



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

**FACULTAD DE INGENIERÍA**

**Detección y reparación de disfluencias  
del habla en el corpus DIME**

T E S I S  
QUE PARA OBTENER EL TÍTULO  
DE: INGENIERO EN COMPUTACIÓN  
P R E S E N T A:

IVÁN MORENO ÁVILA



**Asesor: Dr. Luis Alberto Pineda Cortés**

México, D.F.

2006



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## Agradecimientos

El trabajo que se presenta a continuación pudo llevarse a cabo gracias al apoyo que varias personas me brindaron y que para mi fue muy importante.

En primer lugar quiero agradecerles a mis papás que siempre me han apoyado en todas las decisiones que he tomado, brindando sus consejos y ayuda incondicional por estos más de 23 años. Además, quiero agradecerle a mi hermana que a pesar de nuestras cotidianas luchas por la computadora siempre me dio prioridad para poder avanzar en mi tesis.

A quien quiero agradecerle de una forma muy especial es a Paty, ya que por ella existe esta tesis. Paty fue quien me comentó del proyecto DIME y de las posibilidades de realizar mi tesis en el IIMAS, y por lo tanto, ella fue quien me presentó al Dr. Luis Pineda, quien es mi director de tesis. Además siempre me ha brindado su ayuda cuando la he necesitado, aunque me amenazó que se los iba a cobrar algún día (espero que no me salgan muy caro).

Quiero agradecerle a mi director de tesis el Dr. Luis Pineda por el apoyo que me brindó durante el desarrollo de este trabajo, asimismo por el nuevo conocimiento que pude adquirir al dejarme ser parte del proyecto. El pertenecer a este proyecto y realizar mi tesis en el me dio la oportunidad de tener nuevas experiencias, como fueron las presentaciones de mi trabajo ante personas de otros países y que dominan estos temas, así como la publicación de mi primer artículo y la interacción con personas de otro tipo de estudios, como es el caso de mis compañeros lingüistas.

También quiero agradecer a todos los miembros del proyecto DIME que de diferente manera me ayudaron a la realización de mi tesis. Quiero agradecer a Ivonne López y a Hayde Castellanos por la ayuda que me dieron al momento de definir el esquema de etiquetado de POS, a Sergio Coria por la ayuda que medio para emplear el software que genera los árboles de decisión y en la creación de los mismos, y a Varinia Estrada, Isabel López y Fernanda López, que junto con los mencionados y Patricia Pérez (Paty), llevaron a cabo el proceso de etiquetado del corpus DIME en algunos de los niveles definidos en este trabajo.

Quisiera agradecer a Carlos Rodríguez y a Javier Cuétara por los consejos que me dieron al momento de estar desarrollando mi tesis, al igual quisiera agradecer a James Allen (Rochester University) y a Joaquín Llisterrí (Universidad Autónoma de Barcelona) por los consejos que me dieron cuando realicé la presentación de mi trabajo para a ellos. También quiero agradecerle a Cesar Gamboa por el apoyo técnico que proporciona al IIMAS para poder llevar a cabo el etiquetado, los análisis y las pruebas tanto de mi trabajo como el de los demás estudiante e investigadores del instituto.

Finalmente quiero agradecerle a mi gerente Claudia Veytia y al consultor sinior Hugo García por el apoyo que me brindaron para poder terminar esta tesis, ya que me dieron tiempo para seguir trabajando con mi tesis y llevar a cabo lo necesario para realizar mi proceso de titulación.

Al trabajar en esta tesis pude conocer el mundo de la investigación, el cual me ha dejado muchas experiencias profesionales, de formación y de desarrollo de nuevas habilidades. En este punto aún no estoy seguro de cual va a ser mi siguiente paso a seguir y hacia donde voy a continuar, lo que sí sé es que aún me falta mucho por conocer y aprender. Esta tesis es el resultado de un gran esfuerzo que culmina una etapa muy importante dentro de mi vida (es la cereza del pastel) y comienza otra todavía de mayor peso, mi vida como Ingeniero en Computación de la Universidad Nacional Autónoma de México, y en la cual pondré lo mejor de mi en lo que realice.

Esta tesis se realizó con el apoyo del proyecto 39380-A “DIME II Diálogos inteligentes multimodales en español” bajo la responsabilidad del Dr. Luis A. Pineda, y que se desarrolla en el departamento de Ciencias de la Computación del IIMAS en colaboración con la Universidad de Rochester y es apoyado por el CONACYT y la NSF; además, esta tesis se realizó, en parte, gracias al apoyo de una beca proporcionada por el CONACYT, con el número 7776

# Contenido

<b>AGRADECIMIENTOS .....</b>	<b>I</b>
<b>CONTENIDO .....</b>	<b>IV</b>
<b>LISTA DE FIGURAS .....</b>	<b>VI</b>
<b>LISTA DE TABLAS .....</b>	<b>VII</b>
<b>1 INTRODUCCIÓN .....</b>	<b>1</b>
1.1 EL RECONOCIMIENTO DEL HABLA .....	3
1.2 LA TESIS.....	5
<b>2 FENÓMENOS DEL HABLA ESPONTÁNEA .....</b>	<b>7</b>
2.1 SEGMENTACIÓN .....	7
2.1.1 <i>Elocución</i> .....	9
2.2 REPARACIONES DEL HABLA .....	10
2.2.1 <i>Clasificación de las reparaciones del habla</i> .....	12
2.3 MARCADORES DEL DISCURSO .....	14
2.4 LA RELACIÓN ENTRE LOS FENÓMENOS.....	15
2.4.1 <i>Segmentación y Marcadores del discurso</i> .....	15
2.4.2 <i>Reparaciones del habla y Marcadores del discurso</i> .....	16
2.4.3 <i>Reparaciones del habla y Segmentación</i> .....	17
<b>3 EL CORPUS DIME.....</b>	<b>18</b>
<i>Herramienta de etiquetación</i> .....	20
3.1 ANOTACIÓN DE LOS NIVELES BASE.....	22
3.1.1 <i>Nivel alofónico</i> .....	22
3.1.2 <i>Nivel de sílabas</i> .....	23
3.1.3 <i>Nivel de palabras</i> .....	24
3.2 ANOTACIÓN DE LAS REPARACIONES DEL HABLA .....	25
3.2.1 <i>Estructura de una reparación del habla</i> .....	26
3.2.2 <i>Clasificación de las reparaciones del habla</i> .....	27
3.2.3 <i>Relaciones en la reparación</i> .....	28
3.2.4 <i>Multirreparaciones</i> .....	32
3.3 ANOTACIÓN DE LAS CATEGORÍAS LÉXICAS (POS).....	33
3.3.1 <i>Definición de conjunto de etiquetas</i> .....	33

3.3.2	<i>Etiquetación</i> .....	35
3.4	ANOTACIÓN DE LOS MARCADORES DEL DISCURSO .....	44
3.4.1	<i>Etiquetación</i> .....	45
3.5	ANOTACIÓN DE ÍNDICES DE SEPARACIÓN DE PALABRAS (BREAK ÍNDICES) .....	49
3.5.1	<i>Etiquetación</i> .....	50
3.6	EL ETIQUETADO .....	51
<b>4</b>	<b>DETECCIÓN DE UNA REPARACIÓN .....</b>	<b>54</b>
	<i>Duración de la elocución</i> .....	54
	<i>Número de palabras en una elocución</i> .....	57
	<i>Silencio en una elocución</i> .....	60
	<i>Acto del diálogo</i> .....	62
4.1	DEFINICIÓN DE ÁRBOLES DECISIÓN .....	63
<b>5</b>	<b>CORRECCIÓN DE UNA REPARACIÓN.....</b>	<b>72</b>
5.1	ESTRATEGIA DE CORRECCIÓN .....	72
5.2	ALGORITMO DE CORRECCIÓN .....	79
<b>6</b>	<b>CONCLUSIONES .....</b>	<b>84</b>
	TRABAJO FUTURO.....	86
	<b>APÉNDICE A – ALFABETO FONÉTICO MEXBET .....</b>	<b>88</b>
	<b>REFERENCIAS .....</b>	<b>90</b>

## Lista de Figuras

FIGURA 1.1: ESQUEMA DE UN SISTEMA DE DIÁLOGO .....	2
FIGURA 1.2: RESPUESTA DE UN RECONOCEDOR DE HABLA.....	3
FIGURA 3.1: ESCENARIO DEL CORPUS DIME.....	19
FIGURA 3.2: BARRA DE HERRAMIENTAS DEL <i>SPEECH VIEW</i> .....	20
FIGURA 3.3: SEÑAL DE AUDIO CON ESPECTRO (D03 – UTT17).....	21
FIGURA 3.4: SEÑAL DE AUDIO CON ENERGÍA E INTENSIDAD (D03 – UTT17).....	21
FIGURA 3.5: SEÑAL DE AUDIO CON <i>LABEL WINDOW</i> (D03 – UTT17) .....	22
FIGURA 3.6: NIVEL ALOFÓNICO.....	23
FIGURA 3.7: NIVEL DE SÍLABAS.....	24
FIGURA 3.8: NIVEL DE PALABRAS .....	25
FIGURA 3.9: NIVEL DE REPARACIONES DEL HABLA .....	31
FIGURA 3.10: ETIQUETACIÓN DE UNA MULTIRREPARACIÓN .....	32
FIGURA 3.11: NIVEL DE POS.....	44
FIGURA 3.12: ETIQUETACIÓN DE POS CON MARCADORES DEL DISCURSO .....	44
FIGURA 3.13: NIVEL DE MARCADORES DEL DISCURSO .....	49
FIGURA 3.14: NIVEL DE <i>BREAK INDICES</i> .....	51
FIGURA 3.15: PORCENTAJES DE ELOCUCIONES CON REPARACIÓN POR DIÁLOGO .....	52
FIGURA 4.1: GRÁFICA DE DURACIONES .....	55
FIGURA 4.2: CLASIFICACIÓN DE LAS ELOCUCIONES EN RANGOS DE 2,000 MILISEGUNDOS .....	56
FIGURA 4.3: DISFLUENCIAS SEGÚN EL NÚMERO DE PALABRAS EN LA ELOCUCIÓN.....	57
FIGURA 4.4: NÚMERO DE PALABRAS: REGIÓN 1 – 6 O MENOS PALABRAS.....	58
FIGURA 4.5: NÚMERO DE PALABRAS: REGIÓN 2 – ENTRE 7 Y 15 PALABRAS.....	58
FIGURA 4.6: NÚMERO DE PALABRAS: REGIÓN 3 – MÁS DE 15 PALABRAS.....	60
FIGURA 4.7: GRÁFICA DE LA DURACIÓN DE LOS SILENCIOS .....	61
FIGURA 4.8: CLASIFICACIÓN DE LOS SILENCIOS EN RANGOS DE 500 MILISEGUNDOS .....	62
FIGURA 4.9: EJEMPLO DE ÁRBOL DE DECISIÓN .....	63
FIGURA 4.10: ÁRBOL DE DECISIÓN PARA LA DETECCIÓN DE DISFLUENCIAS DEL HABLA – PRIMERA VARIANTE .	67
FIGURA 4.11: ÁRBOL DE DECISIÓN PARA LA DETECCIÓN DE DISFLUENCIAS DEL HABLA – SEGUNDA VARIANTE	70
FIGURA 5.1: DISTRIBUCIÓN DE LAS REPARACIONES DEL HABLA.....	74
FIGURA 5.2: RELACIÓN DE PALABRAS REPETIDAS POR TIPO DE CORRECCIÓN DE UN <i>MODIFICATION REPAIR</i> .....	76
FIGURA 5.3: DISTANCIA ENTRE PALABRAS REPETIDAS .....	78
FIGURA 5.4: DISTANCIA AJUSTADA ENTRE PALABRAS REPETIDAS.....	79



## Lista de Tablas

TABLA 1.1: FRAGMENTO DE UN DICCIONARIO DE PRONUNCIACIÓN.....	4
TABLA 3.1: TABLA DE CONVERSIÓN.....	25
TABLA 3.2: ETIQUETAS DE REPARACIONES DEL HABLA. ....	28
TABLA 3.3: CATEGORÍAS LÉXICAS BÁSICAS. ....	34
TABLA 3.4: ETIQUETAS PARA LA IDENTIFICACIÓN DE CATEGORÍAS LÉXICAS. ....	35
TABLA 3.5: ETIQUETAS DE MARCADORES DEL DISCURSO. ....	46
TABLA 3.6: ETIQUETAS PARA <i>BREAK INDICES</i> .....	50
TABLA 3.7: NÚMERO DE ELOCUCIONES ETIQUETADAS POR DIÁLOGO.....	51
TABLA 3.8: RELACIÓN DEL NÚMERO DE REPARACIONES POR CADA DIÁLOGO .....	52
TABLA 4.1: PORCENTAJE DE REPARACIONES POR CADA 2,000 MILISEGUNDOS.....	56
TABLA 4.2: RELACIÓN DE DURACIÓN VS. NÚMERO DE PALABRAS .....	59
TABLA 4.3: RESULTADOS DE LA PRUEBA DE VALIDACIÓN CRUZADA PARA LA PRIMERA VARIANTE DEL EXPERIMENTO.....	66
TABLA 4.4: RESULTADOS DE LA PRUEBA DE PORCENTAJE DE DIVISIÓN PARA LA PRIMERA VARIANTE DEL EXPERIMENTO.....	66
TABLA 4.5: RESULTADOS DE LA PRUEBA DE VALIDACIÓN CRUZADA PARA LA SEGUNDA VARIANTE DEL EXPERIMENTO.....	68
TABLA 4.6: RESULTADOS DE LA PRUEBA DE PORCENTAJE DE DIVISIÓN PARA LA SEGUNDA VARIANTE DEL EXPERIMENTO.....	69
TABLA 5.1: NÚMERO DE MULTIRREPARACIONES POR DIÁLOGO.....	73
TABLA 5.2: PORCENTAJE DE MULTIRREPARACIONES CON RESPECTO AL NÚMERO DE REPARACIONES .....	73
TABLA 5.3: COMPARACIÓN DEL NÚMERO DE REPARACIONES DIVIDIDAS Y SIN DIVIDIR.....	74
TABLA 5.4: RESULTADOS DE LA IMPLEMENTACIÓN DEL ALGORITMO DE CORRECCIÓN .....	83
TABLA 6.1: RELACIÓN DE PALABRAS QUE MÁS SE REPITEN.....	85
TABLA - APÉNDICES 1: ALFABETO MEXBET.....	88

## 1 Introducción

d13 – utt001: *¿quieres que traiga algún mueble a la cocina?*

---

Y así es como comienza

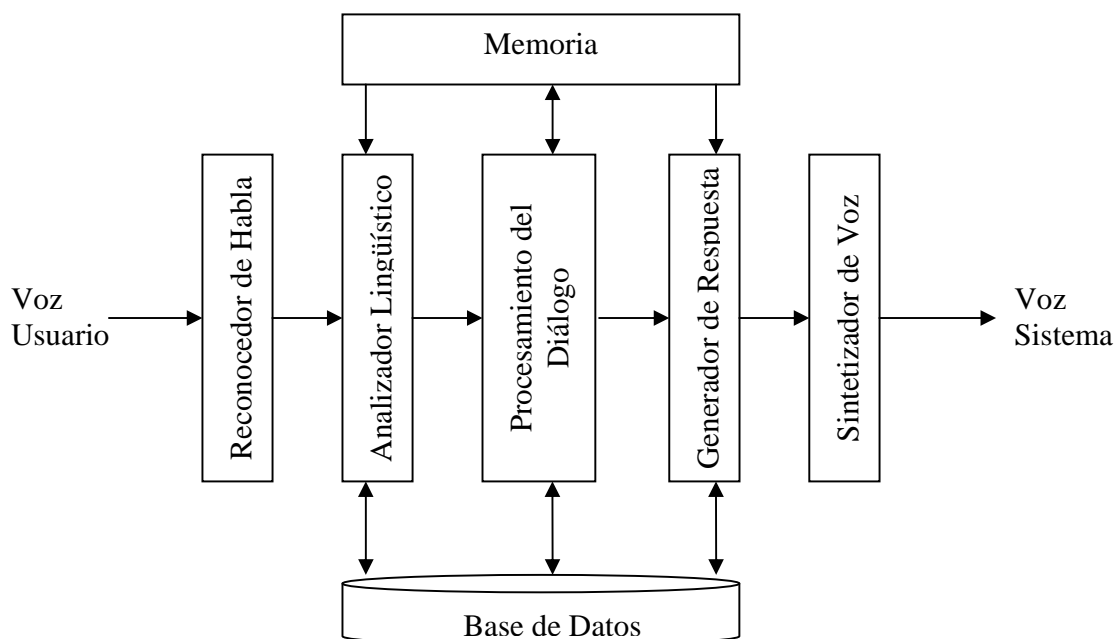
Uno de los principales propósitos del procesamiento del lenguaje natural y del reconocimiento del habla es construir sistemas computacionales capaces de llevar a cabo una conversación en lenguaje natural.

Con el estado actual de la tecnología se pueden construir sistemas de diálogo enfocados a una tarea específica y en los cuales se utilice un lenguaje simple y gramaticalmente bien formado, como son los sistemas telefónicos de atención a clientes.

Los sistemas de diálogo, también conocidos como sistemas conversacionales, son sistemas computacionales que reciben como entrada frases del lenguaje natural expresadas de forma oral y generan como salida frases del lenguaje natural también expresadas en forma oral [24]. El propósito principal de estos sistemas es generar diálogos entre un ser humano y una computadora, simulando una conversación natural orientada normalmente a la situación de una tarea de manera cooperativa.

Estos sistemas se han convertido actualmente en una herramienta para el desarrollo de servicios telefónicos entre los que se pueden encontrar la atención al cliente, la banca telefónica o la venta de productos. Recientemente se han alcanzado desarrollos importantes en otras aplicaciones como son la traducción automática del habla o el uso de las tecnologías lingüísticas para ayudar a las personas con necesidades especiales [17].

Un esquema general de un sistema de diálogo ([24], [37]) se muestra a continuación:



**Figura 1.1: Esquema de un sistema de diálogo**

El reconocedor de habla es la parte del sistema de diálogo que procesa la señal de voz (sonido). El reconocedor genera un conjunto de hipótesis ponderadas de lo que el hablante dice, normalmente tomando como opción preferida la que tenga la mayor probabilidad. El texto generado por el reconocedor es la entrada para el analizador lingüístico, cuya finalidad es encontrar una representación del significado del texto.

Una vez generada esta representación, se procede a procesar el texto. En esta parte del sistema se determina que acciones son las que se deben realizar; el propósito es generar una interacción con una persona de la forma más natural posible. Para lograr este objetivo, se suele realizar confirmaciones de los datos obtenidos del usuario, iniciado de esa manera subdiálogos de comprobación y corrección.

Para llevar a cabo la generación de estos subdiálogos, el generador de respuesta construye frases en texto que corresponden a las acciones hechas al momento de procesar el diálogo. Este texto es la entrada para el sintetizador de voz, el cual genera la misma respuesta pero en forma oral.

Debido a la complejidad del lenguaje natural, los sistemas de diálogo se enfocan principalmente a resolver una tarea específica, como puede ser asistencia telefónica, realizar visitas guiadas, diseño de interiores, etc. El propósito principal es reducir el campo de trabajo del sistema de diálogo para poder resolver la tarea de una manera satisfactoria;

con esto, se evita ambigüedad en el diálogo así como complejidad en los métodos de solución del problema y se pueden aplicar los conocimientos específicos para resolver una tarea en particular.

En la actualidad, los sistemas de diálogo no están restringidos a sólo recibir como entrada la voz para realizar su funcionamiento, también pueden emplear otros medios. A los sistemas que admiten más de una entrada se les denomina sistemas de diálogo multimodales. Las ventajas que presentan este tipo de sistemas se reflejan en la comodidad del usuario para poder realizar una tarea.

Las entradas a un sistema de diálogo multimodal varían desde el empleo de teclado y *mouse*, hasta imágenes, señales y video; de la misma forma, la respuesta de los sistemas multimodales también puede variar, y no solamente utilizar la respuesta oral como medio de comunicación, ya que también pueden ser acciones motoras, gráficas, etc.

## 1.1 El reconocimiento del habla

Como se mencionó antes, la respuesta de un sistema reconocedor de habla es un conjunto de hipótesis ponderadas de lo que se dijo, y generalmente se toma como opción válida la que tenga la mayor probabilidad; un ejemplo del tipo de respuesta se muestra en la Figura 1.2.

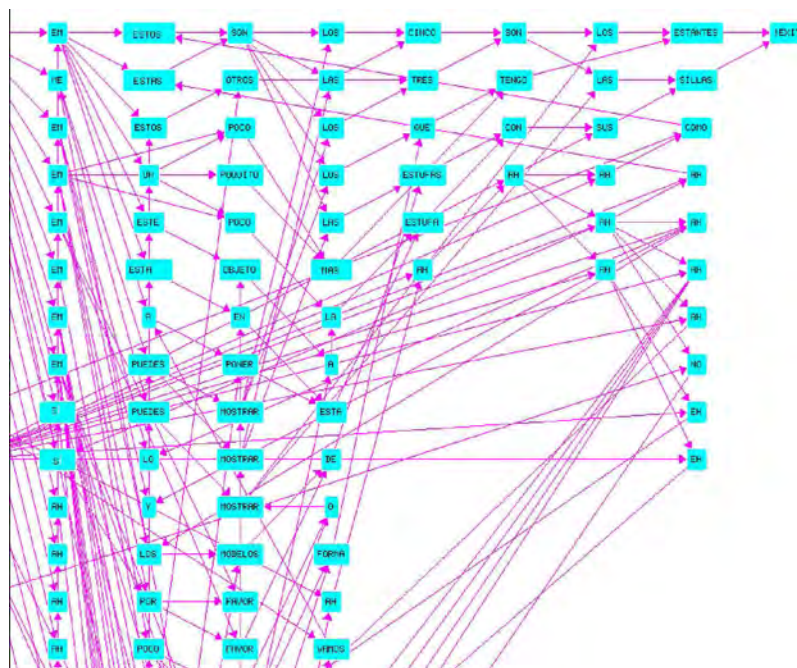


Figura 1.2: Respuesta de un reconocedor de habla

Cada trayectoria representa una hipótesis y de ellas se elegir como respuesta la que tenga mayor ponderación; sin embargo, la hipótesis seleccionada no siempre está formada con las palabras que el hablante dijo.

Una hipótesis se forma principalmente por los siguientes factores: el modelo acústico, el diccionario de pronunciación y el modelo de lenguaje.

El modelo acústico es la abstracción del sonido utilizando símbolos; al conjunto de estos símbolos se le denomina *alfabeto fonético*. Los modelos acústicos se definen en función de una serie de patrones que son comparados con la señal entrante y, de acuerdo al modelo al que más se parezca la señal, se le asigna el símbolo que representa dicho modelo. Posteriormente, por medio del diccionario de pronunciación, se toman los símbolos y se transforman en palabras por medio de símbolos ortográficos, *alfabeto ortográfico*. Los diccionarios de pronunciación codifican a las palabras representadas mediante el alfabeto fonético; un ejemplo de esto se puede observar en la Tabla 1.1.

<b>Representación ortográfica</b>	<b>Representación fonética</b>
actividades	a k t i b i a d e
actividades	a k t i b i a e s
actividades	a k t i b i d a d e s
actividades	a k t i b i d a e s

**Tabla 1.1: Fragmento de un diccionario de pronunciación**

Como parte final, la decisión de que palabras se toman y el orden en que aparecen se rige por el modelo del lenguaje. Los modelos del lenguaje están formados por modelos estadísticos que generan la secuencia de palabras que tienen la mayor probabilidad de ocurrencia.

En un caso ideal, los hablantes se expresan de una manera fluida sin interrupciones ni pausas y de forma gramaticalmente correcta. El reconocedor del habla toma la señal de audio que el hablante genera y se procesa; el resultado obtenido es un texto con exactamente las palabras pronunciadas; esta frase es la que se envía al parse. Sin embargo, en las conversaciones cotidianas, los hablantes expresan interjecciones (*aja, mm, eh*), hacen pausas, cambian o repiten palabras de lo que dicen o simplemente abandonan la elocución

para comenzar de nuevo; a todos estos fenómenos se les llama *disfluencias del habla*. El Ejemplo 1 muestra como son este tipo de fenómenos.

**Ejemplo 1 (d01 – utt33<sup>1</sup> Disfluencia del habla)**

*porque ten <.sil><sup>2</sup> tenemos que mover <.sil> los <.sil> los objetos que acabas de <.sil> bueno el objeto que acabo de colocar éste <.sil>*

La presencia de disfluencias del habla dificulta la tarea del análisis lingüístico debido a que no tienen una regla que los rige y son eventos aleatorios, razón por la cual es de gran utilidad implementar un módulo que detecte y corrija lo mejor posible este tipo de fenómenos. A continuación se muestra la corrección del Ejemplo 1.

**Ejemplo 1 (d01 – utt33 Disfluencia del habla – corregida)**

*porque tenemos que mover el objeto que acabo de colocar éste*

## **1.2 La tesis**

En esta tesis se presenta el fenómeno de las reparaciones del habla, su relación con otro tipo de fenómenos del habla espontánea y un método para realizar la detección y corrección de ellas.

En el siguiente capítulo se da una amplia descripción de los diferentes fenómenos, así como la relación que existe entre ellos, y se muestran que la solución de un fenómeno ayuda a la solución de los demás; para este efecto se utiliza como base empírica el corpus DIME [38].

En el capítulo 3 se presenta una pequeña introducción del corpus DIME; además, se propone un esquema de análisis o etiquetado para las disfluencias del habla; se da la definición de cada uno de los niveles que componen el esquema, además del conjunto de etiquetas que forman parte de cada nivel con ejemplos de uso de cada etiqueta. Asimismo, se da a conocer la porción del corpus etiquetado en los niveles definidos. Los niveles de transcripción que se definen en esta tesis sirven para analizar el comportamiento de las

---

<sup>1</sup> Con esta notación se identifica cada una de las elocuciones, está formado por el número del diálogo (dXX) y por el número de la elocución (uXX).

