerado incorrectas cuando no lo eran. En la Figura 38 se presenta la distribución de las palabras incorrectas en el texto final, después del proceso de normalización.

que no han sido identificadas correctamente en la primera etapa del normalizador, y que tampoco recoge el diccionario de siglas (1,92%), y finalmente palabras combinadas con cifras cuyos dígitos se han convertido en vocales proporcionando palabras diferentes a las de la normalización manual (1,92%). En la Tabla 45 se muestra una serie de ejemplos de estos fenómenos comentados: a la izquierda aparece el texto original, en la segunda columna se muestra la corrección ofrecida por el normalizador TexAFon 2.0, y la columna de la derecha recoge la normalización propuesta en la corrección manual.

Texto original	Normalización manual	Normalización automática
k te han exo	qué te han hecho	que te han hecho
b_d	[emoticono]	be guión bajo de
viendo una peli en el tdt	viendo una peli en el te de te	viendo una peli en el ddt
luna_llena30	luna guión bajo llena treinta	luna guión bajo llena eo

Tabla 45: En negrita, ejemplos de palabras incorrectas que se han corregido mal y cuyo proceso de corrección sólo implicaba una modificación. A la derecha se encuentra la palabra en la intervención original de la conversación virtual, en el centro la corrección realizada manualmente, y a la derecha la propuesta de normalización de TexAFon 2.0

De las palabras marcadas como incorrectas, y cuya corrección no se ha realizado, el 84,6% se corresponde con palabras existentes en la lengua española, pero que en el contexto en el que han sido utilizadas por los usuarios no se consideran correctas. Palabras como ‘ultimo’, ‘que’, ‘ola’ o ‘chiste’ son palabras correctas pero no en contextos como los siguientes:

- a) Pal **ultimo** me deja (Para el **último** me deja)
- b) **Que** comprastes ayre (¿**Qué** compraste, Ayre?)
- c) Estrella **ola** (¡Estrella, **hola!**)
- d) Han desaparecido los **chiste** (Han desaparecido los **chistes**)

En la gran mayoría de estas palabras, concretamente el 69,23% de ellas, lo que las hace incorrectas es la acentuación; no han sido acentuadas cuando en el contexto en el que se encuentran requiere del uso de la tilde diacrítica para ser diferenciadas de otras palabras que se escriben de la misma forma pero cuyo significado es completamente diferente.

Por otra parte, algunos usuarios han escrito varias palabras aglutinadas, elidiendo el espacio en blanco entre ellas. Este fenómeno representa el 3,9% del total de casos de palabras incorrectas no corregidas. En algunos casos, el token original está formado por seis o más palabras juntas, en las que se ha suprimido el espacio en blanco entre cada una de ellas. Además de esta eliminación, la palabra a corregir también presenta otros fenómenos más, como errores de acentuación o de sustitución sistemática de grafías. En estos casos, en los que el normalizador no ha logrado obtener ninguna posible normalización, se ha devuelto la palabra original del texto sin realizar ningún tipo de modificación sobre ella. En 5.1 se muestran unos ejemplos.

(5.1) tunostasconlapeli (Tú no estás con la peli)

Ak de bote (¿A que de bote?)

Cafetero ya te lo contare algundia (Cafetero, ya te lo contaré algún día)

Respecto a los falsos positivos, entre las palabras que eran correctas y han sido modificadas de algún modo se encuentran, en su gran mayoría (59,05%), los nombres propios y nicks de los usuarios del chat. En el caso de los nombres propios, ampliando el diccionario se podría evitar este tipo de correcciones. En el caso de los nicks el problema es más complejo ya que la variabilidad que pueden

presentar estos apodos, sumadas a que las dotes creativas de los propios usuarios son inmensas e impredecibles, hacen que cualquier posible combinación de caracteres pueda convertirse en un sobrenombre. Como ya se ha comentado en el epígrafe 3.1.2, en este dominio, a diferencia de otros correctores (Borrell *et al.*, 2008), no se puede hacer un filtrado inicial para excluir las palabras escritas con mayúsculas, como por ejemplo los nombres propios, y evitar así modificaciones innecesarias, ya que en las conversaciones virtuales los usuarios suelen escribir tanto los nombres propios como los comunes en minúsculas, y cuando emplean la mayúscula lo hacen para expresar rasgos prosódicos.

Otros fenómenos observados dentro de los falsos positivos corresponden a palabras que no recoge el diccionario de trabajo: palabras a las que se le ha añadido algún afijo (17,14%), en su mayoría diminutivos, interjecciones y elementos paralingüísticos (7,62%). Por otro lado, también se han encontrado palabras en otros idiomas cuya frecuencia de aparición en este corpus es reducida (1,9%), y palabras inexistentes (9,52%) como ‘singuanga’, que se han mantenido tal cual en la corrección manual al desconocer qué quería expresar realmente el usuario con ellas.

5.1.3. Discusión

Como ya se ha comentado anteriormente, no existen estudios de normalización de textos de conversaciones virtuales en español. En cambio, sí encontramos algunos trabajos en que se han creado correctores para otros tipos de textos de Internet, como el de Borrell *et al.* (2008) con los correos electrónicos o el de Armenta *et al.* (2003b) para los mensajes cortos de móvil (SMS) y los correos electrónicos. En el primer estudio, el de Borrell *et al.* (2008) no se realizó ningún tipo de evaluación para valorar el funcionamiento de la herramienta creada, en cambio el de Armenta *et al.* (2003b) sí que realizó una evaluación. Desafortunadamente los datos no son directamente comparables, ya que no están presentados como los expuestos en esta investigación, ni proporciona información referente al tamaño del corpus con el que han realizado las pruebas, ni mencionan la ratio de errores que presenta el corpus antes de proceder a su corrección.

Es posible que las desviaciones observadas en los SMS, y sobre todo en los correos electrónicos, sean inferiores en número con respecto a las halladas en el corpus empleado en esta tesis. En los correos electrónicos, a diferencia de las conversaciones virtuales, al no primar la rapidez en su escritura, permiten una mayor elaboración de la redacción, dando la opción a releer y reescribir algún fragmento si fuera necesario.

Aun así, se puede observar un alto grado de corrección de palabras erróneas, rondando el 87%, y un porcentaje muy reducido de falsos positivos, concretamente del 0,18%, gracias al hecho de tratar sólo una desviación por palabra de entrada. En experimentos anteriores, cuando intentaban corregir hasta dos incorrecciones en una misma palabra el porcentaje de falsos positivos se vio incrementado hasta un 0,31% (Armenta *et al.*, 2003b).

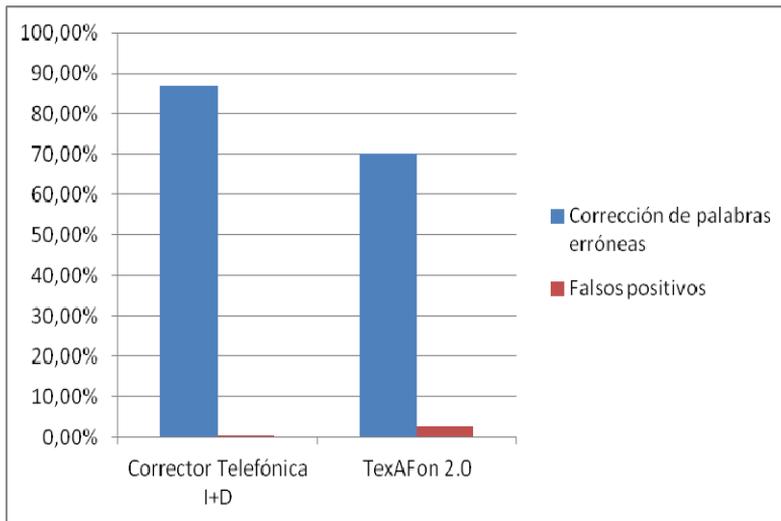


Figura 39: Comparación de la distribución de los resultados obtenidos en la corrección de palabras erróneas y falsos positivos del sistema de Telefónica I+D (Armenta et al. (2003 b) y el desarrollado en esta investigación

En la Figura 39 se puede observar los datos obtenidos por este sistema respecto al porcentaje de palabras incorrectas corregidas correctamente y los falsos positivos, y una comparativa de los mismos datos obtenidos en esta investigación

A pesar de que el porcentaje de palabras corregidas debidamente es ligeramente inferior al obtenido en Armenta *et al.* (2003b), se puede considerar que el empleo del conocimiento lingüístico aportado en los análisis lingüísticos ha jugado un papel favorable ya que se ha conseguido una mejora de casi del 25%, proporcionando un texto final con más de un 85% de palabras correctamente normalizadas y expandidas.

Muy probablemente, TexAFon 2.0 también podría tratar con resultados favorables otros tipos de textos que empleen lenguaje 2.0 como por ejemplo los correos electrónicos, los mensajes de móvil o comentarios de usuarios en foros o páginas de opinión entre otros.

Se considera que los valores obtenidos en la normalización de los mensajes de las conversaciones virtuales son satisfactorios. Aun así es necesario continuar trabajando en la normalización de este tipo de textos para conseguir mejores resultados y ser completamente comprensibles cuando sean reproducidos oralmente por un CTH.

5.2. Síntesis de voz expresiva

La evaluación de la voz sintética expresiva llevada a cabo en esta tesis tiene como objetivo:

- Saber hasta qué punto la aproximación paramétrica presentada en esta tesis es adecuada para la generación de habla expresiva de calidad.
- Conocer la capacidad del sistema paramétrico para generar voz sintética específica de una determinada emoción.

A continuación se muestra el proceso realizado para la evaluación de la voz sintética y los resultados obtenidos.

5.2.1. Procedimiento

El procedimiento que se ha llevado a cabo para evaluar la voz sintética está basado en el realizado en ALBAYZIN, Iberspeech 2012¹² (Montero *et al.*, 2012), ya que permite la comparación de los resultados obtenidos en esta investigación con otros sistemas. Por motivos logísticos se han tenido que realizar ciertas modificaciones con respecto a la evaluación original. Estas diferencias repercuten en el número de enunciados que han sido evaluado: en concreto, en ALBAYZIN se evaluaba un total de 25 enunciados diferentes. Gracias a un sistema aleatorio, cada evaluador sólo puntuaba 5 de estos enunciados. Al no poder realizar este proceso aleatorio de las 25 frases que se emplearon en la evaluación, se seleccionaron cinco, declarativas y con un contenido semántico lo más neutro posible, para agilizar el proceso por parte de los evaluadores. Las oraciones escogidas fueron:

- 1) Dejaron la deuda al cero.
- 2) Yo llevo al niño en el coche.
- 3) No queda fruta los viernes.
- 4) Tengo la llave en el bolsillo.
- 5) El niño sale con sus amigos.

En esta investigación, el modelado paramétrico de las emociones no se ha podido realizar directamente con el conversor de Cereproc, sino que ha sido necesario realizar una simulación del proceso tomando la señal de audio generada por el CTH de Cereproc, y modificándola según la emoción que se deseaba transmitir mediante diversos scripts de Praat, pudiendo introducir cierta degradación en la señal acústica final que haya influido en los resultados.

¹² VII Jornadas en Tecnología del Habla y III Iberian SLTech Workshop (<http://iberspeech2012.ii.uam.es/>)

Se han creado dos audios con la voz sintética del locutor Edu para cada una de las frases: una, seleccionando las unidades exclusivamente del corpus neutro para la versión neutra; y otra, con la selección realizada a partir del corpus oral expresivo del dominio de las conversaciones virtuales para la versión de enunciados emotivos. De este modo las señales acústicas se corresponderán al máximo con las características de cada tipo de corpus.

Ante la imposibilidad de evaluar todas las emociones analizadas en esta investigación, y al tomar como referencia la evaluación propuesta en las VII Jornadas en Tecnología del Habla, únicamente se generaron y evaluaron cuatro emociones (la alegría, el enfado, la sorpresa y la tristeza) más el neutro, las mismas que en la evaluación de IberSpeech 2012.

Para la evaluación se preparó una prueba de percepción, accesible vía [web](http://parles.upf.edu/llocs/moodle/course/view.php?id=302&edit=off&sesskey=EqjzgWlXkF) (<http://parles.upf.edu/llocs/moodle/course/view.php?id=302&edit=off&sesskey=EqjzgWlXkF>), en la que personas voluntarias, y de forma anónima, debían realizar tres pruebas diferentes de percepción. La sucesión de las pruebas no era consecutiva: el evaluador podía acceder a cada una de ellas cuando mejor le conviniera.

En la primera prueba se presentaban 25 audios diferentes, y se debía valorar tres aspectos diferentes para cada estímulo:

1. **El grado de calidad global de la voz sintética (CS):** los evaluadores debían puntuar en una escala del 1 al 5 cada enunciado teniendo en cuenta tanto su inteligibilidad como la naturalidad con la que se realizaban. El 1 corresponde a una calidad muy mala, mientras que el 5 a una muy buena. A pesar de que la voz generada en esta tesis no responde a un proceso realizado íntegramente por el CTH de Cereproc, y por lo tanto el resultado obtenido no es directamente comparable con los realizados en la evaluación de ALBAYZIN, se ha decidido realizar igualmente la consulta, y mantener así la estructura de la evaluación.
2. **Identificación de la emoción que se pretende transmitir (IE):** el evaluador debía indicar qué emoción expresaba el enunciado de la voz sintética. Para ello tenía un desplegable

con seis posibles opciones: alegría, enfado, sorpresa, tristeza, neutro y otro.

3. **La intensidad, la fuerza de esta emoción (FE):** En una escala del 1 al 5, de muy baja a muy elevada, se debía indicar el nivel de intensidad que se creía que el enunciado transmitía.

En la segunda prueba se presentaban nuevamente los mismos 25 audios, y se debía valorar con una puntuación del 1 al 5 la similitud existente entre la voz sintética y la voz original del locutor (SS). Para ello, aparte de los 25 audios modificados, la prueba contaba con tres audios diferentes del corpus neutro del locutor original.

Finalmente, se ha añadido una tercera prueba, no incluida en la evaluación ALBAYZIN, constituida por diez pares de enunciados formados por un enunciado generado por el CTH de Cereproc y otro modificado paramétricamente. En este apartado, el evaluador debía indicar cuál de los dos audios, el estímulo A o el estímulo B, se correspondía mejor con la etiqueta emocional propuesta en la pregunta. De este modo se pretendía valorar qué aproximación, por corpus o predicción paramétrica de prosodia, ofrece mejores resultados en cuanto a la generación de síntesis expresiva.

En la Figura 40 se puede observar un ejemplo de la interfaz empleada para evaluar la voz sintética.

La persona que realizaba la evaluación podía escuchar tantas veces como quisiera cada audio, y debía responder obligatoriamente a cada una de las preguntas. Para cada pregunta se proporcionaba una serie de posibles respuestas en un desplegable.

La evaluación debía llevarse a cabo en un lugar tranquilo, sin ruido de fondo, y empleando auriculares para evitar posibles interferencias que repercutieran negativamente en el correcto desarrollo de la evaluación. Al realizarse la evaluación vía web, y sin supervisión alguna, se ha de presuponer que se llevó a cabo de la manera en que se recomendaba, aunque no es posible asegurarlo en su totalidad.

Los evaluadores podían realizar las diferentes pruebas que conforman la evaluación en momentos diferentes del día, sin que la finalización de una prueba implicara el inicio automático de la

siguiente. A consecuencia de esto, el número de evaluadores es variable en cada una de ellas: se han registrado 24 evaluadores para la prueba 1, 22 para la prueba 2 y 21 para la prueba 3.

(*)Es obligatorio responder a las preguntas señaladas con un asterisco.

Estímulo 1

¿Cuál cree que es la calidad de la voz de la parte superior?*

¿Qué estado emocional cree que transmite el locutor?*

¿Qué intensidad emocional cree que transmite el locutor?*

Estímulo 2

¿Cuál cree que es la calidad de la voz de la parte superior?*

¿Qué estado emocional cree que transmite el locutor?*

¿Qué intensidad emocional cree que transmite el locutor?*

Estímulo 3

¿Cuál cree que es la calidad de la voz de la parte superior?*

¿Qué estado emocional cree que transmite el locutor?*

¿Qué intensidad emocional cree que transmite el locutor?*

Figura 40: Interfaz web usada por los evaluadores para realizar la evaluación de la voz

5.2.2. Resultados

Los resultados obtenidos para cada pregunta han sido normalizados en un rango comprendido entre el 0,0 y el 1,0, tal y como se hizo en la evaluación ALBAYZIN.

A continuación se muestran los resultados obtenidos para cada una de las preguntas realizadas en la evaluación.

5.2.2.1. Calidad de la voz sintética (CS)

Los evaluadores debían puntuar del 1 al 5 la calidad de la señal acústica que escuchaban, es decir, debían valorar la naturalidad y la inteligibilidad de la voz. Las valoraciones para cada una de las emociones generadas se muestran en la Figura 41.

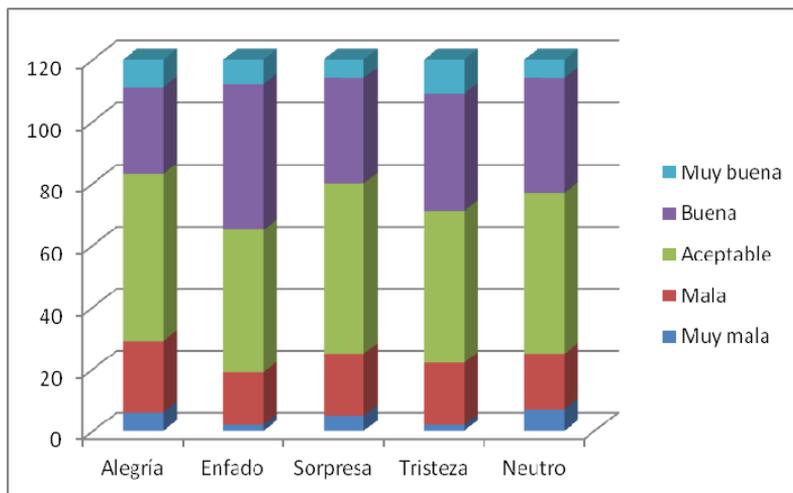


Figura 41: Resultados obtenidos en la prueba de evaluación de la calidad de la voz

La calidad media obtenida de la voz sintética ha sido de 0,55. Tal y como se puede observar en la Tabla 46, la diferencia entre las diferentes emociones es mínima. Aun así, la alegría ha sido la emoción cuya calidad ha sido inferior, y el enfado la que mejor puntuación ha conseguido.

Los valores obtenidos en esta prueba pueden no ser del todo reales, ya que el proceso que se ha llevado a cabo para modelar paramétricamente los enunciados no ha sido una síntesis real, sino

una simulación del proceso que debería desarrollar un CTH que modele la prosodia de un enunciado paramétricamente.

ALEGRÍA	0,52
ENFADO	0,58
SORPRESA	0,53
TRISTEZA	0,57
NEUTRO	0,53

Tabla 46: Resultados obtenidos para la calidad de la voz sintética según el tipo de emoción realizada

5.2.2.2. Similitud con el locutor original (SS)

Los evaluadores han considerado que la similitud entre la voz sintética y la voz original del locutor es baja, otorgándole un valor medio de 0,39. La alegría, una vez más, ha sido la emoción cuyo valor ha sido inferior, con un 0,32 en cambio, la tristeza ha sido la que mejor puntuación ha obtenido con un 0,48, aproximándose más a una similitud media. En la Figura 42 se muestran los resultados obtenidos para cada emoción, mientras que en la Tabla 47 se muestran los valores medios clasificados también por emoción.

Este valor tan bajo puede deberse a que los estímulos no se corresponden con una síntesis ‘real’, sino a una simulación del proceso ya que las modificaciones paramétricas se han llevado a cabo automáticamente mediante scripts de Praat a partir de unos enunciados generados por el sistema de Cereproc.

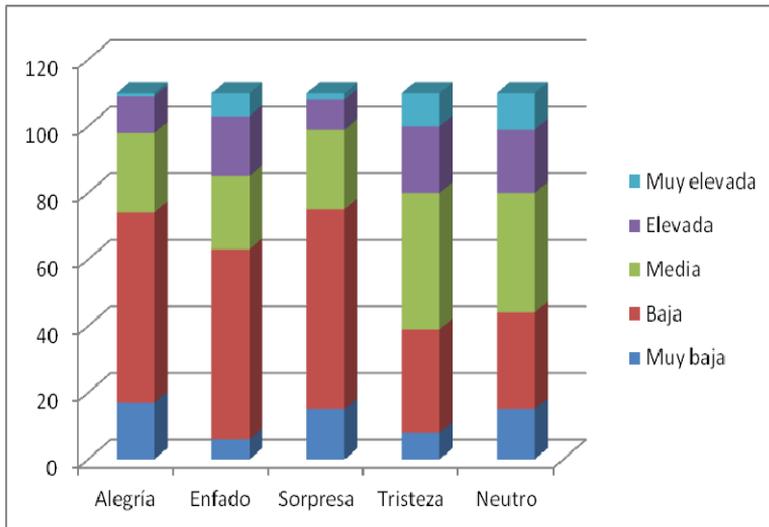


Figura 42: Resultados obtenidos en la prueba de similitud entre la voz original y la sintética

ALEGRÍA	0,32
ENFADO	0,41
SORPRESA	0,32
TRISTEZA	0,48
NEUTRO	0,45

Tabla 47: Resultados obtenidos sobre la similitud de la voz sintética respecto a la voz original del locutor según el tipo de emoción realizada

5.2.2.3. Identificación de las emociones (IE)

Como se puede observar en la Tabla 48, las 4 emociones básicas generadas para la evaluación han sido identificadas por una gran mayoría de los evaluadores como neutras. Las emociones alegría y enfado son las que muestran un menor índice de aciertos, con un 0,08 y 0,09 respectivamente. Ambas han sido identificadas mayoritariamente como neutras o como sorpresa.

	Alegría	Enfado	Sorpresa	Tristeza	Neutro	Otro
Alegría	0,08	0,06	0,13	0,3	0,33	0,09
Enfado	0,1	0,09	0,14	0,1	0,47	0,1
Sorpresa	0,22	0,07	0,18	0,15	0,31	0,08
Tristeza	-	0,16	0,03	0,29	0,48	0,05
Neutro	0,11	0,18	0,05	0,03	0,58	0,06

Tabla 48: Matriz de confusión de los resultados promediados de la prueba referente a la identificación de las emociones

La sorpresa y la tristeza, a pesar de que también han sido identificadas por una gran mayoría de oyentes como neutras, no presentan unos valores tan negativos. La sorpresa, aparte del estilo neutro, también ha sido identificada por un número importante de evaluadores como alegría, y en tercer lugar como sorpresa, con un 0,18. La tristeza es la emoción que mejores resultados ha obtenido, ya que la identificación correcta se encuentra en segundo lugar con 0,29.

El estilo neutro es el que mejor resultado ha obtenido con un 0,58, siendo confundido en algunas ocasiones con la emoción enfado.

El promedio de identificación de las emociones es de 0,24, un coeficiente bastante bajo para poder afirmar que la mayoría de las emociones son realizadas e interpretadas por los oyentes debidamente.

5.2.2.4. Fuerza de las emociones (FE)

La Tabla 49 muestra los diferentes valores obtenidos para cada emoción en cuanto a la intensidad con la que se realizaba cada una de ellas. En la Figura 43 se puede observar la distribución de las valoraciones según la emoción representada.

ALEGRÍA	0,47
ENFADO	0,47
SORPRESA	0,50
TRISTEZA	0,44
NEUTRO	0,43

Tabla 49: Resultados obtenidos sobre la fuerza con la que se realizaba cada emoción

Los resultados obtenidos entre las cuatro emociones y el estilo neutro son bastante similares. La sorpresa es la emoción cuya intensidad ha sido superior con 0,5, mientras que el estilo neutro es la que presenta un valor inferior con 0,43, estableciendo una media de 0,46. Estos resultados se adecúan con el nivel de intensidad analizado en el corpus oral y generado posteriormente en los enunciados sintéticos, ya que las emociones se clasificaron en dos niveles de intensidad distintos: 1 para un nivel medio y 2 para una intensidad alta.

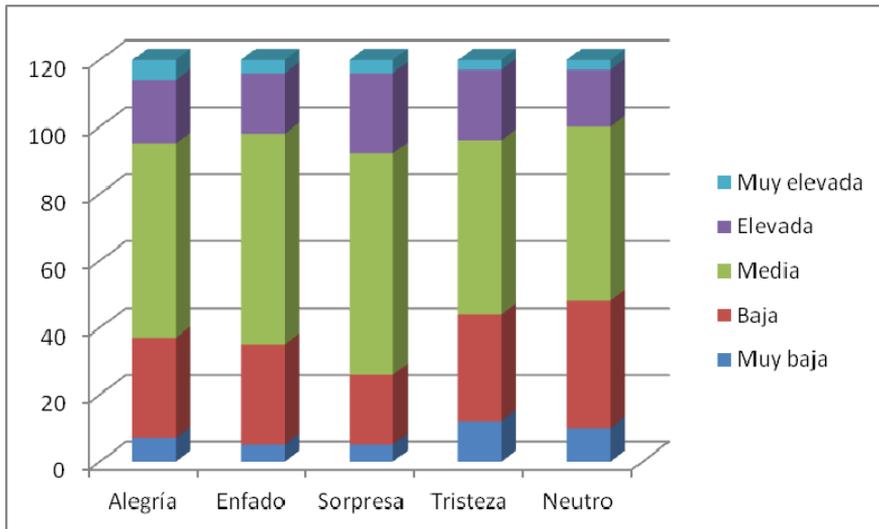


Figura 43: Resultados obtenidos en la identificación del grado de intensidad con la que se realizaba el enunciado

5.2.2.5. Rendimiento emocional

Para calcular el rendimiento emocional del sistema (EP, *Emotional Performance Measure*) se ha empleado la misma fórmula de cálculo (5.1) utilizada en la evaluación de ALBAYZIN 2012, en la que se combinan 4 de los 5 aspectos analizados en esta tesis: la calidad de la voz, la identificación de las diferentes emociones, la intensidad de la emoción y la similitud entre el locutor original y la voz sintética resultante.

$$(5.1) \quad EP = 4 \cdot (CS \cdot IE \cdot SS \cdot FE) / (CD + IE + SS + FE)$$

A diferencia de la evaluación de ALBAYZIN 2012, no se ha realizado una normalización de los valores obtenidos en el rendimiento emocional del sistema con los datos obtenidos para el locutor original, ya que no se disponía de los mismos enunciados

empleados en la generación de síntesis expresiva y en la versión original del locutor.

	CS	IE	FE	SS	EP
Alegría	0,52	0,08	0,47	0,32	0,018
Enfado	0,58	0,09	0,47	0,41	0,025
Sorpresa	0,53	0,18	0,50	0,32	0,039
Tristeza	0,57	0,29	0,44	0,48	0,078
Neutro	0,53	0,58	0,43	0,45	0,119
Sistema general	0,55	0,24	0,46	0,39	0,057

Tabla 50: Resultados obtenidos al calcular el rendimiento emocional del sistema en general y en cada una de las emociones evaluadas

Como se puede observar en la Tabla 50, el rendimiento emocional del sistema es de 0,05775. El estado neutro, seguido a una cierta distancia por la emoción tristeza, son los que han obtenido mejores resultados. Las emociones alegría y enfado, del mismo modo que ocurría en la identificación de las emociones, son las dos emociones que han obtenido una menor puntuación: 0,018 y 0,025 respectivamente.