<sup>2</sup> Este símbolo tiene como propósito representar la aparición de un silencio dentro de la elocución.

reparaciones del habla; sin embargo no todos los niveles fueron utilizados al momento de realizar la detección y corrección de las reparaciones. Estos niveles adicionales pueden utilizarse para una investigación posterior.

En el capítulo 4 se muestra el análisis realizado de las disfluencias del habla para encontrar la información útil que pueda ser utilizada al momento de la detección de las reparaciones. El método seleccionado para llevar a cabo esta tarea es la creación de árboles de decisión. También se reportan cuáles fueron los experimentos que se hicieron, sus características y el resultado obtenido.

En el capítulo 5 se presenta el análisis para la definición de una heurística y su implementación para realizar la corrección de las reparaciones del habla. En esta investigación, se presenta la solución para las reparaciones del habla que se presentaron con mayor frecuencia en el corpus.

Para evaluar el desempeño tanto del el árbol de decisión como de la heurística de corrección se toman como parámetros las variables estadísticas *precisión* y *recall*. La precisión es la relación del número de casos identificados correctamente entre el total de casos que el algoritmo o experimento consideró correctos:

$$precisión = \frac{casos\ identificados\ correctamente}{casos\ identificados\ correctamente + casos\ identificados\ incorrectamente}$$

El *recall* es la relación de los casos que fueron identificados correctamente entre el total de casos que se debían identificar como correctos.

$$recall = \frac{casos\ identificados\ correctamente}{casos\ que\ se\ deberían\ de\ identificar}$$

Por último, se presentan los resultados obtenidos y se sugieren futuros análisis que mejorarían el proceso de detección y corrección de las reparaciones.

## 2 Fenómenos del habla espontánea

d17 – utt148: <.sil> la estufa pegar vamos a quitar <.sil> a <.sil> a intercambiar vamos a poner este fregadero esto <.sil> de este lado de acá y la estufa de este lado de acá <.sil>

---

¿Cómo dice que dijo?

En esta sección se presentan los diferentes fenómenos que ocurren en el habla espontánea y que se relacionan con el fenómeno de las disfluencias del habla, además se muestra que la solución de dichas disfluencias está ligada con la solución de los otros fenómenos.

No sólo las disfluencias del habla pueden provocar problemas al momento de analizar un diálogo; existen también otros fenómenos relacionados como son la segmentación del diálogo en unidades más pequeñas para facilitar su interpretación, así como la presencia de palabras cuya función es la de administrar el diálogo y dejan a un lado su función sintáctica; a este tipo de palabras se les denomina *marcadores del discurso*. A continuación veremos en qué consiste cada uno de estos fenómenos del habla espontánea y la relación entre cada uno de ellos.

### 2.1 Segmentación

En la forma general en la que se lleva a cabo una conversación, los hablantes alternan turnos para continuar con el intercambio de información. En cada turno, el hablante puede realizar varias contribuciones, tanto para realizar alguna tarea, como para mantener el buen entendimiento o el acuerdo entre ellos. En la mayoría de los casos, cada turno concentra una granularidad tan espesa de información, que no puede ser considerado viable como la unidad mínima para el entendimiento; por lo anterior, es necesario hacer una



división de cada turno del hablante para entender lo que realmente se quiere decir; a los fragmentos obtenidos por la segmentación de un turno se le denomina *elocuciones*.

Consideremos el siguiente ejemplo tomado del corpus DIME [38].

**Ejemplo 2 (d03 – [utt2-6])**

*mm que me traiga no que me m[uevas]<sup>3</sup> que muevas el no sé cómo se le llama a las gabinete no sé a esto que está en la pared izquierda sí ah es lo único que está en la pared izquierda sobre el piso*

El texto anterior es un turno de un hablante; el análisis de este texto es sumamente complicado, la tarea de segmentación puede tomar bastante tiempo y volverse pesada; una segmentación posible es como sigue:

**Ejemplo 2 (Segmentado)**

utt2 : mm

utt3 : que me traiga no que me m[uevas] que muevas el

utt4 : no sé cómo se le llama a las gabinete no sé

utt5 : a esto que está en la pared izquierda

utt6 : sí ah es lo único que está en la pared

En este turno se pueden observar los diferentes fenómenos que ocurren, como es la presencia de marcadores del discurso en la elocución 2, las características de los marcadores del discurso son explicadas posteriormente. En la elocución 3 se presentan reparaciones del habla; el hablante corrige la primera cláusula de lo que dice (cambia “que me traiga” por “que me m[uevas]”) y después vuelve a corregir para reestructurar lo que dice (cambia “que me m[uevas] por “que muevas el”). En la elocución 4 existe la repetición de las palabras “no sé” al principio y al final de ella; en este caso no se trata de algún tipo de reparación sino de una reafirmación, lo que indica un error de segmentación. La segunda realización de “no sé” reafirma lo que se dijo anteriormente, “no sé cómo se le llama a las gabinetes”. Por último, la elocución 6 también presenta reparación, el hablante hace una pausa para después continuar con lo que estaba diciendo. La pausa ocurre cuando el

---

<sup>3</sup> El texto entre corchetes no fue dicho por el hablante

hablante dice “ah”, y al igual que la palabra “mm” de la elocución 2, se trata de un marcador del discurso. Aún cuando se cuente con la segmentación del turno del hablante bien definida, la intención final no está todavía clara por todos los fenómenos presentes.

En el ejemplo siguiente se muestra como quedaría el turno eliminando las reparaciones del habla:

### **Ejemplo 2 (Sin reparaciones del habla)**

utt2 : mm

utt3 : que muevas el

utt4 : no sé cómo se le llama a las gabinete no sé

utt5 : a esto que está en la pared izquierda

utt6 : sí es lo único que está en la pared

Existe una estrecha relación entre los problemas de segmentación, reparaciones del habla e identificación de marcadores del discurso; normalmente un marcador del discurso se presenta al inicio de una nueva elocución, por ejemplo, para relacionar la elocución previa y la siguiente; ayudando de esta forma a realizar la segmentación del turno. De la misma forma, el marcador del discurso se relaciona con las reparaciones del habla, y en ocasiones sirve para identificar que una reparación ocurre. A continuación se presentan las características de la elocución.

### **2.1.1 Elocución**

En el lenguaje hablado las elocuciones no son siempre gramaticales; en la expresión oral se presentan frecuentemente oraciones fragmentadas, subordinaciones, errores del habla, reafirmaciones, repeticiones, abandono de oraciones, entre otros, los cuales rompen los esquemas gramaticales normales de la expresión escrita; a continuación se muestran algunos ejemplos de estos fenómenos:

### **Ejemplo 3 (d01 – utt79 Abandono)**

*bueno necesito un gabinete de pared pero a ver*

**Ejemplo 4 (d01 – utt16 Reparación del habla)**

*tendría <sil> tendríamos que girarlo*

**Ejemplo 5 (d01 – utt03 Repetición)**

*ay es que no no recuerdo bien qué era eso*

Las elocuciones ayudan a llevar el acuerdo entre el oyente y el hablante durante un diálogo; por lo tanto, una definición apropiada de los límites de una elocución ayuda al oyente a entender que es lo que el hablante está diciendo [6].

La definición de una unidad de elocución no tiene un consenso bien definido; sin embargo, de manera intuitiva se puede decir que una elocución es una unidad con las siguientes características [6]:

- Tiene una estructura sintáctica y/o semántica
- Define un acto del habla sencillo. Un acto del habla es la acción lingüística que expresa una intención, como puede ser preguntar, ordenar, describir, afirmar, solicitar, etc.
- Es una frase entonativa.
- Es una unidad separada por pausas

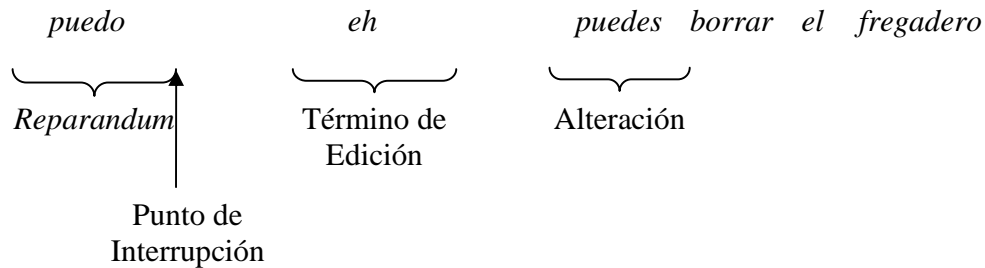
**2.2 Reparaciones del habla**

En la conversación cotidiana frecuentemente el hablante expresa sus ideas e intenciones antes de estar totalmente seguro de que es lo que realmente quiere decir o hacer y consecuentemente, en ocasiones, se requiere corregir la elocución. Las correcciones se dan en varias modalidades, por ejemplo, puede interrumpirse la elocución y corregir o repetir algo de lo que dice; puede hacer una pausa con ayuda de alguna palabra o frase mientras piensa que es lo que va a decir, e incluso puede abandonar lo que decía para expresar una nueva idea; a todos estos tipos de correcciones se les denomina reparaciones del habla, es decir, disfluencias del habla en donde el hablante repara lo que dice.

Las razones por las cuales las reparaciones se presentan pueden ser muy variadas, pero el hecho es que se dan con toda naturalidad en los diálogos hablados. En trabajos previos [34] se ha visto que las reparaciones del habla tienden a formar una estructura

estándar, formada por 3 elementos hablados y 1 tonal. Los elementos hablados son: el *reparandum*, el término de edición y la alteración; el elemento tonal se le llama punto de interrupción. Estos elementos se pueden observar en el Ejemplo 6.

**Ejemplo 6 (d23 – utt31 Estructura de una reparación del habla)**



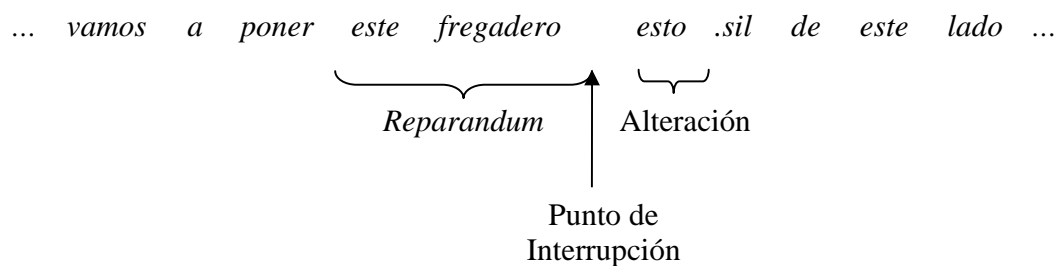
El *reparandum* es la parte de la elocución que el hablante desea reemplazar. Generalmente, la última palabras del *reparandum* es solamente un fragmento ya que el hablante interrumpe lo que decía.

Al final del *reparandum* se presenta el elemento tonal, al que se denomina punto de interrupción; y usualmente está acompañado de una disrupción en la entonación.

El término de edición es la parte que le sigue al punto de interrupción; éste puede realizarse como una interjección como son: “ah”, “mm” o “eh”; o ciertas frases como “es decir”, “bueno” o “perdón”.

La última parte de la reparación es la alteración; ésta es la parte de la elocución que el hablante quiere decir y que reemplaza al *reparandum*; sin embargo, la alteración no puede simplemente sustituir al *reparandum* y al término de edición ya que pueden contener información útil; por ejemplo, cuando en el *reparandum* existe un sustantivo y en la alteración se presenta únicamente el pronombre; en este caso, si se elimina el *reparandum*, se pierde la información exacta de quien o que se está refiriendo, como se muestra en el Ejemplo 7.

**Ejemplo 7 (d17 – utt148)**



Como se puede ver en el Ejemplo 7, si simplemente se sustituyera el *reparandum* por la alteración se perdería la referencia de que es lo que se quiere poner, en este caso el fregadero.

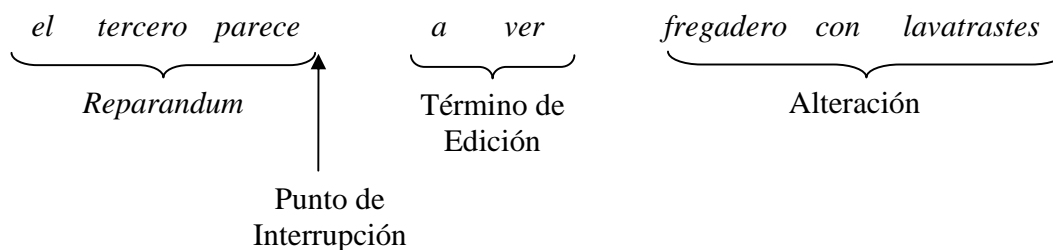
Para mantener la continuación del diálogo, el oyente debe identificar a la reparación para proceder a resolverla; para este fin es necesario identificar la extensión de cada una de las partes que conforman la reparación del habla.

### 2.2.1 Clasificación de las reparaciones del habla

En este trabajo se tomó la clasificación propuesta por Heeman (1997) [35]; ésta se basa principalmente en las similitudes que existen entre el *reparandum* y la alteración; por ejemplo, si en la alteración se repiten palabras del *reparandum* es posible que se agregue nueva información, se trate de componer algún error presente en el *reparandum* o se pretenda expresar más apropiadamente alguna idea. En este esquema se consideran 3 tipos de reparaciones del habla: reinicio, reparación por modificación y reparación por pausa<sup>4</sup>.

El reinicio ocurre cuando se abandona la elocución para empezar de nuevo; en ocasiones, el abandono se presenta con una señal acústica notoria, ya sea en el término de edición o en el principio de la alteración.

#### Ejemplo 8 (d03 – utt55 Reinicio)

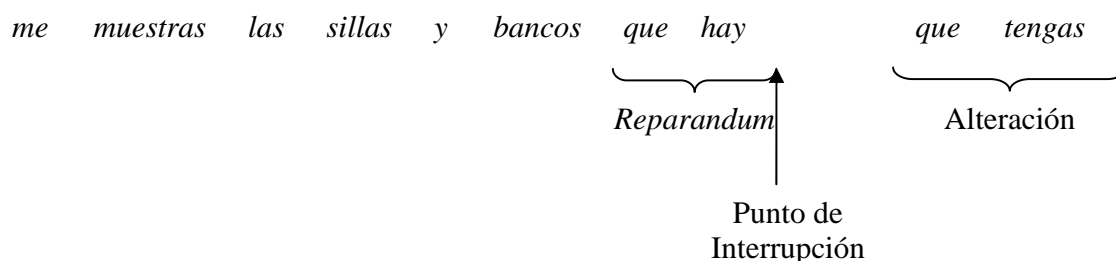


En reparaciones de reinicio no existe relación alguna, o extremadamente baja, entre el *reparandum* y la alteración. En el Ejemplo 8 la elocución se abandona (“el tercero parece”) y se expresa otra idea (“fregadero con lavatrastes”). Esta clase de reparación tiene una fuerte señal de interrupción acústica en la elocución.

<sup>4</sup> Para Heeman estas categorías son denominadas: *fresh start*, *modification repair* y *abridged repair*, respectivamente

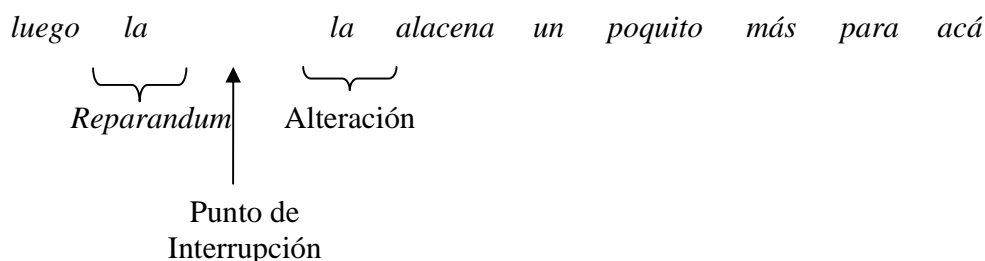
El segundo tipo corresponde a las reparaciones por modificación. En este tipo se encuentran las reparaciones que tienen *reparandum* no vacío.

**Ejemplo 9 (d12 – utt93 Reparación por modificación)**



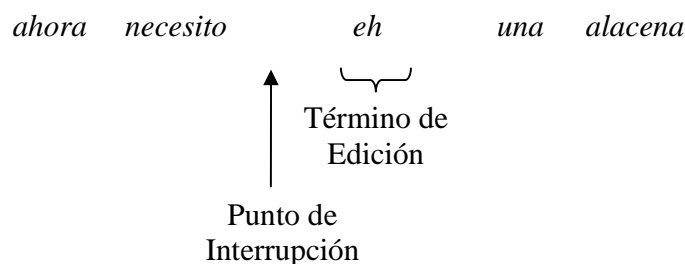
En esta clase de reparaciones existe una relación significativa entre el *reparandum* y la alteración. Esta característica es muy útil para determinar la extensión del *reparandum*. En el Ejemplo 9, el hablante reemplaza “que hay” por “que tengas”, donde la palabra “que” se comparte en el *reparandum* y la alteración, y la palabra “hay” es reemplazada por la palabra “tengas”. En general la reparación por modificación ocurre solamente en una parte de la elocución y frecuentemente la alteración es la repetición exacta del *reparandum*, como se aprecia en el Ejemplo 10:

**Ejemplo 10 (d17 – utt185 Repetición)**



La última clase de reparación es la reparación por pausa. En esta clase se encuentran las reparaciones que únicamente presentan el término de edición y no tienen *reparandum*.

**Ejemplo 11 (d17 – utt185 Por pausa)**



La dificultad que presenta esta clase de reparación es la de identificar cuando se trata de un término de edición si éste es una frase como: “bueno”, “entonces” o “a ver”. Normalmente este tipo de reparación ocurre cuando el hablante piensa en que es lo que va a decir.

### **2.3 Marcadores del discurso**

Los marcadores del discurso son palabras que pierden su categoría gramatical; por ejemplo: “bueno”, “entonces”, “ahora” y “y”; son unidades lingüísticas invariables, que no ejercen una función sintáctica en el marco oracional y poseen el cometido de guiar, de acuerdo con sus distintas propiedades, las inferencias que se realizan en la comunicación; dar al oyente información acerca de la estructura del discurso o ayudar a resolver anáforas ([14], [28], [33], [34])

Los marcadores del discurso pueden clasificarse dependiendo de su función, como sigue [14]:

- Estructurar de la información: Desarrollan la información contenida en un discurso (“pues bien”, “dicho esto”, “por otra parte”, “primero”, “otra cosa”, etc.).
- Conectar: Vinculan a un miembro del discurso con otro anterior (“además”, “incluso”, “es más”, “eso sí”, etc.).
- Reformular: Refrasean lo que ya se ha dicho (“o sea”, “es decir”, “en otras palabras”, “mejor dicho”, “digo”, “o”, “en realidad”, etc.).
- Ejemplificar: Condicionan las posibilidades argumentativas del miembro del discurso en que se incluyen, pero sin relacionarlo con otro miembro anterior (“por ejemplo”, “en particular”, etc.)
- Tienen función conversacional.(“bien”, “buen”, “okey”, “claro”)

Al ser superficialmente como una palabra normal, la identificación de los marcadores del discurso puede complicarse y ser un tanto ambigua.

#### **Ejemplo 12 (d12 – utt25 Marcador del discurso)**

*después .sil me puedes poner .sil l el extractor de aire encima de la .sil de la estufa*

En el Ejemplo 12 se presenta un caso en el que la palabra “después” pierde su categoría léxica de adverbio y queda como un marcador del discurso que sirve como enlace del diálogo. Se puede ver que a esta palabra le sigue un silencio, lo cual indica que en realidad “después” se usa para mantener el nivel de entendimiento y no perder la interacción de la comunicación.

Por otro lado se puede observar que la palabra “después” marca el inicio de una nueva unidad de elocución.

## **2.4 La Relación entre los fenómenos**

Los problemas de identificación de marcadores del discurso, segmentación (identificación de los límites de entonación) y la resolución de las reparaciones del habla están estrechamente relacionados. Para solucionar cada uno de estos problemas se requiere solucionar los otros dos.

A continuación se describe la relación que existe entre cada uno de los fenómenos y la manera en como se ayudan para llegar a la solución de cada uno de ellos.

### **2.4.1 Segmentación y Marcadores del discurso**

Un uso común entre los marcadores del discurso es el de inicializar una nueva unidad de elocución. El Ejemplo 13 muestra un fragmento de un diálogo en el cual el marcador del discurso “ahora” divide el turno en dos unidades.

#### **Ejemplo 13 (d14 – [utt23 – 26] Segmentación y Marcadores del discurso)**

utt23 : s:       ahí está bien ?

utt24 : u:       sí

utt25 :       ahora quiero un refrigerador

utt26 : s:       éstos son los tres tipos de modelos de refrigeradores

La palabra “sí” expresa que se acepta una acción y por lo cual adquiere una función de marcador del discurso. La palabra “ahora”, en este ejemplo, no se refiere a un adverbio de tiempo; más bien, su función es demostrar que el hablante terminó con la intención anterior y que continuará con una nueva idea; por lo que su función es la de marcador del discurso; y en este caso la funcionalidad de “ahora” hace evidente la segmentación del



turno del hablante en dos unidades de elocución. En este ejemplo se puede observar como la identificación de los marcadores del discurso facilita el proceso de segmentación.

## 2.4.2 Reparaciones del habla y Marcadores del discurso

Los marcadores del discurso son de gran ayuda para la detección de las reparaciones del habla. En primer lugar, el término de edición está formado normalmente por marcadores del discurso.

### Ejemplo 14 (d01 – utt85 Reparaciones del habla y Marcadores del discurso)

*pero ahora me hace falta un me parece que es un lavatrastos o un*

Reparandum
Término de Edición
Alteración

↑  
 Punto de Interrupción

En el Ejemplo 14 se puede presentar una confusión al momento de definir las partes de la estructura de una reparación; si se analiza sólo el texto de la elocución, se podría decir que el *reparandum* está formado por las palabras “me hace falta un”; mientras que la alteración está formada por las palabras “me parece que es un lavatrastos” y no habría término de edición; sin embargo, al analiza el audio junto con el texto se pudo observar que en realidad la frase “me parece que es” no tiene una función sintáctica sino una función de administración del diálogo, la función de esta frase es darle tiempo al hablante para poder modificar lo que dice; de esta manera marcar la ocurrencia de una reparación del habla y ayuda a su identificación. En el ejemplo la intención del hablante es pedir un nuevo objeto, en este caso el lavatrastos.

Los marcadores del discurso ayudan a identificar el caso particular de la reparación del habla por pausa, debido a que este tipo de reparación consiste únicamente del término de edición.

### Ejemplo 15 (d19 – utt82 Reparaciones por pausa y Marcadores del discurso)

*puedes girarlo eh en el mismo sentido*

Término de Edición

↑  
 Punto de Interrupción

Un marcador del discurso ayuda a determinar dónde inicia la alteración y dónde termina el *reparandum*.

### 2.4.3 Reparaciones del habla y Segmentación

Hay también una relación estrecha entre la segmentación y las reparaciones del habla. Principalmente, un turno es más fácil de dividir en unidades de elocución si no se presentan disfluencias del habla [34]. Por otro lado, es fácil confundir cuando se trata de un límite entonativo, el cual hace la división en dos diferentes elocuciones, y cuando se trata de un punto de interrupción, que es la parte tonal que identifica cuando ocurre una reparación. La confusión es posible ya que después de cada uno de estos fenómenos le sigue generalmente una pausa.

El ejemplo más claro donde se presenta la confusión entre un punto de interrupción y un límite entonativo es en las reparaciones del tipo reinicio, como ya se dijo; en este tipo de reparaciones el hablante se detiene para comenzar de nuevo provocando confusión al momento de segmentar su turno.

Como se puede ver, las disfluencias del habla no es el único fenómeno que se presenta en el lenguaje natural y que dificulta su procesamiento; el problema de la identificación de marcadores del discurso y la segmentación del turno de un hablante también complica el entendimiento del lenguaje natural. A su vez, estos fenómenos están estrechamente relacionados y estas relaciones ayuda a encontrar su solución. Para realizar el análisis de las reparaciones del habla es necesario considerar las características útiles de todos estos fenómenos. Para definir qué características son útiles, en el capítulo siguiente se presenta el esquema de etiquetación utilizado para el análisis de las reparaciones. El esquema considera diferentes factores tales como la categoría léxica de las palabras, la presencia de los marcadores del discurso, la estructura de las reparaciones del habla y los límites entonativos de las palabras.