En la Figura 44 se muestra la distribución de los valores medios obtenidos en las cuatro pruebas de la evaluación que se han tenido en cuenta para calcular el rendimiento emocional del sistema.

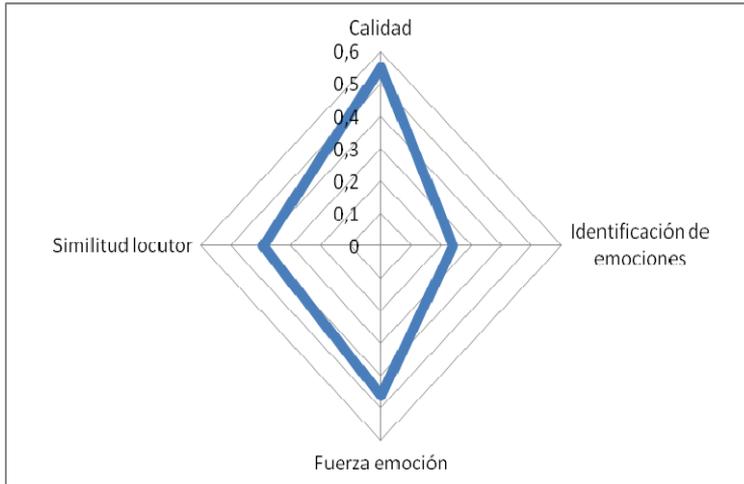


Figura 44: Distribución de los valores medios obtenidos en cuatro pruebas de la evaluación

5.2.2.6. Comparación síntesis por corpus y síntesis paramétrica

La prueba 3 se añadió para poder valorar cuál de las dos aproximaciones, por corpus o paramétrica, proporcionaba mejores resultados en la generación de síntesis expresiva.

La mayoría de los evaluadores opinó que el enunciado modificado paraméricamente representaba mejor la emoción indicada en el enunciado. Tanto la alegría como el enfado o el estilo neutro muestran una alta preferencia hacia los enunciados modificados paraméricamente, siendo precisamente el neutro el que mayor diferencia presenta, con un 0,95.

En cambio, emociones como el enfado o la sorpresa presentan una mejor realización mediante síntesis por corpus. Los datos obtenidos para cada emoción se muestran en la Tabla 51.

EMOCIONES	SÍNTESIS POR CORPUS	SÍNTESIS PARAMÉTRICA
ALEGRÍA	0,29	0,71
ENFADO	0,83	0,17
SORPRESA	0,62	0,38
TRISTEZA	0,26	0,74
NEUTRO	0,05	0,95
TOTAL	0,40	0,60

Tabla 51: Resultados obtenidos en la prueba 3 al comparar los enunciados generados mediante síntesis por corpus y los modelados paraméricamente

5.2.3. Discusión

En la evaluación de ALBAYZIN se valoraron cuatro sistemas diferentes, todos ellos basados en HSMM con ciertas peculiaridades cada uno de ellos. Uno, el sistema B, emplea un modelo único para todas las emociones (Sainz *et al.* 2012), el sistema C utiliza un modelo diferente para cada una de las emociones generadas, el tercero, el sistema D, ‘exagera’ las diferencias observadas entre cada modelo emocional y el modelo neutro, y finalmente, el sistema E construido para evaluar el procesamiento de texto desarrollado para el proyecto Simple4All. Este último modelo no emplea un conversor grafema a fonema, sino que utiliza unos ‘features’ extraídos del texto. Estos últimos tres sistemas están recogidos en Lorenzo-Trueba *et al.* (2012).

Comparando los resultados obtenidos en esta evaluación con otros sistemas de conversión de texto a habla se pueden observar grandes diferencias, sobre todo en algunos aspectos determinados. En el

caso de la calidad de la voz (0,55), a pesar de que los resultados obtenidos del locutor sintético puede que no sean del todo reales, sólo hay un sistema que ha obtenido unos resultados superiores, el sistema C (0,57). El resto de sistemas presentan unos resultados inferiores a los obtenidos en esta investigación.

En lo referente a la similitud observada entre la voz sintética y el locutor original, el resultado obtenido es inferior al resto de sistemas evaluados, a excepción del sistema D, el cual presenta el mismo valor que el sistema desarrollado en esta tesis: 0,39.

En la prueba de identificación de las emociones (IE) es donde se ha observado una mayor diferencia entre los valores obtenidos para la voz de ‘Edu’, y los obtenidos por el resto de sistemas. El promedio alcanzado en esta tesis ha sido de 0,24, un valor inferior al conseguido por los sistemas B, C y D. Estos tres sistemas superan el 50% de identificaciones correctas (Sainz *et al.* 2012 y en Lorenzo-Trueba *et al.* 2012), mientras que la voz desarrollada en esta investigación muestra una mayor confusión en la correcta identificación de las diferentes emociones, especialmente en las referentes a la alegría y el enfado. La Tabla 52 recoge los resultados obtenidos para cada uno de los sistemas analizados en la evaluación de ALBAYZIN 2012 más el obtenido en esta investigación.

	B	C	D	E	TESIS
Alegría	0,38	0,65	0,34	0,46	0,08
Enfado	0,48	0,78	0,82	0,30	0,09
Sorpresa	0,35	0,50	0,37	0,38	0,18
Tristeza	0,56	0,65	0,79	0,38	0,29
Neutro	0,83	0,61	0,76	0,74	0,58
Promedio	0,53	0,65	0,63	0,27	0,24

Tabla 52: Resultados obtenidos en la identificación de las emociones por los diferentes sistemas evaluados en IberSpeech 2012 más los datos obtenidos en esta investigación

Una de las desventajas de trabajar con un corpus emocional en el que las emociones han sido simuladas por un actor es que en algunas ocasiones el locutor puede enfatizar demasiado algunas características acústicas y obviar otras probablemente empleadas en la expresión natural de las emociones.

Tras observar estos resultados, se ha pensado que quizá hubiera sido interesante evaluar el corpus oral emocional grabado por el locutor, y observar si la realización de las emociones era lo suficientemente acertada y fácilmente identificable para los oyentes, o si por el contrario la realización de las emociones no era lo suficientemente clara. Si los resultados obtenidos para el locutor original no fueran lo suficientemente satisfactorios, difícilmente la voz sintética podría mejorarlos.

Respecto a la fuerza de las emociones, la puntuación media otorgada a la voz de 'Edu' ha sido de 0,46. Este resultado es similar a los conseguidos por los otros cuatro sistemas (Sainz *et al.* 2012 y en Lorenzo-Trueba *et al.* 2012), cuyos promedios oscilan entre el 0,41 y el 0,54.

En cuanto al rendimiento emocional del sistema, la voz desarrollada en esta investigación ha obtenido una calificación general de 0,057, un valor inferior a los obtenidos por los otros sistemas evaluados en ALBAYZIN 2012 (Lorenzo-Trueba *et al.* 2012). En la Tabla 53 se muestran los datos sin normalizar de los cinco sistemas.

B	0,095
C	0,151
D	0,136
E	0,080
Tesis	0,057

Tabla 53: Resultados sin normalizar del rendimiento emocional obtenidos por los sistemas evaluados en ALBAYZIN 2012, más la valoración alcanzada en esta investigación

A pesar de que los resultados obtenidos en la evaluación no son los esperados ya que en algunos aspectos, como la identificación de las emociones, se ha obtenido unos resultados considerablemente inferiores a otros sistemas, puede decirse que el rendimiento en general es satisfactorio.

Se ha observado que el empleo de corpus pequeños, como el de esta investigación, que únicamente recoge 25 enunciados para cada una de las emociones, no es suficiente para analizar y modelar paramétricamente la señal acústica. Otros estudios como el de Montero (2003) que emplea 2.000 fonemas por emoción, o el de Iriondo (2008) que recoge un total de 4.638 frases para el corpus emocional formado por cinco emociones, o el empleado en la evaluación de ALBAYZIN que está formado únicamente por cuatro emociones básicas más el neutro y está constituido por más de tres horas de grabación (Montero et al. 2012) son mucho más amplios que el creado para esta investigación, y obtienen mejores resultados.

Sin embargo, ante corpus orales reducidos como el de esta tesis, se observa que la aproximación paramétrica ofrece unos resultados mejores que los enunciados generados a partir de técnicas basadas en síntesis por corpus, ya que ha habido una mayor preferencia por los enunciados emocionales modelados paramétricamente frente a los generados por síntesis por corpus cuando se pedía al oyente que valorase qué señal acústica representaba mejor un determinado estado emocional.

5.3. Resumen

En esta tesis se ha trabajado tanto en la etapa de normalización del módulo de procesamiento lingüístico como en el módulo de generación de habla sintética expresiva para que el CTH, en este caso el de Cereproc, pueda tratar tanto texto normativo como no normativo y que los enunciados orales generados transmitan correctamente una determinada serie de emociones.

Después de realizar una serie de análisis, tanto lingüísticos como prosódicos, y llevar a cabo una adaptación del normalizador de

textos TexAFon y realizar una simulación de la generación de habla sintética expresiva ha sido necesario evaluar ambos procedimientos para determinar si el funcionamiento es óptimo y si se han logrado los objetivos marcados al inicio de la investigación.

En cuanto a la normalización de textos, la evaluación se ha llevado a cabo cotejando los tokens de tres versiones diferentes de un corpus de conversaciones virtuales con el que no se había trabajado previamente: versión original, una versión normalizada manualmente y la versión propuesta por TexAFon 2.0. Los resultados obtenidos a través de sistema basado en reglas son satisfactorios. Se ha logrado normalizar y expandir correctamente el 70,3% de las palabras del texto de entrada consideradas incorrectas, consiguiendo que el porcentaje total de palabras correctas pase del 61,2% al 85,9%, lo que supone una mejora de 24,7 %.

A pesar de estos resultados, es necesario continuar trabajando para conseguir un texto cien por cien correcto, que no dé opciones a errores o ambigüedades por parte del receptor del mensaje. Estas mejoras deberían ir enfocadas a dos aspectos:

1. La mejora de algunas de las reglas y el modelo de lenguaje para corregir debidamente todas aquellas palabras identificadas como incorrectas.
2. Disminuir el número de falsos positivos a partir de una considerable mejora del diccionario general.

En lo referente a la generación del habla sintética expresiva se ha seguido la evaluación realizada en IberSpeech 2012, con ciertas diferencias que repercuten especialmente en una disminución del número de enunciados a evaluar.

En este caso, los resultados obtenidos no son tan satisfactorios como hubiera cabido esperar debido por un lado al reducido tamaño del corpus emocional, el cual recoge muchas más emociones y muestra un número muy inferior de enunciados para cada uno de los estados emocionales y fisiológicos, y por otro lado al hecho de haber realizado el modelado de la señal acústica fuera del sistema de síntesis.

Mientras que los datos relacionados con la calidad de la voz han sido favorables y que la intensidad con la que se han realizado las emociones corresponde con la identificada por los oyentes, los valores obtenidos en cuanto a la similitud de la voz sintética con la del locutor original, la identificación de las emociones transmitidas, y el rendimiento emocional del sistema en general no son tan positivos. Estos tres factores presentan un bajo rendimiento, obteniendo resultados inferiores con respecto a otros sistemas evaluados mediante el mismo procedimiento, sobre todo en lo referente a la correcta identificación de ciertas emociones como son la alegría y el enfado.

Aun así, parece observarse una cierta predilección por los enunciados expresivos modelados paramétricamente frente a los generados mediante síntesis por corpus.

A partir de que los resultados obtenidos para la voz de 'Edu', y de los presentados por Sainz *et al.* (2012) y en Lorenzo-Trueba *et al.* (2012) en la evaluación de ALBAYZIN 2012 queda en evidencia la necesidad de seguir investigando en este ámbito para lograr obtener voz sintética expresiva de calidad y cuya realización de las diferentes emociones sea identificada correctamente por un amplio número de oyentes.

Capítulo 6

CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Con esta investigación se ha pretendido explorar el uso de determinadas aproximaciones basadas en el conocimiento lingüístico en conversores de texto a habla que puedan emplearse para leer en voz alta, de forma lo más natural posible, el texto de las conversaciones virtuales. Este objetivo general se ha focalizado en dos objetivos más concretos: explorar las posibilidades del uso de conocimiento lingüístico para la normalización de los textos de entrada, a partir del análisis computacional de un corpus de textos de conversaciones virtuales; y emplear conocimiento lingüístico para la generación del habla de salida, en forma de modelos prosódicos construidos a partir del análisis computacional de un corpus de habla. A continuación se resumen las principales aportaciones de esta tesis al problema, se muestran las conclusiones extraídas en ambos procesos, y las posibles futuras líneas de investigación.

6.1. Normalización del texto

Las conversaciones virtuales presentan unas peculiares características lingüísticas que la alejan de los textos escritos más convencionales como los artículos periodísticos, o divulgativos e incluso de textos más informales como las cartas que se escribían, no hace tanto tiempo, entre personas conocidas y a las que les unía una relación próxima de amistad o amorosa. Estas diferencias en la forma de escribir, de expresarse, hacen que los sistemas convencionales para el procesamiento del lenguaje natural resulten insuficientes para tratar correctamente este tipo de textos.

Actualmente existen varios estudios que muestran los aspectos lingüísticos de las conversaciones virtuales, pero en ninguno de ellos se hace referencia a cuáles se dan con mayor frecuencia, ni se cuantifica el número de casos. Aparte de saber qué fenómenos lingüísticos se dan en este tipo de textos, también es importante saber con qué frecuencia se suceden, sobre todo si se desea realizar un normalizador-corrector, como es este caso, para tratar automáticamente este tipo de discursos. En esta investigación se aportan los resultados de un estudio de corpus de las conversaciones virtuales en español, que completa las descripciones llevadas a cabo hasta la fecha sobre el tema.

Los resultados de este análisis se han empleado para desarrollar un sistema de normalización de textos basado en reglas, con el que se pretendía explorar las potencialidades de este tipo de aproximación para el procesado de texto no normativo por un CTH. La aproximación seguida ha implicado tratar el texto de las conversaciones virtuales no como texto ‘erróneo’, sino como un texto que debe ser normalizado del mismo modo que otros elementos del texto. Se ha integrado el conocimiento extraído del corpus en un normalizador de texto inicialmente preparado para el procesado de texto normativo. Se han tratado exclusivamente los errores léxicos, dejando de lado incorrecciones gramaticales como elisiones de verbos o cualquier otro tipo de categoría gramatical o las omisiones de los signos de puntuación. A pesar de tratarse de oraciones realmente cortas (el promedio resultante obtenido en esta tesis es de 4,68 palabras por intervención), analizar la ausencia de signos de puntuación para posteriormente ser tratados por el normalizador proporcionaría una mejora en la voz sintética, ya que podría efectuar pausas, realizando así los enunciados de una manera más natural.

Además, disponer de un análisis lingüístico que permite conocer en detalle con qué frecuencia tiene lugar cada fenómeno, puede arrojar mayor conocimiento sobre el tipo de lenguaje que se emplea en las conversaciones virtuales y en Internet en general. A través de análisis detallados y cuantificativos como el presentado en esta tesis se podría observar si este tipo de escritura ha evolucionado de algún modo desde sus inicios, si continúa cambiando, o si estos cambios, en el caso de que los haya, son introducidos por un grupo de personas de unas edades concretas o si por el contrario son más genéricos. Esta información permitiría saber si las reglas empleadas

en esta investigación son lo suficientemente genéricas para tratar este tipo de textos sin realizar modificaciones con asiduidad, o si, por el contrario, este tipo de lenguaje evoluciona a una mayor velocidad y requiere una revisión del procesado del lenguaje más frecuente.

Continuando con los análisis cuantitativos, también sería muy recomendable tratar otros tipos de textos de Internet como los correos electrónicos, los foros o las redes sociales, y comprobar así si realmente existen semejanzas entre estos y las conversaciones virtuales, y en qué grado. De este modo se podría corroborar hasta qué punto este sistema, diseñado especialmente para tratar las conversaciones virtuales, también puede ser aplicable a otros tipos de textos de Internet.

El sistema de corrección basado en reglas propuesto en esta tesis ha ofrecido un resultado satisfactorio en su funcionamiento, proporcionando una mejora satisfactoria del texto de entrada. Aun así, aún queda un gran recorrido para poder tratar correctamente y de forma completamente automática todos los fenómenos lingüísticos presentes en este tipo de textos. Para mejorar el sistema se debería trabajar en dos aspectos diferentes, por un lado en la redefinición de lo que se considera una palabra errónea, y por otro lado, en mejorar algunos de los procesos que se han de llevar a cabo para normalizarlas, como por ejemplo la correcta identificación de emoticonos.

En el caso de qué se considera una palabra correcta y qué una palabra incorrecta, está claro que no es suficiente con remitirse únicamente a la presencia o ausencia de la misma en el diccionario, ya que en la normalización automática no se han detectado correctamente todas las palabras erróneas. Sería aconsejable abordar este aspecto proporcionando otro tipo de información a tener en cuenta como por ejemplo los *nicks* de los usuarios presentes en el canal.

Respecto a las mejoras de la normalización, estas han de realizarse tanto en el proceso de identificación como en los de la propia normalización. Este tipo de lenguaje presenta una gran versatilidad y muestra la gran capacidad creativa por parte de los usuarios, por ejemplo la interjección “¡ay!” puede aparecer escrita de muchas formas posibles: “ai”, “ais”, “aix”, “aish”, “hai” lo que dificulta enormemente la correcta identificación de la etiqueta

correspondiente y su normalización. Es indispensable crear reglas lo suficientemente generales y modulares para poder tratar todas las palabras correctamente. En cuanto a la normalización de las palabras, algunos aspectos importantes a considerar en trabajos futuros podrían ser:

- Ante la imposibilidad de disponer de un diccionario que contemple todas las palabras posibles de una lengua, con todas las palabras derivadas y compuestas, sería pertinente crear un módulo capaz de identificar los lemas de cada una de las palabras y los posibles afijos que las componen. De este modo, posibles diminutivos, aumentativos o verbos con clíticos que aparezcan en el texto, y que no estén recogidos en el diccionario, podrían ser identificados como palabras correctas y no ser modificadas, o tras algún tipo de normalización ser detectadas como una posible normalización de la palabra original.
- Actualmente el modelo de lenguaje creado para seleccionar la mejor normalización de la lista de palabras candidatas está basado en bigramas, pero podría mejorarse aportando información morfosintáctica.
- En cuanto a la relación del procesamiento del lenguaje con los CTH, estos últimos, del mismo modo que identifican y generan los enunciados sonoros en función de la modalidad oracional (enunciativa, interrogativa o exclamativa) del texto, también deberían generar los enunciados según el tipo de emoción que se expresa en el texto de una manera automática, sin que el usuario haya de indicar de ninguna manera el tipo de emoción que desea expresar en cada uno de los diferentes enunciados. Esto implica crear un detector de emociones a partir del texto, proceso en que ya se ha empezado a trabajar.

6.2. Generación de síntesis expresiva

Como ya se ha comentado en más de una ocasión, los CTH comerciales existentes actualmente producen unas voces de gran calidad e inteligibilidad, pero que pueden resultar monótonas ya que no aportan ningún tipo de expresividad, provocando rechazo por parte de ciertas personas. Este es el motivo por el cual se ha observado un interés creciente en cómo aportar expresividad a estas voces, sin perder en calidad e inteligibilidad de la voz.

En esta investigación se ha abordado el problema de dotar de expresividad a las voces sintéticas desde una perspectiva ligeramente diferente a la observada en otros estudios: se ha tomado una situación comunicativa real, como son las conversaciones virtuales. Esto implica que haya un número mayor de emociones, y que éstas puedan aparecer combinadas con otras emociones o actos de habla.

El corpus emocional empleado en esta tesis es considerablemente más pequeño que el utilizado en otros estudios: cada emoción únicamente posee 25 enunciados de una longitud variable. La investigación realizada por Montero (2003) recoge unos 2.000 fonemas para cada emoción, y el corpus emocional empleado por Iriando (2008) está formado por más de 4.000 enunciados.

La señal acústica expresiva no se ha podido obtener directamente del sistema de síntesis de Cereproc, ya que el resultado obtenido en el proceso de modificación de la F0 mediante este sistema no era lo suficientemente satisfactorio. Por ello se ha realizado una simulación del proceso, en el que la señal acústica generada por el CTH de Cereproc ha sido modificada externamente de forma automática para modelarla según las emociones que se deseaban transmitir.

Para conocer con exactitud si la aproximación paramétrica realizada en esta investigación para modelar la voz acústica es lo suficientemente adecuada para la generación de habla expresiva de calidad se ha realizado una evaluación, similar a la propuesta en ALBAYZIN 2012 (Montero *et al.*, 2012). Los resultados obtenidos en esta tesis no son satisfactorios en todos los aspectos, ya que tanto el porcentaje de las identificaciones de las emociones generadas

como el de la similitud existente entre la voz sintética y la del locutor es bajo e inferior a otros estudios similares. Aun así, al tratarse de un corpus reducido, en igualdad de condiciones, los resultados obtenidos mediante el modelado paramétrico son algo mejores que los alcanzados en la aproximación por corpus. El tamaño reducido del corpus emocional y el sistema empleado para el modelado de la señal acústica pueden haber influido en los resultados obtenidos.

A pesar de esto, los resultados en la generación de la síntesis expresiva obtenidos hasta el momento en otros estudios aún son mejorables.

A partir de la investigación desarrollada y de los resultados obtenidos, se muestran una serie de posibles trabajos futuros que permitirían continuar avanzando tanto en el campo de la creación de corpus emotivos como en el de la generación de voz sintética expresiva:

- Actualmente no existe un inventario de emociones universal. Crear un listado lo más amplio posible y consensuado por la comunidad científica sería un primer paso e indispensable para poder trabajar en este ámbito.
- El corpus emocional de esta tesis consta de 525 enunciados, 25 para cada etiqueta emocional. Ampliar el número de enunciados para cada una de las etiquetas proporcionaría:
 - mayor información en el análisis prosódico, ya que algunas emociones, dependiendo del parámetro analizado, contaban con un número muy reducido de casos,
 - y por otro lado una mejora de la generación de la voz tanto si es por síntesis por corpus como mediante modelado paramétrico.
- Sería interesante disponer de más corpus orales de emociones, en los que las emociones no se vean reducidas únicamente a las seis grandes, sino que el inventario fuera

más amplio como el propuesto en HUMAINE, en I3Media o en esta tesis. Teniendo más corpus multilocutor se podrían establecer generalizaciones de los resultados obtenidos en el análisis prosódico de los corpus.

- Ya que se observa una preferencia de los enunciados expresivos realizados mediante el modelado paramétrico por encima de los enunciados generados mediante síntesis por corpus, se podría implementar el modelado paramétrico de la señal de voz en los sistemas de CTH basados en corpus. Esta implementación podría realizarse durante el procesamiento del lenguaje natural, en la asignación de los aspectos prosódicos. La selección y posterior concatenación de los segmentos más favorables para cada enunciado se podría realizar teniendo en consideración los valores obtenidos en la predicción de los diferentes parámetros prosódicos.

REFERENCIAS

Agüero, P. D. (2012). *Síntesis de voz aplicada a la traducción voz a voz*. Tesis Doctoral. Universidad Politécnica de Cataluña. <http://hdl.handle.net/10803/97035>

Aguilar, L.; Alcoba, S.; Carbó, C. y Machuca, M. (2002). "Los marcadores del discurso en la lengua oral informativa". *Actas del Congreso Internacional de Análisis del discurso: lengua, cultura y valores*. Pamplona, 26-28 de noviembre de 2002.

Aida-Zade, K. R.; Ardil, C. y Sharifova, A. M. (2010). "The main principles of text-to-speech synthesis system". *International Journal of Signal Processing*, 6 (1), pp. 13-19. <http://www.waset.org/journals/ijice/v6/v6-1-3.pdf>

Androutsopoulos, J. K. (2000). "Non-Standard spellings in media texts: the case of German fanzines". *Journal of Sociolinguistics*, 4, 4, pp. 514-533.

armenta, A.; Escalada, J. G.; Garrido, J. M. y Rodríguez, M. A. (2003). "Conversor texto a voz multilingüe de Telefónica I+D". *Procesamiento del Lenguaje Natural*, 31, pp. 331-332.

armenta, A.; Escalada, J. G.; Garrido, J. M. y Rodríguez, M. A. (2003 b). "Desarrollo de un corrector ortográfico para aplicaciones de conversión texto-voz". *Procesamiento del Lenguaje Natural*, 31, pp. 65-72. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag65.pdf>
Última consulta: enero de 2013

Aw A.; Zhang M.; Xiao J.; y Su J. (2006). "A phrase-based statistical model for sms text normalization". *Proceedings of COLING/ACL*, pp. 33-40.

Aylett, M. P. y Yamagishi, J. (2008). "Combining Statistical Parametric Speech Synthesis and Unit-Selection for Automatic Voice Cloning". *LangTech*, Roma.

Aylett, M. P.; Andersson, J. S.; Badino, L. y Pidcock, C. J. (2007a). "The Cerevoice Blizzard Entry 2007: Are Small Database Errors Worse than Compression Artifacts?". *Blizzard Challenge Workshop*, Bonn.

- Aylett, M. P. y Pidcock, C. P. (2007b). "The CereVoice Characterful Speech Synthesiser SDK". *AISB*, pp.174-178, Newcastle.
- Aylett, M. P.; Pidcock, C. P. y Fraser, M. E. (2006). "The CereVoice Blizzard Entry 2006: A prototype Database Unit Selection Engine". *Blizzard Challenge Workshop*, Pittsburgh.
- Bagshaw, P. (1998). "Unsupervised training of phone duration and energy models for text-to-speech synthesis". *ICSLP 98*, 2, pp. 17-20. Sydney Convention Centre, Sydney, Australia, 30th November – 4th December 1998.
- Balestri, M.; Facchiotti A.; Quazza, S.; Salz, P.L. y Sandri S. (1999). "Choose the best to modify the least: a new generation concatenative synthesis system". *Eurospeech'99*, 6th European Conference on Speech Communication and Technology, 5, pp. 2291-2294, Budapest.
- Barkova, K.; Haffner, P. y Larreur, D. (1993). "Intensity Prediction for Speech Synthesis in French". *Proceedings of an ESCA Workshop on Prosody*, pp. 280-283. Septiembre 27-29, 1993, Lund, Sweden.
- Barra-Chicote, R.; Yamagishi, J.; King, S.; Montero, J. M.; Macias-Guarasa, J. (2010). "Analysis of statistical parametric and unit selection speech synthesis systems applied to emotional speech". *Speech Communication*, 52(5), pp 394-404, 2010
- Biber, D. (1986). "Spoken and textual dimensions in English: Resolving the contradictory findings", *Language*, 62, 2.
- Black, A.; Zen, H. y Tokuda, K. (2007). "Statistical Parametric Synthesis", *ICASSP*, Hawaii.
- Black, A. y Lenzo, K. (2003). "Optimal Utterance Selection for Unit Selection Speech Synthesis Databases". *International Journal of Speech Technology*, 6(4), pp. 357-363, Kluwer Academic Publishers.
- Black, A. (2003). "Unit Selection and Emotional Speech". *EUROSPEECH 2003*, pp. 1649-1652.
- Blanco, M. J. (2002). "El chat: la conversación escrita", *Estudios lingüísticos 16*. Departamento de Filología Española, Lingüística general y Teoría de la literatura, Universidad de Alicante.

Disponible en: <http://publicaciones.ua.es/LibrosPDF/0212-7636-16/02.pdf> Última consulta: diciembre de 2012

Blondet, M. (2006). “Variaciones de la velocidad de habla en español: patrones fonéticos y estrategias fonológicas. Un estudio desde la producción”, Tesis doctoral, Mérida, Universidad de los Andes

Boersma, P. y Weenink, D. (2009). “Praat: doing phonetics by computer (Version 5.1.05)” [Computer program]. Disponible en: <http://www.praat.org/>

Borrell, C. y Oliver, A. (2008). “How do we working a multilingual classroom? A multilingual environment with Moodle and Apertium”. *CALPIU Conference*. 15 - 17 de diciembre. Roskilde. Denmark. Disponible en: <http://lpg.uoc.edu/aoliverg/uploads/arxiu/borrell-oliver-CALPIU.pdf> Última consulta: diciembre de 2012

Brown, D. (2003). *El código da Vinci*. Umbriel, Barcelona.

Cahn, J. E. (1989). “Generating Expression in Synthesized Speech”, Proyecto Final de Carrera, Massachusetts Institute of Technology.

Campbell, W. N. (1990). “Analog I/O nets for syllable timing”. *Speech Communication*, 9, pp. 57-61.

Campbell, W. N. (1992). “Syllable-based segmental duration”. Bailly, G.; Benoît, C. y Sawallis, T. R. (Eds.) *Talking Machines: Theories, Models and Designs*. Amsterdam: North-Holland. pp. 211-224

Charpentier, F. y Moulines, E. (1989). “Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones”. *Proceedings of Eurospeech 89* (2), pp. 13-19

Choudhury, M.; Saraf, R.; Jain, V.; Sarkar, S. y Basu, A., (2007). “Investigation and modeling of the structure of texting language”. *Proceedings of the IJCAI Workshop on "Analytics for Noisy Unstructured Text Data"*, pp. 63–70, Hyderabad, India.

Cornelius, R. R. (2000). “Theoretical Approaches to Emotion”. *Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, pp. 3–10. Newcastle, Northern Ireland, UK.

Cowie, R. y Cornelius, R. R. (2003). "Describing the emotional states that are expressed in speech". *Speech Communications Special Issue on Speech and Emotion*, 40 (1-2), pp. 5-32.

Cowie, R.; Cowie-Douglas, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W. y Taylor, J. G.; (2001). "Emotion Recognition in Human-Computer Interaction". *IEEE Signal Processing Magazine* 18 (1), pp. 32-88.

Crystal, D. (2002). "El lenguaje e internet", Madrid: Cambridge University Express.

Díaz, A. M. (2005). 'Tipología de errores gramaticales para un corrector automático'. *Procesamiento del Lenguaje Natural*, 35, pp. 409-416.
Disponible en: <http://www.sepln.org/revistaSEPLN/revista/35/50.pdf> Última consulta: enero de 2013

Douglas-Cowie, E.; Cowie, R.; Martin, J. C.; Devillers, L.; Cox, C. y miembros de WP5, (2006). "Humaine: Human-Machine Interaction Network on Emotion", Mid Term Report on Database Exemplar Progress .

Dutoit, T. (1999). "A Short Introduction to Text-to-Speech Synthesis". *TCTS Lab*, Faculté Polytechnique de Mons. Disponible en: http://tcts.fpms.ac.be/synthesis/introtts_old.html Última consulta: noviembre de 2012

Dutoit, T.; Leich, H. (1996). "MBR-PSOLA: Text-to-speech synthesis base don an MBE re-synthesis of the segments database". *Speech Communication*, vol. 13, pp. 435-440
<http://www.assta.org/sst/SST-92/cache/SST-92-PosterSession1-Part2-p46.pdf>

Ekman, P.; Friesen, W. V. y Ellsworth, P. (1982). "What emotion categories or dimensions can observers judge from facial behavior?". P. Ekman (Ed.), *Emotion in the human face* (pp. 39-55). New York, Cambridge University Press.

Esquerra, I. (2006). "Síntesis de habla emocional por selección de unidades." *IV Jornadas en Tecnología del Habla. "Libro de actas de las IV Jornadas en Tecnología del Habla"*, Zaragoza. Disponible en: <http://www.lsi.upc.edu/~nlp/papers/esquerra06.pdf> Última visita: noviembre de 2012

Esquerra, I. y Bonafonte, A. (2004). "Habla emocional mediante métodos de re-síntesis y selección de unidades". *URSI 2004. Actas del XIX simposium nacional de la unión científica internacional de radio*. Universitat Ramon Llull, Barcelona. Disponible en: <http://www.lsi.upc.edu/~nlp/papers/esquerra04.pdf> Última visita: noviembre de 2012

Fordyce, C. S. (1998). "Prosody Prediction for Speech Synthesis Using Transformational Rule-Based Learning", tesis doctoral, Universidad de Boston.

Fujisaki, H. y Nagashima, S. (1969). "A model for the synthesis of pitch contours of connected speech". *Annual Report of Engineering Research Institute, University of Tokyo 2*, pp. 53-60.

Garrido, J. M.; Laplaza, Y.; Marquina, M.; Schoenfelder, C. y Rustullet, S. (2012a). "TexAFon: a multilingual text processing tool for text-to- speech applications". *Proceedings of IberSpeech 2012*, pp. 281-289. Madrid, Spain, November 21-23, 2012.

Garrido, J. M.; Laplaza, Y.; Marquina, M.; Pearman, A.; Escalada, J. G.; Rodríguez, M. A. y Armenta, A. (2012b). "The I3MEDIA speech database: a trilingual annotated corpus for the analysis and synthesis of emotional speech". *LREC 2012 Proceedings*, pp. 1197-1202. Disponible en: <http://www.lrec-conf.org/proceedings/lrec2012/summaries/865.html> Última consulta: diciembre de 2012.

Garrido, J. M. (2012c). "Análisis fonético de los patrones melódicos locales del español: patrones entonativos". *Revista Española de Lingüística* nº 84, pp. 95-125.

Garrido, J. M. (2012d). "GenProso: a parametric prosody prediction module for text-to-speech applications". *Proceedings of IberSpeech 2012*, pp.214-223. Madrid, Spain, Noviembre 21-23. Disponible en: http://iberspeech2012.ii.uam.es/IberSPEECH2012_OnlineProceedings.pdf Última consulta: enero de 2013

Garrido, J. M. (2011). "Análisis de las curvas melódicas del español en habla emotiva simulada". *Estudios de Fonética Experimental*, XX, pp. 205-255

Garrido, J. M. (2010). "A Tool for Automatic F0 Stylisation, Annotation and Modelling of Large Corpora". *Speech Prosody*, Chicago. Disponible en:

<http://speechprosody2010.illinois.edu/papers/100041.pdf> Última consulta: enero de 2013

Garrido, J. M.; Bofias, E.; Laplaza, Y.; Marquina, M.; Aylett, M. y Pidcock, C. (2008). “The CEREVOICE speech synthesizer”. *Actas de las V Jornadas de Tecnología del Habla*, pp.126-129. Bilbao, 12-14 noviembre 2008.

Garrido, J. M. (2001). “La estructura de las curvas melódicas del español: propuesta de modelización”. *Lingüística Española Actual*, XXIII, 2, pp. 173-209

Garrido, J. M. (1996). “Modelling Spanish Intonation for Text-to-Speech Applications”, Tesis doctoral, Universitat Autònoma de Barcelona

Geertz C.; (1993). *“Local Knowledge. Further Essays in Interpretative Anthopology”*. Londres: Fontana.

Hertz, S. R.; Younes, R. J. y Zinovieva N. (1999). “Languages-Universal and Languages –Specific Components in the Multi-Language ETI-Eloquence Text-To-Speech System”. *Proceedings 14th Int. Cong. Phonet. Sciences*, San Francisco. Disponible en: <http://conf.ling.cornell.edu/hertz/ICPhS99.pdf> Última consulta: septiembre de 2012

Hirst, D. J.; Di Cristo, A. y Espesser, R. (2000). “Levels of representation and levels of analysis or intonation”. Merle Horne (Ed.) *Prosody: Theory and Experimen*, pp. 51-87. Kluwer Academic Press.

Holm, B. y Bailly, G. (2000). “Generating Prosody By Superposing Multi-Parametric Overlapping Contours”. *Proceedings of the International Conference on Spoken Language Processing, ICSLP 2000/INTERSPEECH 2000*, Beijing, China.

Hozjan, V.; Kacic, Z.; Moreno, A.; Bonafonte, A. y Nogueiras, A. (2002). “Interface databases: Design and collection of a multilingual emotional speech database”. *Proceedings of the Third Int. Conference on Language Resources and Evaluation (LREC'02)*, pp. 2024-2028. Las Palmas de Gran Canaria.

Iriondo, I. (2008). “Producción de un corpus oral y modelado prosódico para la síntesis de habla expresiva”. Tesis doctoral, Universidad Ramon Llull de Barcelona.

- Iriondo, I.; Alías, F.; Melenchón, J. y Llorca, M. A. (2004). “Modeling and Synthesizing Emotional Speech for Catalan Text-to-Speech Synthesis”. *Affective Dialogue Systems. Tutorial and Research Workshop, ADS*, volume 3068, pp. 197-208. Kloster Irsee, Germany. Springer, Heidelberg.
- Jun, S. A. y Fougeron, C. (2000). “A phonological model of French intonation”. Botinis, A. (ed), *Intonation: Analysis, modeling and technology*, pp. 209-242. Dordrecht, Kluwer Academic Publishers.
- Karlsson, I., (2007). “Diferencias entre lenguaje de chats de habla española y español estándar escrito: aspectos ortográficos, gramaticales y lexicales”, Tesina, Institutionen för kommunikation och information Spanska.
- Klatt, D. H. (1979). “Synthesis by rule of segmental durations in English sentences”. Lindblom, B.; Öhman, S. (Eds.) *Frontiers of Speech Communication*, pp. 287-299. New York: Academic.
- Kenneth, W. C. y Gale, W. (1991). “Probability scoring for spelling correction”. *Statistics and Computing*, 1, pp. 91–103.
- King S., (2011). “An Introduction to statistical parametric speech synthesis”. *Sadhana*, Volumen 36, 5, pp. 837-852.
- Kobus, C.; Yvon F. y Damnati, G. (2008). “Transcrire les SMS comme on reconnaît la parole”. *Actes de la Conférence sur le Traitement Automatique des Langues (TALN’08)*, pp. 128–138, Avignon, France. Disponible en: <http://perso.limsi.fr/yvon/publications/bibs/sources/Kobus08transcrire.pdf> Última consulta: noviembre de 2012
- Llisterri, J.; Machuca, M. J.; Madrigal, N.; Mancini, F.; Massimino, P.; Mota, C.; Riera, M. y Ríos, A. (2004). “Aspectos lingüísticos en el diseño de un conversor de texto en habla en castellano y en catalán: El sistema Loquendo TTS®”. *VI Congreso de Lingüística General*, pp. 521-2. Santiago de Compostela. Disponible en: <http://liceu.uab.cat/~joaquim/publicacions/Llisterri et al 04 Conversor Texto Habla Castellano Catalan Loquendo.pdf> Última consulta: enero de 2013
- Llisterri, J.; Carbó, C.; Machuca, M. J.; De La Mota, C.; Riera, M. y Ríos, A. (2004 b). “La conversión de texto en habla: aspectos lingüísticos”. M. A. Martí y J. Llisterri (Eds.), *Tecnologías del texto*

y del habla, pp. 145–186. Edicions de la Universitat de Barcelona y Fundaci3n Duques de Soria, Barcelona.

Llisterri, J.; Carb3, C.; Machuca, M. J.; De La Mota, C.- Riera, M. y R3os, A. (2003). "El papel de la lingüística en el desarrollo de las tecnolog3as del habla". Casas G3mez, M. (Dir.) Varo Varo, C. (Ed.) *VII Jornadas de Lingüística*, pp. 137-191. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.

Llisterri, J. (2002). "Marcas fonéticas de la oralidad en la lengua de los chats: elisiones y epéntesis consonánticas". *Revista de Investigaci3n Lingüística* 2, 5, pp. 61-100. Disponible en: http://liceu.uab.es/~joaquin/publicacions/Llisterri_02_Oralidad_Chats.pdf Última consulta: septiembre de 2012

Llisterri, J. y Garrido, J. M. (1998). "La ingeniería lingüística en España". *El español en el mundo. Anuario del Instituto Cervantes*, pp. 299-391. Madrid: Instituto Cervantes - Arco/Libros.

Llisterri, J. (1988). "La síntesis del habla: estado de la cuesti3n". *Procesamiento del Lenguaje Natural* 6, pp. 17-41.

L3pez-Ludeña, V.; San-Segundo, R.; Montero, J. M.; Barra-Chicote, R. y Lorenzo, J. (2012). "Architecture for Text Normalization using Statistical Machine Translation techniques". *Proceedings of IberSpeech 2012*, pp. 204-213. Madrid. Disponible en:

http://iberspeech2012.ii.uam.es/IberSPEECH2012_OnlineProceedings.pdf Última consulta: enero de 2013.

L3pez Quero, S. (2003). "El lenguaje de los 'chats'", Port Royal Ediciones, Granada.

Lorenzo-Trueba, J.; Watts, O.; Barra-Chicote, R.; Yamagishi J.; King S. y Montero J. M. (2012). "Simple4All proposals for the Albayzin Evaluations in Speech Synthesis". *Proceedings of IberSpeech 2012*, pp. 204-213, Madrid. Disponible en: http://iberspeech2012.ii.uam.es/IberSPEECH2012_OnlineProceedings.pdf Última versi3n: enero 2013.

Marcoccia, M. (2000). "Les Smileys: une repr3sentation iconique des 3motions dans la Communication M3diatis3e par Ordinateur". C. Plantin, M. Doury y V. Traverso (eds.): *Les 3motions dans les interactions*, Lyon: P.U.L.

Martínez, H. y Rojas, D. (2011). “Prosodia y emociones: datos acústicos, velocidad del habla y percepción de un corpus actuado”. *Lengua y habla*, 15, pp. 59-72.

Mayans, J. (2002). “*Género Chat. O cómo la etnografía puso un pie en el ciberespacio*”. Barcelona, Gedisa.

Mayans, J. (2002b). "De la incorrección normativa en los chats". *Revista de investigación Lingüística*, num. 2, vol. V, pp.101-116. Disponible en el ARCHIVO del Observatorio para la CiberSociedad en: <http://www.cibersociedad.net/archivo/articulo.php?art=43>

Mayans, J. (2000). “Género confuso: género chat”. *Revista TEXTOS de la CiberSociedad*, 1. Temática Variada. Disponible en <http://www.cibersociedad.net>

Möbius, B. y Van Santen, J. (1996). “Modelling Segmental Duration in German Text-to-Speech Synthesis”. *ICSLP 96, The Fourth International Conference on Spoken Language Processing*, pp. 2395-2398. Philadelphia, PA, USA.

Montero, J. M. y Barra-Chicote, R. (2012). “ The Albayzin 2012 Speech Synthesis Evaluation (Albayzin 2012 SS)”. *Proceedings of IberSpeech 2012*, pp. 214-223. Madrid, Spain. Disponible en: <http://iberspeech2012.ii.uam.es/images/PDFs/iberspeech2012synthesis-rbc.pdf> Última consulta: enero de 2013.

Montero, J. M. (2003). “Estrategias para la mejora de la naturalidad y la incorporación de variedad emocional a la conversión texto a voz en castellano”. Tesis doctoral, Universidad Politécnica de Madrid.

Morala, J. R. (2001). “Entre arrobas, eñes y emoticonos”. *II Congreso Internacional de la Lengua Española. El español en la Sociedad de la información*, 2001, Disponible en: http://cvc.cervantes.es/obref/congresos/valladolid/ponencias/nuevas_fronteras_del_espanol/4_lengua_y_escritura/morala_j.htm Última consulta: octubre de 2012.

Moré, J.; Climent, S.; Oliver, A. y Taulé, M. (2005). “Análisis de fenómenos lingüísticos de los mensajes de correo electrónico en catalán desde la perspectiva de la traducción automática”. *Procesamiento del Lenguaje Natural*, núm. 35, pp. 45-51. Disponible en:

<http://www.sepln.org/revistaSEPLN/revista/35/06.pdf>
consulta: noviembre de 2012.

Última

Murray, I. R. y Arnott, J. L. (1995). "Implementation and Testing of a System for Producing Emotion-by-Rule in Synthetic Speech". *Speech Communication*, 16, pp.. 369–390.

Murray, I. R. y Arnott, J. L. (1993). "Toward the simulation of Emotion in Synthetic Speech: A Review of the Literature of Human Vocal Emotion". *Journal of the Acoustic Society of America*, 93(2), pp. 1097-1108.

Navarro, T. (1944). "*Manual de entonación española*". Nueva York, Hispanic Institute on the United States

Nespor, M. y Vogel, I. (1986). "*Prosodic Phonology*", Dordrecht, Foris.

Oliva, K., (1997). "Techniques for accelerating a grammar-checker". *Proceedings of the 5th Conference on Applied Natural Language Processing*, Washington D.C.

O'shaughnessy, D.; Barbeau, L.; Bernardi, D. y Archambault, D. (1988). "Diphone Speech Synthesis". *Speech Communication* 7 pp. 55-65.

Papineni, K.; Roukos, S.; Ward, T. y Wei-Jing, Z. (2001). "*Bleu: a method for automatic evaluation of machine translation*". Technical Report RC22176 (W0109-022), IBM Research Division, Thomas J. Watson Research Center.

Patterson, H. (1996). "Computer-Mediated Groups: A Study of a Culture in Usenet", tesis doctoral, Universitat de Texas. Disponible en: http://www.agm.net/holly/holly_dissert.html Última consulta: diciembre de 2012

Pierrehumbert, J. B. (1980). "*The Phonology and Phonetics of English Intonation*". Bloomington, Indiana University Linguistics Club.

Plutchik, R. (2001). "The nature of emotions". *American Scientist*, 89(4), pp. 344–350.

Pons, C. (2002). "Els xats:) la ludoteca de la llengua". *Interlingüística* 13 (III), pp. 273-281.

- Quazza S.; Donetti L.; Moisa, L. y Salsa, P. L. (2001). "ACTOR®: A Multilingual Unit-Selection Speech Synthesis System". *Proceedings of 4th ISCA Tutorial and Research Workshop on Speech Synthesis*, Atholl, Scotland. Disponible en: http://www.loquendo.com/site_media/brochures/en/tts.pdf Última consulta: enero de 2013.
- Quilis, A. (1993). "Tratado de fonología y fonética españolas", Madrid, Gredos.
- Riedi, M. (1998). "Controlling Segmental Duration in Speech Synthesis Systems." Tesis Doctoral, Zurich.
- Riedi, M. (1995). "A Neural-Network-Based Model of Segmental Duration for Speech Synthesis". *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Technology*, vol. 1, pp. 599-602. Madrid, Spain.
- Rintel, E. S. y Pittan, J. (1997). "Strangers in a strange land: Interaction management on Internet Relay Chat". *Human Communication Research*, 23, 4.
- Rodríguez, A.; Lazaro, P.; Montoya, N.; Blanco, J. M.; Bernadas, D.; Oliver, J. M. y Longhi, L. (1999). "Modelización acústica de la expresión emocional en el español". *Procesamiento del Lenguaje Natural*, 25, pp. 159-166.
- Sacks, H.; Schegloff, E. A. y Jefferson, G. (1974). "A simplest systematics for the organization of turn-taking for conversation". *Language*, 50, 4.
- Sainz, I.; Erro, D.; Navas, E.; Hernández, I.; Sánchez, J. y Saratxaga, I. (2012). "Aholab Speech Synthesizer for Albayzin 2012 Speech Synthesis Evaluation". *Proceedings of IberSpeech 2012*, pp. 204-213. Madrid. Disponible en: http://iberspeech2012.ii.uam.es/IberSPEECH2012_OnlineProceedings.pdf Última consulta: enero de 2013.
- Sanmartín, J. (2007). "El chat. La conversación tecnológica". Madrid, Arco Libros.
- Scherer, K. R. (1999). "Appraisal theory". T. Dalgleish y M. Power (Eds.) *Handbook of Cognition and Emotion*, pp. 637-663. Wiley & Sons, Ltd., New York.

- Schlosberg, H. (1941). "A scale for the judgement of facial expressions". *Journal of Experimental Psychology*, 29, pp. 497-510.
- Shi, Q. y Fischer, V. (2004). "A Comparison of Statistical Methods and Features for the Prediction of Prosodic Structures". *Proceeding of INTERSPEECH 2004- ICSLP, 8th International Conference on Spoken Language Processing*, Jeju Island, Korea.
- Silverman, K.; Beckman, M.; Pitrelli, J.; Ostendorf, M.; Wightman, C.; Price, P.; Pierrehumbert, J. y Hirschberg, J. (1992). "ToBI: a standard for labelling English prosody". *ICSLP 98 Conference Proceedings. The 5th International Conference on Spoken Language Processing*, vol. 2, pp. 867-870. Sydney Convention Centre, Sydney, Australia.
- Sproat, R.; Black, A.; Chen, S.; Kumar, S.; Ostendorf, M. y Richards, C. (2001). "Normalization of non-standard words". *Computer Speech and Language*, 15(3), pp. 287-333. Disponible en: <http://www.cs.columbia.edu/~julia/papers/sproatetal01.pdf> Última consulta: septiembre de 2012.
- Stowell, S. (2012). "Instant R: An Introduction to R for Statistical Analysis". *Jotunheim Publishing*. Disponible en: <http://www.instantr.com/book> Última consulta: agosto de 2012
- 'T Hart, J.; Collier, R. y Cohen, A. (1990). "A Perceptual Study of Intonation. An Experimental - Phonetic Approach to Intonation". Cambridge: Cambridge University Press
- Thorsen, N. (1978). "An acoustical analysis of Danish intonation". *Journal of Phonetics* 6, pp. 151-175.
- Tokuda, K.; Kobayashi, T. y Imai, S. (1995). "Speech parameter generation from HMM using dynamic features". *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'1995)*, volumen 1, pp. 660-663.
- Torres, M. (1999). "Els xats: entre l'oralitat i l'escriptura", *Els Marges: revista de llengua i literatura*, 65 pp. 113-126. Disponible en: <http://perso.wanadoo.es/mtorresv/articles/art-emot.html> Última visita: mayo de 2011
- Toutanova, K. y Moore, R. (2002). "Pronunciation modeling for improved spelling correction". *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 144-151, Philadelphia, PA.

- Tusón, A. (1991). "Las marcas de la oralidad en la escritura". *Signos*, 3. Gijón: Centro de profesores de Gijón.
- Van Santen, P. y Buchsbaum A. (1997). "Methods for optimal text selection". *Eurospeech Proceedings II*, pp.553-556.
- Van Santen, J. (1994). "Assignment of segmental duration in text-to-speech synthesis". *Computer Speech & Language* 8, pp. 95-128.
- Veronis, J. (1988). "Morphosyntactic correction in natural language interfaces". *Proceedings of the 13th International Conference on Computational Linguistics (COLING'88)*, pp. 708-713.
- Violante, L. (2012). "Construcción y evaluación del back-end de un sistema de síntesis de habla en español argentino". Tesis de licenciatura, Universidad de Buenos Aires.
- Wells, J. (1989). "Computer-Code Phonemic Notation of Individual Languages of the European Community", en *Journal of the International Phonetic Association* 19, 1, pp. 31-54.
- Whissel, C. M. (1989). "The dictionary of affect in language". Plutchnik, R. y Kellerman, H. (eds) *Emotion: Theory, Research and Experience*: Vol. 4, New York: Academic press.
- Wilson, A. (1993). "A Pragmatic device in electronic communication". *Journal of Pragmatics*, 19(4).
- Yvon, F. (2008). "Reorthography of SMS messages". *Research Report LIMSI/CNRS*. Disponible en: <http://perso.limsi.fr/yvon/publications/bibs/sources/Yvon08reorthography.pdf> Última consulta: octubre de 2012
- Yus, F. (2001). "Ciberpragmática: El uso del lenguaje en Internet", Barcelona, Ariel
- Yus, F. (2010). "Ciberpragmática 2.0: nuevos usos del lenguaje en Internet", Barcelona, Ariel.

Apéndice A

CONTENIDO DEL CORPUS DE TEXTO

A.1. Estados emocionales y fisiológicos aislados

Son mayoritarios los casos en los que se ha etiquetado un solo estado emocional o fisiológico por enunciado. A continuación se muestran todos los encontrados en el corpus de análisis y la frecuencia de aparición de cada uno de ellos.

Las emociones cuya frecuencia de aparición es igual o superior al uno por ciento, marcadas en las tablas con un sombreado, son las que se han seleccionado para crear el corpus emocional.

Emoción	Nº de casos	Frecuencia aparición
rechazo	949	19,97894737
alegría	425	8,947368421
burla	413	8,694736842
interés	393	8,273684211
enfado	264	5,557894737
afecto	178	3,747368421
alteración	129	2,715789474
sorpresa	112	2,357894737
orgullo	99	2,084210526
tristeza	98	2,063157895
extrañeza	92	1,936842105
diversión	81	1,705263158
admiración	76	1,6
duda	63	1,326315789
resignación	56	1,178947368
decepción	53	1,115789474

ironía	46	0,968421053
indiferencia	45	0,947368421
alegría_2	39	0,821052632
confianza	39	0,821052632
desánimo	36	0,757894737
compasión	31	0,652631579
preocupación	30	0,631578947
desconfianza	22	0,463157895
enfado_2	22	0,463157895
aburrimiento	19	0,4
rechazo_2	19	0,4
impaciencia	18	0,378947368
esperanza	16	0,336842105
cansancio	15	0,315789474
vergüenza	15	0,315789474
afecto_2	13	0,273684211
alteración_2	13	0,273684211
alivio	12	0,252631579
asco	12	0,252631579
interés_2	12	0,252631579
envidia	11	0,231578947
culpa	8	0,168421053
miedo	8	0,168421053
sorpresa_2	7	0,147368421
burla_2	6	0,126315789
admiración_2	5	0,105263158
desánimo_2	5	0,105263158
nostalgia	5	0,105263158
placer	5	0,105263158
ilusión	4	0,084210526
calor	3	0,063157895
extrañeza_2	3	0,063157895
sueño	3	0,063157895
compasión_2	2	0,042105263
complicidad	2	0,042105263
desilusión	2	0,042105263

despreocupación	2	0,042105263
calor_2	1	0,021052632
decepción_2	1	0,021052632
impaciencia_2	1	0,021052632
orgullo_2	1	0,021052632
placer_2	1	0,021052632

A.2. Actos de habla

En el corpus de análisis se han encontrado los siguientes actos de habla:

Acto de habla	Nº de casos	Frecuencia aparición
saludo	199	4,189473684
despedida	77	1,621052632
llamada	11	0,231578947
agradecimiento	7	0,147368421
ánimo	4	0,084210526
amenaza	1	0,021052632
advertencia	1	0,021052632

A.3. Combinación de dos elementos

A menudo, en el habla real las diferentes emociones se entremezclan, pudiendo expresar más de un estado emocional o fisiológico a la vez. A continuación se muestran todas las combinaciones de dos elementos halladas en el corpus de análisis. Estas combinaciones pueden estar formadas por dos estados emocionales, un estado emocional y un estado fisiológico o un estado emocional y un acto de habla.