### 3 El corpus DIME

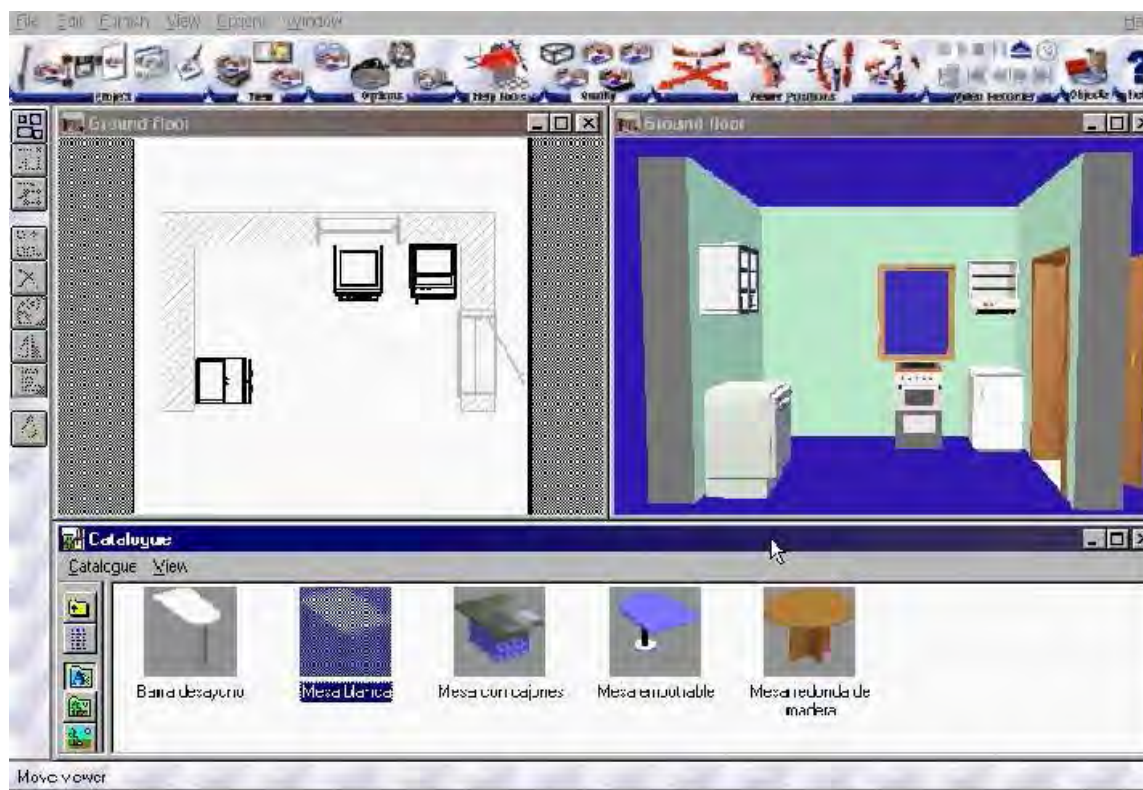
*d03 – utt067: <.bn> ajá no espérame entonces <.bn> el segundo <.bn> el cuarto por favor <.bn>*

---

¿Y ahora cómo la etiqueto?

Para llevar a cabo el estudio de las reparaciones del habla se utilizó el corpus DIME como base empírica. El corpus DIME es una recopilación de conversaciones orientadas hacia la solución de una tarea de diseño, en este caso el diseño de cocinas [38]. El experimento está basado en el escenario del Mago de Oz [5]. En este experimento se genera un ambiente conversacional humano–máquina para obtener la solución a una tarea a través de un diálogo cooperativo. En este escenario un humano toma el papel del mago (sistema) y a diferencia del escenario tradicional del Mago de Oz, el hablante sabe que el mago es realmente un ser humano.

En el experimento, el usuario tenía que resolver dos problemas, en el primero tenía que completar una cocina parcialmente diseñada, y en el segundo tenía que completarla pero sin algún diseño previo. El propósito de la tarea era llevar a cabo la planificación del diseño de una cocina por medio de un diálogo humano–máquina con lenguaje natural. Esta interacción es multimodal porque se utilizan las diferentes modalidades de la percepción: la vista (monitor), el habla (micrófono), el tacto (*mouse*). La interfaz se muestra en la Figura 3.1



**Figura 3.1: Escenario del corpus DIME**

Durante el experimento, toda la información lingüística (archivos de audio) y visual producida tanto por el hablante como por el mago se recopiló de manera automática [29].

El desarrollo de los diálogos fue llevado de acuerdo a las decisiones del hablante, es decir, no existió ningún guión que marcara cuales eran los pasos que tenía que haber dicho o hecho. En consecuencia, el diálogo se llevó de la forma más natural posible y se presentaron todos los fenómenos presentes en las conversaciones cotidianas: la resilabificación<sup>5</sup>, el cambio de acento, la fragmentación de palabras y las reparaciones del habla; por esta razón, el corpus DIME presenta una gran riqueza de fenómenos pragmáticos, sintácticos y lingüísticos.

<sup>5</sup> Este fenómeno se presenta, generalmente, cuando una palabra termina con sonido consonántico y la palabra siguiente comienza con sonido vocálico, de tal forma que estos dos sonidos se unen y forman una nueva sílaba; por ejemplo, si tenemos las palabras “dos horas” textualmente se pueden separar en las sílabas “dos – ho – ras”; pero en el lenguaje oral, el sonido /s/ de la palabra “dos” y el sonido /o/ de la palabra “horas”, normalmente, se juntan y forman una nueva sílaba quedando de la siguiente forma: “do – so – ras”.

Para aprovechar toda esta información empírica, el corpus DIME está transcrito en varios niveles lingüísticos. Los niveles base, y de los cuales parten todos los demás niveles de estudio, son los siguientes: alofónico, de sílabas y de palabras.

Para el estudio de las reparaciones del habla se definieron 4 niveles de transcripción: (1) Nivel de reparaciones del habla, (2) Nivel de identificación de categorías léxicas (POS), (3) Nivel de marcadores del discurso y (4) Nivel de índices de separación de palabras (*Break Indices*). La finalidad de los nuevos niveles de transcripción es analizar la información empírica que tiene el corpus DIME y encontrar las características más relevantes que se presentan en las reparaciones del habla para realizar su detección y corrección.

### Herramienta de etiquetación

Para realizar el proceso de etiquetación se empleó la herramienta *CSLU Toolkit*<sup>6</sup>; específicamente el programa *Speech View* (Figura 3.2).



Figura 3.2: Barra de herramientas del *Speech View*

Por medio de este programa es posible observar las señales de audio al igual que el espectro de frecuencias, como se muestra en la Figura 3.3. Las señales de audio son archivos en formato *wav* de los diálogos generados por el experimento, divididos en elocuciones.

---

<sup>6</sup> <http://cslu.cse.ogi.edu>

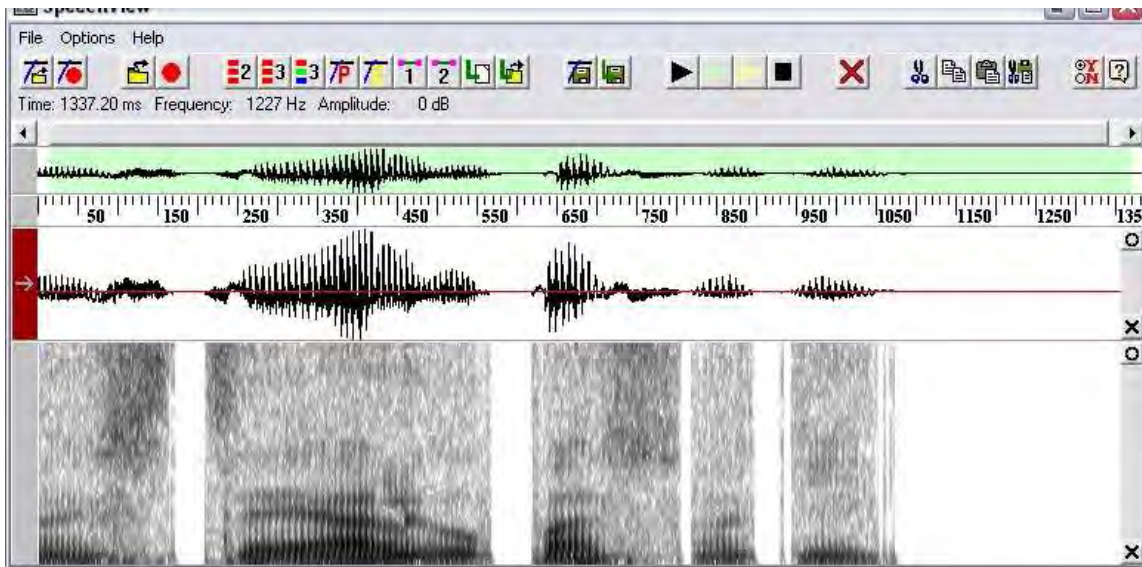


Figura 3.3: Señal de audio con espectro (d03 – utt17)

El programa permite agregar diferentes barras para analizar otros aspectos como son la intensidad y la energía, como se muestra en la Figura 3.4. La barra que se utiliza para realizar la etiquetación de la señal de audio es la *label window*, ésta se ilustra en la Figura 3.5; por medio de esta barra es posible dividir la elocución temporalmente y asociar una etiqueta a cada elocución. En la Figura 3.5 se muestran las palabras de la elocución (ejemplo de etiquetación) alineadas temporalmente a la señal de audio.

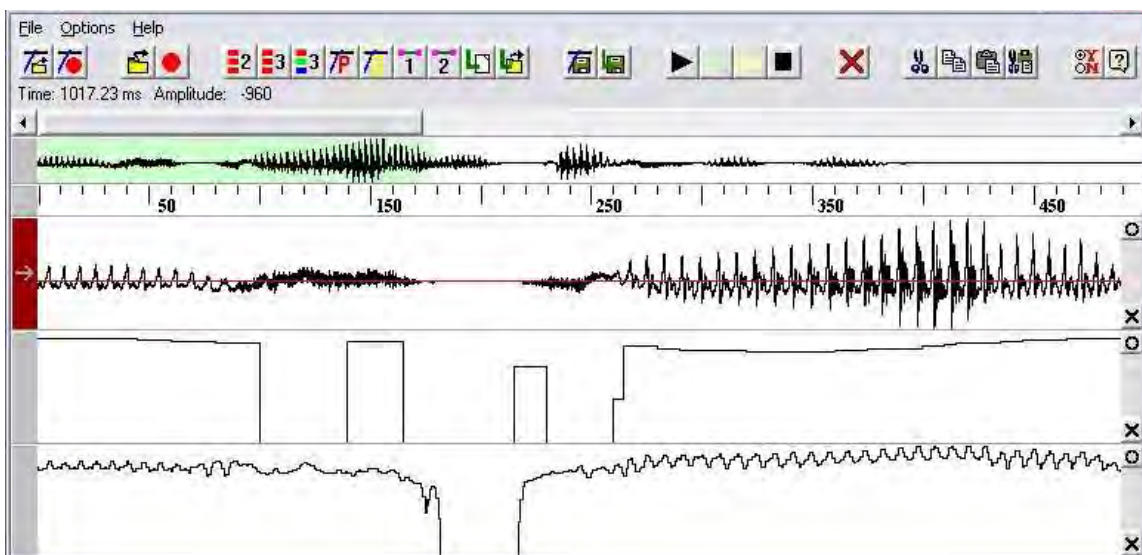


Figura 3.4: Señal de audio con energía e intensidad (d03 – utt17)

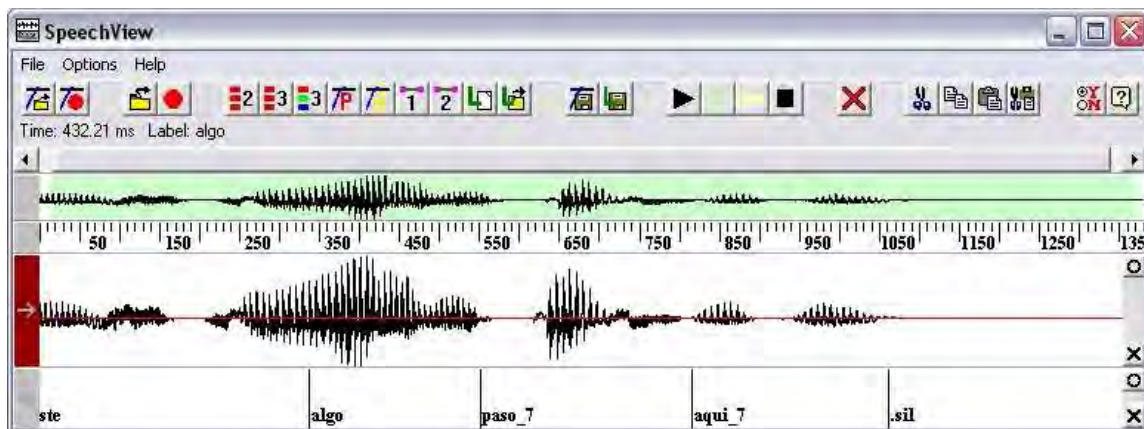


Figura 3.5: Señal de audio con *Label Window* (d03 – utt17)

### 3.1 Anotación de los niveles base

La base de todos los niveles de transcripción del corpus DIME son los siguientes: alofónico, de sílabas y de palabras. A continuación se describen las características de estos niveles de transcripción.

#### 3.1.1 Nivel alofónico

Un fonema es la unidad mínima del lenguaje que tienen una función distintiva; es un sonido del habla que permiten distinguir palabras en una lengua; por ejemplo, los sonidos /p/ y /b/ son fonemas del español porque existen palabras como /pata/ y /bata/ que tienen significado distinto y su pronunciación sólo cambia en relación con esos dos sonidos. Los fonemas no son sonidos con entidad física, sino abstracciones de los sonidos del habla ([13], [19]). Se llama alófono a cada una de las variantes que se dan en la pronunciación de un mismo fonema, según la posición de este en la palabra o sílaba y según el carácter de los fonemas vecinos, sin que las variaciones entre ellos tengan valor diferenciante. Cada sonido corresponde a una determinada forma acústica, pero en las reglas de la lengua se los considera como poseyendo el mismo valor [7], [12].

En nivel alofónico es del que parten todos los demás niveles del corpus. El proceso de etiquetado consiste en la identificación de cada uno de los alófonos que se presentaron en las elocuciones de cada diálogo. El alfabeto utilizado para realizar la transcripción de

este nivel fue MEXBET [26] (versé Apéndice A – Alfabeto fonético MEXBET). Este nivel de transcripción se ilustra en la Figura 3.6

### Ejemplo 16 (d03 – utt35 Etiquetación Alofónica)

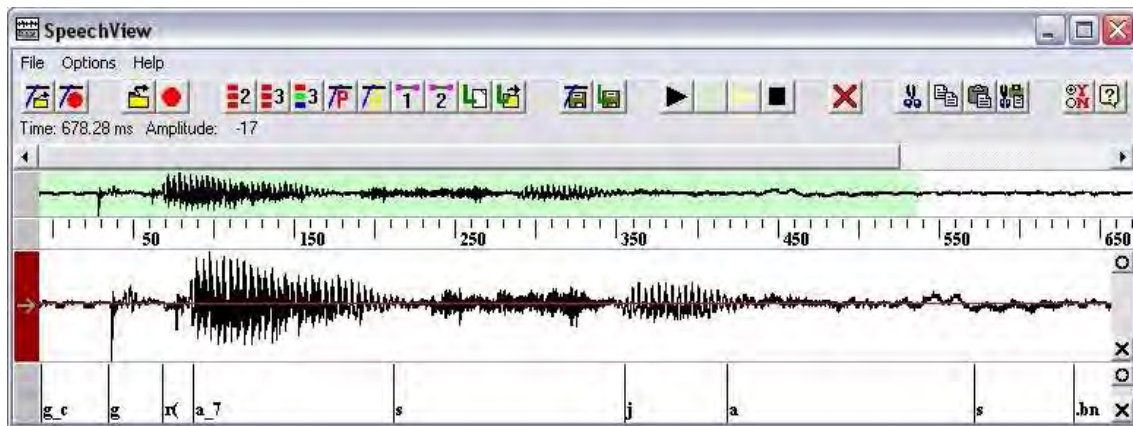


Figura 3.6: Nivel alofónico

### 3.1.2 Nivel de sílabas

Las sílabas son sonidos articulados que constituyen un solo núcleo fónico dentro de la emisión de voz [7]. Este nivel de transcripción toma como base el nivel alofónico; aquí se etiquetan las sílabas formadas por los alófonos contemplando fenómenos como la resilabificación, la sinalefa e incluso la pérdida de alófonos durante la formación del diálogo (eliminación). La etiqueta que se les asigna a cada sílaba es la unión de cada una de las etiquetas de los alófonos que la forman; de esta forma, lo que se etiqueta son sílabas alofónicas, que es lo que realmente el hablante dijo en el diálogo. En la Figura 3.7 se puede observar los niveles alofónico y de sílabas; también se puede ver la forma en que las etiquetas de los alófonos se utilizan para formar la etiqueta de la sílaba. Los alófonos /g\_c/, /g/, /r/, /a\_7/, por ejemplo, forman la sílaba /g\_cgr(a\_7/ (*gra*) mientras que los alófonos /s/, /j/, /a/, /s/ la sílaba /sjas/ (*cias*); estas dos sílabas se realizan en la pronunciación de la palabra “gracias”.



### Ejemplo 17 (d01 – utt35 Etiquetación de sílabas)

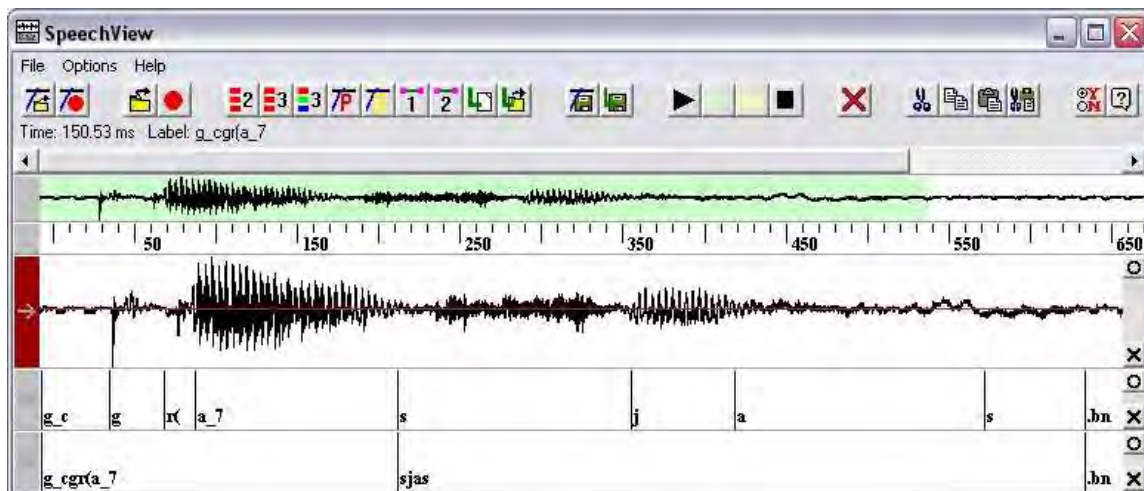


Figura 3.7: Nivel de sílabas

### 3.1.3 Nivel de palabras

La palabra es la unidad formada por uno o varios fonemas, aislable y dotada de significado [15]. De acuerdo al diccionario de la real academia de la lengua española, la palabra es el segmento del discurso unificado habitualmente por el acento, el significado y pausas potenciales inicial y final [7]. Sin embargo, en el lenguaje oral esta definición no se cumple siempre, como se mencionó al explicar el nivel de sílabas, existe fenómenos como la resilabificación y la presencia de sinalefas que complican la definición de una palabra; cuando se mantiene una conversación, las palabras en pocas ocasiones presentan pausas y/o silencios iniciales y/o finales y lo más común es que una palabras esté unida a la que le sigue.

En este nivel de transcripción se etiquetan las palabras presentes en cada elocución, teniendo como base los dos niveles anteriores. Para este nivel, las etiquetas corresponden a la forma ortográfica que debe tener cada palabra, con la excepción de las vocales con acento gráfico, las cuales se etiquetan sin tilde y en su lugar se les agrega el símbolo “\_7”; asimismo, la consonante ñ, se etiqueta con el símbolo “n~”; estos símbolos son las representaciones correspondientes en el alfabeto MEXBET. En la Tabla 3.1 se muestra la conversión de las vocales con acento y la consonante ñ; en la Figura 3.8 se presenta un ejemplo con los tres niveles descritos hasta el momento.

Símbolo ortográfico	Etiqueta
á	a_7
é	e_7
í	i_7
ó	o_7
ú	u_7
ñ	n~

Tabla 3.1: Tabla de conversión

### Ejemplo 18 (d03 – utt101 Etiquetación de palabras)

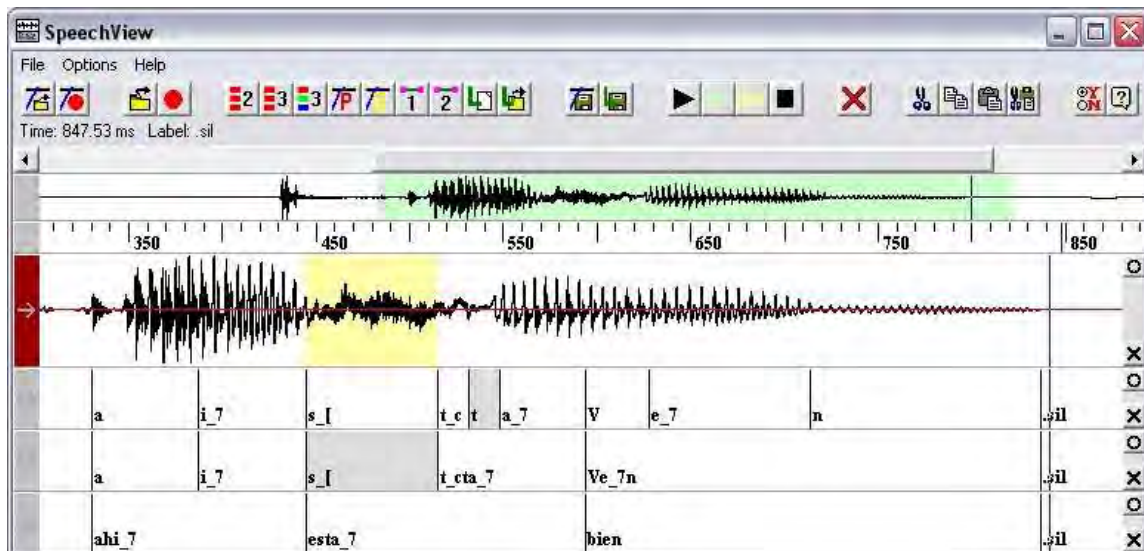


Figura 3.8: Nivel de palabras

Este nivel de transcripción es la base de los niveles de reparaciones del habla, POS y marcadores del discurso definidos en esta tesis, así como del nivel de *Break Indices*.

### 3.2 Anotación de las reparaciones del habla

El nivel de reparaciones del habla está dividido en 3 capas: estructura de la reparación, tipo de reparación y relaciones en la reparación.

### 3.2.1 Estructura de una reparación del habla

Como se mencionó con anterioridad, las reparaciones del habla tienden a establecer una forma estándar consistente en 3 elementos hablados y 1 tonal. En la capa de estructura de una reparación, lo único que se etiqueta son los 3 elementos hablados que son:

- *Reparandum*. La parte de la elocución que el hablante desea cambiar o corregir.
- Término de edición: Pausas como “uh” o “mm”; o frases como “es decir” o “bueno”. Sirve como puente entre el *reparandum* y la alteración, o le permiten al hablante pensar en lo que va a decir. No siempre se presenta.
- Alteración: Es la parte de la elocución que el hablante realmente quiere decir. Es la parte que sustituye al *reparandum*.

El conjunto de etiquetas, grupo de símbolos que se emplean para clasificar los datos, está formado por el nombre de las partes habladas de la estructura de una reparación: *reparandum*, *alteracion* y *termino de edicion*<sup>7</sup>. El proceso de etiquetado consiste en identificar cada una de las partes de la reparación. En el Ejemplo 19 las palabras “las” y “alacenas” forman el *reparandum* y se marcan con la etiqueta *reparandum*, mientras que las palabras “los” y “gabinetes” forman la alteración y son marcadas con la etiqueta *alteracion*.

#### Ejemplo 19 (d12 – utt93 Etiquetación de la estructura de una reparación)

.sil <sup>8</sup>	a	donde	esta_7	las	alacenas	los	gabinetes	.bn <sup>9</sup>
				reparandum		alteracion		

Los silencios (.sil) y ruidos (.bn) que ocurren antes o después de alguna de las partes de la reparación no se etiquetan, como se observa en el Ejemplo 20.

#### Ejemplo 20 (d13 – utt168 Etiquetación de la estructura de una reparación con silencio y ruido)

.sil	un	poquito	más	hacia		.sil	.bn	hacia	acá	.sil
					reparandum			alteracion		

<sup>7</sup> Las etiquetas no llevan acento gráfico porque el software utilizado no lo permite

<sup>8</sup> Esta etiqueta significa silencio en todos los niveles

<sup>9</sup> Esta etiqueta significa ruido en todos los niveles

### 3.2.2 Clasificación de las reparaciones del habla

La etiquetación de esta capa consiste en asignar el tipo de reparación que tiene la elocución en base a su estructura. El conjunto de etiquetas utilizadas corresponden a los nombres de los diferentes tipos de reparaciones del habla.

#### Reinicio (*Fresh Start*)

Ocurre cuando el hablante abandona la elocución y reinicia; el hablante cancela la primera idea y luego la reemplaza por otra. En este tipo de reparación no existe relación entre la alteración y el *reparandum*, como se muestra en el Ejemplo 21.

#### Ejemplo 21 (d01 – utt93 Etiquetación de tipo de reparación)

ay	que	.sil	tengo	que	seleccionar	otro	.sil
reparandum			alteracion				
fresh start							

#### Reparación por modificación (*Modification Repair*)

Esta clase abarca el resto de las reparaciones del habla que tienen *reparandum*. En este tipo de reparación el hablante cambia o repite una parte de la elocución; es muy frecuente que entre el *reparandum* y la alteración existan palabras similares. En el Ejemplo 22 se muestra una reparación de este tipo, así como su etiquetación.

#### Ejemplo 22 (d12 – utt63 Etiquetación de tipo de reparación)

.sil	e_7stos	son	Los	cuatro	tipos	de	gav	de	gabinetes	que	tenemos	.sil
						reparandum	alteracion					
modification repair												

#### Reparación por pausa (*Abridged Repair*)

Este tipo de reparación tiene únicamente el término de edición; es un tipo de pausa que ocurre cuando el hablante necesita tiempo para pensar como continuar la elocución.