Emoción	Nº de casos	Frecuencia aparición
burla - rechazo	57	1,2
alegría 2 - saludo	53	1,115789474
enfado - rechazo	51	1,073684211
alteración - rechazo	33	0,694736842
alteración - enfado	18	0,378947368
alegría - saludo	12	0,252631579
asco - rechazo	11	0,231578947
alteración 2 - rechazo 2	10	0,210526316
alegría 2 - despedida	7	0,147368421
afecto - despedida	6	0,126315789
alegría - sorpresa	6	0,126315789
rechazo - tristeza	6	0,126315789
afecto - saludo	5	0,105263158
desánimo - rechazo	5	0,105263158
desánimo 2 - rechazo	5	0,105263158
afecto - interés	4	0,084210526
alegría - interés	4	0,084210526
burla - interés	4	0,084210526
extrañeza - rechazo	4	0,084210526
indiferencia - rechazo	4	0,084210526
interés - preocupación	4	0,084210526
admiración - alegría	3	0,063157895
alegría - burla	3	0,063157895
alegría - ilusión	3	0,063157895
burla - enfado	3	0,063157895
burla - orgullo	3	0,063157895
decepción - tristeza	3	0,063157895
enfado - impaciencia	3	0,063157895
enfado - orgullo	3	0,063157895
indiferencia - orgullo	3	0,063157895
ironía - rechazo	3	0,063157895
afecto - alegría	2	0,042105263

afecto - burla	2	0,042105263
afecto - rechazo	2	0,042105263
alegría - despedida	2	0,042105263
alegría - orgullo	2	0,042105263
alegría 2 - llamada	2	0,042105263
alteración 2 - enfado 2	2	0,042105263
burla - compasión	2	0,042105263
burla - duda	2	0,042105263
burla - preocupación	2	0,042105263
burla - sorpresa	2	0,042105263
desánimo - tristeza	2	0,042105263
diversión - orgullo	2	0,042105263
diversión - rechazo	2	0,042105263
extrañeza - interés	2	0,042105263
impaciencia - rechazo	2	0,042105263
rechazo - saludo	2	0,042105263
aburrimiento - rechazo	1	0,021052632
admiración - burla	1	0,021052632
admiración 2 - alegría 2	1	0,021052632
afecto - agradecimiento	1	0,021052632
afecto - alegría 2	1	0,021052632
afecto - resignación	1	0,021052632
afecto 2 - saludo	1	0,021052632
afecto 2 - despedida	1	0,021052632
alegría - alivio	1	0,021052632
alegría - llamada	1	0,021052632
alegría 2 - alteración 2	1	0,021052632
alegría 2 - interés	1	0,021052632
alegría 2 - sorpresa 2	1	0,021052632
alteración - burla	1	0,021052632
alteración - interés	1	0,021052632
alteración - rechazo 2	1	0,021052632
alteración 2 - burla 2	1	0,021052632
burla - calor	1	0,021052632

burla - extrañeza	1	0,021052632
cansancio - resignación	1	0,021052632
complicidad - interés	1	0,021052632
confianza - diversión	1	0,021052632
confianza - interés	1	0,021052632
confianza - orgullo	1	0,021052632
diversión - sorpresa	1	0,021052632
enfado - tristeza	1	0,021052632
enfado 2 - rechazo	1	0,021052632
impaciencia - interés	1	0,021052632
impaciencia 2 - interés 2	1	0,021052632
interés - rechazo	1	0,021052632
interés - sorpresa	1	0,021052632
ironía - saludo	1	0,021052632
preocupación - rechazo	1	0,021052632
rechazo - sorpresa	1	0,021052632

A.4. Combinación de tres elementos

En algunos casos aislados se ha observado que un mismo enunciado se han entremezclado hasta tres etiquetas diferentes que hacen referencia al estado emocional o fisiológico, o al acto de habla.

Emoción	Nº de casos	Frecuencia aparición
afecto - alegría - despedida	1	0,021052632
alteración - enfado - rechazo	1	0,021052632
enfado - impaciencia - rechazo	2	0,042105263

Apéndice B

CORPUS DE GRABACIÓN

B.1. Corpus emocional

A continuación se muestran los enunciados que conforman el corpus emocional

B.1.1. Admiración

1. ¡El grito es el cuadro más importante del impresionismo! Es un cuadro que más allá de la estética, busca plasmar el sentimiento interior.
2. ¡Todas las deformidades de ese cuadro vienen del horror del grito que suelta ese hombre!
3. Pero ese tío que tú dices, se ha currado sudando sangre todo lo que ahora tiene.
4. ¡Es más listo!
5. ¡Joder! ¡Cómo molan tus subnombres!
6. ¡Qué buena es Prison!
7. ¡Qué pasada de película!
8. ¡Mi Rosa sí que está buena!
9. Las lanzas, las meninas, el fusilamiento, los cuadros negros... ¡Una pasada todos!
10. ¡Madre mía, fue una bestialidad eso!
11. ¡Hay poca gente que sienta así el arte!

12. ¡Y a mí me encantan los dos!
13. ¡Qué gol de Baraja!
14. Ésta está potente.
15. ¡Es Hool de cojones!
16. ¡Cómo mola el Valencia!
17. ¡Cómo mola el viejo de Goku!
18. ¡El puto Bardak es enorme!
19. ¡Qué grande es el hijo de puta!
20. ¡Cómo molas!
21. ¡Eres la leche!
22. ¡Cómo me gusta tu letra!
23. ¡Qué fenómeno eres!
24. ¡Me encanta el hombre!
25. ¡Qué efectivo!

B.1.2. Afecto

1. Quiero que sepas, que aunque sea a setecientos kilómetros me tienes más a tu lado de lo que parece, "y no me hago líos".
2. Yo también he hecho esa reflexión, tranquilo, intento no hacerme líos.
3. ¡Si estaba allí te daba unos besitos como antídoto de energía!
4. Bueno, pues dale muchos ánimos de nuestra parte y dile que si llego a Santiago le voy a poner una velita.
5. Besos para todos y sobre todo para tu suegra, que seguro que también necesita muchos ánimos.
6. Lo dicho, que me alegro por vosotros.
7. Porque lo entiendes. El amor está por encima de todo eso.
8. Y por eso me sigues queriendo.

9. La más preciosa del mundo.
10. ¡Ay! ¡Qué rica ta mi niña!
11. Cabroncete...
12. Las mujeres nos hacen falta, hay que mimarlas y darle amor y así ser correspondidos.
13. Me alegro, Liadara.
14. Bueno tía, tranqui.
15. ¡Que vaya bien mañana!
16. No discutas tanto campeón.
17. Es que es muy salao.
18. Besillus a todos y a todas.
19. Se nota que se ha recuperado.
20. Liadara, a ti no te hace falta nadie de Andorra.
21. Gracias, ecuatoriano.
22. No te quiero hacer daño.
23. Un besito muy grande.
24. No vale la pena dejarlo.
25. ¡Cuídate, y no trabajes mucho!

B.1.3. Alegría

1. He sido padre.
2. Esta vez es una maravilla.
3. ¡Es estupendo entonces!
4. ¡Qué bueno! Pero la tuya de Paquirrín no la supera nadie, y sin photoshop ¡eh!
5. Estoy rodeado de buenas noticias y eventos mil.
6. El domingo, unos amigos, que fueron mis monitores de scout nos dijeron que serán padres.

7. Y una amiga de la uni también será mamá.
8. Se enteró el sábado... ¡Y ya me lo ha dicho!
9. ¡Va perdiendo el Barça, yuju!
10. ¿Sabes con quién he hablado por el Messenger?
11. Con mi mejor amigo de la infancia, mi compañero de clase de primaria.
12. ¡Javier Fernández! Seguramente quedaremos.
13. Y he tenido la suerte de tener aquí en Valencia la exposición de Sorolla.
14. Mecagüen, me pareció a mí que eras asturiano.
15. No, viene esta noche, a eso de las diez y media al Prat.
16. Este lunes empiezo en Sant Cugat.
17. ¡Ah! Me parece genial.
18. El viernes me voy para Madrid.
19. ¡Mañana, vida nueva!
20. Irá a Andalucía, a verme con Matt e Isma.
21. Y el último proyecto encendido de ayer.
22. He encontrado un chaval que me hará unas fotos.
23. Fotos en blanco y negro al desnudo.
24. No, eso es que llevo casi un año con mono de comprar unos calzoncillos así.
25. Presiento que me viene una época buena.

B.1.4. Alegría nivel 2 más saludo

1. ¡Hola! ¿Qué tal estáis?
2. ¡¡Andalú!! ¡Hola, guapo!
3. ¡¡Buenas!!
4. ¡Chiquitina veintitrés! ¡Buenas, niña!

5. ¡Buenas tardes a todos!
6. ¡¡Hola Robert, guapo!!
7. ¡¡We!! ¡¡¡Muy buenas!!!
8. ¡¡¡¡Hola!!!!
9. ¡Hola, Malota!
10. ¡¡Hola, descarados!!
11. ¡Hola vecina!
12. ¡¡Hola!!
13. ¡¡Hola!! ¿Cómo tamos?
14. ¡¡¡Buenas!!!
15. ¡¡¡¡Hola Mamita!!!
16. ¡¡Holas!!
17. ¡¡Hola!!
18. ¡¡¡¡¡Hola!!!!!!
19. ¡¡Soy todo amor, hola!!
20. ¡¡Hola, chica!!
21. ¡¡Buenas!!
22. ¡¡Buenas, chiquitina!!
23. ¡¡Buenas!!
24. ¡¡¡¡Hola!!!
25. ¡¡Hola Fany!!

B.1.5. Alteración

1. Un hijo de la gran puta.
2. Me cago en su puta madre.
3. A partir de ahora, cada sudaca que vea, un palizón le voy a dar.

4. ¿Es verdad... que Armstrong fue el primero en llegar a la una? ¡Un pene, fue un ruso!
5. ¡El arte mueve el mundo o tendría que movernos!
6. ¡Mentira!
7. ¡Vamos, no me jodas!
8. ¿Y eso qué coño es?
9. ¡Es que mírale!
10. Luego me dice que por qué la tomo con él.
11. ¡Patadas en la boca! ¡Infladlos a capones!
12. ¡Matemos al emo!
13. ¡Eh! ¡Que yo le digo esto después de que ella me llamase cabrón!
14. ¡Yo nunca la he insultao sin razón!
15. ¿Perdona?
16. Y te he mentío a ti, a Rita y a Marcela, porque les he pasao las mismas fotos que a ti.
17. Así que no me vengas de payasadas.
18. ¡Va ganando y punto!
19. Pero te veo por la calle puto sudaca y te reviento la puta cabeza, asqueroso.
20. Tú mejor cómeme la polla.
21. Malota phantom, puta, sé dónde vives.
22. ¡Pégales, amigo!
23. Es que hay una chica que tiene lo mismo.
24. ¡Llevo la ostia esperando alguien que sea artista!
25. Y tú una furcia.

B.1.6. Burla

1. Quién iba a pensar, que un día, una foto suya con los mofletes inflados me serviría para regalarle submarinismo.
2. ¿Le afectará también eso que dicen de que come por los ojos?
3. Los padres le pasan la comida por debajo la puerta.
4. Pareces el que levanta el pañuelo blanco en medio de la batalla,
5. Antes de que secuestre otro avión.
6. Por cierto, ayer le envié una foto nuestra a mi amigo alemán y dijo que tú tenías más barba que él y yo juntos.
7. ¡Anda que no los tienes peludos!
8. ¡Ha sido lo que más me ha recordado a ti!
9. A mí me da paz.
10. El de "yo soy muy vanidoso, que visto de Emidio Tucci".
11. Que quizá cuando los recoja, vuelven con un hueso en la boca.
12. ¡Ostia! Un policía diferente.
13. No sé, pero como lo diga por el cerebro vas apañado.
14. Tú tenías caja con cristales tintados.
15. Ni eso, estaba en la jaula de ensayos número uno.
16. Quizá necesite un psicoanálisis.
17. Sick Boy, te queda bien ese color. Muy acorde a tu desvío sexual.
18. Ya me quedo más tranquila.
19. Luego se tirará unos días sin venir, como cuando ganó el Madrid en el Camp Nou.
20. A lo mejor viene con otro nick, ¿no?
21. ¡Al loro Malota phantom! Eres igual pero algo menos.
22. Sí, al móvil, Chiquitina.

23. Pero es que con tanto catalán... pues me daba miedo.
24. Ese "y punto" me ha convencido.
25. Hombre, es que no me jodas, le pongo mi foto y me come, así, directamente.

B.1.7. Burla más rechazo

1. Así están los Borbones...
2. Que parece que tengan un zapato de esparto en la boca.
3. ¡Uy! Sí que estás enterado tú.
4. De esos que sacan fotos con cara de pena
5. Con el pelo por la cara.
6. Y pantalones de nena.
7. Con sus mírale.
8. Llámale.
9. Papá tengo sez.
10. Patético de Madriz.
11. Simbolito arriba, simbolito abajo.
12. ¡Aix! Eso de chiquito veinticuatro es como gay.
13. ¿Si lo miras incluso lo puedes oír?
14. ¡Qué friki has estado!
15. Mira Luis dieciséis, que manda inventar la guillotina y la estrena él.
16. Ahí sí que hay que ser tonto.
17. ¡Huy! El de veinte años...
18. Porque la bola ya la arrastras ¿verdad?
19. ¡Qué mala!
20. Aquí se desaprende.

21. Ya verás, nos va a salir con hordas de robot.
22. Si ya me cuesta a mí acordarme de la mía... como pa' saber la de nadie.
23. Normal, está lleno de españoles.
24. ¡Sí, claro! ¡Ahora nos da vergüenza!
25. No pierdes el tiempo ¡eh!

B.1.8. Decepción

1. No, acabo mañana.
2. Entonces... está chungui que nos veamos ¿no?
3. La verdad es que sí.
4. El Barça no se llevará nada este año...
5. Mi padre me dijo que estaba horrible...
6. Y me dijo que con ese pelo dónde iba, que quedaba fatal...
7. Siento no ser interesante.
8. Yo soy una persona bastante simple...
9. Pero luego te despiertas y a la mierda.
10. ¡Uf! Eso es algo lejos de mi país.
11. Yo también.
12. ¡Qué faena!
13. Cosas malas, seguro.
14. A este paso...
15. Tía... no veas lo que ha pasao.
16. Resulta que se equivocó de día y viene hoy por la noche...
17. Pues sí... un poco sí, porque yo tenía unas ganas ayer... y saber que aún quedaba un día más...
18. No creo que nadie baje a los Estados Unidos de ahí, la verdad.

19. La historia no siempre es objetiva.
20. Ya veo, ya...
21. Pero como nos vamos escopeteados...
22. Joder. Pues se lo quería regalar yo pero no lo he encontrado...
23. Para la o ene ge teníamos que acabar una presentación y nada... No hemos podido.
24. Pues me lo había parecido.
25. No se ve nada, pero bueno...

B.1.9. Despedida

1. Me las piro.
2. Deu
3. Adiós.
4. Garten Yitan, venga, hasta otra.
5. ¡Ok, hasta otro rato!
6. Ta luego.
7. ¡Deu, Enita!
8. Venga, adiós.
9. Me voy chicos y chicas.
10. ¡Hasta otra!
11. Buenas noches gente.
12. Ciao.
13. Agur.
14. Ok ¡Buenas noches!
15. Chao.
16. ¡Ciao, Ciao!
17. ¡Chao Princes! ¡Chao David!

18. ¡Ale! Me voy.
19. ¡Hasta luego!
20. Chao pulgosa.
21. ¡Adeu, Flor!
22. Venga, chao.
23. ¡Nos vemos demà!
24. Bueno nenín, te deajo, que en nada se me hacen las dos.
25. Me voy otra vez.

B.1.10. Diversión

1. Y lo del final... "cuidado con el barro"
2. Orégano, Anarex treinta, orégano.
3. ¡Alucinógena!
4. Toy colocá tol día.
5. ¿Y camarero? ¡¡Eldelbar!!!
6. Ya pareces andaluz Alex, ¡qué exagerao!
7. Y quizá alguno puede que no salga si entra.
8. Puede, puede...
9. Me parto.
10. Soy gracioso, eso lo tengo.
11. Iglesias...
12. Yo me divierto bastante.
13. Debe ser eso...
14. ¡¿Por qué no te callas?!
15. ¡Ésa fue buena!
16. No creo que hablen de la nariz.
17. Un porro, Amparo.

18. ¡Qué bueno!
19. ¡Qué risa!
20. Que había robao un Banesto.
21. Que qué hago
22. Drogarme todo lo que puedo.
23. La droga es la auténtica salud.
24. ¡Joder! ¡Qué bueno! ¡Hacía tiempo que no veía a nadie escribir así!
25. ¡Me meo!

B.1.11. Duda

1. Es que... no sé exactamente cómo era...
2. No sé si es tu cumpleaños o el suyo...
3. ¿En Sant Pau me dijiste?
4. Parece el de la Cabra Mecánica ¿no?
5. ¿Ése no taba más gordo?
6. Yesi ¿tas?
7. Más ondulado y algo más corto, creo.
8. Supongo que sí.
9. El próximo viernes creo que no tengo ningún problema...
10. Creo que salió el veintinueve de noviembre.
11. No sé, yo no leo nada que diga que quería contigo...
12. Por ahí supongo.
13. ¿Más interesante que él?
14. Supongo que sí.
15. ¿Te conozco?
16. Era por eso ¿no?

17. Pero no sé si se arrancó los ojos por eso.
18. Sí, yo creo que sí.
19. ¿Tú crees que es bueno? Porque si mañana tienes examen... tututu...
20. ¿Al Fernando Alonso?
21. Creo que sí.
22. Que me lo enseñaron en el youtube
23. Pero no me acuerdo
24. Bueno... No sé qué decirte.
25. Ya, pero no sé...

B.1.12. Enfado

1. La gente parece que sólo piensa que internet es solamente para buscar sexo y pasar el rato.
2. ¡Que todos los tíos no somos unos putos salidos, coño!
3. Por unos pocos mierdas, pagamos el resto.
4. ¡Qué capitalista ni que pollas! Aquí somos todos capitalistas, el que quiera y el que no.
5. ¡Putá conexión!
6. El de siempre, prefiero ser rara que ser tan simple como tú, puto manguán.
7. No soy campeón de nada.
8. Soy todo amor, la verdad es que me harías un gran favor si no me trataras.
9. Enior, pues de gay poco.
10. Y no se metan conmigo, que no quiero hablar. ¿Vale?
11. Que no tienes ni dos dedos de frente.
12. Que pases de mí. Si no apporto nada, no me nombres.

13. Primero aprende a leer, que Garnet es nombre de mujer.
Subnormal
14. Rancis, cállate tú, cansina.
15. ¿Qué pasa, te molo o qué?
16. El de siempre, no me toques las pelotas
17. Que confundes el mirar a un madrileño con mirarte al espejo, payasín.
18. ¡Es que mírale al puto ceporro!
19. ¿Qué problema tienes conmigo?
20. Y has empezado a reírte de mí.
21. ¿No sabes callar la puta boca que tienes?
22. ¡Y ella a mí que soy un cabrón así porque sí!
23. No, yo me he argumentao.
24. ¿Por qué no me has llamado, Me visto de mujer?
25. Hablas bien o no te entiendo.

B.1.13. Enfado más rechazo

1. ¡Así que menos rollos!
2. Realmente, Ariadna, me la suda completamente lo que te satisfaga...
3. Todas las ayudas pa' los putos extranjeros de mierda.
4. Porque esto, hoy, es una puta mierda.
5. ¿Te quiere ir por ahí, copia?
6. El de siempre, no me acoses que no me molas.
7. ¿Qué coño hacéis? ¿Por qué maulláis?
8. ¿En qué quedamos?
9. ¿Dieciséis o diecisiete?
10. Mejor emo que una tortuga.

11. ¡Morenito bueno, que te largues del puto privado!
12. El de siempre, asco das tú bastante con la pila de faltas de ortografía que tienes y lo subnormalín que demuestras ser.
13. El de siempre ¡Dios! ¿Es que no tienes ni una idea propia? Me dan ganas de reventarte las pelotas.
14. El problema no es de ellos, es tuyo.
15. ¿No sabes lo que es una broma?
16. ¡Un pene que no! ¡Mira los yanquis, se meten en todos laos!
17. Mikel, por enésima vez. ¡No, gracias!
18. ¡Eh! Que yo no soy na de eso.
19. Nadie quiere un amo. ¡Déjalo ya, porfa!
20. ¿Y qué si es como gay, Enior? Nadie tiene por qué meterse.
21. ¡Es que siempre es igual!
22. ¡Si tienes problema con alguien, solúcionálos y no me metas a mí en tus cosas!
23. No lo has hecho todavía, pero en cuanto te siga el rollo empezarás de nuevo con la misma historia de que la Venus es mu mala etcétera, etcétera.
24. ¿Qué prefieres a caso? ¿Ver la televisión? ¿Seguir las series? ¿Eso qué transmite? ¿Diversión?
25. Éste es un "punky" de mierda.

B.1.14. Extrañeza

1. ¡Jodó! ¿Un gran bajón en Greenbit?
2. ¿Ladilla? Nunca lo había escuchado con sentido de insulto...
3. La otra vez no había nada en el mensaje que se guardó... No sé qué pasaría.
4. ¿Maricomio?
5. ¿Obras en mi calle?

6. El de siempre ¿un privilegio?
7. ¿Y pa' qué me preguntas si soy tío?
8. ¿No te gusta Barcelona?
9. ¿Amigo mío?
10. Me visto ¿te das asco...?
11. ¿Nuestro gran rey?
12. ¿Y ésta qué dice?
13. ¿A quién tengo que dejar?
14. ¿Pempis?
15. ¿Ande tais, peña?
16. ¿Sólo tuyas?
17. ¿Ninguna mía?
18. ¿Y sigues despierto?
19. ¡Qué raro que no me haya llamado!
20. ¿Oye, me has enviado la postal?
21. ¿No te gusta?
22. ¿A Terrassa?
23. ¿Qué chati?
24. ¿Superiores?
25. ¿Nadie habla?

B.1.15. Interés

1. ¿Cómo lo llevas?
2. ¿A dónde se ha ido?
3. ¿Y quién se ha ido con él que yo conozca?
4. Muñeca guerrera ¿de dónde eres?
5. ¿Te enseñan quién fue el Cid Campeador?

6. ¿De dónde eres?
7. ¿De la capi?
8. ¿De qué pueblo?
9. ¿Alguien sabe dónde puedo crear una página para que la gente se pueda descargar series como Aída, eventos de la w w e, etcétera?
10. ¿Qué tal el sábado? ¿Te quedaste mucho rato?
11. ¿A qué cuenta de correo has enviado la foto?
12. ¿De qué parte, chico moreno?
13. Yo molo ¿en qué radio?
14. ¿De quién es el Messenger "Doloresladosa"?
15. ¿Y cuánto tardan en enviarla?
16. ¿Has recibido los correos para pedir muestras?
17. ¿Ya estáis preparados para la aventura?
18. ¿Te ha llamado mi madre?
19. ¿A ti cómo te va?
20. ¿Dónde os casáis, en Barcelona?
21. ¿Por qué siempre pones Generation Simple Plan?
22. ¿Qué son los emos?
23. Pibe ¿qué tienes que estudiar?
24. ¿Para qué quieres saber la contraseña?
25. ¿Ya te has instalado en casa de tu padre?

B.1.16. Orgullo

1. Pero si soy el puto centro de vuestro universo.
2. La mayoría quiere ser como yo.
3. Ya me copian las frases.
4. Aún no he cambiado ningún pañal.

5. El de siempre, todas las mañanas de lunes a viernes.
6. Sí, el de veinte años se está riendo un poco de vosotros.
7. Espero que no os moleste.
8. Primoroso, más chula que un ocho.
9. Somos chulas porque podemos serlo.
10. Soy encantadora.
11. España es uno de los mejores países del mundo.
12. Soy más original.
13. España ha sido la primera potencia del mundo más de un siglo.
14. Por la ambición de España, en el siglo dieciséis, era más fuerte que los Estados Unidos ahora.
15. Autobús y metro me sale gratis.
16. Sí, es gracioso.
17. Soy del Barça.
18. Somos superiores, David.
19. Racista y nazi.
20. Chaval, lo que tú fumas en un día apuesto lo que sea que lo fumo yo en una hora.
21. Es que yo soy aficionado del Real Madrid.
22. Pos claro que confío.
23. Es que toy de vacances.
24. Siempre he querido fotografiar mi cuerpo desnudo.
25. ¡Claro que posaré!

B.1.17. Rechazo

1. Desde luego estarán contentos con vosotros.

2. ¡Jolines, qué soso! No dices na de na... ¡Qué mente de piedra!
3. Sarita, guapa, cuida tus palabras.
4. Hay mucha tía subida en la parra.
5. Todos tiráis la piedra y escondéis la mano.
6. Lo que es penoso es meterse con tu país, sin motivo ni conocimiento de lo que se dice.
7. Lo que es un poco bobo es insultar a reyes que trabajaron por su país.
8. Grey es una egocéntrica.
9. Y algunas de tus bromas no me han hecho mucha gracia...
10. Tú eres muy chulita ¿no?
11. No mola hoy.
12. ¡Mira qué brazos tan peludos!
13. Los reporteros son cansinos.
14. ¡Ostia, qué programa tan malo!
15. Odio lo que estoy estudiando.
16. La gente no tiene mucho aguante.
17. Lo odio. Es un prepotente.
18. La gente sólo quiere pasar el rato y punto.
19. Vaya mierda.
20. ¡Vaya mierda de porros en Madrid!
21. Aquí hay demasiado 'fantasma'.
22. Esto no es callejeros, por favor, no interesa.
23. Vaya estudiante de economía...
24. Porque entre el opá, el aserejé y la Rosa vaya tela...
25. Sí, bueno, es que en lo que se ha convertido eurovisión...

B.1.18. Resignación

1. Y eso miraremos.
2. Que tengo que ponerme el mono de currar y ganarme la vida.
3. Garnet Yitan, los subnormales siempre vienen a mí.
4. No sé... es igual... misterios de la informática...
5. Bueno... ya me calmo.
6. Así es el amor.
7. Por lo menos te ríes un rato...
8. Bueno, así que hoy estudiaré... ¡Qué remedio...!
9. Bueno, al menos os llevaréis la liga.
10. Tengo repaso el sábado por la mañana.
11. Bueno, venga... porque me tienes super tonto por ti...
12. Bueno... pues ya estoy en ello.
13. Pero es curro, tía. Y si me tengo que matar, me mato que luego el dinerito se nota.
14. Bueno, pues a ver qué tonterías suelta...
15. Es que soy un desastre.
16. Es lo que toca.
17. Pero bueno... lo llevo bien.
18. Bueno... va...
19. Pos así nos quedamos...
20. Pero bueno... si tú lo dices...
21. Ya... pero bueno... qué le vamos a hacer...
22. Total en un principio no me iba a presentar porque ya estoy hasta los huevos, pero me convenció pa' que lo hiciera.
23. Y como es el último...
24. Bueno... hay de todo.