**Ejemplo 23 (d19 – utt65 Etiquetación de tipo de reparación)**

.sil	ahora	necesito	.sil	eh	.sil	una	alacena	.sil
				termino de edicion				
abridged repair								

**3.2.3 Relaciones en la reparación**

En este nivel se etiquetan las relaciones que existen entre las palabras de la elocución. La etiquetación se basa en un conjunto reducido de las etiquetas propuestas por Heeman [34]; el conjunto utilizado en la presente tesis se muestra en la Tabla 3.2.

<b>Etiqueta</b>	<b>Descripción</b>
<i>mi</i>	Se usa cuando dos palabras son idénticas.
<i>ri</i>	Se usa cuando una palabra substituye a otra.
<i>xr</i>	Se usa cuando una palabra es eliminada o insertada.
<i>pi</i>	Etiqueta una correspondencia de varias palabras, tal como un reemplazo de un pronombre por una descripción más larga.
<i>srr&lt;</i>	Denota el inicio del <i>reparandum</i> de un <i>fresh start</i>
<i>et</i>	Se usa para marcar un término de edición.

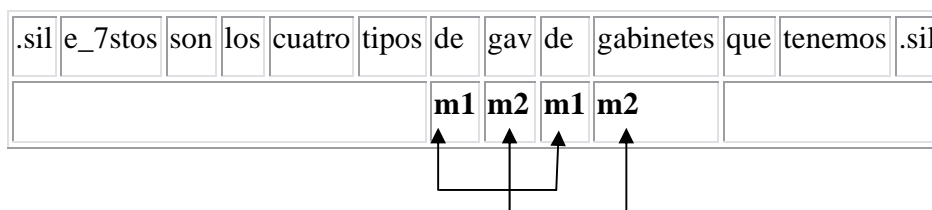
**Tabla 3.2: Etiquetas de Reparaciones del Habla.**

Las etiquetas que no se consideraron con respecto al trabajo de Herman fueron las que corresponden al punto de interrupción. Esto es se debe a que a diferencia del trabajo de Herman, en el presente esfuerzo las etiquetas están alineadas temporalmente, por esta razón el esquema codifica el punto de interrupción implícitamente. En la siguiente parte se presenta la descripción de cada etiqueta, acompañada de un ejemplo de uso.

### 3.2.3.1 Etiquetación de la relación entre palabras

La etiqueta **mi** se utiliza para marcar las palabras correspondientes en el *reparandum* y en la alteración. El índice *i* se utiliza para correlacionar las palabras dentro de la reparación; en este tipo de reparación se reemplaza la palabra del *reparandum* por la palabra de la alteración con el mismo índice, como se muestra en el Ejemplo 24 corregido.

#### Ejemplo 24 (d12 – utt63 Etiqueta *mi* para la etiquetación de palabras correspondientes)

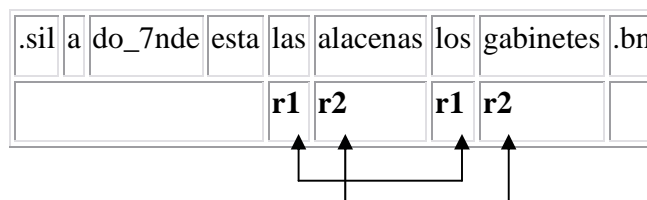


#### Ejemplo 24 (d12 – utt63 Corregido)



La etiqueta **ri** se emplea cuando una palabra en el *reparandum* es reemplazada por otra en la alteración. En este caso el índice *i* permite la coordinación entre las palabras; la palabra en el *reparandum* es sustituida por la palabra en la alteración con el mismo índice. Nótese que la reparación no resulta en una oración gramatical debido a la falta de concordancia.

#### Ejemplo 25 (d12 – utt93 Etiqueta *ri* para la etiquetación de reemplazo)



#### Ejemplo 25 (d12 – utt93 Corregido)



La etiqueta **x<sub>r</sub>** se usa cuando una palabra se inserta en la alteración o se elimina en el *reparandum*; el índice *r* es un número consecutivo que enumera las palabras que se insertan o se eliminan.

**Ejemplo 26 (d12 – utt74 Etiqueta x<sub>r</sub> para la etiquetación de inserción)**

.bn	y	ahora	un	.sil	la	estufa	un	poco	hacia	el	fregadero	.sil
			m1		x1	x2		m1				

↑     ↑

**Ejemplo 26 (d12 – utt74 Corregido)**

.bn	y	ahora	la	estufa	un	poco	hacia	el	fregadero	.sil
-----	---	-------	----	--------	----	------	-------	----	-----------	------

En el ejemplo anterior las etiquetas x1 y x2 se refieren a palabras que se insertan en la alteración.

La etiqueta **p<sub>i</sub>** se utiliza cuando una palabra en el *reparandum* se cambia por varias (una descripción más larga) en la alteración o viceversa; el índice *i* tiene la función de relacionar la descripción corta con la descripción larga; la palabra con el índice *i* en el *reparandum* se reemplaza por las palabras de la alteración con el mismo índice, o varias palabras en el *reparandum* con el índice *i* son reemplazadas por una palabra en la alteración con el mismo índice.

**Ejemplo 27 (d12 – utt40 Etiqueta p<sub>i</sub> para la etiquetación de reemplazo por descripción)**

...	abarca	solamente	.bn	hasta	donde	llega	la	alacena	.bn	perdo_7n	.sil	la	ma_7quina	lavatrastes	.sil
							m1	p1		et		m1	p1		p1

↑     ↑     ↑

**Ejemplo 27 (d12 – utt40 Corregido)**

...	abarca	solamente	.bn	hasta	donde	llega	la	ma_7quina	lavatrastes	.sil
-----	--------	-----------	-----	-------	-------	-------	----	-----------	-------------	------

La etiqueta **srr<** se emplea en las reparaciones del tipo *fresh start* (reinicio) para marcar el *reparandum*, el cual se sustituye íntegramente por la alteración.

### Ejemplo 28 (d01 – utt93 Etiqueta srr< para la etiquetación de *fresh starts*)

Ay	que	.sil	tengo	que	seleccionar	otro	.sil
srr<							

### Ejemplo 28 (d01 – utt93 Corregido)

tengo	que	seleccionar	otro	.sil
-------	-----	-------------	------	------

La etiqueta **et** se utiliza para etiquetar los términos de edición que se presentan en la reparación.

### Ejemplo 29 (d19 – utt65 Etiqueta et para la etiquetación de términos de edición)

.sil	ahora	necesito	.sil	eh	.sil	una	alacena	.sil
				et				

### Ejemplo 29 (d19 – utt65 Corregido)

.sil	ahora	necesito	una	alacena	.sil
------	-------	----------	-----	---------	------

A continuación se muestra una elocución etiquetada y alineada temporalmente, incluyendo todos los niveles del nivel de reparaciones del habla.

### Ejemplo 30 (d01 – utt85 Niveles de etiquetación alineados temporalmente)

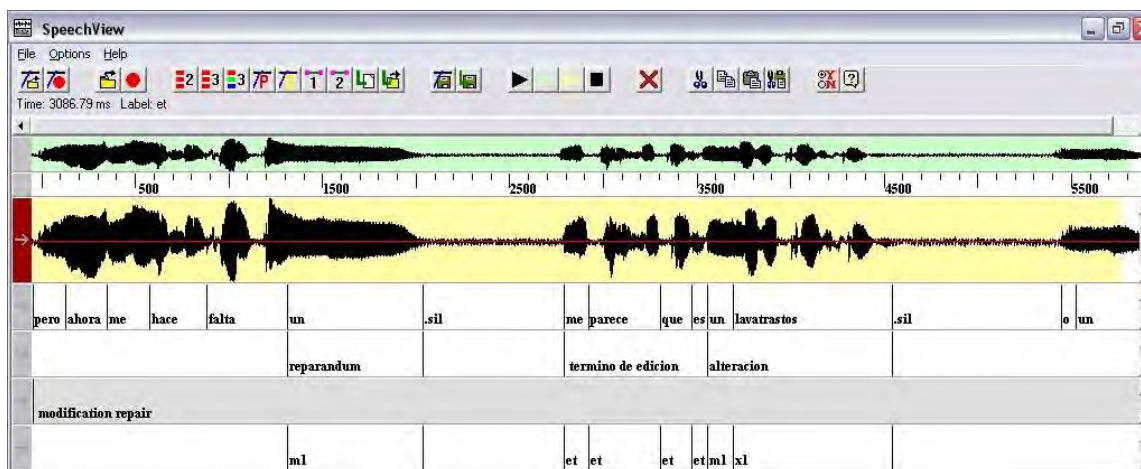


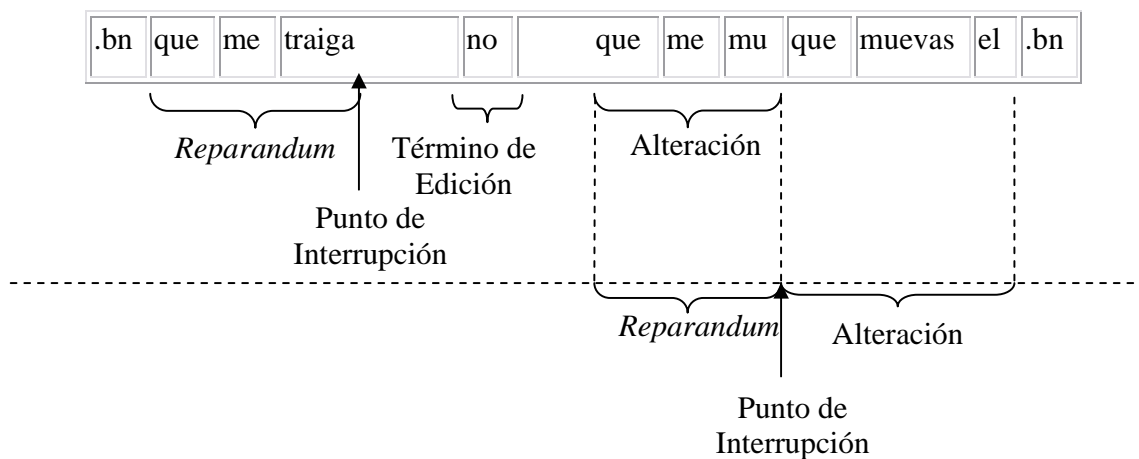
Figura 3.9: Nivel de reparaciones del habla



### 3.2.4 Multirreparaciones

Existen casos en los cuales hay más de una reparación en la misma elocución, como se muestra en el Ejemplo 31.

#### Ejemplo 31 (d03 – utt03 Multirreparación)



#### Ejemplo 31(d03 – utt03 Corregido)

.bn que nuevas el .bn

Cuando se presentan estos casos, se generan las tres capas del nivel de transcripción para cada una de las reparaciones. Un ejemplo se muestra en la Figura 3.10.

#### Ejemplo 31(d03 – utt03 Etiquetación de una multirreparación)

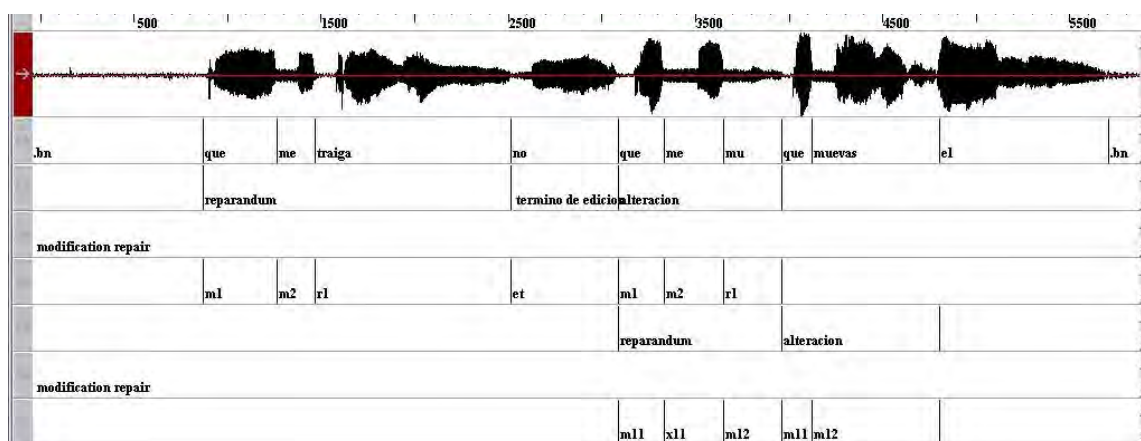
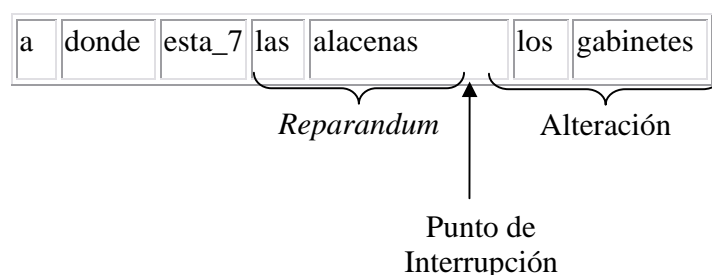


Figura 3.10: Etiquetación de una multirreparación

### 3.3 Anotación de las categorías léxicas (POS)

La identificación de categorías léxicas o POS (*Part-Of-Speech*) puede ser útil para predecir la ocurrencia de una reparación; cuando ocurre una reparación se rompe la estructura normal de la elocución ya que la coherencia gramatical entre las palabras se pierde; por ejemplo, la probabilidad de que un sustantivo sea seguido por un artículo es muy baja, como se ve en el Ejemplo 32. En estos casos, la probabilidad de que haya una reparación es alta; y esta información es útil para predecir la reparación.

#### Ejemplo 32 (d12 – utt93 POS)



#### 3.3.1 Definición de conjunto de etiquetas.

Un conjunto de etiquetas para la anotación de las categorías léxicas puede ser tan complejo y especializado como se desee, de acuerdo a los propósitos para los cuales se necesita realizar este etiquetado; para definir el conjunto de etiquetas para la identificación de las categorías léxicas se realizó un análisis del material empírico del corpus DIME; para este efecto se analizó un diálogo donde se encontraron las siguientes categorías básicas:

Etiqueta	Categoría
N	Sustantivo
V	Verbo
A	Adjetivo
T	Artículo
R	Adverbio
P	Pronombre
S	Preposición

C      Conjunción

---

**Tabla 3.3: Categorías léxicas básicas.**

Con el conjunto de etiquetas básicas se hizo una especialización más profunda ya que es un grupo demasiado general para formar un buen modelo del lenguaje. A pesar de que en cada categoría las diferencias que las definen son muy significativas, se necesita un conjunto de categorías más especializadas porque dentro de una misma categoría básica existe otro tipo de cualidades; por ejemplo, en un relación “verbo” – “verbo” el primer verbo se trata de un verbo auxiliar o modal, el cual carece de agente que realice una acción, mientras que el segundo es un verbo principal, este si tiene una agente que realiza una acción, por lo tanto la relación es “verbo auxiliar o modal” – “verbo”, si se dejará una categoría muy general el modelo sería muy limitado.

La especialización del conjunto de etiquetas se hizo de acuerdo a las características del corpus DIME y tomando como material de apoyo conjunto de etiquetas de otros autores [4], [30], [31] y [34]. Otro apoyo muy importante y fundamental fue la colaboración de los integrantes lingüistas del proyecto, quienes participaron en el análisis del diálogo y en la definición de las etiquetas. El conjunto de etiquetas que resultó de este ejercicio y se adoptó para la presente tesis se muestra en la Tabla 3.4.

Etiqueta	Categoría
N	Sustantivo
V	Verbo
VAM	Verbo Auxiliar – Modal
VC	Verbo con Clítico
A	Adjetivo
AD	Adjetivo Demostrativo
TD	Artículo Determinado
TI	Artículo Indefinido
R	Adverbio
RI	Adverbio Interrogativo
RR	Adverbio Relativo

RN	Adverbio de Negación
RA	Adverbio de Afirmación
P	Pronombre
PD	Pronombre Demostrativo
PR	Pronombre Relativo
PI	Pronombre Interrogativo
PC	Pronombre Clítico
S	Preposición
C	Conjunción

**Tabla 3.4: Etiquetas para la identificación de categorías léxicas.**

### 3.3.2 Etiquetación

Para el nivel de etiquetación de POS se utiliza como base el nivel de transcripción de palabras definido en el corpus DIME.; como el resto de las etiquetas, estas etiquetas están alineadas temporalmente a los niveles de etiquetación. A continuación se ejemplifican cada una de las etiquetas utilizadas.

- **Sustantivo**

Es la parte de la oración que expresa la categoría correspondiente a sustancia, es decir, seres o cosas que pueden ser sujeto u objeto de cualquier acción, estado o accidente expresable con un verbo. Morfológicamente, los sustantivos se caracterizan, según las lenguas, por uno o varios de los siguientes rasgos: género y número determinado intrínsecamente y como declinación (en su caso).

En la tradición gramatical grecolatina se les denominó sustantivos porque designaban «sustancias» (en contraposición al verbo, que representaba acciones). En la actualidad este criterio no se considera semántico, dada la existencia de sustantivos que designan acciones, estados y todo tipo de nociones abstractas (odio, manía, esperanza, etc.) [8].

Para la etiquetación de un sustantivo en el corpus DIME se ocupa la etiqueta **N** y no se considera ninguna otra propiedad. Ejemplo:

**Ejemplo 33 (d12 – utt01 Etiqueta N para sustantivo)**

.sil	quieres	que	desplace	o	traiga	algu_7n	objeto	a	la	cocina	.sil
							N			N	

- Verbo

Es la parte de la oración que expresa la existencia, acción y estado del sujeto; es el único elemento que funciona como núcleo del grupo predicativo de la oración. El verbo indica procesos, o implica la noción de tiempo, mientras que el sustantivo (elemento no verbal por excelencia) indica objetos y no supone temporalidad. No siendo universales estas caracterizaciones semánticas, se define al verbo como el elemento indispensable para la constitución de un enunciado.

Las propiedades que tiene un verbo son: tipo, modo, tiempo y persona [31]. Para la etiquetación de esta categoría sólo son considerados los siguientes tipos: verbo auxiliar o modal y verbo principal. Los verbos auxiliares o modales, como *querer* y *poder*, son los que se combinan con formas nominales de otros verbos para ampliar el esquema aspectual y temporal de la conjugación.

Los verbos principales se etiquetan con la letra **V**.

**Ejemplo 34 (d12 – utt01 Etiqueta V para verbo)**

.sil	quieres	que	desplace	o	traiga	algu_7n	objeto	a	la	cocina	.sil
			V		V						

En ocasiones el verbo está unido a un pronombre clítico; cuando se presente en esta forma el verbo la etiqueta que se utiliza es **VC**, ejemplo:

**Ejemplo 35 (d01 – utt21 Etiqueta VC para verbo)**

podri_7as	juntarlo	un	poco	ma_7s
	VC			

Para los verbos auxiliares – modales la etiqueta a utilizar es **VAM**.

**Ejemplo 36 (d12 – utt33 Etiqueta VAM para verbo auxiliar – modal)**

.bn	me	puedes	mostrar	el	cata_7logo	de	fregaderos	y	ma_7quinas	.sil
		VAM								

Un verbo auxiliar – modal está parcial o totalmente desprovisto de su significado original (no tiene agente) y se utiliza como un mero morfema por otros verbos [18]. Por ejemplo, el verbo *poder* es un auxiliar si la intención no es preguntar sobre la capacidad del agente de realizar una acción, como se muestra en el Ejemplo 36, y por el contrario, si hay un agente que *puede* el verbo desempeña su función normal; otro ejemplo es el verbo *querer* si no se pregunta por la creencia o el deseo, es decir, no hay un agente que *quiere* se etiqueta como verbo auxiliar

- **Adjetivo**

Antiguamente el adjetivo estaba considerado dentro de la categoría del sustantivo (aunque Platón lo incluyó entre los verbos); ya en la Edad Media se distinguieron *nomen substantivum* y *nomen adiectivum*, pero no fue hasta el siglo XVIII que empezó a estudiarse como categoría independiente [8].

El adjetivo es una parte de la oración o categoría de rango secundario (adjunto del sustantivo) denotativa de cualidad, grado de comparación y situación; es una de las dos clases de palabras (la otra es el verbo) que modifica al sustantivo. Los rangos morfológicos que presenta el adjetivo son los de género y número [31].

En el corpus DIME solamente se consideran la distinción entre un adjetivo demostrativo del resto de los adjetivos. Para los adjetivos se utiliza la etiqueta A.

**Ejemplo 37 (d12 – utt102 Etiqueta A para adjetivo)**

.bn	dame	la	blanca	superior	doble	.bn
				A	A	

En este ejemplo la palabra “blanca” es una adjetivo nominalizado, la palabra pierde su función original (adjetival) por una nominal. El artículo que le precede es un morfema del sustantivo que corrobora la afirmación de que es un sustantivo.

En caso de los adjetivos demostrativos se utiliza la etiqueta **AD**.

**Ejemplo 38 (d12 – utt11 Etiqueta AD para adjetivo demostrativo)**

.sil	quieres	que	mueva	.sil	este	objeto	.sil	hacia	aca_7	.sil
					AD					

- **Artículo**

Es el elemento gramatical que puede anteponerse o no a las palabras con función sustantiva, con las que concuerda generalmente en género y número, y cuya presencia puede variar la función del sustantivo.

Según el grado de mayor o menor determinación, el artículo, en español puede ser determinado e indeterminado. La gramática moderna niega el carácter de artículo al indeterminado, considerándolo como indefinido. El artículo es un morfema exclusivo del sustantivo; por ello cuando se antepone a una palabra, la sustantiviza: el cantar, la rica, etc. [8].

La etiqueta que corresponde al artículo determinado es **TD**.

**Ejemplo 39 (d12 – utt06 Etiqueta TD para artículo determinado)**

.sil	hacia	la	derecha	.sil
		TD		

Al artículo indefinido (indeterminado) le corresponde la etiqueta **TI**.

**Ejemplo 40 (d12 – utt65 Etiqueta TI para artículo indefinido)**

.sil	quieres	un	mueble	como	estos	.sil	aqui_7	.sil
		TI						

Existen casos en los que se debe tener cuidado al etiquetar una artículo indefinido, ya que en ocasiones se presentan ocurrencias en las que se trata, en realidad, de un adjetivo o de un pronombre. Como se muestra en el Ejemplo 41 y Ejemplo 42.

**Ejemplo 41 (d12 – utt32 Adjetivo y no artículo indefinido)**

.sil	hay	un	cata_7logo	de	alacenas	.sil	uno	de	estantes	.sil	uno	de	estufas	Y	extractores	.sil	...
		A															

Como Adjetivo

**Ejemplo 42 (d12 – utt32 Pronombre y no artículo indefinido)**

.sil	hay	un	cata_7logo	de	alacenas	.sil	uno	de	estantes	.sil	uno	de	estufas	y	extractores	.sil	...
						P				P							

Como Pronombre

En el Ejemplo 41 se puede ver como la palabra “un” se emplea para hacer notar el número de objetos; en este caso, el número de catálogos de alacenas, mientras que el Ejemplo 42 “uno” se emplea como pronombre de la palabra “cata\_7logo”.

- **Adverbio**

Parte invariable de la oración, que sirve para modificar la significación del verbo, del adjetivo, de otro adverbio e incluso puede aplicarse al sustantivo y a toda una frase.

Hay varias clases de adverbios; los que se consideran para la etiquetación son: adverbios pronominales (los que desempeñan la función de pronombres), interrogativos (“cuándo”, “cuánto”), relativos (“donde”, “como”, “cuando”), adverbio de afirmación (“sí”, “también”, “efectivamente”) y adverbios de negación (“no”, “tampoco”).

Para adverbios interrogativos se utiliza la etiqueta **RI**.

**Ejemplo 43 (d12 – utt53 Etiqueta RI para adverbio interrogativo)**

do_7nde	quieres	que	la	ponga	.sil
RI					

Para adverbios relativos se usa la etiqueta **RR**.



**Ejemplo 44 (d13 – utt79 Etiqueta RR para adverbio relativo)**

mejor	ponme	la	estufa	.sil	eh	donde	esta_7	la	ventana	.sil
						RR				

Para adverbios de afirmación se emplea la etiqueta **RA**.

**Ejemplo 45 (d12 – utt50 Etiqueta RA para adverbio de afirmación)**

.bn	s_7	.sil
	RA	

Y para adverbios de negación la etiqueta **RN** es empleada.

**Ejemplo 46 (d12 – utt90 Etiqueta RN para adverbio de negación)**

.sil	no	.sil
	RN	

Cualquier otro adverbio que no pertenezca a alguna de las clasificaciones mencionadas se etiqueta con la letra **R**, por ejemplo:

**Ejemplo 47 (d17 – utt32 Etiqueta R para adverbios en general)**

.sil	ahí_7	esta_7	bien	.sil
	R		R	

- **Pronombre**

Es una categoría que se define tradicionalmente como la parte de la oración cuya función es suplir al sustantivo, evitando así y por razones de economía una repetición.

Hay varias clases de pronombres; para la etiquetación del corpus DIME se consideran los pronombres demostrativos, interrogativos, relativos y clíticos; cualquier otro tipo de pronombre se considera como pronombre general.

Para los pronombres en general se utiliza la etiqueta **P**.

**Ejemplo 48 (d12 – utt89 Etiqueta P para pronombres en general)**

.sil	quieres	alguno	.sil
		P	

Para los pronombres demostrativos se usa la etiqueta **PD**.

**Ejemplo 49 (d12 – utt42 Etiqueta PD para pronombre demostrativo)**

.sil	e_7ste	junto	a	la	estufa	.sil
	PD					

Para los pronombres interrogativos se emplea la etiqueta **PI**.

**Ejemplo 50 (d12 – utt37 Etiqueta PI para pronombre interrogativo)**

.bn	a	ver	.bn	cua_7l	es	la	diferencia	entre	el	tercero	y	el	cuarto	.bn
				PI										

Para los pronombres relativos se utilizará la etiqueta **PR**.