25. Bueno, otra vez será...

B.1.19. Saludo

1. Buenas tardes ¿Qué tal?
2. ¡Hola! ¿Qué tal mujeres del chat?
3. Hola a todos
4. Muñeca, mu buenas.
5. ¡Hola a todos y a todas!
6. Hola de nuevo.
7. Hola a todas
8. Mu buenas.
9. Buenas a los de siempre.
10. Buenas de nuevo.
11. Nas, amiga.
12. Jelou.
13. Buenos días.
14. Buenas, gente.
15. ¡Holas!
16. ¡Hola! Buenas noches a todos.
17. Buenas noches a todos.
18. Muñeca guerrera, nas tardes.
19. ¡Hola!
20. Garnet Yitan, nas.
21. Buenas tardes.
22. Hola gente de nuevo.
23. Buenos días por la mañana, Gema treinta y nueve.
24. ¡Hola! Buenos días a todos.

25. Hola Anarex.

B.1.20. Sorpresa

1. ¡¿Otra vez?!
2. ¡Hombre! ¡Si está Ariadna por ahí!
3. ¿Pero también eres del Sporting?
4. ¡Ah! ¿No vas a hacer bachiller?
5. ¡Joder, qué rapidez!
6. ¡Wow! ¡Qué tarde te conectas!
7. ¡Qué fuerte! ¡Me he conectado para escribirte un mail de buenos días!
8. Para nada esperaba verte por aquí.
9. ¡Están hablando de Rubí!
10. ¡En te ve tres!
11. ¿Tú otra vez por aquí?
12. ¿Aún estás escribiendo?
13. ¡Ay! ¡Mira! ¡Si está Frodo!
14. ¡Qué me dices!
15. ¿Gratis?
16. ¡No jodas! ¡Estoy escuchando princesa!
17. ¿Dónde están las cámaras?
18. Estoy aborronado.
19. ¡Y estaba valorado en doscientos kilos de euros!
20. ¡Snower habló!
21. ¡Mira quién tenía un Messenger que empezaba por spider no sé qué!
22. ¡Qué claro tan afable has puesto!
23. ¡Qué dices!

24. ¡Wow! ¡Si hay ciento cuatro!

25. ¿Desnudo?

B.1.21. Tristeza

1. Bueno, ahora hemos tenido un gran bajón de gente.
2. No, lo que pasa es que Oscar se ha ido.
3. No podré quedar.
4. Estaré en Figueras, que le hacen una fiesta a los padres de Dani por su aniversario.
5. Mi mejor amiga también estuvo ingresada por anorexia.
6. Además, se le junto todo a la pobre.
7. Tuvo un trombo en la ingle, anemia y anorexia.
8. Y ahora sólo le veo los miércoles, los viernes, sábados y algún domingo.
9. Anoche vino, pero joder, se me hace eterno no verle.
10. Pues disgustado, la verdad.
11. Mis padres sufren el tema de mi homosexualidad.
12. Aquí nunca hay nadie de Andorra...
13. Yo fatal lo llevo.
14. Pues que hoy no viene mi chico...
15. Hasta mañana no le veo.
16. Vamos, que estoy más sola que la una.
17. ¡Qué fin de semana más corto...!
18. No le escucharé.
19. Estaré currando.
20. Se me ha roto el portátil...
21. Me tengo que comprar un pecé.

22. Creo que se pensaba que iba a ser abuela y se ha llevado una decepción...
23. Pues yo... nada de viajes.
24. Yo quiero ser feliz.
25. Pues... complicañlo...

Apéndice C

ANÁLISIS DE LOS PATRONES MELÓDICOS DEL CORPUS DE GRABACIÓN

C.1. Parámetros globales

A continuación se muestran los resultados obtenidos en el análisis de los parámetros globales ('líneas-tipo' y rango) clasificados según la posición del GF (inicial, interior y final) y la emoción expresada.

C.1.1. Grupo fónico inicial

Emoción	Núm. Casos	Valor inicial P media	Valor final P media	Pendiente P media	Valor inicial V media	Valor final V media	Pendiente V media	Rango inicial P V	Rango final P V
Admiración	2	247.20	20.82	-212.61	115.38	84.03	-23.58	131.81	-63.20
Afecto	8	125.38	84.27	-37.82	101.15	10.55	-22.98	24.23	16.27
Alegría	1	193.24	185.08	-4.09	138.59	132.36	-3.13	54.64	52.71
Alteración	1	227.98	172.69	-27.50	139.05	93.31	-22.75	88.92	79.37
Burla	7	145.91	156.48	109.65	65.75	124.66	235.70	80.15	31.81
Burla – rechazo	2	153.63	174.20	-0.59	136.05	129.81	-17.70	17.57	44.39
Decepción	3	127.63	99.33	-48.57	108.20	86.33	-34.64	19.42	13.00
Diversión	1	149.58	241.17	97.43	60.73	191.66	139.28	88.85	49.51
Duda	2	114.81	118.32	-9.52	92.87	77.24	-11.49	21.94	41.07
Enfado	2	155.15	271.72	97.15	133.59	137.79	-53.51	21.56	60.26
Enfado – rechazo	6	161.25	181.96	86.70	119.19	119.76	29.48	42.05	62.20
Extrañeza	3	171.76	98.62	-77.38	120.28	82.55	-40.97	51.47	16.06
Interés	2	42.899	222.43	273.78	112.53	112.54	8.90	-69.64	109.89
Orgullo	2	129.99	133.59	-14.60	104.22	102.55	-39.37	25.77	80.85
Rechazo	1	168.66	104.68	-28.94	120.84	85.87	-15.82	47.81	18.81
Resignación	14	134.89	109.77	-47.86	114.06	81.32	-48.90	20.83	33.29
Tristeza	2	140.33	84.12	-104.03	116.96	76.46	-51.16	23.37	7.65
Neutro	1946	149.52	109.80	-22.92	112.97	85.06	-12.05	36.54	27.62

C.1.2. Grupo fónico interior

Emoción	Núm. Casos	Valor inicial P media	Valor final P media	Pendiente P media	Valor inicial V media	Valor final V media	Pendiente V media	Rango inicial P V	Rango final P V
Admiración	1	94.46	182.86	111.90	110.49	97.61	-16.30	-16.03	85.25
Afecto	1	117.35	101.94	-10.00	93.41	73.47	-12.95	23.93	28.46
Burla	1	134.94	238.49	170.06	101.57	220.21	194.85	33.36	18.27
Duda	1	88.72	150.01	30.49	83.11	127.06	21.86	5.61	22.95
Enfado	1	150.82	210.56	55.31	102.51	176.81	68.79	48.30	33.74
Enfado – rechazo	2	169.61	138.71	-27.35	139.78	86.80	-62.62	29.83	51.91
Resignación	2	123.12	102.61	-34.78	104.00	86.54	-29.55	19.12	16.07
Neutro	1823	137.50	106.15	-18.17	108.13	80.72	-14.78	29.37	28.02

C.1.3. Grupo fónico final

Emoción	Núm. Casos	Valor inicial P media	Valor final P media	Pendiente P media	Valor inicial V media	Valor final V media	Pendiente V media	Rango inicial P V	Rango final P V
Admiración	3	160.76	156.78	3.46	120.04	96.83	-7.86	40.71	59.94
Afecto	9	118.43	96.07	-16.83	96.89	72.18	-19.78	21.54	23.89
Alegría	3	168.73	114.34	-72.20	131.14	87.62	-50.64	37.59	26.72
Alteración	1	189.96	168.08	-10.57	131.90	124.15	-3.74	58.05	43.92
Burla	9	181.67	104.03	-47.34	140.92	85.19	-37.83	40.74	18.84
Burla – rechazo	3	193.93	119.84	-56.50	141.58	102.15	-29.95	52.34	17.68
Decepción	4	109.88	107.63	-10.42	95.71	80.03	-15.32	14.17	27.59
Diversión	1	199.58	101.45	-90.02	140.53	80.79	-54.80	59.04	20.65
Duda	5	140.26	84.60	-51.35	100.66	70.30	-21.32	39.60	10.08
Enfado	5	245.76	43.15	-183.99	125.20	71.71	-43.00	120.56	-28.56
Enfado – rechazo	6	188.12	92.08	-69.97	138.48	88.71	-27.52	49.64	8.99
Extrañeza	2	132.52	122.50	-3.69	116.59	86.77	-21.36	15.93	35.72
Orgullo	4	159.31	104.61	-41.15	125.27	87.12	-27.24	34.04	17.48
Rechazo	3	161.10	92.25	-44.89	114.90	85.87	-18.58	46.19	6.38
Resignación	14	136.79	99.20	-28.26	104.75	88.26	-12.19	32.04	10.38
Tristeza	2	155.97	75.58	-57.05	100.02	70.98	-18.30	55.94	4.60
Neutro	404	141.82	91.15	-37.55	109.07	69.21	-29.32	32.74	22.98

C.2. Patrones iniciales

A continuación se muestran todos los patrones iniciales tanto del corpus neutro como del corpus emocional de los grupos entonativos más frecuentes, los enunciativos formados por una, dos o tres sílabas. En el caso del corpus emocional, los patrones aparecen clasificados según las diferentes emociones representadas.

C.2.1. Una sílaba

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VI0_PM0	75
	VI0	35
	PI0	30
	VM0	24
	PM0	20
	PI0_PM0	12
	PI0_VF0	9
	VI0_VF0	8
	VI0_PFO	6
	VI0_PM0_VF0	6
	PI0_PFO	4
	VI0_PI0_PM0_VF0	4
	VI0_VM0	4
	PI0_VM0	3
	VI0_PM0_PFO	3
	0	2
	PI0_PM0_VF0	2
	PI0_VM0_VF0	2
	VM0_PFO	2
	VM0_VF0	2

	P+I0	1
	P+I0_P+M0	1
	P+I0_PM0	1
	PF0	1
	PM0_PF0	1
	VI0_P+M0	1
ADMIRACIÓN	0	1
AFECTO	VI0_PF0	1
ALEGRÍA	PI0_PF0	1
	PM0_VF0	1
	VI0_PM0_PF0	1
ALTERACIÓN	VI0	1
BURLA	PI0	2
	0	1
BURLA-RECHAZO	PI0	1
DECEPCIÓN	PM0	2
	PM0_VF0	1
	VF0	1
DUDA	0	3
	VI0_PM0_VF0	1
ENFADO	PI0_PM0_VF0	1
	PI0_VM0_VF0	1
	VF0	1
ENFADO-RECHAZO	VI0_PM0_PI0_VF0	1
EXTRAÑEZA	VI0_PM0_VF0	1
INTERÉS	VI0_PM0_VF0_VM0	1
ORGULLO	0	2
	VI0	1
RECHAZO	VI0_PM0	3
	0	1
	PI0	1

	PM0	1
	VI0	1
RESIGNACIÓN	VI0_PM0	1
TRISTEZA	VI0_PM0_VF0	1
	PI0_VF0	1
DESPEDIDA	VM0	1
SALUDO	VM0_PF0	1

C.2.2. Dos sílabas

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VI0_PM0	32
	PI0	30
	PI0_PI1	29
	VI0_PI1	26
	VI0	23
	PM0_PI1	19
	PM0	15
	VI0_PF0	13
	PI0_VI1	12
	VI0_PM0_PI1	12
	VI0_PM0_VF1	12
	VM0	12
	VM0_PI1	12
	PI0_PM1	11
	VI0_PM0_VM1	9
	VI0_VF1	9
	PI0_VM1	8
	PM0_VF1	8
	VI0_PM1	8

VM0_VF1	7
0	6
PI0_PM0	5
PI0_PM0_VF1	5
PM0_PM1	5
PM0_VM1	5
VIO_PM0_PF0_VF1	5
VIO_PM0_PI1_VF1	5
PI0_PF0_VM1	4
PI0_PI1_VF1	4
PI0_PM0_VF0	4
PM0_PI1_VF1	4
VIO_PM0_PM1	4
VM0_PF0	4
PI0_PF0	3
PI0_PF1	3
PI0_PI1_VM1	3
PI0_VF1	3
PI0_VI1_VM1	3
PM0_PF0	3
PM0_PI1_VM1	3
PM0_PI1_VM1_VF1	3
VIO_PF0_PI1_VF1	3
VIO_PM0_VI1	3
VIO_PM1_VF1	3
VIO_VI1	3
VIO_VM0_PI1	3
VM0_PI1_VF1	3
PI0_PF0_VF1	2
PI0_PM0_VI1	2
PI0_VI1_VF1	2
PM0_PF0_VM1	2

PM0_VF0	2
PM0_VI1	2
PM0_VI1_VF1	2
PM0_VI1_VM1	2
VIO_PF0_PI1	2
VIO_PF0_VI1_PM1	2
VIO_PI1_VF1	2
VIO_PM0_PF0	2
VIO_PM0_PF1	2
VIO_PM0_VF0_VF1	2
VIO_PM0_VI1_VF1	2
VIO_PM0_VM1_VF1	2
VIO_VI1_PM1	2
V-I0_VM0_VI1_PM1	2
VIO_VM1	2
VM0_PF0_VF1	2
VM0_VM1	2
P+I0_P+I1	1
P+I0_P+M0_PI1	1
P+I0_PM0_PF0_VF1	1
P+I0_PM0_PIO_VM1	1
P+I0_PM0_PI1	1
P+M0_P+I1_VF1	1
PIO_PF0_VI1_VF1	1
PIO_PI1_VI1_VM1_V-F1	1
PIO_PI1_VM1_VF1	1
PIO_PM0_VF0_VF1	1
PIO_PM0_VF0_VI1_PM1	1
PIO_PM0_VF0_VM1_PF1_PM1_P+F1	1
PIO_PM0_VI1_PF1	1
PIO_PM0_VM1	1
PIO_PM0_VM1_VF1	1

PI0_PM1_VF1	1
PI0_VF0	1
PI0_VF0_VF1	1
PI0_VF0_VI1_PM1	1
PI0_VM0	1
PI0_VM0_VF1	1
PI0_VM0_VI1_PM1	1
PM0_PF0_VI1	1
PM0_PF0_VI1_PM1_VF1	1
PM0_PF1	1
PM0_VF0_PI1_VF1	1
VIO_P+M0_P+I1_VF1	1
VIO_P+M0_PI1	1
VIO_PF0_PF1	1
VIO_PF0_VF1	1
VIO_PF0_VM1	1
VIO_PI0_PM0_VF0	1
VIO_PI0_PM0_VF0_VF1	1
VIO_PI1_PM1_VF1	1
VIO_PI1_VM1	1
VIO_PI1_VM1_VF1	1
VIO_PM0_PF0_VI1_VF1	1
VIO_PM0_PF0_VM1_VF1	1
VIO_PM0_PI0_VM0	1
VIO_PM0_PI1_VM1	1
VIO_PM0_VF0	1
VIO_PM0_VF0_PI1	1
VIO_PM0_VF0_PI1_PF1	1
VIO_PM1_PF1	1
VIO_VI1_PM1_P+F1	1
VIO_VM0_PM1_VF1	1
VII	1

	VM0_PM1	1
	VM0_VF0	1
	VM0_VF0_PI1	1
	VM0_VF0_PM1	1
	VM0_VI1_PM1	1
AFECTO	PI0_PI1_VF1	1
	PI0_VM1	1
	VIO_PI1	1
	VIO_PM0	1
	VIO_VF1	1
	VM0_VM1	1
ALEGRÍA	VM0_PI1	1
	VM0_VF1	1
ALTERACIÓN	VIO_PI1	1
	VM0	1
BURLA	PM0_PI1_VM1	1
	VIO	1
	VIO_PM0_PM1	1
	VIO_VI1_PM1	1
BURLA-RECHAZO	PI0	1
	PI0_PM1	1
	PI0_VM0	1
DECEPCIÓN	PI0_VF0_PI1	1
	PI0_VF1	1
	VIO_PI1_VM1	1
	VIO_VF1	1
	VIO_VI1_PM1	1
DIVERSIÓN	VIO_PM0	2
	PI0_PI1	1
	VIO_VM0_PI1_VF1	1
DUDA	VMO_VM1	1
ENFADO	VIO_PI1	1

	VIO_VF1	1
ENFADO- RECHAZO	VIO	1
	VIO_PI1	1
	VIO_PM0_PI1_VF1	1
	VIO_PM1	1
	VIO_VM1	1
INTERÉS	PI0_VI1	1
ORGULLO	PI0	1
	PI0_PFO_VM1	1
	PI0_VF0	1
	PM0_PI1	1
	VIO_PM0_PM1	1
	VM0	1
	VM0_PFO_PI1	1
RECHAZO	PI0_VF0	1
	PM0_VF1	1
	VIO_PM0_PI1_VM1	1
RESIGNACIÓN	VIO_PM0	2
	VM0_PI1	2
	PI0	1
	PI0_PFO_VF1	1
	PM0_VI1_PM1	1
	VIO	1
	VIO_PI1	1
TRISTEZA	VIO_PI1	1
	VIO_PM0_PI1_P+M1_PF1	1
	VIO_PM0_VI1_PM1	1
DESPEDIDA	VIO_PI1_PF1	1
	VIO_PM0_VF1	1
	VIO_VF1	1
SALUDO	VIO_PM0	3
	VIO_P+I1_VF1	1

	VIO_PI1	1
	VIO_PM0_PI1_VF1	1
	VIO_PM0_VF1	1
	VIO_PM1_VF1	1
	VIO_VI1_PF1	1
	VM0_PI1	1

C.2.3. Tres sílabas

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VIO_PI1	17
	VM0_PI1	14
	VIO_VF2	12
	PI0_PM1	10
	VIO_PM0	10
	PI0_PI1_VF2	9
	PM0_PI1_VI2	9
	PI0_PI1_VI2	6
	PI0_VI1	6
	VIO_PI1_VM2	6
	VM0_PI1_VF2	6
	0	5
	PI0	5
	PI0_VI1_PM1	5
	VIO_PI1_VF2	5
	PI0_VF2	4
	VIO_PF0	4
	VIO_PM0_PI1_VM2	4
	VIO_PM0_PM1_VI2	4
	VIO_PM1_VI2	4

VI0_VI1_PM1_VM2	4
VM0_PI1_VM2	4
PI0_PI1	3
PI0_PM1_VF2	3
PI0_VF0	3
PI0_VI1_VF2	3
PI0_VM2	3
PM0_PI1	3
PM0_PI1_VI2_VF2	3
PM0_PI1_VM2	3
VI0_PI1_VI2	3
VI0_PM1	3
VI0_PM1_VF2	3
VI0_PM1_VM2	3
VI0_VM0_PI1	3
VM0	3
VM0_PI1_VI2	3
VM0_PI1_VI2_VF2	3
VM0_PM1_VI2	3
PI0_PI1_VI2_VF2	2
PI0_PI1_VM2_VF2	2
PI0_PI2_VF2	2
PI0_PM0_VF2	2
PI0_PM0_VI1	2
PI0_PM1_VI2	2
PI0_PM1_VM2	2
PI0_VF0_VF2	2
PI0_VF1	2
PI0_VI1_PM1_VI2	2
PI0_VI1_VI2	2
PI0_VI2	2
PI0_VI2_VM2	2

PM0_PI1_VF2	2
PM0_PM1_VF2	2
PM0_VM1	2
VI0	2
VI0_PM0_PI1	2
VI0_PM0_PI1_VF2	2
VI0_PM0_PI2_VM2	2
VI0_PM0_PM1	2
VI0_PM0_PM1_VF1	2
VI0_PM0_VI1_PM1	2
VI0_PM0_VI1_VF2	2
VI0_PM0_VI2_VF2	2
VI0_VF0_PI1_PM1_VI2_VM2	2
VI0_VI1_PF1_VI2	2
VI0_VI1_PM1	2
VI0_VI1_PM1_PI2	2
VI0_VI1_PM1_VI2	2
VM0_PI1_PM1_VI2	2
VM0_PM1	2
VM0_PM1_PI2	2
VM0_VF2	2
P+I0_P+I1_PF2	1
P+I0_P+I1_PM1_VF1_VM2	1
P+I0_P+M1_VI2	1
P+I0_PM0	1
P+I0_PM0_PI1_VM2	1
P+I0_VI1_VM2	1
P+M0_PI1	1
PI0_P+M1_PF2	1
PI0_PFO_VI1_VF2	1
PI0_PFO_VI2	1
PI0_PFO_VI2_VF2	1

PI0_PI1_VF1	1
PI0_PI1_VF1_VF2	1
PI0_PI1_VM1_VF2	1
PI0_PI1_VM2	1
PI0_PI2	1
PI0_PM0	1
PI0_PM0_VF0	1
PI0_PM0_VF0_PI1_PF1	1
PI0_PM0_VF0_PM1_PI2	1
PI0_PM0_VF0_VF2	1
PI0_PM0_VF1	1
PI0_PM0_VI1_PM1_PM2	1
PI0_PM0_VI2	1
PI0_PM0_VI2_VF2	1
PI0_PM1_VF1	1
PI0_PM1_VI2_VF2	1
PI0_PM1_VI2_VM2	1
PI0_PM1_VM2_VF2	1
PI0_VF0_PI1	1
PI0_VF0_PI1_VM2	1
PI0_VF0_PI2	1
PI0_VF1_VM2	1
PI0_VI1_PM1_VF2	1
PI0_VM1	1
PI0_VM1_VI2_PM2_VF2	1
PM0	1
PM0_P+I1_PI2	1
PM0_PFO_VF2	1
PM0_PFO_VI1_VF2	1
PM0_PFO_VM1_VM2	1
PM0_PF1	1
PM0_PF1_VI2	1

PM0_PI1_VF1_VF2	1
PM0_PI2	1
PM0_PI2_VF2	1
PM0_PM1	1
PM0_PM1_VI2	1
PM0_PM1_VI2_VM2	1
PM0_PM2	1
PM0_VF0_VF2	1
PM0_VF0_VM2	1
PM0_VI2	1
PM0_VI2_VM2	1
VI0_P+M0_PI1_VF2	1
VI0_P+M0_PM1_VM2_VF2	1
VI0_PF0_PI1	1
VI0_PF0_PI1_VF1_PI2	1
VI0_PF0_VI1	1
VI0_PF0_VI1_PM1_PI2	1
VI0_PF0_VI1_VF2	1
VI0_PIO_PM0_P+F0_PI1_VM1	1
VI0_PI1_PI2_VM2	1
VI0_PI1_PM1_VI2_VF2	1
VI0_PI1_VF1	1
VI0_PI1_VI2_VF2	1
VI0_PI1_VI2_VM2	1
VI0_PI2	1
VI0_PM0_PF0_VF1	1
VI0_PM0_PF0_VF2	1
VI0_PM0_PF0_VI1_PM1_PI2	1
VI0_PM0_PF0_VI2_VF2	1
VI0_PM0_PF1	1
VI0_PM0_PF1_VM2_VF2	1
VI0_PM0_PIO_VI2	1

VI0_PM0_PI1_VF1_VF2	1
VI0_PM0_PI1_VI2	1
VI0_PM0_PI1_VI2_PM2	1
VI0_PM0_PI1_VM1	1
VI0_PM0_PI1_VM1_VF2	1
VI0_PM0_PI2	1
VI0_PM0_PI2_VF2	1
VI0_PM0_PM1_VI2_PM2	1
VI0_PM0_PM1_VM2	1
VI0_PM0_VF0	1
VI0_PM0_VF0_PI1	1
VI0_PM0_VF0_PM1	1
VI0_PM0_VF0_PM1_PI2	1
VI0_PM0_VF2	1
VI0_PM0_VI1	1
VI0_PM0_VI1_PM1_VF2	1
VI0_PM0_VI1_PM1_VI2	1
VI0_PM0_VI2	1
VI0_PM0_VM1_PM2	1
VI0_PM0_VM1_VF2	1
VI0_PM0_VM2	1
VI0_PM1_PF2	1
VI0_PM1_PI2_VF2	1
VI0_PM1_VF1	1
VI0_PM1_VI2_VF2	1
VI0_PM1_VI2_VM2	1
VI0_PM1_VM2_VF2	1
VI0_PM2	1
VI0_VF0_PI1	1
VI0_VF0_PM1	1
VI0_VF0_PM1_PM2	1
VI0_VI1_PM1_P+F1_P+M1_PM2_VF2	1

	VI0_VI1_PM1_PF1_VF2	1
	VI0_VI1_PM1_VF2	1
	VI0_VM0_PF0	1
	V-	1
	I0_VM0_PF0_PI1_P+M1_P+I1_PM1_VM2_VF2	1
	V-I0_VM0_PF0_PM1_VM2	1
	VI0_VM0_PM1_PI2_VF2	1
	VI0_VM2	1
	VM0_PF0	1
	VM0_PF0_PI1_VM2	1
	VM0_PF0_VF1_PI2	1
	VM0_PF0_VI1_PF1	1
	VM0_PF1_VI2_PM2	1
	VM0_PF1_VM2	1
	VM0_PI1_PF1_VI2_VF2	1
	VM0_PI1_PF1_VM2	1
	VM0_PI1_PM1	1
	VM0_PI1_PM1_VI2_PM2	1
	VM0_PI1_PM1_VM2	1
	VM0_PM1_VI2_VF2	1
	VM0_PM1_VI2_VM2	1
	VM0_PM1_VM2	1
	VM0_PM1_VM2_VF2	1
	VM0_VF0_PM1_VM2	1
	VM0_VF0_PM1_VM2_VF2	1
	VM0_VI1_PM1	1
ADMIRACIÓN	VM0_PI1_PM1_VF2	1
AFECTO	PM0	1
	VI0_PM0_VI1	1
	VM0_VF0_PI1	1
ALEGRÍA	VI0_VI1_PM1_PI2_VM2	1

	VM0_VI2_PM2_VF2	1
ALTERACIÓN	VI0_PM0_VM1	1
	VI0_VM2	1
BURLA	VI0_VF0_PM1	1
	VM0_PI1_VM2	1
BURLA- RECHAZO	PI0_VF0	1
DIVERSIÓN	VI0_PF0_VF2	1
	VI0_PM0_PM1	1
DUDA	VI0_PI1_VI2_VF2	1
	VI0_PM0_VI1_VF2	1
	VI0_PM1_VF2	1
	VI0_VM0_PI2	1
	VM0_PM1_PM2	1
ORGULLO	PI0_PM0_VF0_PF2	1
RECHAZO	VI0_VF1	1
RESIGNACIÓN	VI0	1
	VI0_VF2	1
	VM0_VF0_PI1_VI2	1
TRISTEZA	PI0_VM0	1
DESPEDIDA	PM0_VM1	1
	VI0_PF0_VM2	1
SALUDO	VI0_PM0_PM1_VF1	1
	VM0_PI2	1
	VM0_PM1	1
	VM0_VF0_PM2	1

C.3. Patrones interiores

A continuación se muestran todos los patrones interiores tanto del corpus neutro como del corpus emocional de los grupos entonativos

más frecuentes, los enunciativos formados por una, dos o tres sílabas. En el caso del corpus emocional, los patrones aparecen clasificados según las diferentes emociones representadas.

C.3.1. Una sílaba

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VI0_PM0	169
	0	149
	PM0	107
	PI0	95
	VF0	47
	VI0	28
	PF0	27
	VI0_VF0	22
	VI0_PM0_VF0	20
	PI0_PFO	17
	PM0_VF0	17
	PI0_VF0	13
	VI0_PM0_PFO	13
	VM0	12
	PI0_VM0	10
	VI0_PFO	10
	PI0_VM0_VF0	9
	PI0_PM0	8
	PI0_P+M0	4
	PM0_PFO	4
	VI0_VM0	3
	PI0_PM0_VF0	2
	VM0_V-F0	2
	P+F0	1
	P+I0_PM0_PI0_VF0	1

	P+IO_PM0_VF0	1
	P+M0	1
	PIO_P+M0_PFO	1
	PIO_VM0_PFO	1
	VI0_P+M0	1
	VI0_P+M0_P+F0	1
	VI0_PIO_P+M0_PFO	1
	VI0_PIO_PM0_VF0	1
	VI0_PIO_VM0_V-F0	1
	VI0_PM0_P+F0	1
	VI0_VM0_PFO	1
	VI0_VM0_V-F0	1
	VM0_PFO	1
ADMIRACIÓN	VI0_PM0	2
AFECTO	PIO	2
	0	1
	PFO	1
	PIO_VM0	1
	PIO_VM0_PFO	1
	VI0_PFO	1
ALEGRÍA	0	3
	VI0_PM0	3
	PIO	1
	PIO_PM0	1
ALTERACIÓN	VM0	2
	PIO	1
	PM0	1
	VI0	1
	VI0_PFO	1
BURLA	0	4
	PIO	3
	PM0	1

	VFO	1
	VI0_PFO	1
	VI0_PM0	1
BURLA-RECHAZO	0	1
	PI0	1
DECEPCIÓN	0	6
	PI0	4
	PM0	1
	VFO	1
	VI0_PM0	1
DIVERSIÓN	0	2
	PFO	2
	PI0	1
	VFO	1
DUDA	0	2
	PI0	1
	PM0	1
	PM0_VFO	1
	VFO	1
	VI0	1
	VI0_PFO	1
ENFADO	0	6
	PI0	3
	PM0	1
	VI0	1
	VI0_PM0	1
	VI0_PM0_VFO	1
ENFADO-RECHAZO	0	3
	PFO	2
	PM0	2
	VI0_PM0	2
	VM0	2

	PI0	1
	PI0_PM0_VF0	1
EXTRAÑEZA	VF0	1
	VI0_PM0	1
ORGULLO	0	5
	PI0	2
	VI0	2
	PI0_VM0_VF0	1
	PM0	1
	PM0_VF0	1
RECHAZO	0	4
	PI0_P+M0_PFO	1
	VI0	1
	VI0_PFO	1
RESIGNACIÓN	0	1
	PI0_VF0	1
	PM0	1
	PM0_VF0	1
	VI0_PM0	1
TRISTEZA	VF0	2
	0	1
	PFO	1
	PM0	1
	VI0	1
	VI0_PM0	1
	VI0_PM0_VF0	1
	VM0	1
SALUDO	PI0	1
	VI0_PM0	1

C.3.2. Dos sílabas

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VI0_PM0	226
	0	188
	PI1	135
	PM0	126
	PI0	114
	VI0_PM0_VF1	59
	PF0	57
	VI0_PM0_VM1	49
	VI1	49
	VI0_PI1	45
	VI0_PM0_PI1	42
	PM0_VF1	37
	VF1	37
	PI0_VM1	35
	PI0_VI1	32
	PM1	29
	PM0_PI1	28
	VI0	28
	VM1	28
	VI0_PFO	25
	PI0_VF1	23
	PI1_VF1	22
	PM0_VM1	22
	VI0_PM0_VI1	21
	VI0_PM0_PM1	19
	PI0_PI1	18
	VF0	18
	VI0_PM0_PI1_VF1	16
	PF0_VF1	15

PI0_PM1	15
PM0_VI1	15
PF0_VM1	14
VM0	13
VIO_PM0_PFO	12
PI0_VI1_VF1	11
PM0_PM1	11
VIO_PM0_PFO	11
PI0_VF0	10
PI0_VM1_VF1	10
PM0_PI1_VF1	10
VM0_VF1	10
PI0_PFO	9
PI0_VM0	9
VIO_PM0_VI1_VF1	9
VIO_PM1	9
VIO_VF1	9
VIO_VM1	9
PF1	8
PI0_PI1_VF1	8
PI1_VM1	8
VIO_PFO_PI1	8
VIO_PI1_VF1	8
VIO_PM0_PFO_VM1	8
VIO_PM0_VF0	8
PI0_PM0	7
PI0_PM0_VI1	7
PI0_VF0_VF1	7
PM0_VF0	7
VFO_VF1	7
VIO_PFO_VM1	7
VIO_PM0_PI1_VM1	7

VI1_VF1	7
PM0_PF1	6
PM0_VI1_VF1	6
VIO_PF0_VF1	6
PF0_PM1	5
PF0_VI1	5
PIO_PF0_VF1	5
VIO_PM0_PF0_VF1	5
VIO_PM0_PF0_VI1	5
VIO_PM0_VM1_VF1	5
VI1_PM1	5
VI1_VM1	5
VM0_PI1	5
PIO_PM0_VM1	4
PIO_VI1_VM1	4
PM1_VF1	4
VIO_PF0_VI1	4
VM0_VM1	4
P+M0_PI1	3
PF0_PI1	3
PIO_PF0_VI1	3
PIO_PF1	3
PIO_PM0_VF1	3
PIO_VI1_PM1	3
PM0_PF0_VF1	3
PM0_PI1_VM1	3
PM0_VI1_PM1	3
PM0_VI1_VM1	3
PM0_VM1_VF1	3
VF0_VM1	3
VIO_PI1_PM1	3
VIO_PM0_PF0_VI1_VF1	3

VI0_VI1_PM1	3
P+M0	2
PF0_PF1	2
PF0_PI1_VF1	2
PIO_P+F0_PI1	2
PIO_P+M0_PI1	2
PIO_PF0_VM1	2
PIO_PF0_VM1_VF1	2
PIO_PI1_VM1	2
PIO_PM0_VF0	2
PIO_PM0_VM1_VF1	2
PI1_PF1	2
PI1_PM1_VF1	2
PI1_VM1_VF1	2
PM0_PF0	2
PM0_PF0_VI1	2
PM0_VF0_VF1	2
VF0_PI1	2
VI0_PF0_PI1_VF1	2
VI0_PM0_PF0_VM1_VF1	2
VI0_PM0_PM1_VF1	2
VI0_VF0	2
VI0_VF0_PI1	2
VI0_VM0_PI1	2
VM0_PF1	2
VM1_VF1	2
P+I0	1
P+I0_PM0	1
P+I0_PM0_PF0_VM1	1
P+I0_PM0_PI1_VM1_VF1	1
P+I0_VF0_VF1	1
P+I0_VM0	1

P+M0_P+F0_PF1	1
P+M0_PI1_PM1	1
P+M0_VM1	1
P+M1	1
PF0_PI1_P+M1	1
PF0_VI1_VF1	1
PF0_VI1_VM1_PF1	1
PF0_VM1_VF1	1
PIO_P+F0_PF1	1
PIO_P+IO_PM0_VF0_VF1	1
PIO_P+I1_PF1	1
PIO_P+M0	1
PIO_P+M0_P+F0_VI1	1
PIO_P+M0_P+IO_PM0_PF0_VM1_VF1	1
PIO_P+M0_P+IO_PM0_PIO_VM0_VF1	1
PIO_P+M0_P+IO_PM0_VF0	1
PIO_P+M0_PM1	1
PIO_PF0_VI1_PM1	1
PIO_PF0_VI1_VM1	1
PIO_PI1_P+M1_PF1	1
PIO_PI1_VM1_PF1	1
PIO_PI1_VM1_VF1	1
PIO_PM0_P+I1_PF1	1
PIO_PM0_VF0_VF1	1
PIO_PM0_VI1_PM1	1
PIO_PM0_VI1_VF1	1
PIO_PM1_P+F1	1
PIO_VF0_PI1_PM1_P+F1_P+M1	1
PIO_VF0_VM1	1
PIO_VI1_PM1_PF1	1
PIO_VI1_VM1_PF1	1
PIO_VM0_PF0	1

PI0_VM0_VF1	1
PI1_P+M1_P+I1_PM1	1
PI1_PM1	1
PM0_PF0_VM1	1
PM0_PI1_VM1_VF1	1
PM0_VF0_PM1	1
PM0_VF0_VI1	1
PM0_VF0_VM1	1
PM1_PF1	1
V-I0	1
VI0_P+F0_P+I1_PM1	1
VI0_P+F0_VI1	1
VI0_P+M0_PF1	1
VI0_PF0_PF1	1
VI0_PF0_PI1_P+M1	1
VI0_PF0_PM1	1
VI0_PF0_VI1_PM1	1
VI0_PF0_VI1_PM1_VF1	1
VI0_PI0_PM0_VF0_VF1	1
VI0_PI1_PF1	1
VI0_PI1_PM1_VF1	1
VI0_PI1_VM1	1
VI0_PM0_P+F0_PI1	1
VI0_PM0_P+F0_PM1	1
VI0_PM0_P+F0_VM1	1
VI0_PM0_P+I0_PM0_VF0	1
VI0_PM0_P+I0_PM0_VI1	1
VI0_PM0_PI0_VM0_V-F0_PI1	1
VI0_PM0_PI1_VM1_VF1	1
VI0_PM0_VF0_PI1_VF1	1
VI0_PM0_VF0_PM1	1
VI0_PM0_VF0_VF1	1

	V-I0_PM0_VF1	1
	VI0_PM0_VI1_PM1	1
	VI0_PM0_VI1_VM1	1
	VI0_PM1_VF1	1
	VI0_V-F0_VM1	1
	VI0_VI1	1
	VI0_VI1_V-F1	1
	VI0_V-I1_VF1	1
	V-I0_VM0	1
	V-I0_VM0_PIO_PM1	1
	VI0_VM0_PII_VM1_VF1	1
	VI1_VM1_PF1	1
	VI1_VM1_V-F1	1
	VM0_PFO	1
	VM0_PFO_PII_VF1	1
	VM0_PFO_PM1	1
	VM0_PII_PF1	1
	VM0_PII_VF1	1
	VM0_PM1	1
	VM0_VF0_PII	1
	VM0_VF0_PII_PF1	1
	V-M0_VF1	1
	VM0_V-I1_VF1	1
ADMIRACIÓN	PI1	1
	PI1_VF1	1
	PM0_VI1_PM1	1
	VI0_PFO	1
	VI0_PII	1
	VI0_PM0_VI1_VF1	1
AFEECTO	0	6
	PFO	2
	PI1	2

	VIO_PM0 PIO_PI1 PIO_VF0_PI1 PM0_PI1_P+M1_P+F1 PM0_VI1 VF1 VIO_PI1 VIO_PM1_VF1 VI1	2 1 1 1 1 1 1 1 1
ALEGRÍA	PI1 PM0_PF1 VIO VIO_PM0 PIO_VF1 PIO_VI1_VF1 PM0_PM1 PM0_VF1 PM1 VF1 VIO_PF0_VF1 VIO_PF1 VIO_PI1 VIO_PI1_VF1 VIO_PM0_P+F0 VI1	5 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1
ALTERACIÓN	PM0 0 PF0 PF0_VF1 PI1 VIO_PI1 VIO_PM0_PF0	4 1 1 1 1 1 1

	VI0_PM0_VI1	1
	VI0_PM0_VM1_VF1	1
	VM0_PI1_PM1	1
	VM0_VM1	1
BURLA	0	4
	PI1	3
	VF1	2
	VI0_PF0	2
	VI0_PM0	2
	PF0	1
	PF0_VF1	1
	PF1	1
	PI0_VI1_VM1	1
	PI1_VF1	1
	PM0	1
	PM0_PM1	1
	VF0_PI1	1
	VI0	1
	VI0_PI1	1
	VI0_PM0_PI1	1
	VI0_PM0_PI1_VF1	1
	VI0_PM0_VF0_PI1_VM1	1
	VI0_VI1	1
	VM0	1
	VM0_PI1_VM1	1
BURLA- RECHAZO	0	3
	P+I0_PM1	1
	PI0	1
	PI0_P+F0_PF1	1
	PI0_VF1	1
	PI1	1
	PM0_VF0	1

	PM1	1
	VF0	1
	VF1	1
	VI0	1
	VI0_PF0	1
	VI0_PI1_VF1	1
	VI1_PM1	1
DECEPCIÓN	0	2
	PM0_VI1	2
	PM1	2
	VF1	2
	VI0	2
	VI0_PI1	2
	VI0_PM0	2
	VI1	2
	PI0_VF0	1
	PI0_VF1	1
	PI0_VI1	1
	PI0_VM1	1
	PI1	1
	PM0_VM1	1
	VI0_PF0	1
	VI0_PM0_PM1	1
	VI0_PM1	1
	VI1_PM1	1
DIVERSIÓN	PF0	1
	PM1	1
	VI0_PI1	1
	VI0_VI1_PM1_VF1	1
	VM0	1
DUDA	0	2
	PI0	2

	PI0_VF0_PM1	1
	PI0_VM1	1
	PI1	1
	PM0_VF1	1
	VF1	1
	VI0_PI1_PM1	1
	VI1	1
	VI1_PF1_P+I2_P+M2_PF2	1
ENFADO	0	3
	PI1	3
	PF0	2
	VI0_PM0	2
	PI0_VF1	1
	PI1_VF1	1
	PM0	1
	PM0_VI1	1
	VI0	1
	VI0_P+I1	1
	VI0_P+M1	1
ENFADO- RECHAZO	0	5
	VI0_PI1	3
	PF0	1
	PI0	1
	PI0_VF1	1
	PI1	1
	PM0_P+I1	1
	PM0_VM1	1
	VI0_PM0	1
	VI0_VF1	1
	VM0_VF1	1
EXTRAÑEZA	PI0_PM0	1
	VI0	1

ORGULLO	0	4
	PI0	2
	PI0_VI1	2
	VI0_PI1	2
	PI0_VM0_PI1	1
	PI1	1
	PI1_VF1	1
	PM0_VI1_PM1	1
	VI0_PM0	1
	VI0_PM0_PFO_VF1	1
	VI0_PM0_VFO_VF1	1
	VI0_PM1	1
	RECHAZO	PI0
PI0_PI1		2
PI1		2
VF1		2
0		1
PFO		1
PI0_VI1_VM1		1
PI0_VM1		1
PM0		1
PM0_PI1		1
PM0_VF1		1
PM0_VI1		1
PM1		1
PM1_VF1		1
VI0		1
VI0_PM0		1
VI0_PM0_P+M1		1
VI0_PM0_PI1		1
VI0_PM0_PI1_VF1		1
VI0_PM0_PM1_VF1		1

	VI0_PM0_VM1_VF1	1
	VI0_PM1	1
	VI1_VM1	1
	VM1	1
RESIGNACIÓN	0	3
	PM0_VI1	2
	VI0_PM0	2
	PI0	1
	PI0_VF1	1
	PI1	1
	PM0_VI1_PM1_VF1	1
	VF0_PI1	1
	VF0_PI1_VF1	1
	VM1	1
SORPRESA	PM0_VM1	1
TRISTEZA	0	6
	PI0	2
	PI1	2
	PM0	2
	VF1	2
	PI0_PI1	1
	PM0_VF1	1
	PM0_VI1_PM1	1
	PM1	1
	VI0_PFO	1
	VI0_PI1	1
	VI0_PM0	1
	VI0_PM0_VF0_PM1	1
	VI0_PM0_VM1	1
DESPEDIDA	PM0_PI1	1
	PM0_PI1_VM1_VF1	1
	PM0_VI1_VF1	1

	VI1_VF1	1
SALUDO	PI0_VF0_VF1	1
	PI0_VI1	1
	VI0_PM0	1
	VI0_PM0_P11_VM1	1
	VI1_PM1	1

C.3.3. Tres sílabas

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VI0_PM0	103
	0	96
	PI1	73
	PM0	62
	PI0	61
	PI1_VF2	35
	VI0_PM0_PI2	33
	VI0_PM0_P11_VF2	32
	VI0_PM0_P11_VM2	32
	VI0_P11	31
	PM0_PI2	30
	VF2	30
	PI1_VI2	29
	PI1_VM2	28
	VI0_PM0_P11	27
	PI0_PI2	26
	PI2	25
	VI0_PM0_P11_VI2	24
	PI0_P11	19

PM0_PI1	19
PM0_VF2	19
VM2	19
PI0_PI1_VF2	18
PM0_PI1_VM2	18
PI0_PI1_VI2	17
PI0_PI1_VM2	17
PI1_VI2_VF2	17
PM1	17
VI0_PI1_VF2	17
VI0_PM0_PI2_VF2	17
PM0_PI1_VI2	16
VI0_PFO	16
VI0_PM0_PM1_VF2	16
PFO	15
PI0_PM2	15
VI0_PM0_VI2	15
VFO_VF2	14
VI0_PM0_PM1	14
VI0_PM0_PM1_VM2	14
VI0_VF2	14
PM0_VM2	13
PM2	13
VFO	13
VI0_PM0_PFO	13
VI0_PM0_PI1_VI2_VF2	13
VI0_PM0_PM2	13
PM0_PI1_VF2	12
PI0_VF2	11
PI1_PI2	11
VI0_PM0_VF2	11
PI0_VI1	10

PI0_VI2	10
VI0	10
PF0_VM2	9
PI0_PI1_VI2_VF2	9
PI0_PM1	9
PI0_PM1_VF2	9
PI0_VF1	9
PI0_VI1_VF2	9
PM0_PM2	9
VF1	9
VI0_PM0_PI1_VM2_VF2	9
VI0_PM0_VM2	9
VI2	9
PF0_VI2	8
PI0_VF0	8
PI0_VI2_VF2	8
PI1_VF1	8
PI2_VF2	8
PM0_PI1_VI2_VF2	8
PM0_VF1	8
PM0_VI1	8
PM0_VI1_VF2	8
PM0_VI2	8
PM1_VF2	8
VI0_VM2	8
PF0_VF2	7
PI1_VM2_VF2	7
PM0_PF2	7
VI0_PI1_PI2	7
VI0_PI1_VM2	7
VI0_PM0_PM1_VI2	7
VI0_PM0_VI1_VF2	7

VI1	7
VI1_VF2	7
PI0_VM2	6
PM0_PF1	6
PM0_PI1_VF1_VF2	6
PM0_PM1	6
PM1_VI2	6
VI0_PF0_PI2	6
VI0_PM0_PI1_VF1_VF2	6
VI0_PM0_PI2_VM2	6
VI0_PM0_VF1	6
VI0_PM0_VI1	6
PF0_PI1	5
PF0_PI1_VF2	5
PI0_PI1_VF1_VF2	5
PI0_PM0	5
PI0_VF1_VF2	5
PI0_VI1_VM2	5
PM0_PI1_VM2_VF2	5
PM0_PI2_VM2	5
VI0_PI1_VI2_VF2	5
VI0_PM0_PF0_VM2	5
VI0_PM0_VF0	5
VI0_PM0_VI1_VM2	5
VI0_PM0_VI2_VF2	5
VI0_PM1_VM2	5
VI1_PM1	5
PI0_PF0_VM2	4
PI0_PF1_VF2	4
PI0_PI1_VF1	4
PI0_PM0_VI1_VF2	4
PI0_PM1_VI2	4

PI0_VM0	4
PI0_VM1	4
PI1_PF2	4
PI1_VI2_PM2	4
PI1_VI2_VM2	4
PM0_PI1_VF1	4
PM0_PI2_VF2	4
PM0_PM1_VM2	4
PM0_VI1_PM1	4
PM0_VM1	4
PM1_VM2	4
VI0_PF0_PI1_VI2_VF2	4
VI0_PF0_PM1	4
VI0_PF0_PM2	4
VI0_PF0_VF1	4
VI0_PF0_VM2	4
VI0_PI1_VM2_VF2	4
VI0_PM0_PF0	4
VI0_PM0_PF0_VF1_VF2	4
VI0_PM0_PF1	4
VI0_PM0_PM1_VF1	4
VI0_PM0_VF0_VF2	4
VI0_PM0_VM2_VF2	4
PF0_VI1_VF2	3
PF1	3
PI0_PF0	3
PI0_PI1_VM2_VF2	3
PI0_PI2_VF2	3
PI0_PI2_VM2	3
PI0_PM0_VF1	3
PI0_PM0_VI1	3
PI0_PM0_VI2	3

PI0_PM0_VI2_VF2	3
PI0_PM0_VM2	3
PI0_PM1_VI2_PM2	3
PI0_PM1_VI2_VF2	3
PI0_VF0_VF2	3
PI1_PI2_VM2	3
PM0_PF1_VF2	3
PM0_PI1_VI2_PM2	3
PM0_PM1_VI2	3
PM0_VF0	3
PM0_VF1_VF2	3
PM0_VI1_PM1_VM2	3
PM0_VI1_VM2	3
PM0_VI2_VF2	3
PM0_VM1_VF2	3
VF0_PI1	3
VI0_PF0_PI1	3
VI0_PF0_PI1_VF2	3
VI0_PF0_VF2	3
VI0_PI1_VI2	3
VI0_PM0_P+I1	3
VI0_PM0_PF0_VI2	3
VI0_PM0_PF0_VI2_VF2	3
VI0_PM0_PM1_VI2_PM2	3
VI0_PM0_PM1_VI2_VF2	3
VI0_PM0_PM1_VM2_VF2	3
VI0_PM0_PM2_VF2	3
VI0_PM1	3
VI0_PM1_VF2	3
VI2_PM2	3
P+I1_PM2	2
P+I1_VF2	2

	PF0_PI1_VI2	2
	PF0_PI1_VM2	2
	PF0_PI1_VM2_VF2	2
	PF0_PI2	2
	PF0_VF1	2
	PF1_VM2	2
	PF2	2
	PI0_P+M0_PI2	2
	PI0_PFO_VF1_VF2	2
	PI0_PFO_VI2_VF2	2
	PI0_PF1_VM2	2
	PI0_PM0_VF0	2
	PI0_PM1_VF1_VF2	2
	PI0_PM1_VM2	2
	PI0_PM1_VM2_VF2	2
	PI0_VF0_PI1	2
	PI0_VF0_VM2	2
	PI0_VI1_PM1_VI2	2
	PI0_VI1_VF1	2
	PI0_VI2_VM2	2
	PI1_PI2_VF2	2
	PI1_PM1_VM2	2
	PI1_VF1_VF2	2
	PI1_VF1_VM2	2
	PI2_VM2	2
	PM0_PFO_VF1	2
	PM0_PFO_VI2	2
	PM0_PI1_VF1_VM2	2
	PM0_PI1_VI2_VM2	2
	PM0_PM1_VF1	2
	PM0_PM1_VF2	2
	PM0_VF0_VF2	2

PM1_P12_VF2	2
PM2_VF2	2
VF0_VM2	2
VF1_VF2	2
VF1_VM2	2
VI0_PF0_PF1_VM2	2
VI0_PF0_PF2	2
VI0_PF0_P12_VF2	2
VI0_PF0_PM1_VI2	2
VI0_PF0_VI1	2
VI0_PF0_VI2_VF2	2
VI0_PF2	2
VI0_P11_PF1	2
VI0_P11_PF1_VM2	2
VI0_P11_PM1_VF2	2
VI0_P11_PM2	2
VI0_P11_VI2_VM2	2
VI0_PM0_PF0_VF2	2
VI0_PM0_PF0_VM1_VF2	2
VI0_PM0_PF1_VI2	2
VI0_PM0_P11_VF1	2
VI0_PM0_P11_VF1_VM2	2
VI0_PM0_P11_VI2_PM2_VF2	2
VI0_PM0_P11_VI2_VM2	2
VI0_PM0_PM1_VF1_VF2	2
VI0_PM0_VI1_PM1	2
VI0_PM0_VI1_VF1	2
VI0_PM0_VM1	2
VI0_PM1_VI2	2
VI0_VI2	2
VI1_PM1_VF2	2
VI2_PF2	2

VI2_VF2	2
VM0_PI1_VM2	2
VM0_PM1_VM2	2
P+I0_PI1_VM2	1
P+I0_PM0	1
P+I0_PM0_VF0	1
P+I1_PF2	1
P+I1_PI2	1
P+I1_PM1_PI2	1
P+I1_PM1_PM2_VF2	1
P+M0_P+M1_VF2	1
P+M0_PF2	1
P+M0_PI1	1
P+M0_PI2_VM2	1
P+M0_PM1_VM2	1
P+M0_VF2	1
P+M0_VI1_VF2	1
P+M1_VF1_VF2	1
PF0_P+I1_PM1_VF1	1
PF0_P+I1_PM1_VF2	1
PF0_PF1	1
PF0_PF2	1
PF0_PI1_VF1_PM2_VF2	1
PF0_PI1_VI2_VF2	1
PF0_PI2_VF2	1
PF0_PI2_VM2	1
PF0_PI2_VM2_VF2	1
PF0_PM1	1
PF0_PM1_VF1	1
PF0_PM1_VF2	1
PF0_PM1_VM2_VF2	1
PF0_PM2	1

PF0_VF1_VF2	1
PF0_VI1	1
PF0_VI1_PM1	1
PF0_VI1_PM1_VF2	1
PF0_VI1_V-M1_PI2	1
PF0_VI2_PM2_VF2	1
PF0_VI2_VF2	1
PF0_VM1	1
PF0_VM2_VF2	1
PF1_VF2	1
PI0_P+F0_PF2	1
PI0_P+F0_PM1	1
PI0_P+I1_P+F1_PF2	1
PI0_P+I1_PI2	1
PI0_P+I1_PM1	1
PI0_P+I1_VI2_VF2	1
PI0_P+M0_P+M1_PI2	1
PI0_P+M0_PF1_VI2_VF2	1
PI0_P+M0_PM1_PM2	1
PI0_P+M0_PM2	1
PI0_P+M1_PF2	1
PI0_P+M1_PI2_VF2	1
PI0_PF0_P+M1	1
PI0_PF0_VF1	1
PI0_PF0_VF2	1
PI0_PF0_VI1_PI2	1
PI0_PF0_VI1_VF2	1
PI0_PF0_VI1_VI2	1
PI0_PF0_VI1_VM2	1
PI0_PF0_VI2	1
PI0_PF0_VI2_PM2	1
PI0_PF0_VM1	1

PI0_PF1_P+I2_PM2	1
PI0_PF1_VI2	1
PI0_PF2	1
PI0_PI1_VF1_VM2	1
PI0_PI1_VI2_PM2	1
PI0_PI1_VM1	1
PI0_PI1_VM1_PF1_VF2	1
PI0_PI1_VM1_VF2	1
PI0_PI1_VM1_VI2_PM2	1
PI0_PM0_P+I1	1
PI0_PM0_VF0_PI1_PM2	1
PI0_PM0_VF0_PI1_VF1	1
PI0_PM0_VF1_VF2	1
PI0_PM0_VI1_PM1	1
PI0_PM0_VI1_PM1_VF2	1
PI0_PM0_VI1_VM2	1
PI0_PM0_VI2_PF2	1
PI0_PM0_VI2_V-F2	1
PI0_PM0_VM1_VM2	1
PI0_PM1_VF1	1
PI0_PM1_VF1_PI2	1
PI0_PM1_VF1_VI2	1
PI0_PM1_VF1_VI2_PM2	1
PI0_PM2_VF2	1
PI0_VF0_P+I2	1
PI0_VF0_PI1_VF2	1
PI0_VF0_PI1_VI2	1
PI0_VF0_PM1_PI2	1
PI0_VF0_VI1_PI2	1
PI0_VF0_VI2_PM2	1
PI0_VF0_VM1_PF2	1
PI0_VI1_PM1	1