**Ejemplo 51 (d12 – utt09 Etiqueta PR para pronombre relativo)**

.sil	a	la	mitad	del	espacio	que	hay	entre	la	ventana	y	la	pared	.sil
						PR								

Para los pronombres clíticos se usará la etiqueta **PC**.

**Ejemplo 52 (d12 – utt53 Etiqueta PC para pronombre clítico)**

do_7nde	quieres	que	la	ponga	.sil
			PC		

- **Preposición**

Es un parte invariable de la oración que sirve de nexo entre un elemento sintáctico cualquiera y su complemento.

Para la etiquetación de una preposición se utiliza la etiqueta **S**.

**Ejemplo 53 (d12 – utt49 Etiqueta S para preposiciones)**

.sil	en	la	pared	.sil	de	la	izquierda	.sil
	S				S			

Las contracciones ‘del’ y ‘al’ se etiquetan como preposiciones porque su función principal es de subordinar al elemento posterior.

**Ejemplo 54 (d12 – utt9 Etiqueta S para contracciones “del” y “al”)**

.sil	a	la	mitad	del	espacio	que	hay	entre	la	ventana	y	la	pared	.sil
				S										

- **Conjunción**

Es la parte de la oración o clase de palabras cuya función es unir dos frases o dos miembros de una misma frase. La conjunción cumple la función de enlazar miembros gramaticalmente, aunque a veces signifique contrariedad o separación de sentido entre unos y otros.

Para la conjunción se emplea la etiqueta **C**.

**Ejemplo 55 (d12 – utt35 Etiqueta C para conjunciones)**

.bn	e_7ste	es	el	cata_7logo	de	fregaderos	.sil	y	ma_7quinas	lavatrastes	.sil
								C			

En ocasiones el pronombre ‘que’ funciona como conjunción y no como pronombre relativo como se muestra en el Ejemplo 56.

**Ejemplo 56 (d12 – utt53 Etiqueta C para “que” con función de conjunción)**

do_7nde	quieres	que	la	ponga	.sil
		C			

Esto es porque la palabra ‘que’ no hace referencia a alguien, algo o algún lugar y solamente sirve como nexos. En contraste tenemos:

**Ejemplo 57 (d12 – utt9 Ejemplo de “que” como pronombre relativo)**

.sil	a	la	mitad	del	espacio	que	hay	entre	la	ventana	y	la	pared	.sil
						PR								

aquí, ‘que’ esta haciendo referencia a un algo (el espacio).

- **Silencio y ruido**

En el caso de presencia de silencio o ruido, se mantienen las etiquetas que vienen desde el nivel de palabras, “.sil” y “.bn”, respectivamente.

**Ejemplo 58 (d13 – utt11 Etiquetación de un silencio)**

po_7nmelo	.sil	aqui_7	.sil	en	esta_7	esquina	.sil
	.sil		.sil				.sil

**Ejemplo 59 (d13 – utt77 Etiquetación de un ruido)**

.sil	e_7ste	.sil	y	estos	otros	tres	.bn	son	los	modelos	de	extractores	que	tenemos	.sil
							.bn								

Para concluir, en la Figura 3.11 se muestra la etiquetación de las categorías léxicas de una elocución completa.

### Ejemplo 60 (d13 – utt68 Nivel de etiquetación POS)

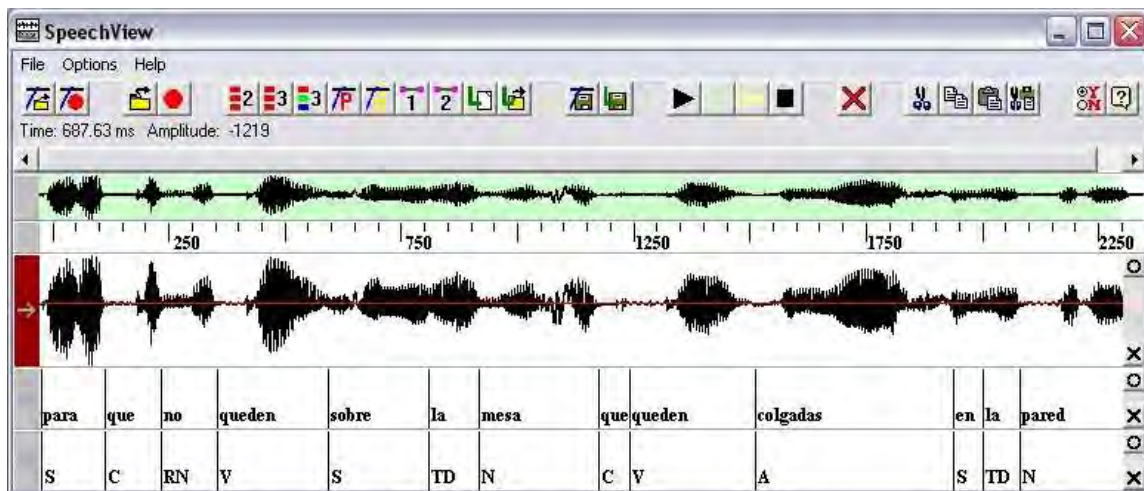


Figura 3.11: Nivel de POS

En el caso de que alguna palabra funcione como un marcador del discurso no se asigna una etiqueta de categoría léxica, un ejemplo se muestra en la Figura 3.12.

### Ejemplo 61 (d01 – utt01 Nivel de etiquetación POS con marcadores del discurso)

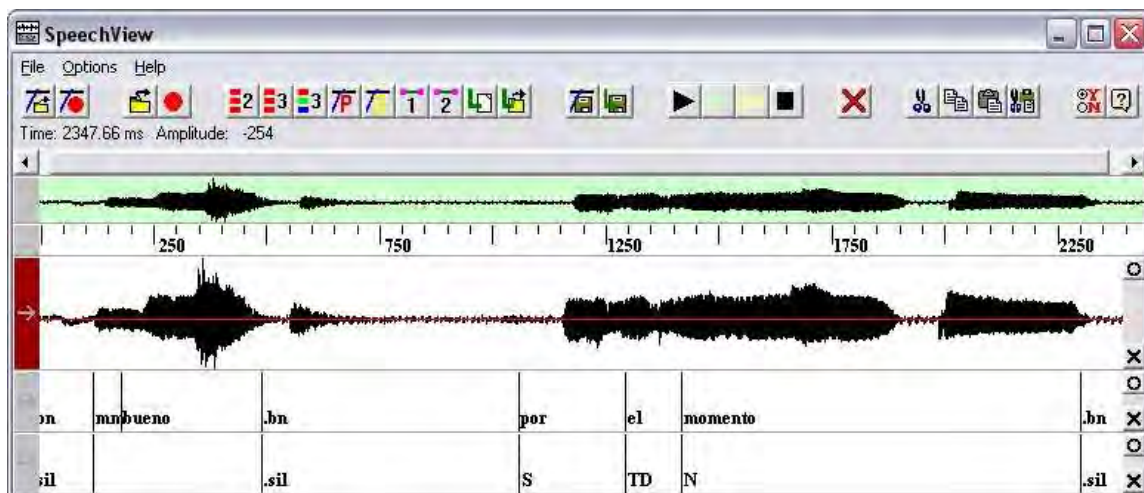


Figura 3.12: Etiquetación de POS con marcadores del discurso

## 3.4 Anotación de los marcadores del discurso

Como se mencionó anteriormente, en el lenguaje natural existen palabras o frases que pierden su sentido sintáctico y su función se enfoca principalmente en el buen

entendimiento del diálogo, a este tipo de palabras o frases se les denomina marcadores del discurso.

Una de las funciones principales de los marcadores del discurso es guiar las inferencias que se realizan en la comunicación, y que dan al oyente información acerca de la estructura del discurso, o ayudan a resolver anáforas [28], [33], [34].

La relación entre las reparaciones del habla y los marcadores del discurso es muy estrecha, en muchas ocasiones el término de edición se trata de marcadores del discurso; además ayudan a señalar el límite de una elocución en el lenguaje hablado.

### 3.4.1 Etiquetación

El propósito de la etiquetación de los marcadores del discurso es analizar el comportamiento que presentan en las elocuciones con disfluencias.

Como ya se mencionó, los marcadores del discurso son palabras que en apariencia tienen la forma de alguna categoría léxica pero en realidad tienen una función discursiva más que una función sintáctica; por lo anterior, las etiquetas que se utilizan para este nivel están relacionadas con las etiquetas definidas para el nivel de identificación de categorías léxicas. Cuando un marcador del discurso tiene alguna forma de categoría léxica se etiqueta la categoría léxica que le correspondería pero anteponiéndole las letras **MD**. Ejemplo:

#### Ejemplo 62 (d12 – utt25 Marcadores del discurso)

.sil	despue_7s	.sil	me	puedes	poner	.sil	l	el	extractor	de	aire	encima	de	la	.sil	de	la	estufa	.sil
	MDR																		

En este ejemplo la palabra “despue\_7s” es formalmente un adverbio, por lo cual se tendría que etiquetar con **R**, pero al funcionar como enlace del diálogo pierde ese significado y se trata de un marcador del discurso, por lo tanto la etiqueta que le corresponde es **MDR**.

La Tabla 3.5 muestra las etiquetas más utilizadas para los marcadores del discurso.

Etiqueta	Categoría
MDN	Sustantivo



**Ejemplo 67 (d03 – utt83 Marcadores del discurso – conjunción)**

.bn	pues	va	quedar	bien	pegada	.bn
	MDC					

Las etiquetas **MDI** y **MDK** son empleadas para dos tipos de palabras que no se presentan en la etiquetación de categorías léxicas; corresponden a interjecciones (**MDI**) y a *acknowledgment* (**MDK**).

- **Interjección**

Es una voz que forma por sí sola una oración, siempre exclamativa, cargada de contenido emocional.

Para ésta se utiliza la etiqueta **MDI**.

**Ejemplo 68 (d12 – utt30 Marcadores del discurso – interjección)**

.bn	eh	me	puedes	mostrar	los	tipos	de	muebles	que	tengo	.sil
	MDI										

- ***Acknowledgment***

El *acknowledgment* es una marca de entendimiento en el diálogo. Es una especie de contra recibo que se presenta cuando el oyente le da a entender al hablante que el mensaje ha sido recibido.

Para éste se emplea la etiqueta **MDK**.

**Ejemplo 69 (d12 – utt67 Marcadores del discurso – *acknowledgment*)**

.bn	okey	.bn
	MDK	

**Casos especiales**

- Un caso especial es la palabra “este”; esta palabra se utiliza muy frecuentemente y puede clasificarse en varios tipos de acuerdo a su función; por esta razón, se decidió



que tuviera su propia etiqueta a nivel de marcadores del discurso. La etiqueta que se utiliza es **MDeste**.

**Ejemplo 70 (d03 – utt17 Marcadores del discurso – palabra “este”)**

este	algo	paso_7	aqui_7	.sil
MDeste				

- Existen marcadores del discurso que están formados por más de una palabra; en este caso, el marcador se etiqueta con cada una de las etiquetas de las palabras que lo forman.

**Ejemplo 71 (d12 – utt74 Marcadores del discurso – varias palabras)**

.bn	y	ahora	un	.sil	la	estufa	un	poco	hacia	el	fregadero	.sil
MDC_MDR												

En el ejemplo anterior se presenta un marcador del discurso que está formado por dos palabras, las cuales corresponden a una conjunción y a un adverbio respectivamente; para etiquetarlo se utilizan las etiquetas correspondientes de cada marcador del discurso separadas por un guión bajo.

En este nivel de transcripción solamente se etiquetan las palabras que son marcadores del discurso. Una elocución etiquetada en este nivel queda de la siguiente forma:

### Ejemplo 72 (d03 – utt87 Nivel de etiquetación marcadores del discurso)

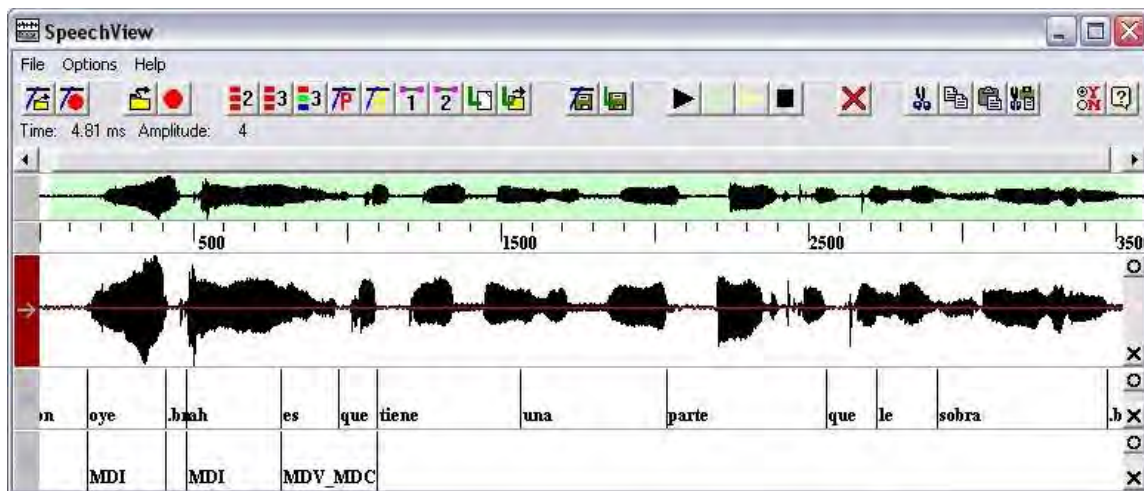


Figura 3.13: Nivel de marcadores del discurso

### 3.5 Anotación de índices de separación de palabras (*Break Indices*)

Una parte importante de la segmentación, y que a su vez ayuda a la detección de cada unidad de elocución, es la identificación de los límites de las palabras y frases entonativas. Un esquema que se utiliza para analizar estos límites es el definido en ToBI (*Tone and Break Index*) [3].

ToBI fue desarrollado para cubrir las necesidades de descripción fonética y de anotación de corpus, con un enfoque en el desarrollo de sistemas de tecnologías del habla [27]. El esquema se basa en el inglés americano, pero se ha desarrollado una versión para el español (Sp-ToBI) ([36], [22]).

Se trata de una transcripción de patrones entonativos partiendo de un análisis de la curva melódica representada como evolución temporal de la frecuencia fundamental [27]. El esquema consiste en 4 niveles de etiquetación.

- Representación ortográfica de la elocución.
- Nivel de índices de disyunción (*Break Indices*).
- Nivel tonal.
- Nivel misceláneo.

Este esquema se basa en el modelo fonológico desarrollado por Pierrehumbert (1980) [25]. Su modelo describe la entonación inglesa como una serie de niveles altos y

bajos en el ámbito de la frecuencia fundamental, donde el nivel más bajo de análisis es el nivel de palabras y la palabra tonal se marcan con un alto o un bajo acento.

El nivel de Break Indices sirve como indicador del límite que marca la frontera entre las unidades melódicas, define grados de cohesión entre las mismas; utiliza la siguiente clasificación:

- 0.- No se percibe separación entre palabras: sinalefa. Puede observarse, por ejemplo, una resilabificación.
- 1.- Separación normal entre palabras.
- 4.- Final de frase entonativa
- 2 y 3.- Se percibe separación y existen indicios claros.

### 3.5.1 Etiquetación

El esquema de etiquetado del nivel de break indices para el corpus DIME es una adaptación del TOBI. Esta adaptación fue elaborada por Isabel López [16] como parte de su tesis de licenciatura.

El esquema emplea las 5 etiquetas del modelo original asignándole características especiales para las etiquetas 2 y 3 como se muestra en la Tabla 3.6.

Etiqueta	Descripción
0	Sinalefa
1	Separación normal de palabra
2	Punto de interrupción de una reparación del habla
3	Frases entonativas intermedias
4	Final de frase entonativa

**Tabla 3.6: Etiquetas para *Break Indices***

Como se puede observar en la Tabla 3.6, en este nivel de etiquetado se tiene una etiqueta especial para el punto de interrupción de una reparación del habla. El propósito de este nivel es de ayudar, junto a los demás niveles previamente descritos, a determinar las características útiles para poder detectar que se realiza una reparación. El Ejemplo 73 muestra una elocución etiquetada en este nivel.

### Ejemplo 73 (d17 – utt39 Nivel de etiquetación de *break indices*)

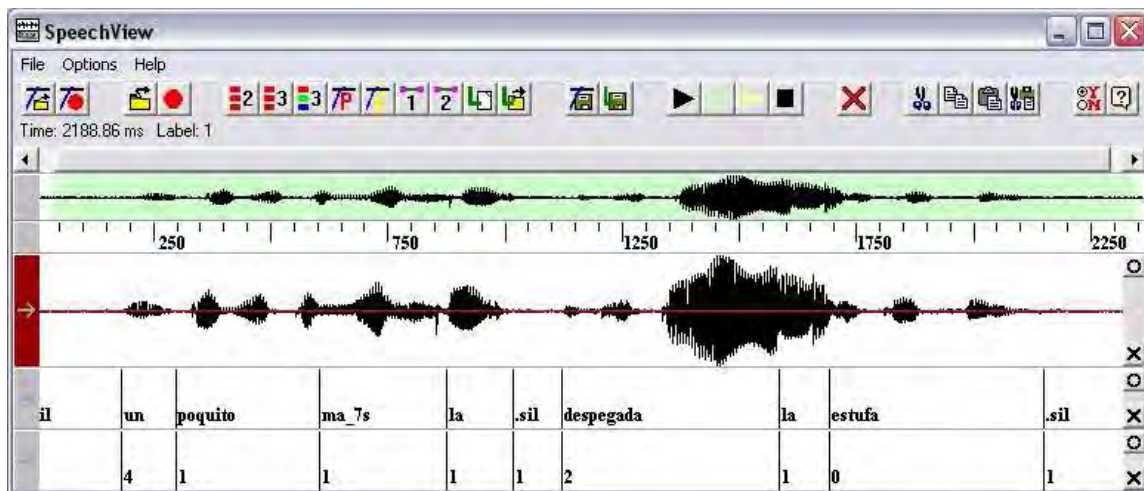


Figura 3.14: Nivel de *break indices*

## 3.6 El etiquetado

En este trabajo se etiquetaron 8 diálogos en cada uno de los niveles descritos en las secciones anteriores. En la Tabla 3.7 se puede observar el número de elocuciones que conforman a cada uno de estos diálogos, así como el número de elocuciones etiquetadas en cada nivel.

Diálogo	Elocuciones	Reparaciones del Habla	Marcadores del Discurso	POS
d01	116	26	49	115
d03	168	19	93	161
d12	117	12	36	115
d13	191	6	62	189
d15	90	5	0	83
d17	237	19	104	236
d19	105	7	47	105
d23	81	11	40	81
<b>Total</b>	<b>1105</b>	<b>105</b>	<b>431</b>	<b>1085</b>

Tabla 3.7: Número de elocuciones etiquetadas por diálogo

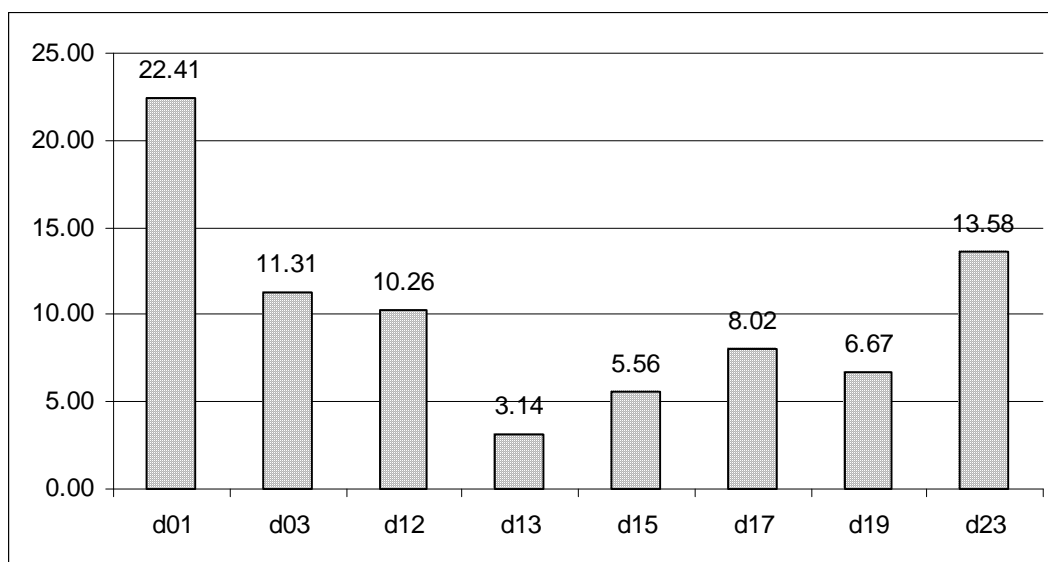
La diferencia que existe entre el número total de elocuciones por diálogo y el número de elocuciones etiquetadas del mismo diálogo se debe a que no todas las elocuciones presentan disfluencias o reparaciones. En el caso de POS, existen elocuciones que sólo están formadas por ruido, y por lo tanto no se deben etiquetar estos casos.

El objetivo del análisis del material empírico, proporcionado con el corpus DIME, fue determinar que información es útil para detectar las disfluencias del habla y a su vez definir un método para corregir las reparaciones más comunes.

Como era de esperarse, no todas las elocuciones presentaron disfluencias del habla sólo el 9.5% del material etiquetado tiene al menos una reparación del habla. En la Tabla 3.8 se puede observar cuantas elocuciones por diálogo presentaron reparaciones.

<b>Diálogo</b>	<b>Elocuciones</b>	<b>Reparaciones</b>	<b>Porcentaje de reparaciones</b>
d01	116	26	22.41
d03	168	19	11.31
d12	117	12	10.26
d13	191	6	3.14
d15	90	5	5.56
d17	237	19	8.02
d19	105	7	6.67
d23	81	11	13.58
<b>General</b>	<b>1105</b>	<b>105</b>	<b>9.50</b>

**Tabla 3.8: Relación del número de reparaciones por cada diálogo**



**Figura 3.15: Porcentajes de elocuciones con reparación por diálogo**

Aunque las disfluencias del habla se presentan en menos del 10% de las elocuciones, se presentan características muy similares que ayudan a la detección y corrección. El análisis para obtener estas características se describe en los siguientes capítulos.

## 4 Detección de una reparación

d19 – utt026: <.sil> a ver <.sil> en este <.sil> en este lugar <.sil>  
d19 – utt091: <.sil> bueno creo que es todo lo que <.sil> lo que necesito  
<.sil> para mi cocina <.sil>

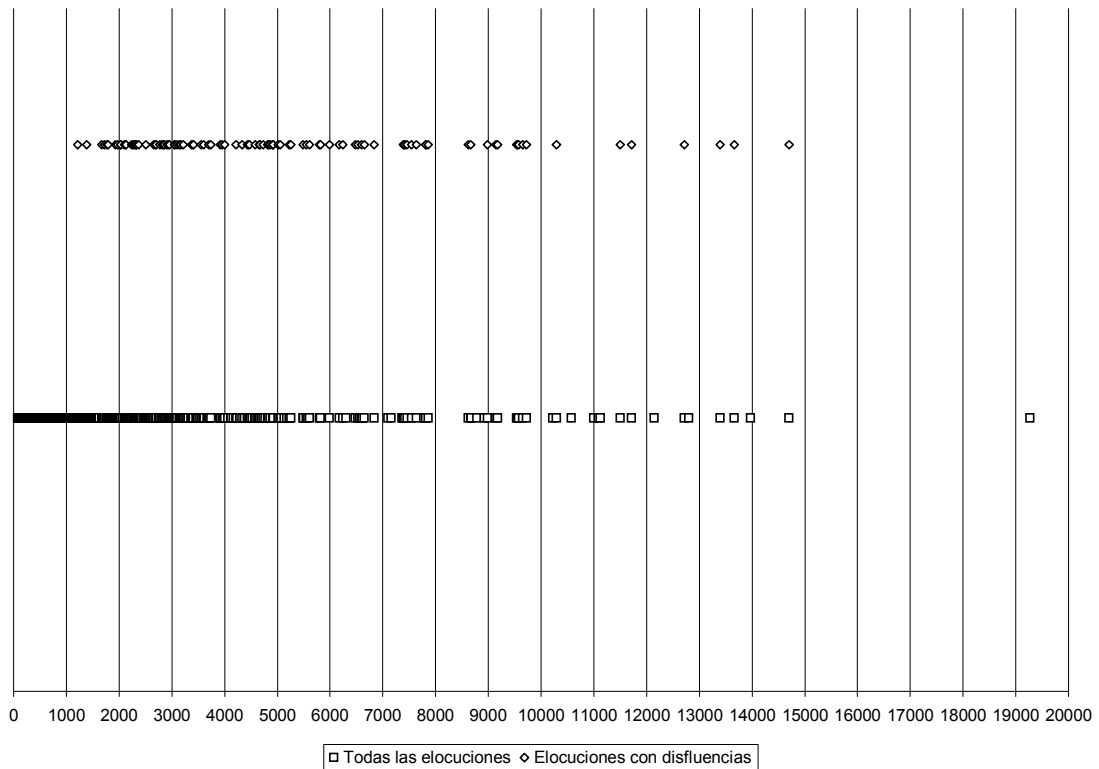
---

Cómo que si se parecen

En este capítulo se describe el análisis realizado para detectar las reparaciones del habla. El análisis se basó en la obtención de atributos útiles para la clasificación de una elocución. La clasificación sólo consiste de 2 conjuntos: con o sin disfluencias. El análisis de la información empírica dio como resultado 4 atributos útiles para la detección: duración de la elocución, número de palabras en la elocución, presencia de silencio en la elocución y tipo de acto del diálogo. El método seleccionado para realizar la detección de una reparación fue la creación de árboles de decisión. Para la creación del árbol se consideraron las 4 variables que se mencionaron.