PI0_VI1_VF1_PM2_PF2	1
PI0_VI1_VI2	1
PI0_VI1_VI2_PM2	1
PI0_VI1_VM1_PF2	1
PI0_VI2_PF2	1
PI0_VI2_PM2	1
PI0_VM0_PI1_PI2	1
PI0_VM0_VF0_PM2	1
PI0_VM0_VF2	1
PI0_VM0_VI0_PF0	1
PI0_VM0_VI1_PI2	1
PI0_VM0_VI1_PM1_VF1_PI2	1
PI0_VM0_VM1_PI2	1
PI0_VM1_VF2	1
PI0_VM1_VI2_PM2	1
PI0_VM2_VF2	1
PI1_P+M1	1
PI1_PF1	1
PI1_PF1_VF2	1
PI1_PF1_VI2	1
PI1_PF1_VI2_VF2	1
PI1_PF1_VM2_VF2	1
PI1_PI2_VM2_VF2	1
PI1_PM1_VF1	1
PI1_PM1_VM2_VF2	1
PI1_PM2	1
PI1_VI2_PM2_VF2	1
PI1_VI2_VM2_PF2	1
PM0_P+I1_PI2	1
PM0_P+M1_PI2	1
PM0_PF0_P+I1_PM1_VF1_VF2	1
PM0_PF0_VM2	1

	PM0_PFI_VI2	1
	PM0_PFI_VI2_PM2	1
	PM0_PII_VF1_PFI	1
	PM0_PII_VF1_PII_VF2	1
	PM0_PII_VM1	1
	PM0_PII_VM1_PFI	1
	PM0_PII_VM1_VM2	1
	PM0_PII_P+M2	1
	PM0_PII_P+M2_PII_PM2_VF2	1
	PM0_PM1_VF1_VF2	1
	PM0_PM1_VI2_PM2	1
	PM0_PM1_VI2_VM2	1
	PM0_PM1_VM2_VF2	1
	PM0_PM2_VF2	1
	PM0_VF0_PII	1
	PM0_VF0_VI2_PM2	1
	PM0_VF0_VM2	1
	PM0_VF1_VM2	1
	PM0_VI1_PII	1
	PM0_VI1_PM1_VI2_VF2	1
	PM0_VI1_VM1_PII	1
	PM0_VI2_PM2_VF2	1
	PM0_VI2_VM2	1
	PM0_VM1_VM2	1
	PM0_VM2_VF2	1
	PM1_P+F1_P+M1_PII_PM2_VF2	1
	PM1_PFI	1
	PM1_PM2	1
	PM1_VF1	1
	PM1_VF1_VF2	1
	PM1_VI2_VF2	1
	PM1_VI2_VM2	1

PM1_VM2_VF2	1
VF0_PI1_VF2	1
VF0_PI1_VI2_PF2	1
VF0_PI1_VM1_VF2	1
VF0_PI1_VM2	1
VF0_PI2	1
VF0_PM1	1
VF0_PM2	1
VF0_VF1	1
VF0_V-F1	1
V-F0_VF2	1
VF0_VI2	1
VF1_PI2	1
VF1_PI2_PF2	1
VI0_P+F0_VI2_VF2	1
VI0_P+M0	1
VI0_P+M0_P+F0_PI1	1
VI0_P+M0_P+M1_VF1	1
VI0_P+M0_PF1	1
VI0_P+M0_PI1_VF2	1
VI0_P+M0_VF1_VM2	1
VI0_P+M0_VF2	1
VI0_P+M0_VM2_VF2	1
VI0_PF0_P+I1_PM1_PI2_VF2	1
VI0_PF0_P+M1_VI2_VF2	1
VI0_PF0_PI1_VM2	1
VI0_PF0_PM1_VF2	1
VI0_PF0_PM1_VM2	1
VI0_PF0_VF1_VF2	1
VI0_PF0_VI1_PM1	1
VI0_PF0_VI1_VF2	1
VI0_PF0_VI2	1

VI0_PFO_VI2_PFO	1
VI0_PFO_VM1_VF2	1
VI0_PFO	1
VI0_PII_PFO	1
VI0_PII_PII2_VF2	1
VI0_PII_PM1_VF1_VF2	1
VI0_PII_PM1_VI2	1
VI0_PII_PM2_VF2	1
VI0_PII_VF1	1
VI0_PII_VF1_VF2	1
VI0_PII	1
V-I0_PM0	1
VI0_PM0_P+F0	1
VI0_PM0_P+F0_PFO	1
VI0_PM0_P+F0_VI2	1
VI0_PM0_P+F0_VI2_VF2	1
VI0_PM0_P+I1_PII	1
VI0_PM0_P+I1_PII_VF2	1
VI0_PM0_P+I1_PM2	1
VI0_PM0_P+I1_VI2	1
VI0_PM0_P+M1_PM2_VF2	1
VI0_PM0_PFO_VF1	1
VI0_PM0_PFO_VI1	1
VI0_PM0_PFO_VI1_PFO_VI2	1
VI0_PM0_PFO_VI1_PM1_PM2	1
VI0_PM0_PFO_VI2_PM2_VF2	1
VI0_PM0_PFO_VM1_VM2	1
VI0_PM0_PFO_VM2_VF2	1
VI0_PM0_PFO_VF2	1
VI0_PM0_PFO_VI2_VF2	1
VI0_PM0_PFO_VM2	1
VI0_PM0_PIO_VFO	1

VI0_PM0_PIO_VF0_VM2	1
VI0_PM0_PIO_VM0_PI1_VM1_VF2	1
VI0_PM0_PI1_VF1_PI2_VM2	1
VI0_PM0_PI1_VI2_PM2	1
V-I0_PM0_PI1_VM2	1
VI0_PM0_PM1_P+F1_P+M1_PM2	1
VI0_PM0_PM1_P+F1_P+M2_PF2_PM2_VF2	1
VI0_PM0_PM1_VF1_VM2	1
VI0_PM0_PM1_VI2_VM2	1
VI0_PM0_VF0_PI1	1
VI0_PM0_V-F0_PI1	1
VI0_PM0_VF0_PM1_VI2_VF2	1
VI0_PM0_VF0_VI2	1
VI0_PM0_VF0_VI2_PF2	1
VI0_PM0_VF1_VF2	1
VI0_PM0_VI1_PF1_VF2	1
VI0_PM0_VI1_PF2	1
VI0_PM0_VI1_PM1_VF1	1
VI0_PM0_VI1_PM1_VF2	1
VI0_PM0_VI1_PM1_VI2	1
V-I0_PM0_VI1_PM1_VM2	1
VI0_PM0_VI1_PM2	1
VI0_PM0_VI1_VI2_PM2	1
VI0_PM0_VI1_VM1_PM2	1
VI0_PM0_VI2_PM2	1
VI0_PM0_VI2_VM2	1
VI0_PM0_VM1_PI2_VF2	1
VI0_PM0_V-M1_VI2	1
VI0_PM0_VM1_VM2	1
VI0_PM1_PI2	1
VI0_PM1_PM2	1
VI0_PM1_VF1	1

VI0_V-F0_VI2	1
VI0_VF1	1
VI0_V-I1_PM1	1
VI0_VI1_PM1_VF2	1
VI0_V-I1_VM1_PM2	1
VI0_V-I1_VM2	1
VI0_VI2_PM2_VF2	1
VI0_VM0_PF0_PI1_V-F2	1
VI0_VM0_PI1	1
VI0_VM0_PI1_VF2	1
V-I0_VM0_PI1_VI2	1
V-I0_VM0_VI1_PM1_VI2_VF2	1
VI0_VM1_PF2	1
VI0_VM2_PF2	1
VI1_PM1_PI2_VM2	1
VI1_PM1_VI2	1
VI1_PM1_VI2_VF2	1
VI1_PM1_VM2	1
VI1_VF1	1
VI1_VF1_PI2	1
VM0_PF0	1
VM0_PF0_PI2	1
VM0_PF0_VM1_VF2	1
VM0_PI1	1
VM0_PI1_VI2	1
VM0_PI1_VI2_VF2	1
VM0_VF2	1
VM0_V-I1_PF1_VM2	1
VM0_V-I1_VF1_VF2	1
VM0_V-I2_VM2	1
VM1_PI2	1
VM1_VF2	1

	VM1_VM2	1
ADMIRACIÓN	PI0_VM1	1
	PI2	1
	VI0_PM0_PI1_VF2	1
	VI0_VM0_PM1_VI2	1
	VI1_V-I2	1
AFECTO	VI0_PM0	2
	PI0	1
	PI0_VF1	1
	PI0_VF2	1
	PI0_VI2	1
	PI0_VM2	1
	PI1_PI2_VF2	1
	PM0	1
	VI0	1
	VI0_PF0_VI1_PM1	1
	VI0_PF0_VM2	1
	VI0_PM0_PF1	1
	VI0_PM0_PI1_VM2	1
	VI0_PM0_PI2	1
ALEGRÍA	PI1_VM2	2
	0	1
	P+F0_P+I1_PM1_VI2_VM2	1
	P+I1	1
	PF0_VI2	1
	PI0_PI1_VM2	1
	PI0_VM0	1
	PM0_VI2	1
	PM1_VM2	1
	PM2	1
	VI0_P+M1	1
	VI0_PI1_VF2	1

	VI0_P11_VI2	1
	VI0_PM0	1
	VI0_PM0_P11_VM2	1
	VI0_PM0_VF1	1
	VI0_PM0_VM1_V-F1_VM2	1
	VI0_VM2	1
	VI1	1
	VI2_PF2	1
	VM0_P11_VI2	1
		1
ALTERACIÓN	0	1
	PF0	1
	PI0_P+I1	1
	PI0_VF2	1
	PI2	1
	PM0_VI2	1
	VI0	1
	VI0_P11	1
	VI0_P11_PM1_VF2	1
	VI0_PM0_PF0_P+I1	1
	VI0_PM0_P11_VF2	1
BURLA	0	2
	PI1	2
	PI1_P12	2
	PM0	2
	P+I0_PM0_P10_VF1_VM2	1
	P+M1_P12	1
	PF0_P12_VM2	1
	PI0_PM0_VF2	1
	PI1_VF2	1
	PI2	1
	VF2	1

	VI0_PFO_P12	1
	VI0_PM0	1
	VI0_PM0_P+F0_VF1	1
	VI0_PM0_PI1	1
	VI0_PM0_PI1_VI2	1
	VI1_PM1_VI2	1
	VM0	1
BURLA- RECHAZO	VF0	2
	VI0_PM0_VF2	2
	PI0_PM1_VI2_PM2	1
	PI0_VM2	1
	PI1	1
	PI1_PM1_VM2	1
	PI1_VM2	1
	PM0_VM1	1
	PM1_VI2_VF2	1
	VI0_PFO_PI1_VI2	1
	VI0_PM0	1
	VI0_VI1_PI2	1
DECEPCIÓN	PI0	1
	PI0_VI1_PM2	1
	PI0_VI2	1
	PI0_VI2_PM2	1
	PI1	1
	PI1_VF1_PI2	1
	PI1_VI2_PM2	1
	PM0_VM2_VF2	1
	VI0_PM0_VI2	1
	VI0_PM1_PI2_VF2	1
	VI0_VI1_PFO_VI2	1
	VI1_VI2	1
	VI2	1

	VM0	1
	VM1	1
DIVERSIÓN	PI0_VI1_VM2	1
	PI1_PF2	1
	PI1_VI2_PF2	1
	PM0_VI2	1
	VF0_PI1_PM2	1
	VI0_PF0	1
	VI0_PM0_VF0_PM1_VF2	1
	VI0_PM1_VI2	1
DUDA	0	3
	PI1_VI2	1
	PM0_VI2	1
	VF2	1
	VI0_PM0_VI2_VF2	1
	VM0_PI1_PM1_VM2	1
ENFADO	VI0_PM1	2
	0	1
	PI0_PI1	1
	PI0_PI1_VF1	1
	PI0_VF1_VF2	1
	PI0_VM0_VF1	1
	PI1	1
	PI1_VF1	1
	PM0_VI2	1
	PM1	1
	PM1_VM2	1
	VF2	1
	VI0_PM0_PI1_VI2	1
	VI0_PM0_VF2	1
	VI0_PM0_VI1	1
	VI2_PI2_PM2_VF2	1

ENFADO- RECHAZO	PI0_P+F1_PF2	1
	PI0_PM0_VF0_PM1_VF2	1
	PI0_PM2_VF2	1
	PI1_VI2	1
	PM0_PI1_VF2	1
	PM1_VF2	1
	VF2	1
	VI0_PI1	1
	VI0_PI1_PF1_VF2	1
	VI0_PM0_PF0	1
	VI0_VF0_PM1	1
EXTRAÑEZA	0	1
ORGULLO	PI1_VI2	2
	PF0_P+I1_PF1	1
	PI0_P+M0_PI2	1
	PI0_P+M1_PI2	1
	PI1_VF2	1
	PI2	1
	PM0_PM1_VF2	1
	PM0_VI2	1
	PM1	1
	PM1_VI2	1
	PM1_VI2_PM2	1
	VF2	1
	VI0_PF1	1
	VI0_PM0	1
	VI0_PM0_PM1_VF2	1
	VI0_PM0_VF1	1
	VI0_PM0_VI1_PM1	1
	VI0_PM1	1
	VM0_PF0_PI1	1
	VM1_V-F1_VM2	1

RECHAZO	VI0_PM0	2
	PI0	1
	PI0_PI1_VM2	1
	PI1	1
	PI1_PI2	1
	PI1_PM1	1
	PM2_VF2	1
	VI0_PI1	1
	VI0_PM0_PI2_VM2	1
	VI0_VF2	1
VI0_VI2_PM2	1	
RESIGNACIÓN	PI1_VF2	2
	PM0	2
	0	1
	PI0_VF1	1
	PI1	1
	PI1_VI2	1
	PM0_PFO_VI2	1
	PM0_VI1	1
	PM1	1
	VI0_PM0_PI2	1
TRISTEZA	PF1	1
	PI0	1
	PI0_PI1_VI2_VF2	1
	PI0_VF1	1
	PI0_VI2	1
	PI1	1
	PI1_VF2	1
	PI2	1
	PM0_VM1_VM2	1
	PM1	1
	VI0_PI1	1

	VI0_PM0_PM1_VM2	1
	VI0_VI2	1
DESPEDIDA	PM2_VF2	1
	VI2_VF2	1
SALUDO	0	2
	PF0_PI2	1
	PI1	1
	VI0_PM0_VF0_PI1	1

C.4. Patrones finales no finales de enunciado

A continuación se muestran todos los patrones finales no finales de enunciado tanto del corpus neutro como del corpus emocional de los grupos entonativos más frecuentes, los enunciativos formados por una, dos o tres sílabas. En el caso del corpus emocional, los patrones aparecen clasificados según las diferentes emociones representadas.

C.4.1. Una sílaba

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VI0_PM0_PFO	103
	VI0_PM0_VF0	57
	VI0_PFO	50
	PI0_VF0	44
	PFO	41
	0	36
	PI0_PFO	33
	VFO	31
	PM0_PFO	28
	PI0_VM0_VF0	27
	PM0_VF0	20

VIO_PM0	18
VIO_VF0	15
PIO_VM0	14
PM0	11
PIO_VM0_PFO	9
PIO_PM0_VF0	9
VIO_PM0_P+F0	5
VM0_PFO	4
VM0	4
PIO_PM0	3
VIO_PIO_VM0_PFO	3
VIO_VM0_PFO	3
VIO_PM0_PIO_VF0	3
PIO_P+F0	2
VIO_PIO_PM0_VF0	2
VIO_PM0_PIO_VM0_VF0	2
VIO_VM0	2
VM0_VF0	2
P+F0	1
PIO_PM0_P+F0	1
P+IO_PM0_PIO_VIO_VIO_PM0_P+F0	1
PIO_VIO_VM0_PFO	1
PIO_VM0_P+F0	1
VIO_PM0_PIO_VM0_PFO	1
VIO_PM0_VF0_VIO_PM0_P+F0	1
VIO_PM0_VF0_VM0_PFO	1
VIO_VIO_PM0_P+F0	1
V-IO_VIO_VIO_PM0_P+F0	1
VM0_PFO_PM0_P+F0	1
P+IO_PM0_PIO_VM0_VF0	1
P+IO_PM0_VIO_VF0	1
PIO_VM0_V-F0	1
VIO_V-F0	1
VIO_V-M0	1
V-M0	1

BURLA	VF0	2
	PI0_VM0_PF0	1
DECEPCIÓN	VI0_PM0_P+F0	1
	VI0_PM0_VF0	1
DUDA	0	1
ENFADO	0	1
	VI0_PM0_VF0	1
ENFADO- RECHAZO	PI0_VF0	1
EXTRAÑEZA	PM0	1
ORGULLO	0	2
RECHAZO	VF0	1
RESIGNACIÓN	PF0	1
	VI0_PF0	1
	VI0_PM0_PF0	1

C.4.2. Dos sílabas

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	PI0_VI1_VM1	37
	0	32
	PM0_VI1_VM1	25
	PI0_VI1_VF1	24
	VI1_PM1	23
	PI0_VM1	20
	VI0_PM0_VI1_VF1	19
	PM0_VI1_VF1	19
	VI0_PM0_PM1	18
	PI0_PM1	17
	VI0_PM0_PF1	17
	PI1_PM1	16
	VI0_PM0_VI1_VM1	15
	PM1	13
	VI1_PM1_PF1	13

PIO_PF1	12
PM0_VI1_PM1	12
VI0_PI1_PM1	12
VI0_PM0_VI1_PM1	12
PI1_PF1	11
VF0_PM1	10
VI0_PI1_PF1	10
PM0_VM1	10
VI0_PM1	9
PIO_VF0_VM1	9
VI0_PM0_VI1_PM1_VF1	9
PF0_PM1	8
PIO_VI1_PF1	8
PIO_VI1_VM1_PF1	8
PM0_PM1	8
PM0_VI1_PF1	8
VI0_PM0_VI1_PF1	8
VI1_PF1	8
PIO_VI1_PM1	7
PM0_VI1_VM1_PF1	7
VF0_PI1_PM1	7
VI0_PM0_VI1_VM1_PF1	7
VI0_PM1_PF1	7
VI0_PI1_VM1	7
PIO_VF0_VF1	7
PIO_VI1_PM1_VF1	7
VI0_VM1	7
VI0_VI1_PM1_PF1	6
VI0_PI1_PM1_VF1	6
VI0_PM0_PI1_VM1	6
VI0_PM0_VF0_VM1	6
VI0_PM0_VF1	6
VI0_PM0_VM1	6
PIO_PI1_VM1	6
PIO_VM1_VF1	6

PI1_VM1	6
VF0_PF1	5
VIO_PM0_VI1_PM1_PF1	5
VIO_VI1_PM1	5
VM0_PM1_PF1	5
VIO_PM0_VF0_VF1	5
VIO_PM0_VM1_VF1	5
VIO_PM1_VF1	5
VIO_VI1_PM1_VF1	5
PF0_VI1_VM1	5
PF0_VM1	5
PIO_VM0_VM1	5
PM0_VM1_VF1	5
VF1	5
PIO_PM0_VI1_PF1	4
PIO_VM0_VI1_PM1_P+F1	4
PI1_VM1_PF1	4
VIO_PF1	4
VIO_VF0_PM1	4
PM0_VF0_VF1	4
VF0_VF1	4
VI1_VM1	4
PF0	3
PF1	3
PIO_VF0_PM1	3
PM0_PF1	3
PM0_VF0_VM1_PF1	3
PM1_PF1	3
VF0_PI1_PF1	3
VF0_PM1_PF1	3
VIO_PF0_PM1	3
VIO_PM0_PF0_VI1_PM1	3
VIO_PM0_VI1_PM1_P+F1	3
VIO_PM1_P+F1	3
VIO_VI1_PF1	3

VI1_PM1_P+F1	3
VM0_PI1_PF1	3
VIO_PF0_PM1_VF1	3
VIO_PM0_PM1_VF1	3
VIO_VF0_PI1_VM1	3
VM0_PI1_VM1	3
PIO_PF0_VI1_VM1	3
PIO_PM0_VM1	3
PIO_PM1_VF1	3
PIO_VF1	3
PI1_VM1_VF1	3
PM0_PM1_VF1	3
PM0_VF0_VM1	3
PM1_VF1	3
VM1	3
PF0_PF1	2
PF0_VI1_PF1	2
PIO_PF0_VI1_PF1	2
PIO_PM0_VI1_PM1	2
PIO_VI1_VI1_PM1_P+F1	2
PIO_VM0_VI1_PM1	2
PIO_VM0_VM1_PF1	2
PIO_VM1_PF1	2
PI1	2
PM0_PI1_VM1_PF1	2
VIO_PM0_VI1_VI1_PM1_P+F1	2
VIO_PM0_VM1_PF1	2
VIO_VF0_PF1	2
VIO_VI1_PM1_P+F1	2
VI1_PM1_VF1_PM1	2
VM0_PF0_PF1	2
VM0_PI1_PM1	2
VM0_PM1	2
VM1_PF1	2
VF0_PI1_VM1	2

VIO_PM0_PF0_VI1_VM1	2
VIO_PM0_PI1_VF1	2
PI1_PM2_VF2	2
PF0_VI1_VF1	2
PIO_PI1_VF1	2
PIO_PI1_VM1_VF1	2
PIO_PM0_VI1_VM1	2
PIO_VF0_PM1_VF1	2
PIO_VI1_VM1_V-F1	2
PI1_PM1_VF1	2
PI1_VF1	2
PM0_VF1	2
PM0_VI1_PM1_VF1	2
PM0_VI1_V-M1	2
VF0_VM1	2
VI1	2
PM1_VI2_VM2_PF2	1
PF0_P+M1	1
PF0_VM1_PF1	1
PIO_PF0	1
PIO_PI1_P+F1	1
P+IO_PM0_PM1_P+F1	1
PIO_PM0_VF0_PM1	1
PIO_PM0_VF0_VI1_PF1	1
P+IO_PM0_VF0_VI1_PM1_P+F1	1
PIO_PM0_VF0_VM1_PF1_PM1_P+F1	1
P+IO_PM0_VIO_VI1_PM1	1
PIO_PM1_P+F1	1
PIO_VF0_PF1	1
PIO_VF0_PI1_PF1	1
PIO_VF0_PI1_P+F1	1
PIO_VF0_PI1_VM1_PF1	1
PIO_VF0_PM1_PF1	1
PIO_VI1_PM1_VF1_PM1	1
PIO_VM0_VM1_PF1_P+M1	1

PI1_PM1_P+F1	1
PM0_PF0	1
PM0_PF0_VI1_PF1	1
PM0_PF0_VI1_VM1_PF1	1
PM0_PF0_VM1_PF1	1
P+M0_PI1_PM1	1
PM0_VF0_PM1	1
PM0_VF0_PM1_PF1	1
PM0_VF0_PM1_P+F1	1
PM0_VF0_VI1_PM1	1
PM0_VM1_PF1	1
VF0_PM1_P+F1	1
VF0_VM1_PF1	1
VI0_PF0_PF1	1
VI0_PF0_PI1_P+M1	1
VI0_PF0_VI1_PM1	1
VI0_PI0_PM1_P+F1	1
VI0_PI1_PM1_P+F1_P+M1	1
VI0_PI1_VM1_PF1	1
VI0_PM0_PF0_VI1_VM1_PF1	1
VI0_PM0_P+F1	1
VI0_PM0_PI1_P+M1	1
VI0_PM0_PI1_VM1_PF1	1
VI0_PM0_VF0_PF1	1
VI0_PM0_VF0_PM1	1
VI0_PM0_VF0_PM1_P+F1	1
VI0_PM0_VF0_VI1_PF1	1
VI0_PM0_VF0_VI1_PM1	1
VI0_PM0_VF0_VM1_PF1	1
VI0_PM0_VI1_VM1_P+F1	1
VI0_VF0_PM1_PF1	1
VI0_VM0_PF0_VI1_PM1	1
VI0_VM0_PI1_PM1	1
VI0_VM0_PM1_PF1	1
V-I0_VM0_VI1_PM1_P+F1	1

VIO_VM1_PF1	1
VM0_PF0_PM1	1
VM0_PI1	1
VM1_PF1_P+M1	1
VM1_PF1_PM1_P+F1	1
VI1_PM2_VF2	1
VF0_PI1_PM1_VF1	1
VF0_PM1_VF1	1
VF0_PM1_V-F1	1
VF0_VI1_PM1_VF1	1
VIO_PF0_PI1_VM1	1
VIO_PF0_VI1_VF1	1
VIO_PI1_VF1	1
VIO_PM0_PF0_VM1_VF1	1
VIO_PM0_PIO_VF0_VM1	1
VIO_P+M0_PI1_VM1_VF1	1
VIO_PM0_PI1_VM1_VF1	1
VIO_PM0_PI1_VM1_V-F1	1
VIO_PM0_VF0_V-F1	1
VIO_VF0_PI1_VF1	1
V-IO_VM0_PF0_PM1_VF1	1
VIO_VM0_PI1_VF1	1
V-IO_VM0_PM1_VF1	1
VI1_PM1_VF1	1
VI1_PM1_VI1_VM1	1
VM0_PF0_PM1_VF1	1
VM0_VF0_PI1_PM1_VF1	1
VM0_VF0_PI1_VM1	1
0	1
PF0_PI1_VM1	1
PF0_PM1_VF1	1
PF0_PM1_VF1_VM1	1
PF0_VF1	1
PF0_VI1_PM1_VF1	1
PIO_PF0_VM1	1

	PIO_PF0_VM1_VF1	1
	P+IO_PM0_PIO_VM0_VI1_PM1_VF1	1
	PIO_PM0_VI1_VF1	1
	PIO_P+M0_VI1_VM1	1
	PIO_PM0_VI1_V-M1	1
	PIO_PM1_VF1_V-M1	1
	PIO_VF0_PII_PM1_VF1	1
	PIO_VF0_PII_VF1	1
	PIO_VI1_V-M1	1
	PIO_VM0_VF0	1
	PIO_VM0_VF1	1
	PIO_VM0_VI1_PM1_VF1	1
	P+I1_PM1_VF1	1
	PM0_PF0_VI1_VM1	1
	PM0_PII_VF1	1
	PM0_VF0_PII_VM1	1
	PM0_VI1	1
	PM0_VM1_V-F1	1
	VI0_PM0_PF0_VI1_PM1_VF1	1
	VI0_PM0_PIO_VF0_PII_VF1	1
	VI0_PM0_VF0_PII_VF1	1
	VI0_PM0_VF0_PII_VM1	1
	VI0_PM0_VI1_PM1_VF1_VM1	1
	VI0_VF1	1
	VI1_VF1	1
	VM0_V-F1	1
	VM0_VI1	1
	VM0_VM1	1
	VM1_VF1	1
ADMIRACIÓN	VI0_PM0_VI1_VF1	1
AFECTO	PM0_VI1_PM1	1
	PM0_VM1_PF1	1
	VI1_PM1_P+F1	1
	VI0_PM0_VF0_VF1	1
	PIO_VM1	2

	PI1_VM1	1
ALEGRÍA	VIO_PI1	1
ALTERACIÓN	VM1	1
BURLA	PM0_PI1_VM1_PF1_PM1	1
	VIO_PI1_VF1	1
BURLA- RECHAZO	VIO_PM1	1
DECEPCIÓN	VM1	1
DUDA	VIO_PM0_VM1_PF1	1
	PM1_VF1	2
ENFADO	VM0_PI1_PM1_VF1	1
ENFADO- RECHAZO	PIO_PM0_VI1_VM1_PF1	1
	VIO_VF0_PM1	1
	VIO_PI1_VM1	1
	PF0_VF1	1
INTERÉS	PM0_PI1_VI1_PM1_P+F1	1
	VIO_PM0_PF0_VI1_VM1	1
RESIGNACIÓN	PM0_VM1_PF1	1
	VIO_PM0_PF1	1
	VIO_PM0_VM1_PF1	1
	VIO_PF0_VM1_VF1	2
	VIO_PIO_PM1_VF1	1
	VIO_PI1_VF1	1
	VIO_PM0_PF0_VF1	1
	VIO_PM0_PI1_VM1	1
	VIO_PM0_VM1_VF1	1
TRISTEZA	VIO_PM0_VI1_PM1	1
	PM0_VM1	1
DESPEDIDA	VIO_PI1_VF1	1

C.4.3. Tres sílabas

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VF0_P12_P2	2
	0	2
	PI0_VM1_VM2	2
	PM0_VF1_VF2	2
	PM0_VF1_VM2	2
	PF0_VM1_VI2_PM2	1
	PI0_PF0_VI2_PM2_P2	1
	PI0_PM0_VI1_VI2_PM2	1
	PI0_VF0_P11_VF1_PM2	1
	PI0_VF1_VI2_PM2_P+F2	1
	PI0_VI1_PM1_PM2	1
	PI0_VI1_VI2_PM2_P2	1
	PI0_VI2_VM2_P2	1
	PM0_VF1_PM2	1
	PM0_VI1_VM1_PM2	1
	PM0_VI2_P2	1
	PM0_VM1_PM2	1
	PM0_VM1_VI2_P2	1
	PM0_VM2_P2	1
	VF0_VF1_PM2	1
	VF0_VI1_PM2	1
	VF1_P2	1
	VF1_P12_PM2	1
	VF1_PM2	1
	VI0_P11_P2	1
	VI0_P12_P2	1
	VI0_P12_PM2	1
	VI0_PM0_P11_VI2_VI2_PM2_P+F2	1
	VI0_PM0_VF1_PM2_P+F2	1
	VI0_PM0_VI2_P2	1
	VI0_PM0_VI2_P+F2	1
	VI0_VF1_P12_PM2	1

VI0_VI1_PM1_PM2	1
VI0_VI2_PM2	1
VI0_VM1_PI2_PM2	1
VI1_PF2	1
VI1_PI2_PF2	1
VI1_PI2_PM2	1
VI1_PM1_PF1	1
VI1_PM2	1
VI2_PM2	1
VI2_PM2_PF2	1
VI0_PM0_PM1_VI2_VF2	1
VI0_PM0_VF1_VF2	1
VI0_PM0_VI1_VF2	1
VM0_PI1_PI2_VM2	1
VM0_PI1_VM2	1
VM0_VF1_PI2_VM2	1
PI0_PI2_VM2	1
P+I0_PM0_VF0_PI1_VM2	1
PI0_PM0_VI2_VF2	1
PI0_PM2_VF2	1
PI0_VF0_VM2	1
PI0_VF1_VF2	1
PI0_VI1_VF2	1
PI0_VI1_V-M2	1
PI0_VI2_VF2	1
PI0_VI2_VM2	1
PI0_VM1_VF2	1
PI0_VM1_VI2_V-M2	1
PI1_VF2	1
PI1_VI2_VF2	1
PM0_PI1_VF2	1
PM0_VI1_VF2	1
PM0_VI2_VM2	1
PM0_VM1_VF2	1
VI0_PI1_VM1_PM2_VF2	1

VI0_PM0_VF1_VI2_PM2_VF2	1
VI0_PM0_VI2_PM2_VF2	1
VI0_PM0_VI2_VM2	1
VII_VM2	1

C.5 Patrones finales de enunciado

A continuación se muestran todos los patrones finales no finales de enunciado tanto del corpus neutro como del corpus emocional de los grupos entonativos más frecuentes, los enunciativos formados por una, dos o tres sílabas. En el caso del corpus emocional, los patrones aparecen clasificados según las diferentes emociones representadas.

C.5.1. Una sílaba

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VF0	145
	VI0_VF0	105
	PI0_VM0	69
	VM0	69
	VI0_VM0	57
	PI0_VM0_VF0	48
	PI0_VF0	46
	VM0_VF0	41
	VI0_PM0_VF0	26
	0	13
	PI0_VM0_PF0	13
	VI0_V-M0	10
	PM0_VF0	9
	V-F0	9

	VIO_PM0_PFO	9
	VIO_V-F0	8
	VIO_VM0_PFO	7
	PM0	6
	VIO_VIO_PM0_P+F0	5
	VM0_PFO	5
	VIO_VM0_V-F0	4
	PIO_PFO	3
	VIO	3
	VIO_PM0	3
	VIO_PM0_P+F0	3
	V-M0	3
	PFO	2
	PIO_VM0_V-F0	2
	VIO_PFO	2
	VIO_PM0_VF0_VM0	2
	VIO_PM0_VIO_VM0	2
	VM0_V-F0	2
	PIO_P+F0	1
	PIO_P+M0_P+I0_PM0_PIO_VM0_PFO	1
	PIO_P+M0_P+I0_PM0_VIO_VIO_PM0_P+F0	1
	0	1
	PIO_VM0_PIO_VM0	1
	PM0_PFO	1
	VIO_P+F0	1
	VIO_PIO_VIO_PM0_P+F0	1
	VIO_PIO_VM0_PFO	1
	VIO_PM0_PIO_VF0	1
	VIO_PM0_VF0_VM0_PFO	1
	V-I0_VIO_PM0_P+F0	1
	VIO_V-M0_VF0	1
ADMIRACIÓN	VIO_PFO	1
AFECTO	PM0_VF0	1
	VF0	3
ALEGRÍA	VIO_PM0_PFO	1

	PI0_VM0	1
	V-F0	1
	VI0_VF0	1
ALTERACIÓN	VI0_VM0	1
	VI0_V-M0	1
	VM0_V-F0	1
BURLA	VI0_VF0	2
	VI0_PFO	1
	VI0_PM0_VF0	1
	PI0_VM0_VF0	1
BURLA-RECHAZO	VF0	3
	VI0_PFO	1
	VI0_PM0_PFO	1
	VI0_PM0_VF0	1
DECEPCIÓN	PFO	1
	PI0	1
	PM0_P+F0	1
	PM0_PFO	1
	VI0_PFO	1
	VI0_PM0_VF0	1
	PI0_VM0	1
DIVERSIÓN	PFO	1
	VI0_PM0_PFO	1
DUDA	VI0_PFO	3
	PI0_PFO	1
	VI0_PM0_PFO	1
	PM0_VF0	1
ENFADO	VI0_PM0_PIO_VF0	1
	VI0_PM0_VF0	1
	VI0_PM0_VF0_VM0	1
	PI0_PM0_VF0	1
	PI0_VM0_VF0	1
	VF0	1
ENFADO-RECHAZO	VM0	1

	PI0_PFO	1
	0	1
	PI0_VF0	1
	VF0	1
ORGULLO	VI0_VF0	1
	PM0_PFO	1
	VI0_PFO	1
	VI0_PM0_VF0	1
RECHAZO	VF0	1
	VF0	2
	VI0_VF0	2
	VI0_PM0_PFO	1
	VI0_PM0_VF0	1
	PI0_VF0	1
RESIGNACIÓN	PM0_VF0	1
SORPRESA	VI0_PM0	1
	VI0_PFO	2
	VI0_PM0_PFO	1
	VI0_PM0_VF0	1
	PI0_PM0_VF0	1
	PI0_VF0	1
TRISTEZA	VM0	1
	VI0_VF0	2
DESPEDIDA	VI0_PM0_VF0	1
	PM0_PFO	1
SALUDO	VI0_PM0_PFO	1

C.5.2. Dos sílabas

EMOCIÓN	PATRÓN	N° CASOS
NEUTRO	PI0_VI1_VM1	110
	VM1	104
	VI1_VM1	94
	PI0_VF0_VM1	75
	PI0_VM1	74
	VI0_VM1	70
	VF0_VM1	64
	PI0_VI1_VF1	49
	PI0_VM0_VM1	48
	VM0_VM1	46
	PM0_VI1_VM1	45
	PI0_VF0_VF1	41
	VI1_VF1	41
	VF1	35
	VI0_VF1	29
	VF0_VF1	26
	0	23
	PI1_VM1	22
	VI0_PM0_VM1	18
	PM0_VM1	17
	PM1	16
	PM0_VI1_VF1	16
	VM0_VF1	16
	VI0_PM0_VI1_VM1	14
	PI0_VF1	14
	PI0_VM0_VF1	14
	VI0_PM0_VI1_VF1	13
	PM0_VF0_VM1	13
	VI0_PM0_VF0_VM1	10
	PI0_VM1_VF1	10
	PI0_VF0_VM1_PF1	8

VI0_PM0_VF0_VF1	8
PF0_VM1	8
PM0_VF0_VF1	8
VI0_VF0	8
PF1	7
VI1_PM1	7
VI1_PM1_P+F1	7
VM1_PF1	7
VI0_PM0_VM1_VF1	7
VF0	7
VI1	7
VI1_V-M1	7
PI0_VI1_VM1_PF1	6
VI0_VI1_PM1_P+F1	6
VM1_VF1	6
PI0_PF1	5
PI0_VM1_PF1	5
PM0_VI1_VM1_PF1	5
VI0_PM0_PI1_VM1	5
VM0_PI1_VM1	5
PM0_VF1	5
VI0_VI1	5
VI0_VI1_V-M1	5
VI0_V-M1	5
PI0_VI1_PM1	4
VF0_PM1	4
VI0_PM0_PF1	4
VI0_PM0_PM1	4
VI0_PM0_VI1_PM1	4
VI0_PM0_VI1_VM1_PF1	4
VI0_PM0_VM1_PF1	4
VF0_PI1_VM1	4
PF0_VI1_VM1	4
PI0_VF0_PI1_VM1	4
PI0_VI1_PM1_VF1	4

VM0_VI1	4
V-M1	4
PI0_PM1	3
PI0_VI1_PF1	3
PM0_VF0_PF1	3
PM0_VI1_PM1	3
PM0_VM1_PF1	3
VI0_PM0_PI1_VM1_PF1	3
VI1_VM1_PF1	3
VI0_PM0_VF1	3
VI0_PM0_VI1_V-M1	3
PI0_PF0_VM1	3
PI0_PI1_VM1	3
PI0_VF0_PI1_VF1	3
PI0_VF0_VI1_V-M1	3
PI0_VI1	3
PI0_VI1_V-M1	3
PI0_VM0_V-M1	3
VF0_V-M1	3
VI0_VM0_V-M1	3
PF0_VM1_PF1	2
PI0_VF0_PM1_PF1	2
PI0_VF0_VI1_PM1_P+F1	2
PI0_VI1_VI1_PM1_P+F1	2
PI1_VM1_PF1	2
VF0_PI1_PM1	2
VF0_VI1_PM1_P+F1	2
VI0_PM0_VF0_VM1_PF1	2
VI1_PF1	2
VI1_VI1_PM1_P+F1	2
VM0_VI1_PM1_P+F1	2
VF0_PI1_VF1	2
VI0_PI1_VM1	2
VI0_PM0_PF0_VM1	2
VI0_PM0_VF0_V-M1	2

VII_PM1_VF1	2
PIO_PM0_VI1_VF1	2
PIO_VF0_P11_VM1_VF1	2
PIO_VM0_VI1	2
PIO_VM0_VI1_V-M1	2
PI1_VF1	2
PI1_VM1_VF1	2
PM0_P11_VM1	2
VF0_VI1_V-M1	2
VI0_PM0_VF0_P11_VM1	2
VI0_V-I1	2
VI0_VM1_V-F1	2
VM0	2
VM0_VF0	2
VM0_V-M1	2
PIO_P11_VM1_PF1	1
PIO_PM0_VF0_VM1_PF1	1
PIO_VF0_PF1	1
PIO_VF0_P11_PM1	1
PIO_VF0_VI1_P11_VM1_PF1	1
PIO_VI1_P11_VM1_PF1	1
PIO_VI1_PM1_PF1	1
PIO_VM0_VF0_P11_VM1_PF1	1
PIO_VM0_VI1_PM1	1
PIO_VM0_VI1_PM1_P+F1	1
PIO_VM0_VM1_PF1	1
PI1_PF1	1
PI1_VI1_VI1_PM1_P+F1	1
PM0_VF0_P11_PM1	1
PM0_VF0_VI1_PM1	1
PM0_VI1_PF1	1
PM0_VI1_PM1_PF1	1
VF0_P11_P+M1	1
VF0_PM1_PF1	1
VF0_VM1_PF1	1

VI0_PF1	1
VI0_PI1_PF1	1
VI0_PM0_PF0_VM1_PF1	1
VI0_PM0_PI1	1
VI0_PM0_VF0_PI1_VM1_PF1	1
VI0_PM0_VI1_PM1_P+F1	1
VI0_PM1	1
VI0_PM1_PF1	1
VI0_PM1_P+F1	1
VI0_VF0_PI1_PM1	1
VI0_V-I1_VI1_PM1_P+F1	1
VI0_VM0_PF0_P+I1_PM1	1
VI0_VM0_PF0_P+I1_P+M1	1
VI0_VM0_PF0_VM1_PF1	1
VI1_PM1_PF1	1
VI1_VM1_PF1_PM1	1
VI1_V-M1_VF1_VI1_PM1_P+F1	1
VM0_PF0_P+I1	1
VM0_PI1_VM1_PF1	1
VM0_P+M1	1
V-M0_VF0_PM0_P+F0_P+I1	1
VM1_PF1_PM1_P+F1	1
VF0_PM1_VF1	1
VI0_PF0_PI1_P+M1_PI1_VM1_VF1	1
VI0_PM0_P+F0_VI1_VM1	1
VI0_PM0_PI1_VM1_VF1	1
VI0_PM0_VF0_V-F1	1
VI0_PM0_VF0_VI1	1
VI0_PM0_VF0_VI1_V-F1	1
VI0_PM0_V-F1	1
VI0_PM0_VI1	1
VI0_PM0_V-M1_V-F1	1
VI0_VM0_PF0_PI1_P+M1_P+I1_PM1_PI1_VM1	1
VI0_VM0_PF0_P+I1_PM1_VF1_VM1	1

VII_PM1_P+F1_P+I1_PM1_VF1	1
VM0_PFO_VI1_VM1	1
VM0_PFO_VM1	1
VM0_PI1_PM1_VF1	1
VM0_PI1_VM1_VF1	1
VM0_PM1_VF1	1
PF0_VF1	1
PF0_VI1_VF1	1
PIO_PI1_VF1	1
PIO_PI1_VM1_VF1	1
PIO_PM0_VFO_VM1	1
P+IO_PM0_VI1_VF1	1
P+IO_PM0_VI1_VM1	1
PIO_P+M0_VI1_VM1	1
PIO_PM0_VM1	1
PIO_PM0_VM1_VF1	1
PIO_VFO_VI1	1
PIO_VFO_V-M1	1
PIO_VI1_PM1_VI1_VM1	1
PIO_VI1_V-F1	1
P+IO_VI1_VM1	1
PIO_VI1_VM1_V-F1	1
PIO_VM0	1
PIO_VM0_PFO_VF1	1
PIO_VM0_VFO	1
PIO_VM0_VFO_PI1_VM1	1
PIO_VM0_VFO_V-M1	1
PIO_VM0_VI1_PM1_VF1	1
PIO_V-M1	1
PM0_PI1_VM1_VF1	1
PM0_VFO	1
PM0_VFO_PI1_VM1	1
PM0_VI1_PM1_VF1	1
PM0_VM1_VF1	1
PM1_VF1	1

	VF0_V-F1	1
	VF0_VI1	1
	VI0_VF0_V-M1	1
	VI0_V-II_V-F1	1
	VI0_VI1_V-F1	1
	VI0_VM0	1
	VI0_V-M1_VF1	1
	VII_V-F1	1
	VM0_VI1_V-M1	1
	VM0_VM1_V-F1	1
ADMIRACIÓN	VI0_PM0_PFO_VF1	1
	VI0_PM0_PI1_VM1_VF1	1
	VI0_PM0_PI1_VM1_VII_V-F1	1
AFECTO	VM1	2
	PI0_VM0_VI1_PM1_PF1	1
	VII_PM1_P+F1	1
	VI0_PM0_PF1	1
	VI0_PM0_PI1_VM1	1
	VI0_PM0_VF1	1
	VI0_PM0_VI1_VM1	1
	VI0_PM0_VM1_VF1	1
	VM0_PM1_VF1	1
	PF0_VI1_VM1	1
	PI0_VM1	1
	PI1_VM1	1
	PM0_VI1_VF1	1
	PM0_VM1	1
	PM0_VM1_VF1	1
	VI0_VI1	1
	VI0_VI1_V-F1	1
	VI0_V-M0_VI1	1
ALEGRÍA	VI0_VI1_PM1	1
	VI0_PM0_VI1_VM1	1
	PF0_VI1_V-F1	1
	PI0_VF1	1

	PI1_VM1	1
	PM0_VI1_VM1	1
	VF0_V-I1	1
	VI0_VM1	1
	VI1_VF1	1
	VI1_V-M1	1
	VM1	1
ALTERACIÓN	VI0_P11_VM1_PF1	1
	PF0_VI1_VM1	1
	PI0_PM0_VI1_VM1	1
	PI0_VI1	1
	PI1_VM1	1
	VF1	1
	VI0_VF1	1
	VI0_VM1	1
BURLA	VI0_VM1	4
	PF0_VM1	2
	VF0_PM1	1
	VM0_P11_PM1	1
	VI0_PF0_VM1	1
	PM0_VI1_VM1_V-F1_V-M1	1
	PM1_VF1	1
	VF1	1
	VI1_VM1	1
	VM0_VF1	1
	VM1	1
BURLA- RECHAZO	VM1	2
	PI0_VM1_PF1	1
	VI0_PF0_PF1	1
	VI0_VM0_P11_PM1	1
	VI1_PF1	1
	VI0_PF0_VM1_VF1	1
	PI0_VM1	1
	PI1_VF1	1
	PM0_VM1	1

	VI0_VM1	1
DECEPCIÓN	VI0_PI1_VM1	2
	VM1	2
	PM0_PM1	1
	VI1_PM1_P+F1	1
	VI0_PI1_VF1	1
	VI0_PM0_VF1	1
	VI0_PM0_VI1_VM1	1
	VI0_PM0_VM1	1
	PI0_PFO_VM1	1
	PI0_VF1	1
	PI0_VI1_VM1	1
	PI0_VM1	1
	PM0_VM1_VF1	1
	VF0	1
DIVERSIÓN	VI0_PM0_VM1	2
	PI1_VM1	2
	VM0_PI1_VM1_PF1	1
	VI0_PFO_VM1	1
	VI0_PI1_VM1	1
	VI0_PM0_PI1_VM1	1
	PI0_VFO_VM1	1
	PI0_VM1	1
	PM0_VFO_VM1	1
DUDA	PFO_VI1_VF1	2
	VI0_PM0_PI1_VM1_PF1	1
	VM1_PF1	1
	VI0_PM0_VI1_VF1	1
	VI0_PM0_VI1_VM1	1
	VI0_VF0_PM1_VF1	1
	VI0_VI1_PM1_PI1_VM1_VF1	1
	PFO_VI1_VM1	1
	PI0_VI1_VM1	1
	PI0_VM0_VM1	1
	PM0_VM1_VF1	1

ENFADO	VI0_VM1	2
	VM1	2
	VI0_PM0_VF0_PM1	1
	VI0_PM0_VF0_VF1	1
	VI0_PM0_VM1	1
	PF0_VF1	1
	PI0_VM1	1
	PM0_VI1_VM1	1
	VM0_VF1	1
ENFADO- RECHAZO	VM1	2
	PM1	1
	VI0_PM0_PI1_VM1_VF1	1
	PF0_VM1	1
	PI1_VF1	1
	PI1_VM1	1
	PM0_PF0_VM1	1
	PM0_PI1_VM1	1
	VM0_VM1	1
EXTRAÑEZA	VI0_PI1_PM1	1
	PF0_VM1	1
ORGULLO	VF1	3
	PI0_VM1	2
	VI0_VF1	2
	PI0_VF0_VM1_PF1	1
	PM1	1
	VI0_PM0_PM1	1
	VI0_PM0_VI1_PM1	1
	VF0_PI1_VM1	1
	VI0_PI1_VM1	1
	VI0_PM0_VF1	1
	VI0_PM0_VM1	1
	PI0_PF0_VM1	1
	PI1_VF1	1
	PM0_VI1_VF1	1
	VM1	1

RECHAZO	VI0_VM1	2
	PM0_VF0_PF1	1
	VI0_PI1_VM1	1
	VI0_PM0_PF0_VM1	1
	VI0_PM0_PI1_VM1	1
	VI0_PM0_VF1	1
	VI0_PM0_VM1	1
	PI0_PF0_VF1	1
	PI0_VM1	1
	PI1_VM1	1
	PI1_VM1_VF1	1
	PM0_PI1_VM1	1
	VI1_VF1	1
	VI1_VM1	1
	VM1	1
RESIGNACIÓN	VF1	3
	VI0_PM0_PF1	2
	VI1_PF1	1
	VI0_PF0_VM1	1
	VI0_PI1_VF1	1
	VI0_PM0_VM1_VF1	1
	PF0_VM1	1
	PI0_PM0_VF0_VM1	1
	PI0_VF0_PI1_VM1_VF1	1
	PI0_VF0_VM1	1
	PI0_VM1_VF1	1
	PM0_VM1	1
SORPRESA	PF0_VM1	1
TRISTEZA	VI0_PM0_VM1	2
	PI0_PM0_VI1_PM1	1
	PI0_VI1_PF1	1
	PM0_VM1_PF1	1
	VI0_PI1_VF1	1
	VI0_PM0_PI1_VM1	1
	VI0_PM0_VF1	1

	PF0_VI1_VF1	1
	PF0_VI1_VM1	1
	PI0_VM1	1
	PI0_VM1_VF1	1
	PI1_VM1	1
	PM0_VM1_VF1	1
	VIO_VM1	1
	VM1	1
DESPEDIDA	0	2
	VIO_PM0_PI1_VM1_VF1	1
	PI0_PI1_VF1	1
	PI0_PI1_VM1	1
	PI0_PM0_VM1	1
	PI1_VF1	1
	PM0_PI1_VM1	1
SALUDO	PI0_VF1	3
	PF1	1
	VIO_PM0_VI1_PF1	1
	VIO_VI1_PM1_P+F1	1
	VIO_PM0_PF0_VM1	1
	VIO_PM0_VF0_VM1	1
	PF0_VM1_VF1	1
	PI0_VF0_VI1	1
	PI0_VI1_PM1_VF1	1
	PI1_VF1	1
	PM0_PF0_VM1_VF1	1
	PM0_VF1	1
	PM0_VM1_VF1	1
	VF1	1
	VI1_VM1	1

C.5.3. Tres sílabas

EMOCIÓN	PATRÓN	Nº CASOS
NEUTRO	VII_VM2	12
	PI0_VI1_VM2	8
	VF0_VM2	7
	PI0_VI1_VF2	5
	VF1_VM2	5
	PI0_VF0_VM2	4
	PI0_VF1_VM2	4
	PI0_VM1_VM2	4
	VM2	4
	VI2_PM2_P+F2	3
	PI0_VF0_VF2	3
	PI1_VI2_VM2	3
	VII_VF2	3
	PI0_PM2	2
	PI0_VI2_VF2	2
	PI0_VM1_VF2	2
	PM1_VM2	2
	VI2_VM2	2
	VM0_VF2	2
	VM0_VM2	2
	VM1_VF2	2
	VM1_VM2	2
	PI0_PF2	1
	PI0_VF0_VI2_PM2_P+F2	1
	PI0_VI1_PM1_PF2	1
	PI0_VI1_VI2_PM2_P+F2	1
	PI0_VI1_VM2_PF2	1
	PI1_PI2_P+F2	1
	PM0_VF1_PI2_VM2_PF2	1
	PM0_VI2_VM2_PF2	1
	VF1_PM2_P+F2	1

VI0_PM0_PF0_VI1_PM1_VI2_PF2	1
VI0_PM0_PI1_VM1_PF1_PM2	1
VI0_PM0_VF0_VI2_PM2_P+F2	1
VI0_PM0_VM2_PF2	1
VI1_PM2_PF2	1
VI1_VI2_PM2_P+F2	1
VI1_VM2_PF2	1
VF0_PI1_VM2_VF2	1
VI0_PM0_PI2_VM2	1
VI0_PM0_PM1_VI2_VM2	1
VI0_PM0_PM1_VI2_V-M2	1
VI0_PM0_VF1_VM2	1
VI0_PM0_VF2	1
VI0_PM0_VI1_VF2	1
VI0_PM0_VI1_VM2	1
VI0_PM0_VM1_VM2	1
VM0_PF0_VI2_VF2	1
VM0_VI1_PM1_VI2_VF2	1
0	1
PF0_VI1_VM2	1
PI0_PF0_VF1_VM2	1
PI0_PM0_VI1_VF2	1
PI0_VF0_PI1_VI2_VM2	1
PI0_VF0_VM1_V-M2	1
PI0_VF1_VF2	1
PI0_VF2	1
PI0_VI2_VM2	1
PI0_VI2_V-M2	1
PI1_VF1_VM2	1
PI1_VI2_VF2	1
PI1_VM1_VM2	1
PI1_VM2_VF2	1
PM0_VF0_VF2	1
PM0_VF1_VF2	1
PM0_VF1_VM2	1

	PM0_VI1_VF2	1
	PM0_VI1_VM2	1
	PM0_VI1_V-M2	1
	PM0_VM1_VF2	1
	PM0_VM1_VI2	1
	PM0_VM2	1
	PM0_VM2_VF2	1
	VF0_VF2	1
	VF0_V-M2	1
	VF1_VI2_V-M2	1
	VI0_P11_VF1_P12_PM2_VF2	1
	VI0_V-M2	1
	VI1_V-M2	1
	VI2	1
	VI2_VF2	1
AFECTO	PI0_VF1_VF2	1
BURLA	VF2	1
BURLA- RECHAZO	VI0_PM0_PF0_VM2	1
DIVERSIÓN	PM0_VM1_VM2	1
ENFADO_ RECHAZO	VI1_V-M3	1
RECHAZO	VI0_VM2	1
RESIGNACIÓN	PM0_VI1_VM2_PF2	1
	VI0_PM2	1