### Duración de la elocución

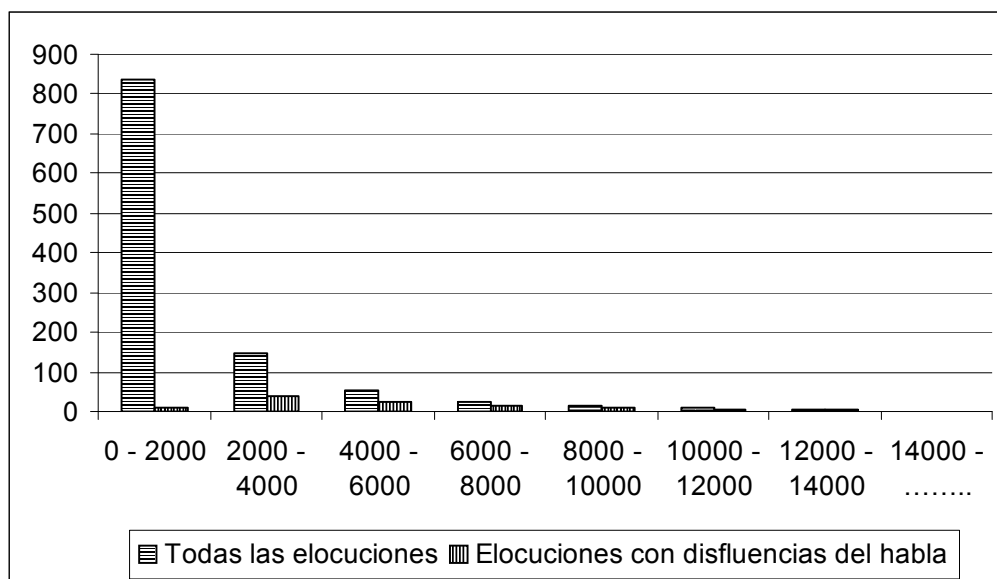
Como primera parte del análisis se graficó la duración de cada una de las elocuciones, distinguiendo su duración con y sin disfluencias; la Figura 4.1 muestra la gráfica obtenida (las unidades están en milisegundos).



**Figura 4.1: Gráfica de duraciones**

Como se puede observar, las elocuciones con disfluencias tienen una duración mayor. También se puede ver que la mayoría de las elocuciones con disfluencias tienen una duración mayor de 2,000 milisegundos. En base a esto, se generaron rangos de 2000 milisegundos y se contabilizó el número de elocuciones que pertenecen a cada rango. El resultado se muestra en la Figura 4.2.





**Figura 4.2: Clasificación de las elocuciones en rangos de 2,000 milisegundos**

Como se puede percibir, el rango de elocuciones con menos de 2,000 milisegundos es el mayor y corresponde al 76.72% de las elocuciones, y sólo el 0.96% de estas presentan disfluencias. Por otro lado, en los rangos restantes son mucho más frecuentes las disfluencias. En la Tabla 4.1 se puede observar el porcentaje de aparición de disfluencias en cada un de los rangos restantes.

Rango	Porcentaje de disfluencias
2000 - 4000	25.50%
4000 - 6000	50.00%
6000 - 8000	66.67%
8000 - 10000	76.92%
10000 - 12000	37.50%
12000 - 14000	50.00%
14000 - .....	50.00%

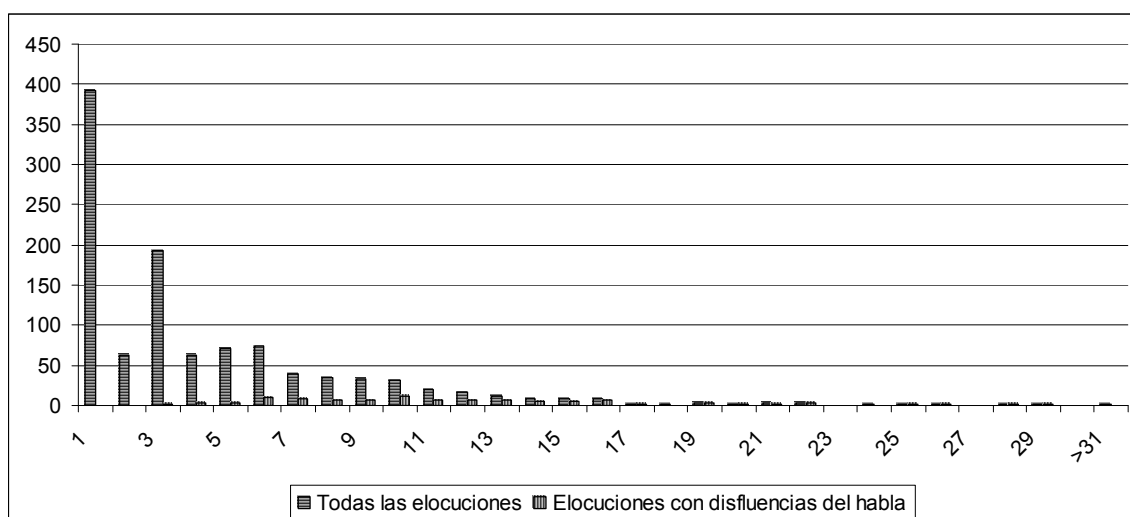
**Tabla 4.1: Porcentaje de reparaciones por cada 2,000 milisegundos**

Si se considera como un solo rango todas las elocuciones con más de 2,000 milisegundo, el porcentaje de elocuciones que presentan disfluencias es del 38.19%. Como se puede apreciar, la duración de una elocución ayuda a clasificarla; esta característica

permite realizar la clasificación más rápido porque al estar la mayor parte de la información de un solo rango (elocuciones menores a 2,000 milisegundos) el análisis se puede enfocar a un conjunto más específico de datos.

### Número de palabras en una elocución

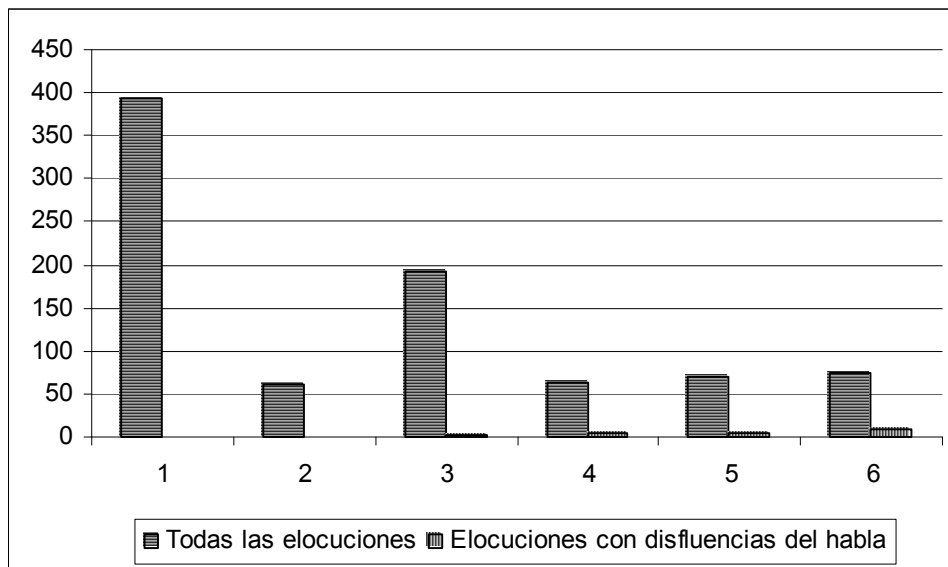
Cuando un hablante realiza una disfluencia normalmente debe agregar más palabras para repararla; esto lo hace al repetir palabras o al cambiar de idea y comenzar de nuevo. Para medir este parámetro se graficó el número de palabras que presenta cada elocución indicando si se presenta una reparación o no, esto se puede apreciar en la Figura 4.3.



**Figura 4.3: Disfluencias según el número de palabras en la elocución**

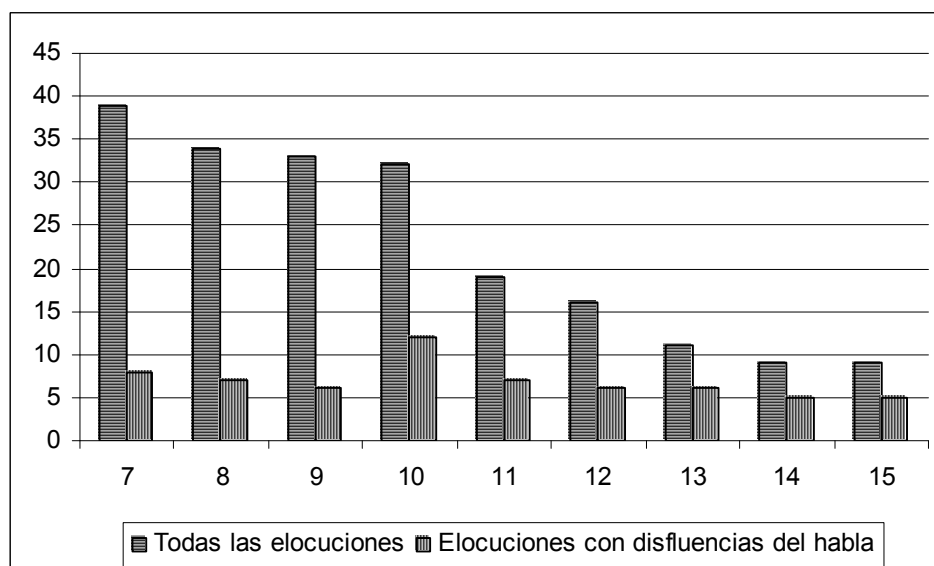
En esta gráfica se definieron 3 regiones: elocuciones con 6 o menos palabras (R1), entre 7 y 15 (R2) y elocuciones con más de 15 palabras (R3).

La región R1 se muestra en la Figura 4.4. Esta región contiene el 78.46% de los datos etiquetados y, como se esperaba, las elocuciones con al menos una disfluencia del habla son tan sólo el 2.34%.



**Figura 4.4: Número de palabras: región 1 – 6 o menos palabras**

La región R2 está formada por el 18.52% del material empírico, como se muestra en la Figura 4.5. Esta es la región crítica debido a que el 30.69% de las elocuciones de esta región presentan disfluencias; el porcentaje de aparición de disfluencias no es tan grande para considerar que el rango está predominado por ellas, pero a la vez es un porcentaje bastante aceptable para decir lo contrario.



**Figura 4.5: Número de palabras: región 2 – entre 7 y 15 palabras**

Para observar que comportamiento presentan las elocuciones en este rango se obtuvo el promedio de la duración de las elocuciones con disfluencias y el de las elocuciones sin disfluencias para elocuciones con el mismo número de palabras. El resultado se muestra en la Tabla 4.2.

<b>Número de Palabras</b>	<b>Duración promedio sin disfluencia [ms]</b>	<b>Duración promedio con disfluencia [ms]</b>
7	2461	2939
8	2818	2978
9	2902	4456
10	3507	3622
11	4289	5063
12	5378	7039
13	4902	5789
14	5266	5497
15	6834	6243

**Tabla 4.2: Relación de duración vs. número de palabras**

Con excepción del caso de elocuciones con 15 palabras, en todos los casos la duración de una elocución con disfluencias fue mayor que sin disfluencias. El cálculo de las duraciones promedio se hizo también en los rangos R1 y R3, como resultado se obtuvo que en el 57% de los casos la duración de una elocución con disfluencias es mayor que la duración de una sin disfluencias, el 13% presentó el caso contrario y en el 30% restante ambas duraciones fueron iguales; este último resultado se debe a que en todos esos casos no se presentaron elocuciones sin disfluencias. Con esta observación se confirma aún más que la duración de una elocución es de gran ayuda para la detección de una disfluencia del habla.

La última región R3 de la Figura 4.6 contiene el 3.02% de las elocuciones; en esta región el 69.70% de las elocuciones presentan al menos una disfluencia; aunque es una región pequeña, este dato nos confirma que entre más palabras tiene una elocución, la probabilidad de que presente una disfluencia se incrementa.

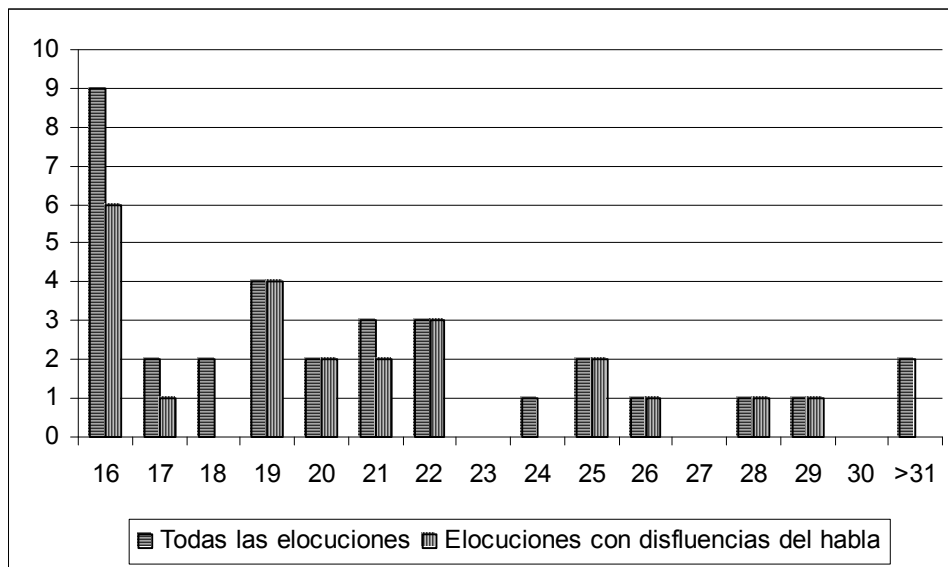
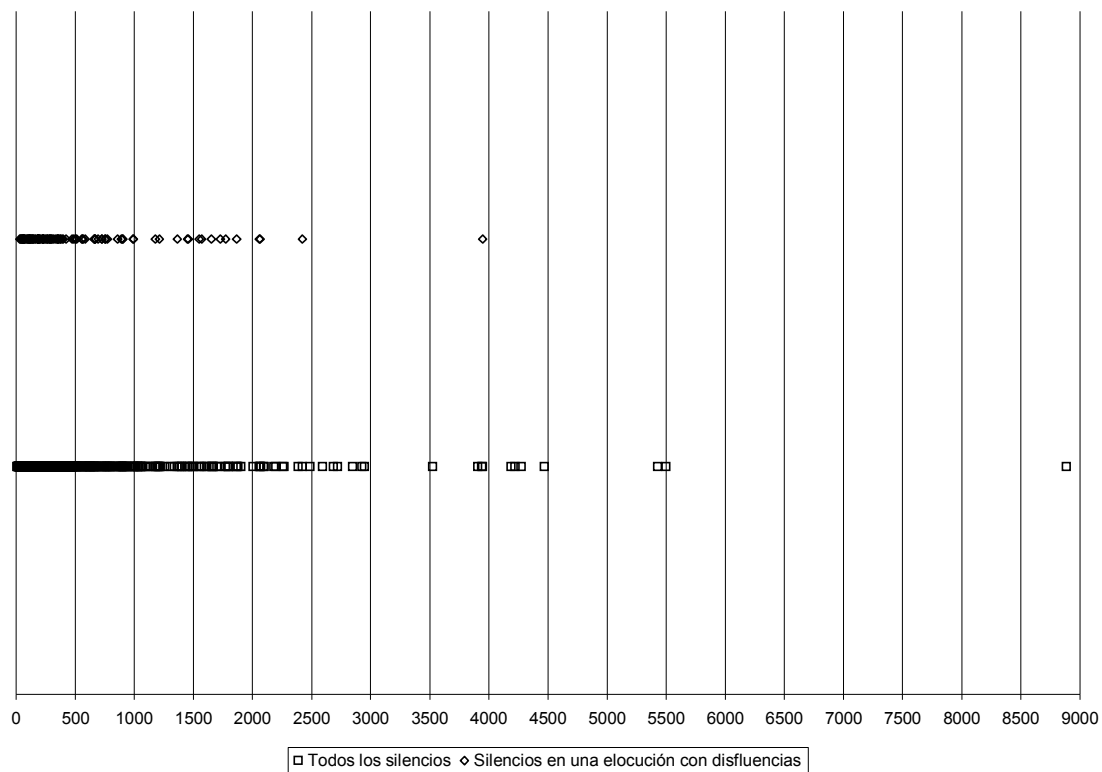


Figura 4.6: Número de palabras: región 3 – más de 15 palabras

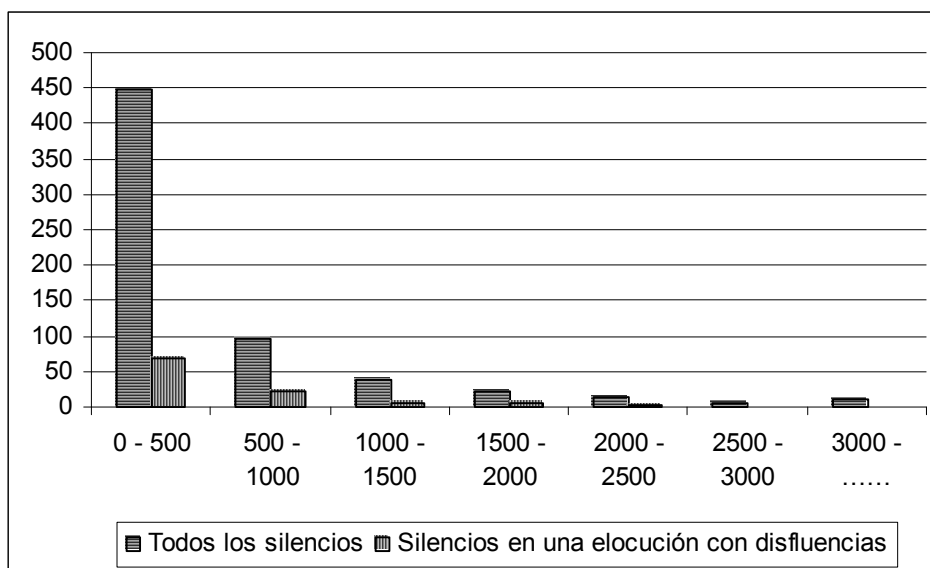
### Silencio en una elocución

Con una idea similar a la que se planteó con la duración de una elocución, se graficó la duración de los silencios que se presentan en las elocuciones. Se consideró la hipótesis de que la duración de un silencio dentro de una elocución con disfluencia del habla es más larga que la duración de un silencio en una elocución que no presenta disfluencia; el fundamento de la hipótesis es que el hablante al expresar disfluencias comúnmente utiliza pausas para realizar la reparación. La gráfica obtenida se muestra en la Figura 4.7 (la unidades de medida están en milisegundos).



**Figura 4.7: Gráfica de la duración de los silencios**

Sin embargo los resultados no apoyan esta hipótesis; en general, la duración de un silencio en una elocución con disfluencia no es más larga que en una sin disfluencia, como se puede observar en la Figura 4.7. Tomando en cuenta rangos de 500 milisegundos, como se muestra en la Figura 4.8, se observa que en ningún de ellos predominan las elocuciones con disfluencias del habla; por esta razón la duración de un silencio no facilita la identificación de una disfluencia.



**Figura 4.8: Clasificación de los silencios en rangos de 500 milisegundos**

Sin embargo, la presencia de un silencio si es relevante para la detección. Una característica que tiene las elocuciones con reparaciones del habla es que comúnmente después del punto de interrupción sigue un silencio; el 85.71% de las elocuciones que tiene al menos una reparación del habla presenta un silencio, por lo tanto, esta característica es útil para detectar que hay una reparación. Asimismo, al existir presencia de silencios dentro de la elocución contribuye a que la duración de dicha elocución aumente.

### **Acto del diálogo**

Como parte final del análisis se observó que tipo de acto de habla o diálogo es relevante en las elocuciones para identificar las disfluencias; un acto del diálogo es la acción lingüística que expresa una intención. Una elocución puede expresar más de un tipo de acto del diálogo a la vez. El esquema de etiquetado para los actos del diálogo es una adaptación aumentada para el corpus DIME del esquema DAMSL [1], denominado DIME-DAMSL. Para llevar a cabo este análisis se tomó el acto del diálogo más evidente; en particular se encontró que las disfluencias del habla están muy relacionadas con 2 tipos de actos de diálogo: directivas de acción (64%) y afirmaciones (30%). Por lo tanto, la identificación del tipo de acto del diálogo del hablante es relevante para facilitar la clasificación de las elocuciones.

#### 4.1 Definición de árboles de decisión

En esta sección se presenta un método para identificar de las disfluencias del habla basado en la creación de árboles de decisión con la ayuda de las variables útiles definidas en la sección anterior.

El árbol de decisión es un diagrama que representa, de manera jerárquica, condiciones y acciones; sirve para modelar funciones que se forman a través un conjunto de variables y basándose en el valor de cada una de ellas (preguntas), determinan la acción a tomar (toma una decisión o realiza un clasificación). Al responder una pregunta se llega a una decisión, y es jerárquico porque la toma de una decisión lleva a otra; además, por cada pregunta pueden existir más de dos opciones ([21], [11]).

A cada pregunta del árbol se denomina nodo; al primer nodo se le llama raíz y a partir de él se generan los siguientes nodos. Se llama nodo padre a aquél que genera al menos otro nodo; a los nodos generados por un nodo padre se les denomina nodos hijos; un padre puede tener varios hijos pero un hijo sólo tiene un padre; un nodo que no tiene hijos se le llama hoja. Cada hoja representa una decisión final o clasificación. Un ejemplo de árbol de decisión se muestra en la Figura 4.9.

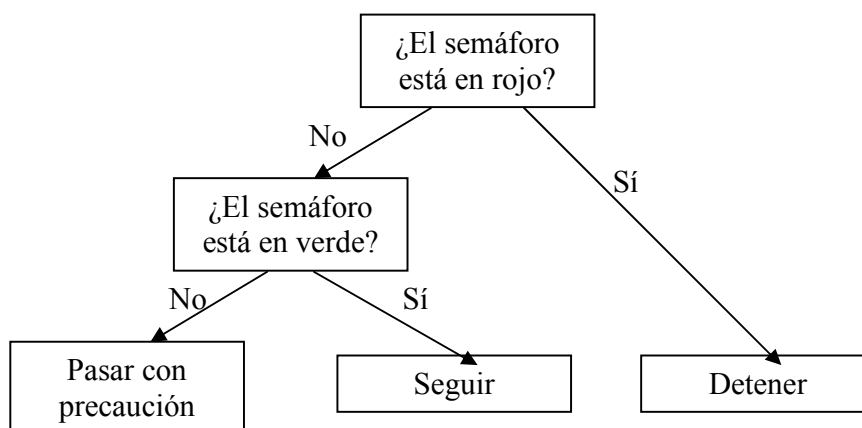


Figura 4.9: Ejemplo de árbol de decisión

El experimento realizado en esta tesis consistió en construir árboles de de decisión estilo CART<sup>10</sup>. Este tipo de árboles son denominados árboles de partición binaria recursiva [2]; son binarios porque cada padre es dividido en exactamente dos hijos y es recursivo

<sup>10</sup> <http://www.salford-systems.com/cart.php>



porque el proceso se realiza a cada hijo. El algoritmo tiene la posibilidad de generar más de dos hijos cuando las características de alguna variable no permiten hacer una división binaria.

El algoritmo de creación de árboles estilo CART consiste, en primer lugar, en buscar las preguntas o reglas de partición. La búsqueda es por fuerza bruta; esto es, el algoritmo considera todas las posibles particiones que se puedan generar de las diferentes variables y los casos a clasificar. Por ejemplo, si hay 10 tipos de clasificación y se emplean 4 variables características, se generarían 40 reglas de partición, porque por cada una de las variables se generan tantas reglas de partición como clasificaciones se quieran realizar.

La segunda parte es la elección del orden de las reglas. La elección se basa en un criterio de calidad de la división. El criterio por defecto es la regla GINI, es una medida de que tan bien la regla separa las clases contenidas en un nodo padre [20].

Una vez que se ha encontrado la mejor partición, el proceso se repite en los nodos hijos continuando recursivamente hasta que sea imposible la partición o esta se haya detenido; no es posible seguir con la partición si en el nodo sólo existe un caso o si los casos que le pertenecen son una copia exacta de otro nodo. La partición se detiene sin en el nodo hay muy pocos casos clasificados, por defecto el límite es de 10 casos.

La última parte consiste en la etapa de prueba, la cual determina el mejor árbol. Esta etapa puede ser llevada a cabo por dos tipos de pruebas: porcentaje de división (*percentage split*) y validación cruzada (*cross-validation*) [32].

La prueba de porcentaje de división es el método más simple y consiste en definir un porcentaje del material que va a ser utilizado para la generación del árbol y el material restante será para pruebas. Esta prueba es mayormente recomendada cuando se tiene suficiente material.

La prueba de validación cruzada es recomendada cuando no se tiene mucho material. Consiste en dividir el material en  $n$  partes iguales, con  $n-1$  partes se crea el árbol y con la parte restante se realiza la prueba. Esto se realiza con cada una de las  $n$  partes.

El experimento que se realizó tomó como variables los atributos más característicos de las disfluencias del habla que fueron definidos en la etapa de análisis: duración de la elocución, número de palabras, presencia de un silencio y tipo de acto del diálogo. La clasificación sólo se basa en dos conjuntos: tiene o no tiene disfluencia del habla.

Como material se tomaron las 105 elocuciones con disfluencias del habla y se tomaron otras 105 elocuciones sin disfluencias. Las elocuciones sin disfluencias fueron tomadas aleatoriamente de todos los diálogos con la misma distribución que las elocuciones con disfluencias; en otras palabras, si en el diálogo  $X$  hay 5 elocuciones con disfluencias del habla, se seleccionaron 5 elocuciones sin disfluencias del mismo. El propósito de este proceso fue mantener el material lo más balanceado posible.

Se realizaron dos variantes del experimento; en la primera se tomaron las elocuciones sin disfluencias tanto del sistema como del usuario y en la segunda se tomaron únicamente del usuario.

Las pruebas que se hicieron fueron:

- Validación cruzada con  $n$  igual a: 2, 3, 5, 10, 15 y 20
- Porcentaje de división del: 50%, 66%, 70%, 75% y 90%

El software empleado para crear los árboles de decisión y llevar a cabo estos experimentos fue WEKA<sup>11</sup>.

Los resultados obtenidos en la primera variante del experimento se muestran en la Tabla 4.3 y Tabla 4.4 y el árbol generado en la Figura 4.10.

Tipo de Prueba	Casos		Kappa	Precisión	Recall
	correctamente clasificados (%)				
2	87.62	0.7524	N <sup>12</sup>	0.954	0.79
			S <sup>13</sup>	0.821	0.962
3	86.67	0.7333	N	0.914	0.81
			S	0.829	0.924
5	87.14	0.7429	N	0.924	0.81
			S	0.831	0.933
10	86.67	0.7333	N	0.923	0.8
			S	0.824	0.933
15	85.71	0.7143	N	0.895	0.81
			S	0.826	0.905

<sup>11</sup> <http://www.cs.waikato.ac.nz/ml/weka>

<sup>12</sup> Significa "No tiene disfluencia"

<sup>13</sup> Significa "Tiene disfluencia"

20	85.71	0.7143	N	0.903	0.8
			S	0.821	0.914

**Tabla 4.3: Resultados de la prueba de validación cruzada para la primera variante del experimento**

Tipo de Prueba	Casos			Precisión	Recall
	correctamente clasificados (%)	Kappa			
50%	85.71	0.7117	N	1	0.706
			S	0.783	1
66%	86.11	0.715	N	0.96	0.727
			S	0.809	0.974
70%	87.30	0.7372	N	0.955	0.75
			S	0.829	0.971
75%	84.91	0.6831	N	0.941	0.696
			S	0.806	0.967
90%	90.48	0.7407	N	1	0.667
			S	0.882	1

**Tabla 4.4: Resultados de la prueba de porcentaje de división para la primera variante del experimento**

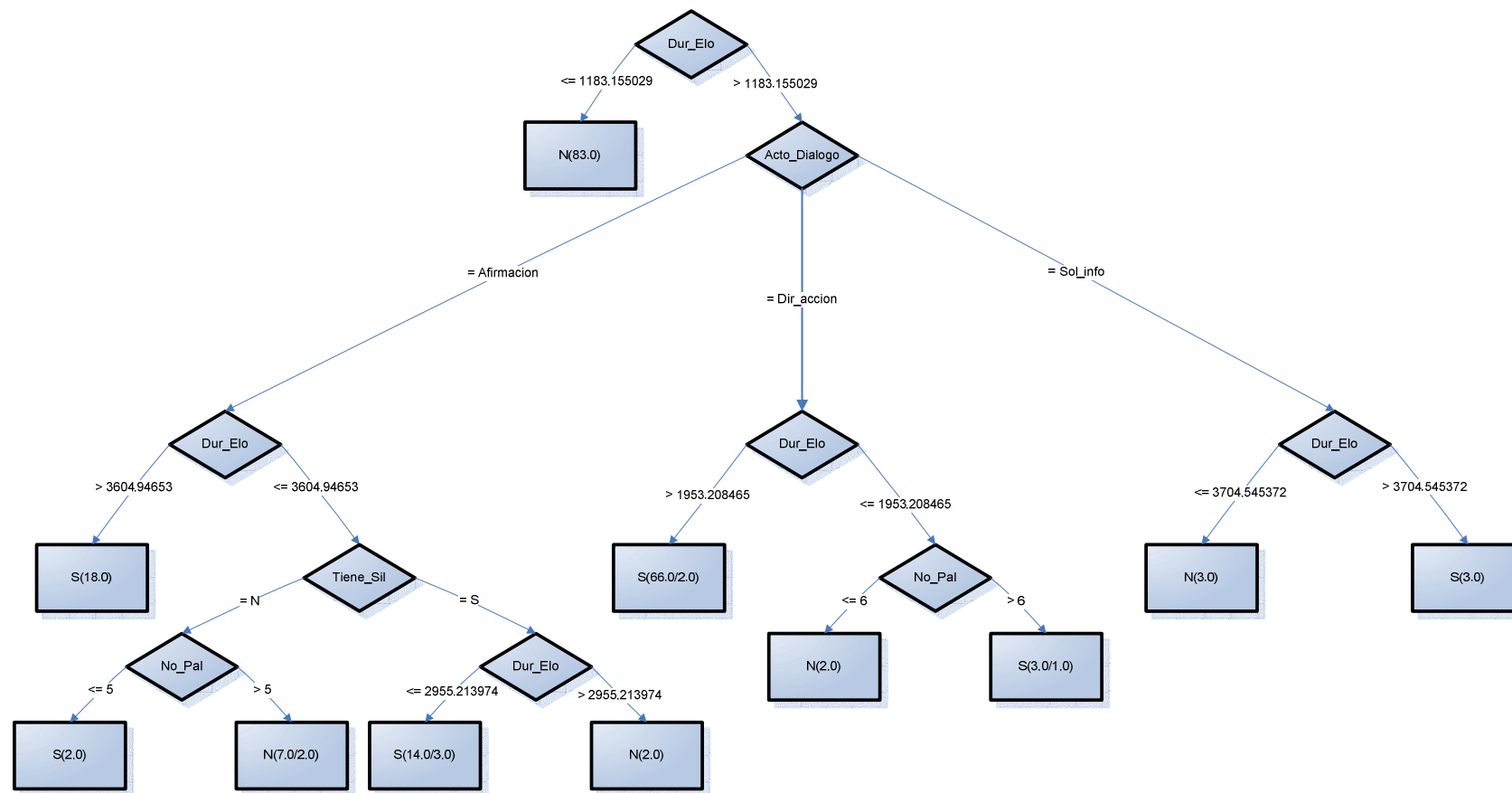


Figura 4.10: Árbol de decisión para la detección de disfluencias del habla – primera variante

Los resultados obtenidos con cada prueba fueron bastante semejantes y muy aceptables; de acuerdo con lo esperado, la variable más importante del árbol es la duración de una elocución. La duración es el nodo principal del árbol y del cual se clasificaron 83 casos correctamente, además se puede confirmar aún más su importancia al observarse que varios nodos del árbol vuelven a referirse a la duración. En segundo término queda el tipo de acto de diálogo, por el cual se hace la segunda división del árbol; además se puede observar que el árbol se parte en más de dos caminos porque los valores de esta variable no son binarios sino un conjunto de datos. Al final, se consideran la presencia de un silencio y el número de palabras.

Después de este resultado se realizó la segunda variante del experimento. Los resultados obtenidos se muestran en la Tabla 4.5 y Tabla 4.6 y el árbol generado en la Figura 4.11.

Tipo de Prueba	Casos		Kappa	Precisión	Recall
	correctamente clasificados (%)				
2	83.33	0.6667	N	0.907	0.743
			S	0.782	0.924
3	84.29	0.6857	N	0.867	0.81
			S	0.821	0.876
5	86.67	0.7333	N	0.897	0.829
			S	0.841	0.905
10	86.67	0.7333	N	0.874	0.857
			S	0.86	0.876
15	89.05	0.781	N	0.927	0.848
			S	0.86	0.933
20	86.67	0.7333	N	0.889	0.838
			S	0.847	0.895

**Tabla 4.5: Resultados de la prueba de validación cruzada para la segunda variante del experimento**

Tipo de Prueba	Casos			Precisión	Recall
	correctamente clasificados (%)	Kappa			
50%	80.95	0.6104	N	0.914	0.653
			S	0.757	0.946
66%	76.39	0.5049	N	0.857	0.563
			S	0.725	0.925
70%	79.37	0.5473	N	0.833	0.6
			S	0.778	0.921
75%	77.36	0.4709	N	0.769	0.526
			S	0.775	0.912
90%	85.71	0.6897	N	0.75	0.857
			S	0.923	0.857

**Tabla 4.6: Resultados de la prueba de porcentaje de división para la segunda variante del experimento**

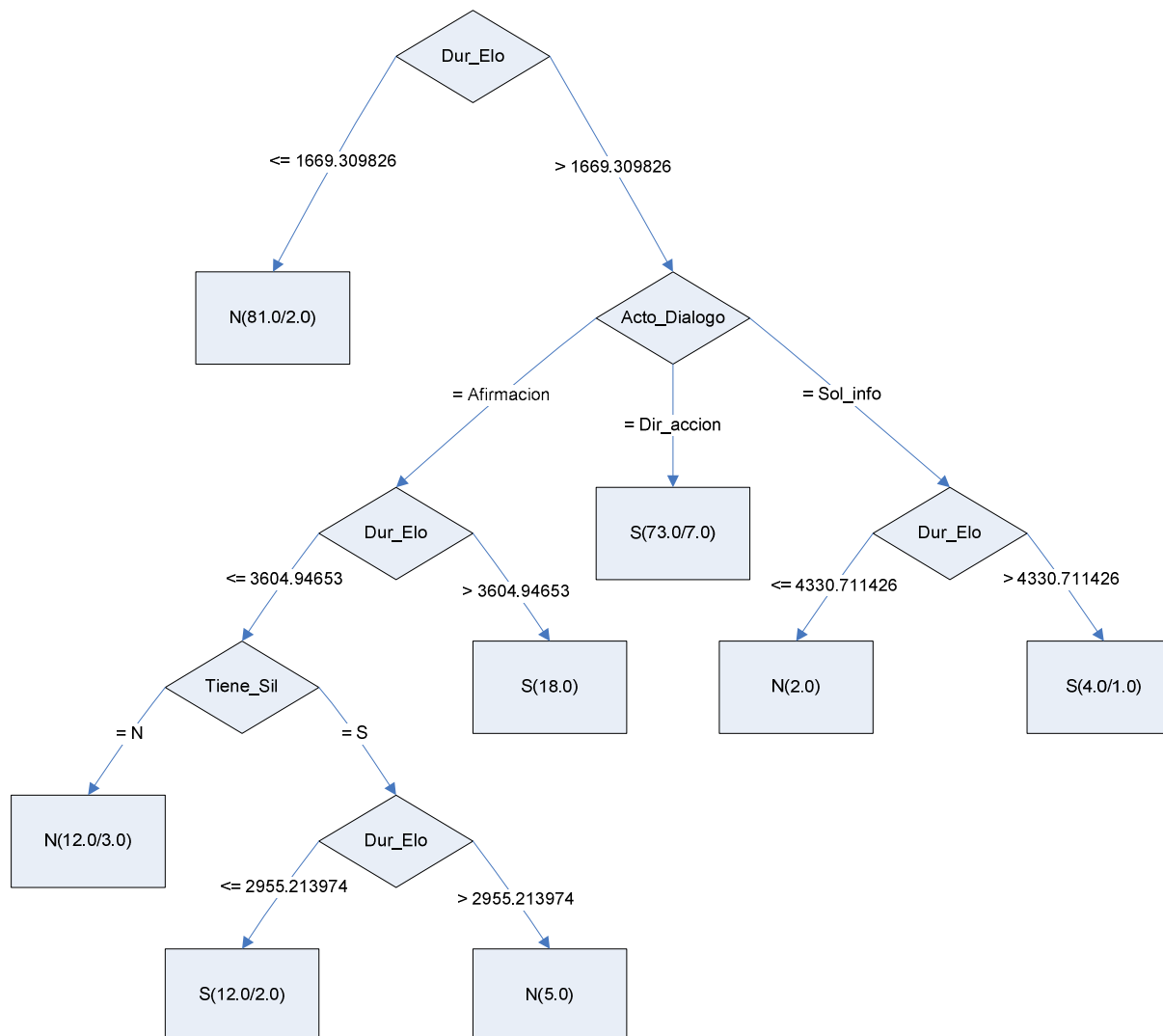


Figura 4.11: Árbol de decisión para la detección de disfluencias del habla – segunda variante

A diferencia de la primera versión, las pruebas con porcentaje de división no salieron tan bien como las pruebas de validación cruzada. Los resultados obtenidos con porcentaje de división indican que la información no es suficiente ya que estas pruebas tienen un buen comportamiento cuando el material de pruebas es suficiente. Por otro lado, las pruebas de validación cruzada en esta variante son similares a los resultados obtenidos en la primera variante.

El árbol obtenido en esta ocasión es muy similar al que se obtuvo en la primera variante; conserva una estructura muy semejante con la diferencia de que es más pequeño. Se observa que la duración de una elocución sigue siendo la variable más importante para la clasificación, seguida por el tipo de acto del diálogo, como fue en el caso anterior. Una característica muy importante en este segundo experimento es que el número de palabras fue eliminado por el algoritmo, confirmando que no es una variable necesaria. La razón por la que no es importante el número de palabras se debe a que una elocución con disfluencias y  $X$  número de palabras, en promedio, su duración es mayor a la de una elocución sin disfluencias y el mismo número de palabras; y por lo tanto, esta característica es considerada en todas las preguntas referentes a la duración.

Los resultados obtenidos en cada experimento fueron bastante aceptables, principalmente con las pruebas de validación cruzada; la mayoría de las elocuciones fueron clasificadas correctamente, en promedio el 86.7%; además de confirmar que las variables seleccionadas son realmente características de las disfluencias del habla, principalmente la importancia de la duración de la elocución y como está estrechamente relacionada con otras variables, como es el caso del número de palabras,

Una vez que se detecta la disfluencia, lo que se prosigue es a corregir la elocución. En el capítulo siguiente se describe el análisis y método empleado para realizar la corrección de las reparaciones que se presentan con más frecuencia.



## 5 Corrección de una reparación

d01 – utt011: *entonces <.sil> por el momento quisiera mover este esta alacena al al lado de <.sil> de la estufa <.sil> que está aquí*

---

Y casi siempre así hablamos

Una vez que el reconocedor del voz haya procesa una elocución y se detecta una disfluencia por medio del árbol de decisión, se procede a intentar corregirla. El objetivo en este capítulo es determinar qué tipo de reparaciones son las que ocurren con más frecuencia y definir una heurística para su corrección.

### 5.1 Estrategia de corrección

En la comunicación oral, la presencia de reparaciones del habla es muy común, además es muy frecuente que el hablante se corrija en más de una ocasión, a esto se le denomina multirreparación. Para realizar el estudio de cada tipo de reparación del habla, como se mencionó en el capítulo 3, las reparaciones presentes en una multirreparación fueron etiquetadas por separado. El número de multirreparaciones por diálogo se puede observar en la Tabla 5.1.

<b>Diálogo</b>	<b>Elocuciones</b>	<b>Multirreparaciones</b>
d01	116	8
d03	168	7
d12	117	3
d13	191	1
d15	90	1
d17	237	6

d19	105	1
d23	81	7
<b>General</b>	<b>1105</b>	<b>34</b>

**Tabla 5.1: Número de multirreparaciones por diálogo**

El porcentaje de multirreparación, con respecto a las reparaciones que se tienen en la porción del corpus DIME etiquetado, es del 32.38%. El número de reparaciones que estaban involucradas en una multirreparación variaba desde 2 hasta 5. En la Tabla 5.2 se puede ver el porcentaje en cada uno de los diálogos.

<b>Diálogo</b>	<b>Reparaciones</b>	<b>Multirreparaciones</b>	<b>Porcentaje de Multirreparaciones</b>
d01	26	8	30.77
d03	19	7	36.84
d12	12	3	25.00
d13	6	1	16.67
d15	5	1	20.00
d17	19	6	31.58
d19	7	1	14.29
d23	11	7	63.64
<b>General</b>	<b>105</b>	<b>34</b>	<b>32.38</b>

**Tabla 5.2: Porcentaje de multirreparaciones con respecto al número de reparaciones**

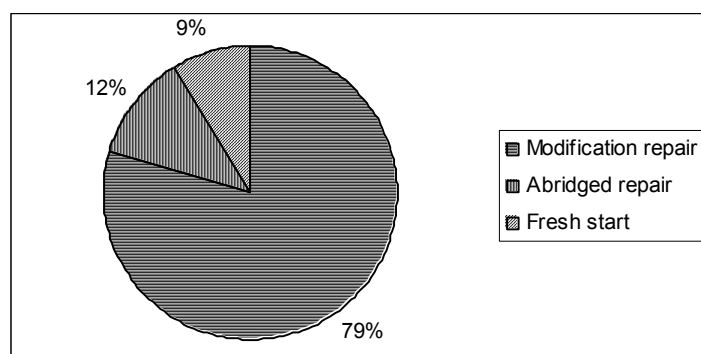
Para el análisis de la corrección, las reparaciones que forman una multirreparación fueron consideradas como elocuciones individuales, en consecuencia el número de elocuciones consideradas incrementó de 105 a 151. La Tabla 5.3 muestra en que cantidad se incrementó el número de reparaciones por diálogo considerando la división de las multirreparaciones como unidades separadas.

<b>Diálogo</b>	<b>Reparaciones sin división</b>	<b>Reparaciones divididas</b>
d01	26	42
d03	19	27
d12	12	15
d13	6	7

d15	5	6
d17	19	28
d19	7	8
d23	11	18
<b>General</b>	<b>105</b>	<b>151</b>

**Tabla 5.3: Comparación del número de reparaciones divididas y sin dividir**

Este análisis realizado para la corrección se enfocó en la definición de una heurística que pudiera ser implementada en un algoritmo para corregir el tipo de reparación del habla que se presentó en la mayoría de los casos. La distribución de cada tipo de reparación del habla se muestra en la Figura 5.1.

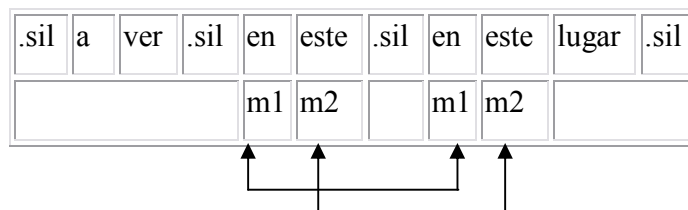


**Figura 5.1: Distribución de las reparaciones del habla**

Las reparaciones del tipo *modification repair* tiene casi el 80% de aparición, razón por la cual, el análisis se enfocó en este tipo de reparación. Estas reparaciones se dividen en tres tipos de correcciones: por repetición, simple y compleja.

La corrección por repetición, como su nombre lo dice, sólo consiste de elocuciones en las cuales la reparación sólo está formada de palabras repetidas, como se muestra en el Ejemplo 74.

**Ejemplo 74 (d19 – utt26 *Modification repair* – corrección por repetición)**



**Ejemplo 74 (d19 – utt26 *Modification repair* – corrección por repetición – corregida)**

a	ver	en	este	lugar
---	-----	----	------	-------

En las reparaciones de corrección simple se presentan en su mayoría palabras repetidas entre el *reparandum* y la alteración, y sólo unas cuantas palabras son sustituidas o eliminadas después de ellas. Un ejemplo de este tipo se muestra en el Ejemplo 75.

**Ejemplo 75 (d03 – utt92 *Modification repair* – corrección simple)**

.bn	entonces	el	primero	.bn	el	tercero	.bn
		m1	r1		m1	r1	

**Ejemplo 75 (d03 – utt92 *Modification repair* – corrección simple – corregida)**

entonces	el	tercero
----------	----	---------

El tercer tipo, la corrección compleja, está formado por todas las elocuciones que no tienen palabras repetidas entre el *reparandum* y la alteración, y si las presentan pasa de ser percibidas al haber demasiadas palabras que son sustituidas, eliminadas o insertadas antes y/o después de las palabras repetidas. El Ejemplo 76 muestra reparación donde no existen palabras repetidas y en el Ejemplo 77 se muestra un caso donde hay palabras repetidas pero la corrección involucra demasiadas palabras sustituidas antes y después de ellas.

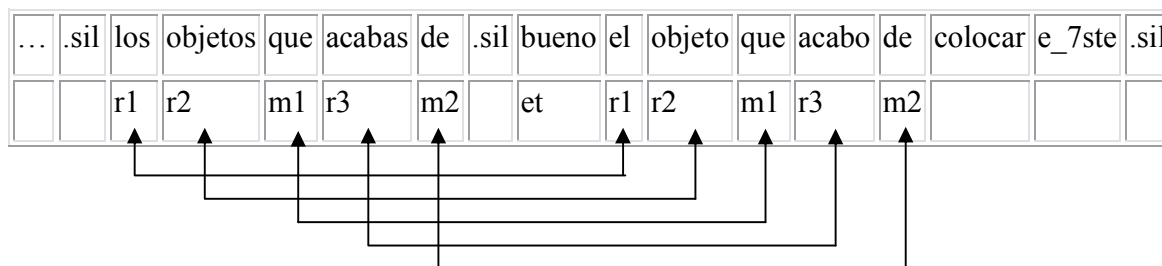
**Ejemplo 76 (d12 – utt93 *Modification repair* – corrección compleja)**

.sil	a	donde	esta_7	las	alacenas	los	gabinetes	.bn
				r1	r2	r1	r2	

**Ejemplo 76 (d12 – utt93 *Modification repair* – corrección compleja – corregida)**

a	donde	esta_7	Los	gabinetes
---	-------	--------	-----	-----------

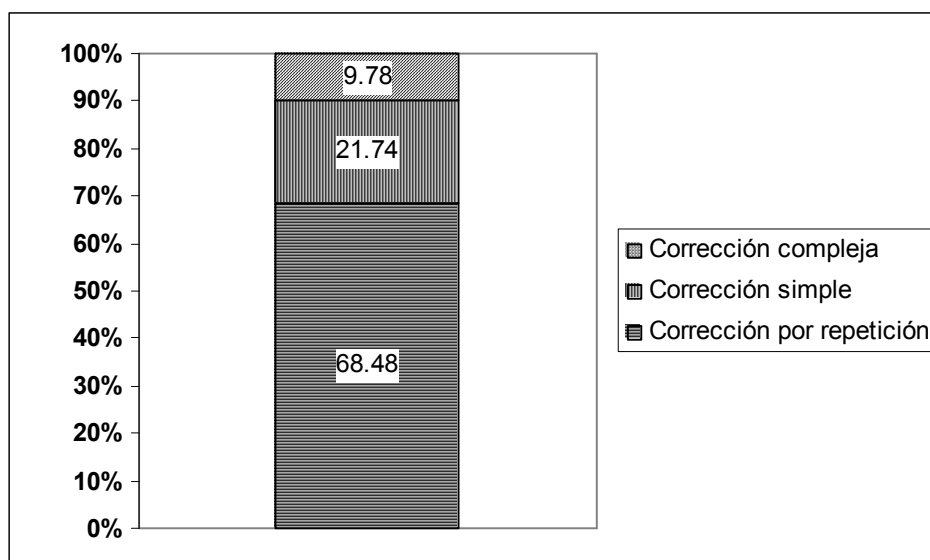
**Ejemplo 77 (d01 – utt33 *Modification repair* – corrección compleja)**



**Ejemplo 77 (d01 – utt33 *Modification repair* – corrección compleja – corregida)**

porque	tenemos	que	mover	el	objeto	que	acabo	de	colocar	e_7ste
--------	---------	-----	-------	----	--------	-----	-------	----	---------	--------

La característica principal de las elocuciones con reparación del tipo *modification repair*, es la presencia de la repetición de palabras. El 76.67% de las elocuciones presentaron repeticiones de palabras, en la Figura 5.2 se muestra en que proporción se presentaron en los diferentes tipos de corrección definidos anteriormente.



**Figura 5.2: Relación de palabras repetidas por tipo de corrección de un *modification repair***

Como se puede observar en la Figura 5.2, un poco más del 90% de los casos con repetición de palabras se encuentran en los tipos: corrección por repetición y corrección simple; por lo mismo el algoritmo se enfoca en la solución de estos dos casos.

La definición de la heurística para la corrección de las reparaciones se basa principalmente en la identificación de las palabras repetidas en la elocución, al ser la una característica muy constante de los *modification repairs*,

Para mejorar la efectividad de la heurística se consideró otro factor más: la distancia medida en número de palabras<sup>14</sup> que separa a una palabra repetida de su similar. El fundamento es que cuando ocurre una repetición de palabras, la aparición de la palabra repetida es casi inmediata o muy cercana, como se muestra en el Ejemplo 78.

**Ejemplo 78 (d17 – utt215 Repetición de palabra)**

.sil	entonces	.sil	quiero	una	.sil	una	superior	.sil	doble
				m1		m1			

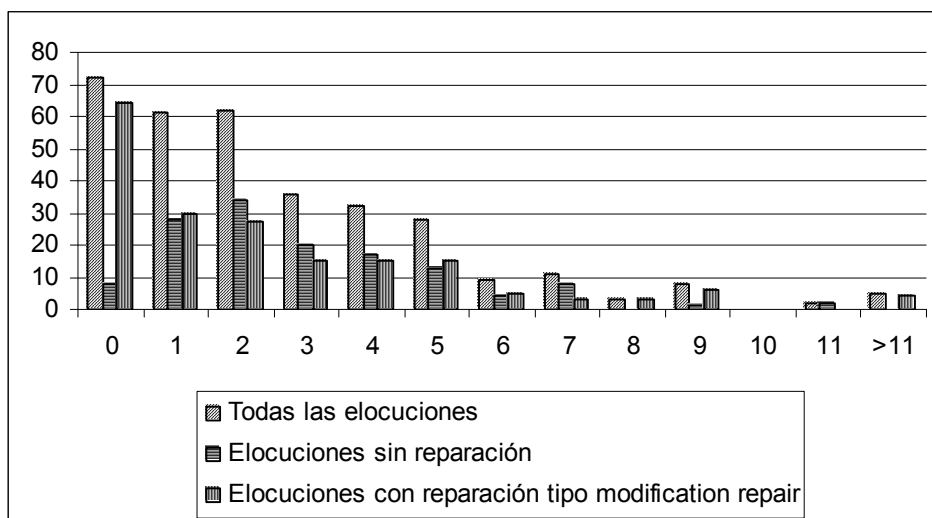
**Ejemplo 78 (d17 – utt215 Repetición de palabra – corregida)**

entonces	quiero	una	superior	doble
----------	--------	-----	----------	-------

Asimismo, en las reparaciones de corrección simple se presentan las palabras repetidas a pocas palabras de distancia, como se puede observar en el Ejemplo 75.

Para observar el comportamiento de la distancia entre palabras repetidas se obtuvieron todas las distancias de la porción de corpus etiquetado, teniendo como resultado la distribución de palabras mostrada en la Figura 5.3.

<sup>14</sup> No son considerados como palabras los silencios (.sil) y los ruidos (.bn)



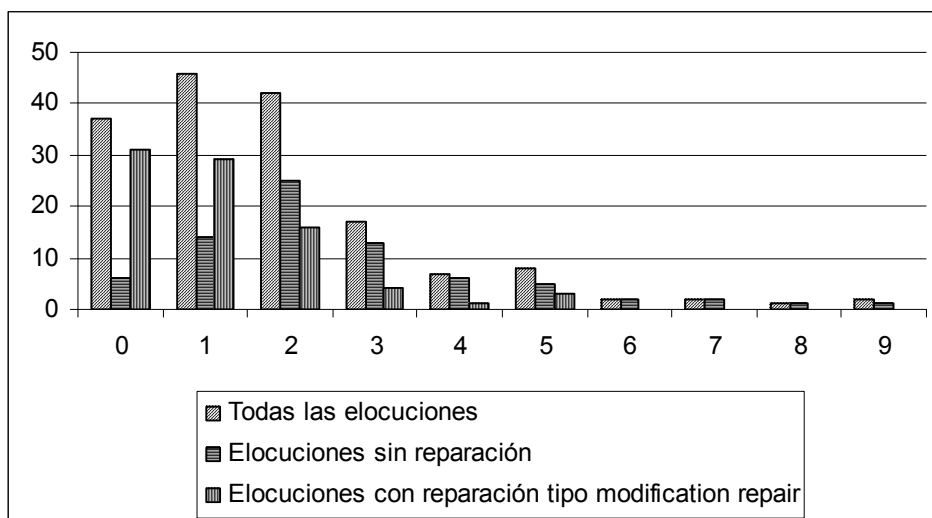
**Figura 5.3: Distancia entre palabras repetidas**

La gráfica de la Figura 5.3 representa todas las distancias de todas las palabras repetidas de las elocuciones; para realizar un mejor análisis se ajustó la distancia de las palabras, es decir, si dos o más palabras son adyacentes y se repiten no se cuenta. Esto se puede observar en el Ejemplo 79.

**Ejemplo 79 (d01 – utt102 Ajuste de distancias)**

		bueno	hay	que	borrar	.bn	hay	que	borrar	la	accio_7n	anterior
Distancia			2	2	2							
Distancia ajustada			0	0	0							

Las distancias se volvieron a graficar pero en esta ocasión solamente se consideró una distancia por elocución, la que tenga el menor valor. La gráfica resultante se muestra en la Figura 5.4.



**Figura 5.4: Distancia ajustada entre palabras repetidas**

Como se observa en la figura anterior, la mayoría de las elocuciones con reparaciones del habla se encuentran entre los primeros rangos, el 71.43% de las reparaciones tiene palabras repetidas con una separación de entre 0 y 1 palabra, y el 90.48% la tiene entre 0 y 2: además, se puede observar que las elocuciones sin reparaciones son pocas y tienden a dominar los rangos de más de 2 palabras de separación.

En conclusión, la heurística para la corrección de una reparación del habla del tipo *modification repair* está formada por la detección de las palabras repetidas y de la distancia, medida en número de palabras, que las separa.

## 5.2 Algoritmo de corrección

En esta sección se define el algoritmo para la corrección de las reparaciones más común, y de acuerdo a los datos que se tienen son los *modification repairs* y en particular las elocuciones con repetición de palabras. Los dos tipos de correcciones que se atacan son corrección por repetición y corrección simple.

El algoritmo de solución se basa en la heurística definida en la etapa de análisis, que consiste en la determinación de las palabras repetidas y en la distancia, medida en número de palabras, que separa a una palabra de su igual.

El algoritmo de corrección propuesto en este trabajo consta de 5 puntos que resuelven de una manera muy eficaz el problema de la corrección; a continuación se describe el algoritmo:



1. Se toma como entrada del proceso el texto que el sistema reconocedor del habla produce y se eliminan todos los silencios (.sil) y todos los ruidos (.bn) que la elocución presente.

**Ejemplo 80 (d01 – utt35 Punto 1 de la corrección – corrección por repetición)**

eh igual con la con la estufa

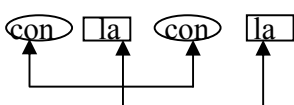
**Ejemplo 81 (d03 – utt95 Punto 1 de la corrección – corrección simple)**

entonces el Primero el tercero

2. Se identifican todas las palabras que se repiten a lo largo de la elocución.

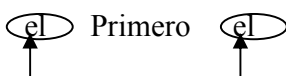
**Ejemplo 80 (d01 – utt35 Punto 2 de la corrección – corrección por repetición)**

eh igual con la con la estufa



**Ejemplo 81 (d03 – utt95 Punto 2 de la corrección – corrección simple)**

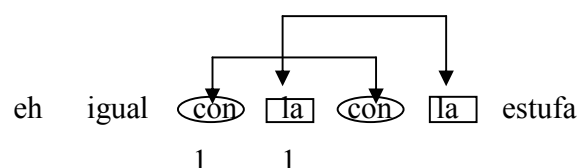
entonces el Primero el tercero



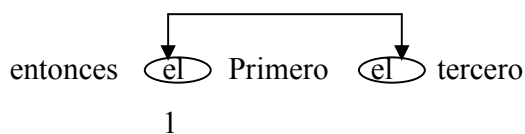
3. Se obtienen las distancias de cada una de las palabras repetidas. Se asocia la distancia a la primera aparición de cada palabra repetida.

**Ejemplo 80 (d01 – utt35 Punto 3 de la corrección – corrección por repetición)**

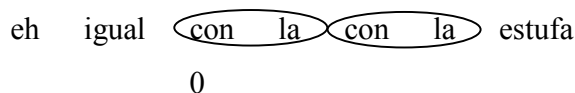
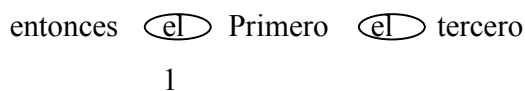
eh igual con la con la estufa



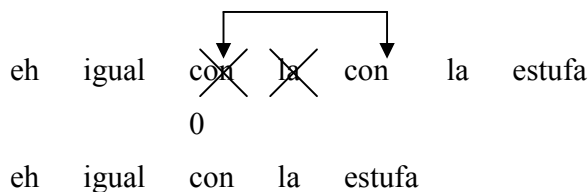
1 1

**Ejemplo 81 (d03 – utt95 Punto 3 de la corrección – corrección simple)**

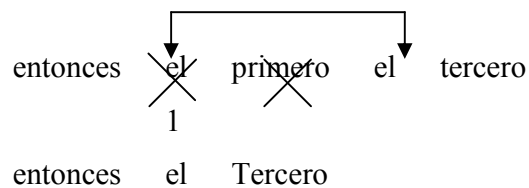
4. Se ajusta la distancia; este punto consiste en agrupar todas las secuencias de palabras que aparecen dentro de la elocución para considerarlas como una sola unidad y se recalcula la distancia entre estas unidades. La distancia se asocia a la primera palabras de la primera aparición de la unidad. Cuando no existen secuencias de palabras, esta etapa no tiene ningún efecto en la distancia original; esto sucede normalmente en correcciones simples como es el caso del Ejemplo 81.

**Ejemplo 80 (d01 – utt35 Punto 4 de la corrección – corrección por repetición)****Ejemplo 81 (d03 – utt95 Punto 4 de la corrección – corrección simple)**

5. Se remueven las palabras que se encuentren desde la palabra que tiene distancia menor o igual a 2 hasta encontrar la otra palabra repetida.

**Ejemplo 80 (d01 – utt35 Punto 5 de la corrección – corrección por repetición)**

**Ejemplo 81 (d03 – utt95 Punto 5 de la corrección – corrección simple)**



Con este algoritmo se resuelven los tipos de corrección más abundantes de las reparaciones del tipo *modification repair* que son la corrección por repetición, como es el caso del Ejemplo 80, o la corrección simple, como es el Ejemplo 81.

El pseudocódigo que se muestra a continuación es el que se implementó para llevar a cabo la corrección de las reparaciones:

```

elocucion = nuevoArreglo(Toma la elocución del reconocedor)
elocucion = quitaRuidosYSilencios(elocucion)
N = obtieneNumeroDeElementos(elocucion)
distancias = nuevoArreglo(N)
Desde i = 1 hasta N has
    palabra = elocucion[i]
    Si tieneRepetición(elocucion,palabra) entonces
        distancias[i] = obtenDistancia(elocucion,palabra,i)
    Si no
        distancias[i] = null
distancias = ajustaDistancia(distancia)
indicesDistancia2oMenor = obtieneIndices(distancias)
M = obtieneNumeroDeElementos(indicesDistancia2oMenor)
Desde i = 1 hasta M has
    indice = indicesDistancia2oMenor[i]
    palabra = elocucion[indice]
    Has
        palabras = quitaElemento(elocucion,indice)
        indice = indice + 1
        Si indice > N entonces
            Salir
    Mientras (elocucion[indice] != palabra) Y (indice < N)
Regresa elocucion
  
```

Los resultados obtenidos durante la ejecución de este algoritmo se muestran en la Tabla 5.4 (la suma de los porcentajes es el 100% del material etiquetado y con reparaciones del habla).

		A	B
		Se corrigió	No se corrigió
1	Se debía corregir	55%	18%
2	No se debía corregir	4%	23%

**Tabla 5.4: Resultados de la implementación del algoritmo de corrección**

El algoritmo recibe cualquier tipo de reparación (*fresh start*, *abridged repair* o *modification repair*) porque el árbol de decisión sólo clasifica las elocuciones en tiene o no tiene reparación. El material que se considera que no se debe de corregir está formado principalmente por elocuciones que tienen reparación del tipo *fresh start* y *abridged repair*. El resto de las elocuciones que no se deben de corregir son las elocuciones con *modification repair* de corrección compleja, porque no fueron consideradas en la definición de la heurística; por otro lado, el material que debe de ser corregido está formado por las elocuciones con *modification repair* de corrección por repetición y corrección simple.

En la Tabla 5.4 se muestra que 78% de los casos fueron correctamente procesados (celdas 1A y 2B) y sólo el 22% (celdas 2A y 1B) fue procesado inadecuadamente por la heurística. Del total de las elocuciones que se debieron corregir, el 75.34% de ellas fueron correctamente corregidas, las restantes no fueron corregidas o se corrigieron mal. La precisión obtenida en experimento fue de 0.94.

El resultado obtenido es bastante aceptable y muy prometedor para realizar la corrección, además de que se verifica la efectividad de la heurística y se valida que la repetición de palabras es una característica muy importante en las reparaciones del tipo *modification repair*.

## 6 Conclusiones

d12 – utt116: *hemos terminado la tarea*

---

Por ahora, aún faltan cosas por resolver

El fenómeno de las disfluencias del habla es muy amplio y complejo al presentar muchas variantes; sin embargo, al detectar las características representativas de las disfluencias y enfocar el análisis a la solución de las reparaciones más comunes, se puede generar un sistema que sea eficiente y capaz de resolver el problema.

Con el análisis de la información se observó que en las disfluencias del habla se presentan cuatro características significativas que son: duración, número de palabras, presencia de silencio interno y tipo de acto de diálogo; con ayuda de estas características se crearon árboles de decisión con el propósito de clasificar las elocuciones en dos conjuntos: tiene o no tiene disfluencia.

En el análisis realizado se encontró que la duración de una elocución es la característica más útil e importante para determinar si hay una disfluencia del habla, e incluso provoca que algunas otras características dejen de ser importantes, como es el caso del número de palabras en el segundo árbol. Otra característica importante es la identificación del acto de diálogo del hablante, porque en la mayoría de los casos, las reparaciones tienden a presentarse en las directivas de acción y en las afirmaciones.

Se encontró que las reparaciones más comunes en el corpus DIME son del tipo *modification repair*; con un 79%; por lo tanto, el análisis de corrección de reparaciones se enfocó en este tipo. Para facilitar la solución de este tipo de reparaciones se elaboró una

clasificación de los tipos de correcciones que se pueden realizar, los cuales son: simple, compleja y por repetición.

Además, se observó que en las reparaciones del tipo *modification repairs* se presenta, en gran medida, la ocurrencia de repetición de palabras, asimismo, la distancia entre estas palabras se encuentra en un rango de entre 0 y 2 palabras (casi el 90% de las elocuciones); esto ayudó a realizar el modelo de corrección.

El modelo de corrección se basa en la detección de las palabras repetidas y el número de palabras que las separa.

Una observación interesante es que normalmente se repiten las mismas palabras de acuerdo a los datos que se tienen, el 50% de las palabras que se repiten en una reparación está formado por sólo 6 palabras, como se muestra en la Tabla 6.1.

Palabra	Número de veces que se repitieron	Porcentaje de aparición
la	22	19.30%
el	10	8.77%
de	10	8.77%
que	6	5.26%
a	5	4.39%
hacia	4	3.51%
<b>Total</b>	<b>57</b>	<b>50%</b>

**Tabla 6.1: Relación de palabras que más se repiten**

Al analizar las palabras repetidas desde el punto de vista de su categoría léxica, se observa que en su mayoría son preposiciones y artículos definidos; en el corpus, el 56.14 % de las palabras repetidas son preposiciones o artículos definidos.

El algoritmo para realizar la corrección de las reparaciones es satisfactorio; al realizar el experimento, el 78% del material fue manejado correctamente por la heurística y se corrigió el 75.34% de los casos que se tenían que corregir.

El resultado de esta tesis es un primer acercamiento al fenómeno de las disfluencias del habla; sin embargo, la solución es sencilla y eficaz, resuelve gran parte del problema de las disfluencias de una manera satisfactoria al tener un porcentaje promedio de clasificación

del 86.7% y de corrección del 75.34%. De esta manera, se demuestra que al enfocar el trabajo al caso más común se puede obtener una solución bastante aceptable. A su vez, la solución fue desarrollada de tal manera que puede ser implementada junto con el sistema reconocedor del habla para que solucione las reparaciones al momento de hacer el reconocimiento.

## Trabajo futuro

Estos experimentos aún son preeliminares; sin embargo, se espera tener más datos para verificar que las ideas y métodos que fueron desarrollados y presentados en esta tesis sigan siendo válidos y poder llevar a cabo otro tipo de análisis.

Aún falta definir una solución para corregir los casos restantes, los *fresh start*, los *abridged repairs* y las correcciones complejas de los *modification repairs*. Es difícil encontrar patrones para realizar su detección con la finalidad de robustecer el árbol de decisión, o definir una heurística para su solución con el material actual; por eso es importante tener el resto del corpus etiquetado en los niveles que se definieron; si se cuenta con más material el árbol no sólo detectaría la disfluencia del habla, además, se podría llevar a cabo la clasificación de los tipos de reparación del habla.

Otros tipos de análisis pueden desarrollarse para mejorar la detección de las difluencias. Por ejemplo:

- Un estudio de los alargamientos de la vocal o consonante final de una palabra, y el estudio de las palabras fragmentadas puede indicar el fin de un *reparandum*. Estos dos fenómenos suelen ocurrir al final de un *reparandum* porque es el momento en el cual, el hablante se interrumpe y expresa una reparación del habla. Incluso, el alargamiento de una vocal o consonante puede considerarse un marcador del discurso porque en ese tiempo el hablante piensa en lo que quiere decir.
- Se puede considerar la inserción de nuevas variables como la cadena de tonos que se genera con INTSINT ([9], [27]). Este nivel de etiquetación forma parte de la gama de niveles que ya cuenta el corpus DIME. INTSINT (*IN*ternational *T*ranscription *S*ystem for *INT*onation) tiene por objeto capturar las distinciones superficiales de la lengua para construir patrones entonativos distintivos para la comparación lingüística de sistemas prosódicos ([10], [27]). Con este esquema se

puede identificar si existe algún cambio brusco en la entonación o identificar patrones de etiquetas de tal forma que se pueda detectar la presencia de una disfluencia por medio del tono, basado en la detección del punto de interrupción que se presenta como una distorsión durante la elocución.

- Se puede realizar un análisis más profundo de los marcadores del discurso; en esta tesis, los marcadores del discurso no fueron contemplados como variable útil para la generación del árbol de decisión o como una característica representativa y útil para robustecer la heurística de corrección. La razón de esta decisión es que sólo el 22.52% de las elocuciones con reparación presentaron marcadores del discurso como término de decisión, que es la relación más importante entre los marcadores del discurso y las reparaciones del habla; esto no significa que esta relación tan importante no exista, simplemente se necesita contar con más casos para llevar a cabo un mejor análisis. Este análisis puede determinar que tipo de marcadores del discurso se presentan más en las reparaciones del habla; por ejemplo, en las reparaciones de tipo *abridged repair*, el 83.33 % de los términos de edición tiene al menos un marcador del discurso del tipo interjección (MDI) y el 86.67% de estos casos sólo eran MDI. Al hacer un modelo de detección de marcadores del discurso se puede mejorar la detección de las reparaciones del habla, en particular la detección y corrección de las reparaciones de tipo *abridger repair*, que pueden ser corregidas incluso en el instante en que se detecta el marcador del discurso, lo único que se hace es remover esta palabra.



## Apéndice A – Alfabeto fonético MEXBET

El alfabeto MEXBET es el resultado del trabajo realizado por Javier Cuétara sobre el estudio de la fonética de la Ciudad de México [26]. Para el estudio se tomó como base empírica el corpus DIME, el alfabeto resultante se muestra a continuación:

<b>Consonantes</b>	Labiales	Labiodental	Dentales	Alveolares	Palatales	Velares
Oclusivos sordos	p_c / p		t_c / t		k_j	k_c / k
Oclusivos sonoros	b_c / b		d_c / d			g_c / g
Africado sordo					tS_c / tS	
Africado sonoros					dZ_c / dZ	
Fricativos sordos		f	s_[]	s		x
Fricativos sonoros	V		D	z	Z	G
Nasales	m		n_[]	n	n~	N
Vibrantes				r(/ r		
Lateral				l		
<b>Vocales</b>				Anteriores	Media	Posteriores
Cerradas				i / i_7 j		u / u_7 w
Medias				e / e_7 E / E_7		o / o_7 O / O_7
Abierta				a_j / a_j_7 a / a_7		a_2 / a_2_7

**Tabla - Apéndices 1: Alfabeto MEXBET**

La clasificación de este alfabeto se basa en el punto de articulación (columnas) y el modo de articulación (renglones). El modo de articulación se refiere a la postura que

adoptan los órganos que producen los sonidos; y el punto de articulación es el lugar donde hacen contacto los órganos que intervienen en la producción del sonido.

El nivel de transcripción de alófonos del corpus DIME se basa en los 54 alófonos mostrados en la tabla anterior.

## Referencias

- [1] Allen, J.F., Core, M.G.: Draft of DAMSL: Dialog Act Markup in Several Layers Annotation Scheme. Department of Computer Science, Rochester University, October, 1997.
- [2] An overview of the CART methodology,  
[http://www.researchmethods.org/CART\\_Overview.pdf](http://www.researchmethods.org/CART_Overview.pdf)
- [3] Beckman, M., Diaz Campos, M., Tevis and J, Morgan, T. (2000). Intonation across Spanish, in the Tones and Break Indices framework, *Forbus* (14), pp. 9 – 36.
- [4] D. Farwell, S Helmreich & M. Casper. SPOST: a Spanish Part-of-Speech Tagger.  
[http://crl.nmsu.edu/Publications/farwell/far\\_etal95.html](http://crl.nmsu.edu/Publications/farwell/far_etal95.html)
- [5] Dählback, N., Jönsson, A and Ahrenberg, L (1993). Wizard of Oz Studies – Why and How, *Knowledge-based Systems*, 6(4), páginas. 258 – 266.
- [6] David R. Traum and Peter A. Heeman, Utterance Units in Spoken Dialogue, *Dialogue Processing in Spoken Language Systems*, Notas de la conferencia en inteligencia artificial, páginas 125–140. Springer-Verlag, Heidelberg, 1997.
- [7] Diccionario de la Real Academia Español, <http://www.rae.es>
- [8] Enciclopedia Salvat.
- [9] Hirst, Daniel., Di Cristo, A., and Espesser, R. Levels of representation and levels of analysis for the description of intonation systems, In M. Horne (ed) *Prosody: Theory and Experiment*, Kluwer-Dordrecht, 2000
- [10] Hirst, Daniel. The symbolic coding of duration and timing: an extension to the INTSINT system. *Proceedings Eurospeech '99*. septiembre 1999, Budapest.  
<http://www.lpl.univ-aix.fr/~hirst/articles/1999%20Hirst.pdf>
- [11] [http://ciat-library.ciat.cgiar.org/documentos\\_electronicos\\_ciat/articulos\\_ciat/Manual\\_Arboles.pdf](http://ciat-library.ciat.cgiar.org/documentos_electronicos_ciat/articulos_ciat/Manual_Arboles.pdf)
- [12] <http://es.wikipedia.org/wiki/AI%C3%B3fono>
- [13] <http://es.wikipedia.org/wiki/Fonema>
- [14] [http://es.wikipedia.org/wiki/Marcadores\\_del\\_discurso](http://es.wikipedia.org/wiki/Marcadores_del_discurso)
- [15] <http://es.wikipedia.org/wiki/Palabra>
- [16] <http://leibniz.iimas.unam.mx/~luis/DIME/dimex-index.html>

- [17] <http://liceu.uab.es/fds/fds2005.html>
- [18] <http://www.amerschmad.org/spanish/gram/conjuga.htm>
- [19] <http://www.edufuturo.com/educacion.php?c=2754>
- [20] <http://www.geog.ntu.edu.tw/course/%AA%C5%B6%A1%AA%BE%C3%D1%BA%DE%B2z/ch06.files/1.htm>
- [21] <http://www.ilustrados.com/publicaciones/EpyuVuuEylbYOjNAoa.php>
- [22] <http://www.ling.ohio-state.edu/~tobi/sp-tobi/spanish.html>
- [23] <http://www.seminarioabierto.com/homiletica20.htm>
- [24] [http://www.ugr.es/~rlopezc/sistemas\\_dialogo.htm](http://www.ugr.es/~rlopezc/sistemas_dialogo.htm)
- [25] J. B. Pierrehumbert, *The Phonology and Phonetics of English Intonation*, Disertación doctoral, Massachusetts Institute of Technology, 1980.
- [26] Javier O. Cuétara Prieto. *Fonética de la ciudad de México. Aportaciones desde las tecnologías del habla. Tesis para obtener el grado de maestría. Universidad Nacional Autónoma de México*, 2004.
- [27] Joaquim Llisterri. *La representación fonética suprasegmental de corpus orales. Universidad Autónoma de Barcelona.*  
[http://liceu.uab.es/~joaquim/language\\_resources/spoken\\_res/Repres\\_fon\\_supraseg.html](http://liceu.uab.es/~joaquim/language_resources/spoken_res/Repres_fon_supraseg.html)
- [28] Julian Hirschberg y Diane Litman, *Empirical Studies on the Disambiguation of Cue Phrases. Computational Linguistics*, 19(3): 501 – 530, 1993.
- [29] L. A. Pineda, A. Massé, I. Meza, M. Salas, E. Schwarz, E. Uruga, y L. Villaseñor. *The DIME Project. En el proceedings del 2º Mexican Internacional Conference on Artificial Intelligence (MICAI-2002)*, páginas 166 – 175, Mérida, Yucatán, México, Abril, 2002.
- [30] M. Civit & M. A. Martí. *Design Principles for Spanish Treebank. En proceedings of the First Workshop on Treebanks and Linguistics Theories (TLT2002)*, páginas 61 – 77, septiembre 2002.
- [31] M. Civit. *Criterios de Etiquetación y Desambiguación Morfosintáctica de Corpus en Español. Tesis Doctoral. Universidad de Barcelona*, 2003.
- [32] *Manual de weka.* <http://metaemotion.com/diego.garcia.morate/>

- [33] Maria Vittoria Calvi y Giovanna Mapelli, Los marcadores bueno, pues, en fin, en los diccionarios de español e italiano. *Artifara*, n. 4, sezione Monographica, <http://www.artifara.com/rivista4/testi/marcadores.asp>
- [34] P. A. Herman. *Speech Repairs, Intonational Boundaries and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialog*. Tesis Doctoral. Universidad de Rochester, 1997.
- [35] Peter A. Heeman, Kyung-ho Loken-Kim, y James F. Allen. Combining the Detection and Correction of Speech Repairs. En el Proceedings del 4rd International Conference on Spoken Language Processing (ICSLP-96), páginas 358–361, Philadelphia, Octubre 1996, también aparece en *International Symposium on Spoken Dialogue*, 1996, páginas 133-136.
- [36] Pilar Prieto, (ed.), *Teorías de la entonación*, Barcelona, Ariel. 2003.
- [37] R. López-Cózar, A. J. Rubio, P. García, J. E. Díaz-Verdejo. *Interacción Hombre-Máquina mediante Sistemas Automáticos de Diálogo*. <http://giig.ugr.es/interaccion2000/trabajos/articulos/Lopez-cozar.doc>
- [38] Villaseñor, L., Massé, A. & Pineda, L. A. (2001). *The DIME Corpus*, Memorias 3º. Encuentro Internacional de Ciencias de la Computación ENC01, Tomo II, C. Zozaya, M. Mejía, P. Noriega y A. Sánchez (eds.), SMCC, Aguascalientes, Ags. México, Septiembre, 2001.